

Creació i manteniment d'un servei que analitzi el respecte a la privacitat dels usuaris a la xarxa

Grau en Enginyeria Informàtica

Especialització en Tecnologies de la Informació

Pol Mesegue Molina

Director

Pere Barlet Ros

Departament d'Arquitectura de Computadors

Tutor de GEP

Jorge Enrique Esteban Pérez

28 de Juny de 2021

Resum

El projecte ha consistit en crear i mantenir un servei que, de forma automatitzada, analitza els continguts d'una pàgina web donada, comprovant si conté elements que vulneren d'alguna manera la privacitat dels usuaris. Els resultats dels anàlisis es mostren de manera visual a través de diferents gràfics generats dinàmicament a partir de la informació obtinguda. Per a cada un dels dominis analitzats s'assigna un nivell d'intrusió en una escala del 0 al 99, per tal que qualsevol usuari que faci ús del servei pugui entendre i comparar el nivell de intrusió de les pàgines web que visita.

Resumen

El proyecto ha consistido en crear y mantener un servicio que, de forma automatizada, analiza los contenidos de una página web dada, comprobando si contiene elementos que vulneran de alguna manera la privacidad de los usuarios. Los resultados de los análisis se muestran de manera visual a través de diferentes gráficos generados dinámicamente a partir de la información obtenida. Para cada uno de los dominios analizados se asigna un nivel de intrusión en una escala del 0 al 99, para que cualquier usuario que haga uso del servicio pueda entender y comparar el nivel de intrusión de las páginas web que visita.

Abstract

The project consisted of creating and maintaining a service that automatically analyzes the contents of a given web page, checking if it contains elements that in any way violate the privacy of users. The results of the analysis are displayed visually through different graphs dynamically generated from the information obtained. For each of the domains analyzed, an intrusion level is assigned on a scale from 0 to 99, so that any user using the service can understand and compare the level of intrusion of the web pages they visit.

Índex de continguts

Índex de continguts.....	4
1 Contextualització.....	7
1.1 Conceptes propis.....	7
1.1.1 Online Resource Mapper (ORM).....	7
1.2 Actors implicats	9
1.3 Abast.....	10
1.3.1 Objectius.....	10
1.3.2 Requeriments no funcionals	10
2 Estudi del Problema	11
2.1 Solucions existents	11
2.2 Justificació de l'elecció	12
3 Arquitectura del servei.....	13
3.1 Disseny inicial	15
3.2 Disseny final	16
3.3 Arquitectura lògica final	18
4 Desenvolupament	19
4.1 Instal·lació de requisits.....	19
4.2 Backend	19
4.3 Frontend.....	20
4.3.1 Decisions de disseny.....	21
5 Resultat Final	31
5.1 Pàgina principal	31
5.1.1 Estadístiques	31
5.1.2 Inserir domini	31
5.1.3 TOPs.....	33
5.1.4 Diagrama de barres i taula de domini	34
5.2 Pàgina de domini.....	34
5.2.1 Actualitzar domini i nivell d'intrusió	34
5.2.2 Graf de recursos	35
5.2.3 Mapa de recursos.....	38
5.2.4 Diagrama de barres i taula de recurs	39
6 Gestió del Projecte	40
6.1 Metodologia	40
6.1.1 Eines de Treball	40

6.1.2	Seguiment	40
6.2	Planificació Temporal	41
6.2.1	Descripció de les Tasques.....	41
6.2.2	Estimacions.....	46
6.2.3	Desviacions.....	50
6.3	Gestió del Risc	50
6.3.1	Obstacles i riscos	50
7	Pressupostos	53
7.1	Identificació i estimació de costos	53
7.1.1	Recursos Humans	53
7.1.2	Recursos Materials	55
7.1.3	Recursos Indirectes	56
7.1.4	Contingències	57
7.1.5	Imprevistos.....	57
7.1.6	Pressupost Final	57
7.2	Control de Gestió	58
8	Informe de sostenibilitat.....	60
8.1	Projecte posat en producció	60
8.1.1	Dimensió ambiental	60
8.1.2	Dimensió econòmica	60
8.1.3	Dimensió social.....	60
8.2	Vida útil	60
8.2.1	Dimensió ambiental	60
8.2.2	Dimensió econòmica	61
8.2.3	Dimensió social.....	61
8.3	Riscos.....	61
8.3.1	Dimensió ambiental	61
8.3.2	Dimensió econòmica	61
8.3.3	Dimensió social.....	62
9	Conclusions i Competències tècniques.	63
9.1	Conclusions	63
9.2	Treball futur.....	63
9.3	Competències Tècniques	64
10	Referències.....	66

Índex de Figures

Figura 1: Diagrama UML de la base de dades.	14
Figura 2: Arquitectura inicial del servei.....	15
Figura 3: Arquitectura final del servei.....	16
Figura 4: Arquitectura lògica del servei.....	18
Figura 5: Preparació d'una sentència SQL per a evitar SQLInjection.	20
Figura 6: Enllaç de recursos Bootstrap i d3js.	21
Figura 7: Esbós del disseny de la pàgina web	23
Figura 8: Estadística de l'ús de tracking.	24
Figura 9: Estadística de l'ús de cada mètode de tracking.	25
Figura 10: Captura del resultat final de la pàgina principal del servei.....	26
Figura 11: Captura del resultat final de la pàgina d'un domini del servei.	27
Figura 12: Exemple d'ús d'Ajax.	28
Figura 13: Formulari de cerca de domini.	31
Figura 14: Pàgina d'espera.	32
Figura 15: Taula amb diferents pestanyes per a cada TOP.	33
Figura 16: Diagrama de barres i taula de domini.....	34
Figura 17: Nivell d'intrusió i botó d'actualitzar domini.....	34
Figura 18: Graf de domini.....	35
Figura 19: Graf amb colors segons si els recursos són third-party.	36
Figura 20: Graf amb colors segons el tipus d'element que és cada node.....	37
Figura 21: Mapamundi de recursos carregats.	38
Figura 22: Diagrama de barres i taula amb informació sobre un node.	39
Figura 23: Diagrama de Gantt.	49

Índex de Taules

Taula 1: Nivell d'intrusió per a cada tipus de tracking.	29
Taula 2: Estimacions.....	48
Taula 3: Sous i costos de cadascun dels rols del projecte.....	53
Taula 4: Costos de cadascuna de les tasques del projecte.	55
Taula 5: Costos materials.	56
Taula 6: Costos indirecte.	56
Taula 7: Costos amb contingència aplicada.	57
Taula 8: Costos d'imprevistos.	57
Taula 9: Costos totals del projecte.....	58

1 Contextualització

Aquest Treball de fi de Grau “Creació i manteniment d'un servei que analitzi el respecte a la privacitat dels usuaris a la xarxa” pertany als estudis de Grau en Enginyeria Informàtica de la Facultat d'Informàtica de Barcelona, de la Universitat Politècnica de Catalunya.

Està desenvolupat per l'estudiant Pol Mesegue Molina, de l'especialització en Tecnologies de la Informació, i dirigit pel professor Pere Barlet Ros, del departament d'Arquitectura de Computadors.

1.1 Conceptes propis

- **Cookie:** Petit arxiu de dades creat pel lloc web visitat i emmagatzemat en el navegador o dispositiu utilitzat pels usuaris. (1)
- **Privacitat:** Dret d'un usuari a protegir les seves dades en la xarxa i decidir quina informació és visible per a la resta.
- **Recurs *third-party*:** Recurs obtingut des d'un domini diferent del que visita l'usuari.
- **Tracker:** Entitat de seguiment o rastrejadora integrada en un lloc web per a identificar als seus usuaris, recol·lectant informació sobre aquest i analitzant el seu comportament.

1.1.1 Online Resource Mapper (ORM)

L'eina que analitza les pàgines web, anomenada Online Resource Mapper (2) , ha estat desenvolupat per diversos estudiants de la UPC durant els seus respectius Treballs de fi de Grau i Treballs de fi de Màster, liderats per Ismael Castell, estudiant de Doctorat.

ORM és una eina creada amb l'objectiu de mapejar la relació entre cada URL amb els recursos carregats en obrir-lo. L'eina llista tots els recursos carregats pels llocs web inspeccionats. La informació obtinguda es pot utilitzar per a diversos estudis, com ara trobar els recursos més utilitzats, així com el nombre de llocs web que carreguen un recurs específic.

ORM funciona executant finestres de Firefox que obren les pàgines webs a les que apunten els dominis que estiguin carregats en una Base de Dades. A continuació realitza una sèrie de proves i anàlisis, en els que obté tots els recursos i URL de la pàgina web, i per cada un d'ells comprova si conté alguna mena de rastreig.

Actualment, ORM detecta els següents tipus de tracking:

- **Very long-living Cookies:** Cookies amb un nivell d'expiració superior a 1 any. El període d'expiració està marcat per la GDPR en màxim un any (3), per tant aquest tipus de Cookie seria il·legal a la Unió Europea.
- **Long-living Cookies:** Aquelles Cookies amb un temps d'expiració d'entre 3 mesos i 1 any, que no són *third-party*.
- **Third-party Cookies:** Una *Third-Party Cookie* és una Cookie que un lloc web col·loca al disc dur d'un usuari des d'un domini diferent del visitat, amb un període d'expiració menor a 3 mesos.
- **Tracking Cookies:** Aquelles Cookies que són Long-Living Cookies i que a més són *third-party*.
- **Session Cookie:** Cookie *first-party* amb temps d'expiració menor a 3 mesos.
- **JavaScript Cookies:** Cookie creada, llegida o modificada amb *JavaScript*.
- **WebGL fingerprinting (4):** Aquesta tècnica utilitza l'element *Canvas* del navegador, que bàsicament és un element per a dibuixar gràfics. *JavaScript* pot crear un element *Canvas* nou i escriure-hi formes o textos personalitzats en aquesta àrea. Per dur a terme el *fingerprinting*, s'imprimeix una forma personalitzada en aquesta àrea i després el resultat es converteix en un identificador únic amb un codi hash. Atès que la memòria intermèdia d'impressió depèn de les capacitats de la resolució de la targeta gràfica i del monitor, proporciona una identificació única bastant bona de l'usuari. Aquest mètode permet accedir a la informació de hardware gràfic de l'usuari.
- **Mouse fingerprinting (5):** Aquesta tècnica consisteix en capturar el moviment del ratolí de l'usuari fent servir *JavaScript*. La forma de moure el ratolí es pot considerar un element bastant únic i identificador d'un usuari. També dona informació de com navega i en quins anuncis s'atura l'usuari.

- **Canvas fingerprinting (6):** Aquesta tècnica consisteix en generar àrees definibles (alçada i amplada) utilitzades per a la creació d'elements gràfics, logos i botons amb text en *JavaScript*. Així, depenent del sistema operatiu, el navegador web, la targeta gràfica, els controladors de la targeta gràfica i les fonts preinstal·lades en client, el text es visualitza de manera diferent, el que genera un fingerprint. Diferenciem entre *Canvas* de més de 32x32 píxels, que no se sol fer servir per res més que per a fingerprinting, i *Canvas* de 16x16, que encara que també es poden fer servir amb finalitats de tracking, en aquesta categoria també entrarien els les imatges de tipus icona.
- **Font fingerprinting (7):** Les tècniques de Font fingerprinting es basen en mesurar les dimensions dels d'elements HTML mostrats en pantalla omplint-los de fragments de text o de glifos Unicode.

1.2 Actors implicats

- **Desenvolupador:** El treball ha estat desenvolupat per Pol Mesegue, qui farà les tasques de documentació i implementació del servei web final que mostra tota la informació obtinguda per ORM, així com la posta en marxa i configuració dels servidors on s'allotjarà la pàgina web, base de dades i des d'on s'executarà ORM.
- **Director:** El Director és Pere Barlet Ros, professor agregat del Departament d'Arquitectura de Computadors de la UPC i director científic del Barcelona Neural Networking Center (BNN-UPC). S'ha encarregat de guiar al desenvolupador quan aquest ha tingut problemes o dubtes.
- **Altres estudiants:** L'eina ORM ha estat desenvolupada conjuntament amb altres estudiants de la UPC, qui han dissenyat, programat i implementat diferents mètodes de detecció de *tracking*.
- **Usuaris d'Internet:** Seran els principals beneficiats del servei, ja que podran consultar quines pàgines web respecten la privacitat i quins mètodes de *tracking* fan servir aquelles pàgines web que més utilitzen.
- **Serveis Web:** Focus principal de l'ORM, seran els objectes d'anàlisi.

1.3 Abast

El projecte pretén assolir l'objectiu de mantenir un servei web dinàmic que vagi mostrant la informació recopilada per ORM, des de diferents estadístiques globals, fins a informació detallada per a cada web analitzada. Aquesta informació es mostrarà de manera que sigui comprensible per a tota mena d'usuaris, sigui quin sigui el seu nivell de coneixements. També mantenir sempre en marxa el ORM, per tal de que un usuari pugui analitzar o actualitzar la informació d'una pàgina web sota demanda, així com per anar actualitzant de forma automàtica aquelles pàgines web que fa temps que no han sigut analitzades per ORM.

1.3.1 Objectius

- **Mantenir el servei operatiu i actualitzat:**
 - Mantenir els servidors on s'executa la base de dades, web i ORM segurs i actualitzats.
 - Poder consultar una pàgina web sota demanda.
 - Que si una web consultada sota demanda no ha estat analitzada, s'analitzi al moment.
 - Actualitzar regularment aquelles pàgines web que fa temps que no s'analitzen.
- **Divulgar la necessitat de la privacitat a la xarxa:**
 - Fer el *frontend* amigable per tal que els usuaris que facin servir la pàgina web, sigui quin sigui el seu nivell tècnic, la puguin fer servir i veure com estan sent objecte de diferents mètodes de *tracking*.
 - Donar informació en llenguatge comú sobre què significa cada mètrica que s'analitzi, per tal que els usuaris la puguin entendre i avaluar.

1.3.2 Requeriments no funcionals

- **Usabilitat:** Cal que el sistema sigui fàcil de fer servir per als usuaris que volen consultar les mètriques analitzades per una pàgina web en concret.
- **Escalabilitat:** El sistema ha de permetre analitzar milers de pàgines web.
- **Disponibilitat:** Idealment el sistema ha d'estar disponible el major temps possible.

2 Estudi del Problema

Actualment bona part de les pàgines web més populars estan plenes d'elements rastrejadors, *Cookies* i altres mètodes de *tracking* que pretenen desanonimitzar als usuaris o bé seguir-los per tal de crear-ne un perfil ja sigui polític, econòmic o social. Amb la informació que recopilen sobre nosaltres poden mostrar-nos publicitat dirigida, vendre les nostres dades a tercers o bé utilitzar-les amb qualsevol altre propòsit no necessàriament declarat per la pròpia pàgina.

La majoria d'usuaris no són conscients de quins mètodes de rastreig es fan servir, com evitar-los o bloquejar-los, quina informació és la que es pot recavar sobre nosaltres ni quines implicacions té que això sigui una pràctica generalitzada i habitual.

El propòsit del treball és conscienciar als usuaris d'això, mostrant, per a diferents pàgines webs, quin respecte per a la privacitat de les dades tenen, i que l'usuari es pugui formar al respecte.

A espanya s'aplica el Reglament General de la Protecció de Dades (GDPR) (8) així com la Llei Orgànica de Protecció de Dades Personals (LOPD) (9), que tenen l'objectiu de vetllar per la privacitat dels usuaris. Cal, però, veure fins a quin punt les pàgines webs que es troben regulades per aquestes normatives rastregen igualment els usuaris fent servir diferents mètodes de rastreig.

2.1 Solucions existents

Actualment existeixen sistemes que avaluen si una pàgina web és *confiable* o no, en el sentit de si és un web de *phishing* o conté *malware*. Hi ha centenars d'extensions de navegadors, antivirus i webs com Virustotal (10) que s'encarreguen d'avaluar aquests casos.

Però en el nostre cas no serveixen, ja que les pàgines web que rastregen a l'usuari no estan catalogades necessàriament com a perilloses per aquests serveis.

Hi ha extensions de navegadors que poden bloquejar alguns elements de tracking, ja sigui fent servir llistes negres d'URL que és conegut que rastregen o bloquejant l'ús de

JavaScript. Igualment, aquestes extensions, tot i que molt útils, no compleixen l'objectiu de formar a l'usuari en matèria de privacitat.

També hi ha repositoris d'alternatives a les webs i serveis més comuns, els quals és *vox populi* que no respecten als usuaris. Un exemple podrien ser privacytools.io o prism-break.org. Però aquestes pàgines web, només contenen informació per fugir dels serveis que rastregen oferint alternatives de codi lliure i confiabls, però no necessàriament donen la informació concreta de quines pràctiques duu a terme cada pàgina web.

2.2 Justificació de l'elecció

Havent vist que no hi ha una solució concreta que analitzi pàgines web i mostri quins mètodes de tracking fa servir, creiem necessària l'existència del nostre servei. És important oferir un lloc on els usuaris puguin conèixer i entendre quines pàgines web els rastregen i de quina manera, conscienciant-se de per què és important que es respecti la privacitat a la xarxa.

3 Arquitectura del servei

L'arquitectura del servei ha patit algunes modificacions al llarg del desenvolupament del projecte. Inicialment i durant la fase de proves, s'executava una sola instància d'ORM en el mateix servidor en el que hi havia funcionant la base de dades i el servidor web. Finalment es va passar d'aquest disseny en el que tot s'executava en un mateix servidor a un altre disseny amb tres servidors i dues instàncies d'ORM independents.

La forma en que funciona ORM també ha patit variacions. Inicialment, per executar-lo, calia indicar l'interval de IDs de dominis que es volien analitzar. Aquestes IDs de domini depenen de l'ordre en que el domini s'ha introduït a la base de dades. Per tant, analitzar diversos dominis que no haguessin sigut introduïts a la base de dades de manera seqüencial era un problema, doncs s'havien de llançar o bé diferents instàncies d'ORM paral·leles, amb el problema de rendiment i escalabilitat que això suposa, o bé anar llançant instàncies una darrere de l'altre conforme van acabant. Aquesta última opció requeriria mantenir un control a nivell de sistema operatiu que vigilés el número d'instàncies d'ORM executant-se i mantenir una cua d'instàncies a llançar-se.

Per tant l'ORM i la base de dades s'han modificat per a solucionar aquest problema. Ara cada domini té associat a la base de dades un bit de prioritat, i si ORM s'executa amb el paràmetre *--priority-scan* només analitzarà aquells dominis amb el bit de prioritat amb valor 1.

Altrament, si no s'afegeix aquet paràmetre, l'ORM va comprovant a la base de dades aquells dominis que fa, per defecte, 30 dies que no han sigut analitzats.

Així doncs actualment és mantenen dues instàncies d'ORM sempre en marxa, una mirant els dominis amb bit de prioritat a 1, per a analitzar dominis sota demana dels usuaris, i una altra instància que mira quins dominis tenen una data d'últim anàlisis superior als 30 dies.

En qualsevol cas, les tecnologies emprades no s'han modificat, i són les següents:

- Un servidor de base de dades MySQL, des de el que ORM recull la llista de dominis a analitzar i on emmagatzema tota la informació que obté durant els anàlisis. És la mateixa base de dades que després el servei web utilitza per a

recollir la informació, i on afegeix els nous dominis a analitzar que els usuaris han introduït.

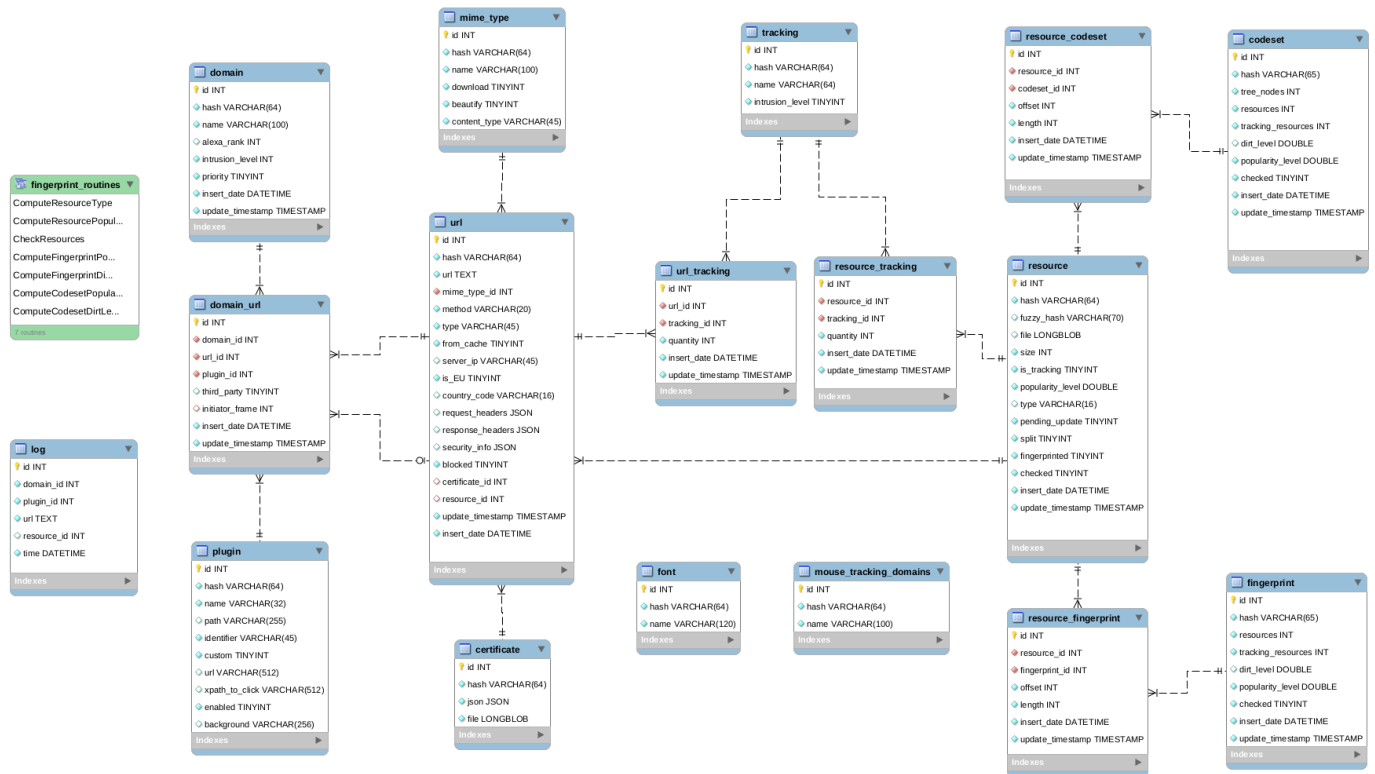


Figura 1: Diagrama UML de la base de dades.

- Apache HTTP server com a servidor.
- PHP com a llenguatge de programació per al *backend*.

3.1 Disseny inicial

Com s'ha comentat anteriorment, un principi es va apostar per una arquitectura d'un sol servidor que executés una instància d'ORM, el servidor de base de dades MySQL i la pròpia pàgina web.

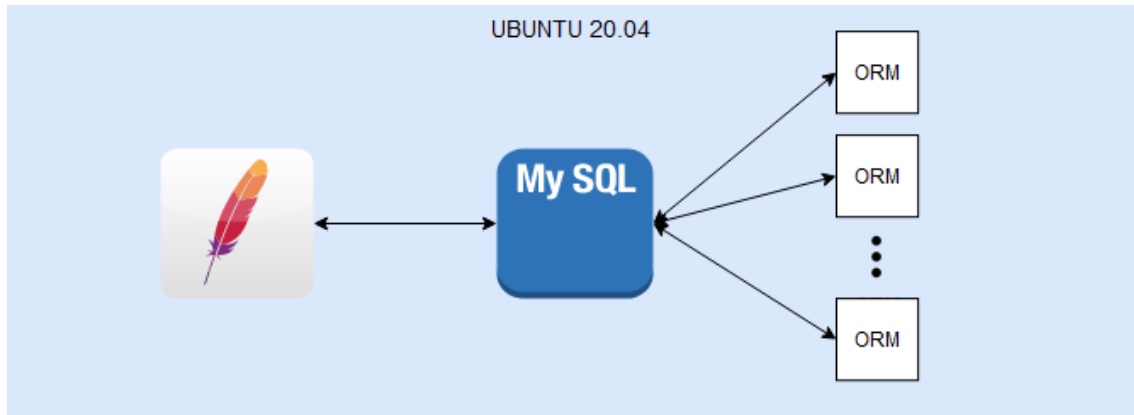


Figura 2: Arquitectura inicial del servei.

Durant la fase de proves en aquest entorn, van aparèixer els problemes anteriorment esmentats: si es volia afegir un nou domini per a ser analitzat, primer calia afegir-lo a la base de dades, després obtenir la ID del domini, i executar una instància de ORM amb un interval que inclogués aquella ID.

Aquesta versió s'executava en un servidor de Kimsufi amb les següents característiques:

- Cpu: Intel Atom N2800 2c/4t a 1,86 GHz
- Ram: 4GB DDR3 1066 MHz

3.2 Disseny final

Finalment, degut a les mancances descrites anteriorment, s'ha optat per el disseny següent:

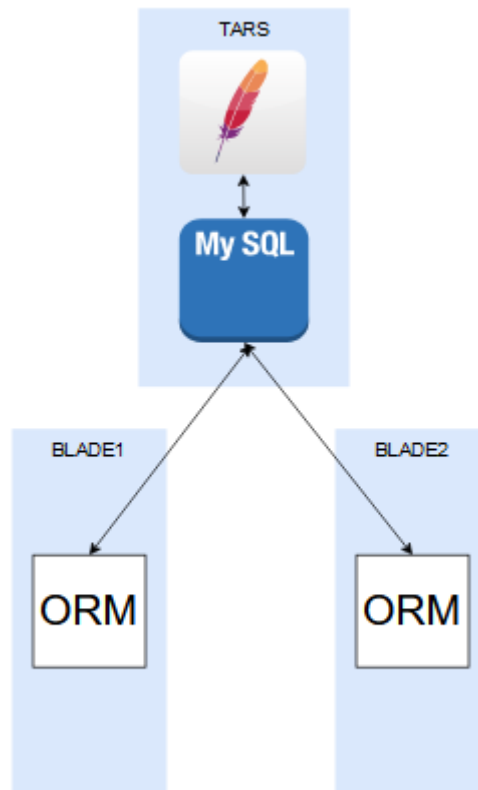


Figura 3: Arquitectura final del servei.

El servidor Tars conté el servidor web Apache HTTP, i la base de dades MySQL.

El servidor Blade1 ORM amb el paràmetre *--priority-scan*, analitzant només aquells dominis amb el bit de prioritat a 1.

El servidor Blade2 executa ORM analitzant aquells dominis que fa més de 30 dies que no s'actualitzen.

Blade1 i Blade2 es comuniquen a través del port 3306 amb Tars per a poder accedir a la base de dades.

Les característiques hardware dels servidors són:

Tars:

- CPU: 2 x Intel Xeon E5450 4c/4t
- RAM: 32GB DDR3 1066 MHz

Blade1:

- CPU: 2 x Intel Xeon L5640 6c/12t
- RAM: 6GB DDR2 1066 MHz

Blade2:

- CPU: 2 x Intel Xeon L5640 6c/12t
- RAM: 6GB DDR2 1066 MHz

Mentre els servidors de la solució final no estaven disponibles, calia igualment un lloc per a proves. Per això es va fer servir un servidor CPX21 de Hetzner (11) amb 3vCPU i 4GB de RAM, que executava una arquitectura intermèdia entre la solució final i la proposta inicial, amb el servidor Apache HTTP, una base de dades i una instància de ORM amb el paràmetre de *--priority-scan*.

3.3 Arquitectura lògica final

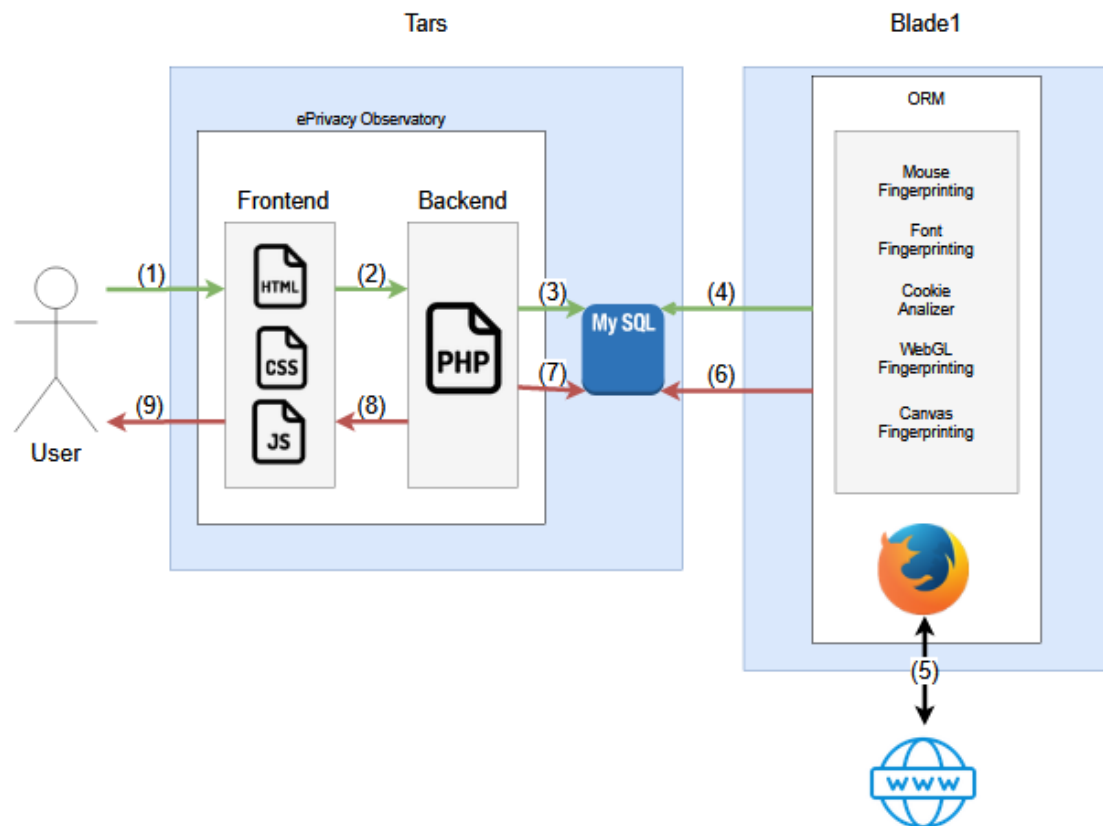


Figura 4: Arquitectura lògica del servei.

A continuació es mostra com es desenvolupa una petició d'analitzar un domini sota demanda per part d'un usuari.

1. L'usuari introdueix en el formulari de la pàgina web el domini que vol analitzar.
2. Mitjançant una petició POST, el domini introduït per l'usuari passa al *backend*.
3. El *backend* tracta el domini introduït, afegint-lo a la base de dades amb el bit de prioritat a 1.
4. L'ORM del servidor BLADE1, consulta la base de dades i veu que hi ha un domini amb el bit de prioritat a 1.
5. ORM analitza la pàgina web associada al domini, i n'extreu tota la informació.
6. ORM desa a la base de dades la informació extreta, i deixa el bit de prioritat del domini a 0.
7. El *backend* va fent *polling* a la base de dades, esperant a veure el bit de prioritat del domini a 0.
8. Un cop això passa, s'envia al *frontend* la informació respectiva al domini, per a que aquest generi els gràfics i mostri tota la informació.
9. L'usuari finalment pot veure la pàgina de domini interactuar amb ella.

4 Desenvolupament

4.1 Instal·lació de requisits

Els tres servidors descrits anteriorment funcionen amb GNU/Linux Ubuntu 20.04 LTS.

ORM funciona amb Python3.7 o superior, i necessita de pip (12) per a gestionar les llibreries i dependències necessàries. Per a computar els *fuzzy_hashes* (13) i poder comparar recursos, cal instal·lar TLSH (Trend Micro Locality Sensitive Hash). Finalment necessitem del navegador Firefox i de GeckoDriver, un Proxy per a poder utilitzar Webdriver, una interfície de control remot per a navegadors desenvolupat per Selenium (14). Cal notar que ORM s'executa en un servidor sense interfície gràfica, amb la qual cosa cal instal·lar també un servidor de visualització X11. Com no ens interessa realment poder veure les finestres que obre Firefox i tampoc volem instal·lar un escriptori gràfic als servidors, fem servir Xvfb (15), que a diferència d'altres servidors de visualització, realitza totes les operacions gràfiques a la memòria virtual sense mostrar cap sortida en pantalla.

Aquests requisits s'hauran de satisfer a blade1 i blade2, els servidors que executaran ORM.

A tars s'ha instal·lat un *stack* LAMP (Linux, Apache, MySQL i PHP) convencional.

A tots els servidors s'han realitzat tasques per a augmentar-ne la seguretat, instal·lant el *firewall* UFW (16) permetent només connexió a l'exterior en els ports 80, 443 i 22, i el 3306 només entre els tres servidors. S'instal·la i configura també *fail2ban* (17) per impedir atacs de força bruta contra els servidors. No s'ha optat per un sistema d'identificació amb claus privades *ssh* per petició de l'administrador dels servidors, aliè al projecte.

4.2 Backend

Com s'ha comentat a l'apartat d'arquitectura, s'ha optat per un *stack* LAMP en el costat del servidor. S'ha escollit aquest *stack* perquè són unes tecnologies ja conegudes per al desenvolupador, amb les quals ja ha realitzat altres projectes. A part, és un *stack*

completament de codi obert amb abundant informació a Internet, amb una solvència tècnica demostrada al llarg dels anys.

S'han pres una sèrie de mesures per evitar el que es coneix com *SQLInjection* (18), que consisteix en intentar alterar les bases de dades o obtenir-ne informació introduint sentències SQL en els formularis de les pàgines web.

Com l'únic formulari que hi ha és per analitzar i cercar dominis d'internet, el que s'ha fet és impedir l'enviament del formulari si no es compleix una expressió regular amb un patró de domini: `((?:[a-z\d](?:[a-z\d-]{0,63}[a-z\d])?|*)\.)+[a-z\d][a-z\d-]{0,63}[a-z\d]`.

Igualment, en el servidor, totes les sentències SQL que fan servir paràmetres introduïts per usuaris es tracten amb *mysqli*, un controlador de base de dades relacional per a PHP. Mitjançant *mysqli*, creem una sentència SQL preparada amb *prepare*, enllacem una o més variables a un paràmetre mitjançant *bind_param* i executem la consulta amb *execute*. No cal netejar els arguments en absolut, doncs *mysqli* ho fa per nosaltres.

```
$domain = htmlspecialchars(stripslashes(trim($_GET["domain_url"])));

$sql = "SELECT id FROM domain WHERE name = ?";
if ($stmt = mysqli_prepare($link, $sql)) {
    mysqli_stmt_bind_param($stmt, "s", $param_domain);
    $param_domain = $domain;
    if (mysqli_stmt_execute($stmt)) {
        mysqli_stmt_store_result($stmt);
        if (mysqli_stmt_num_rows($stmt) != 1) {
            header('Location: ./error.php?code=3');
            mysqli_stmt_close($stmt);
            die();
        }
    }
}
```

Figura 5: Preparació d'una sentència SQL per a evitar *SQLInjection*.

4.3 Frontend

El *frontend* s'ha realitzat amb *Bootstrap* (19), un *framework* de *css* de codi obert que simplifica el disseny de pàgines webs amb plantilles de disseny amb tipografies, formularis, botons, graelles i altres elements de disseny basats en *html*, *css* i *JavaScript*.

Per a realitzar els gràfics, que veurem més endavant, s'ha fet servir *d3.js* (20), una llibreria *JavaScript* per a produir gràfics dinàmics i interactius.

Tant per *Bootstrap* com per *d3.js* s'ha optat per no instal·lar-los directament en el servidor. En el seu lloc s'ha enllaçat, a les pàgines on fa falta *Bootstrap* o *d3.js*, els respectius repositoris que les dues eines mantenen amb tots els recursos necessaris per a funcionar.

```
#Origen dels scripts de d3js
<script src="https://d3js.org/d3.v4.min.js"></script>

#Origen del full d'estil de Bootstrap
<link href="https://cdn.jsdelivr.net/npm/bootstrap@5.0.1/dist/css/bootstrap.min.css" rel="stylesheet" integrity="sha384-+0n0xVW2eSR5OomGNYDnhzAbDsOXxcvSN1TPprVMTNDbiYZCxYbOO17+AMvyTG2x" crossorigin="anonymous">

#Origen dels scripts de Bootstrap.
<script src="https://cdn.jsdelivr.net/npm/bootstrap@5.0.1/dist/js/bootstrap.bundle.min.js" integrity="sha384-gtEjrD/SeCtmISKJkNUaakMoLD0//ElJ19smozuHV6z3Iehds+3U1b9Bn9Plx0x4" crossorigin="anonymous"></script>
```

Figura 6: Enllaç de recursos *Bootstrap* i *d3js*.

Aquesta decisió ha sigut presa degut a que d'aquesta manera no cal instal·lar res extra al servidor, i en el cas d'un canvi de versió només caldria actualitzar el *link* del document. No havent d'instal·lar res hem guanyat temps de desenvolupament, ja que amb els canvis d'arquitectura i de servidors, s'hauria d'haver realitzat la instal·lació fins tres vegades.

Igualment, en un futur estaria bé tenir-ho instal·lat, ja que si per algun motiu JsDeliver, el CDN que fa servir *Bootstrap*, o el repositori de *d3js* deixen d'estar operatius, la nostra web deixaria de funcionar correctament. Per tant la instal·lació disminuiria el grau de dependència de la web.

4.3.1 Decisions de disseny

Abans de començar a programar el *frontend* es va fer un esbós de com es volia veure el resultat final. Era important tenir una pàgina principal que contingués informació prou impactant com per a retenir als usuaris a la pàgina i animar-los a consultar aquelles webs

que fan servir. En un principi es va decidir que la primera informació a mostrar fos una taula amb diferents pestanyes que mostressin els següents TOPs: TOP pàgines més intrusives, TOP pàgines més populars segons la classificació Alexa (21), i TOP pàgines amb més elements *third-party*. Cada domini havia de tenir associat un nivell d'intrusió, el càlcul del qual ampliarem més endavant.

Al costat d'aquesta taula hi havia d'haver informació detallada sobre el tipus de *tracking* que tingués la web prèviament seleccionada, així com una taula secundària amb informació general sobre la pàgina que també recull ORM. Aquesta informació no necessàriament està relacionada amb el tracking, com pot ser la IP de la pàgina, si fa servir HTTPS o en quin país s'allotja, entre d'altres.

Després, per a cada domini, s'havia de poder accedir a una nova pàgina que contingués un graf de nodes que mostrés les dependències i relacions de cada URL i recurs de la pàgina web. Així doncs, cada node conté informació sobre la seva URL, si conté tracking i de quin tipus, si és *third-party*, si fa servir SSL i quina és la seva IP. Paral·lelament s'havia de poder veure un mapamundi on consultar per a cada país quants recursos carrega de la pàgina web.

Finalment, a la part dreta, s'havia de mostrar de nou informació relativa al tracking de la pàgina web i informació general sobre aquesta.

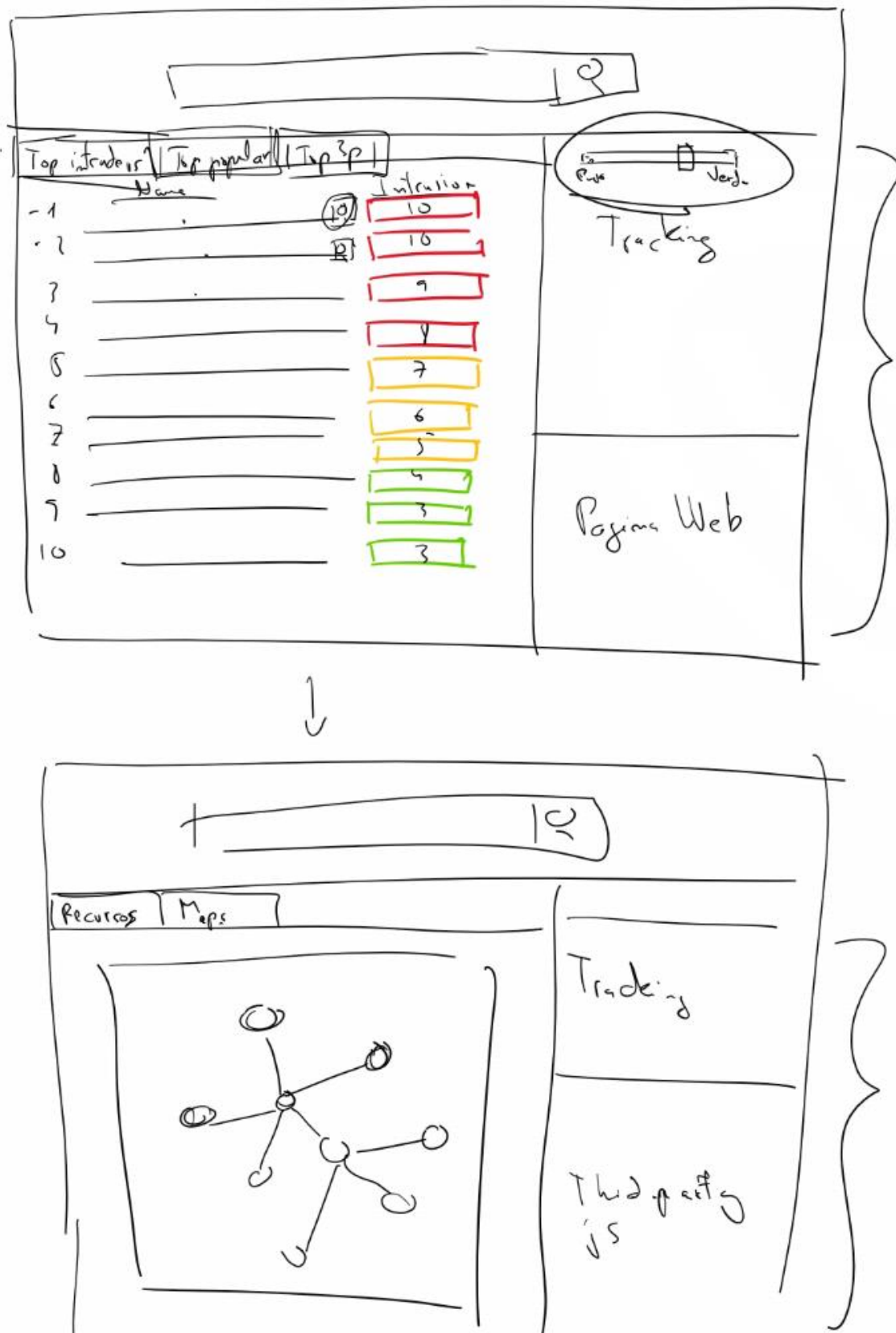
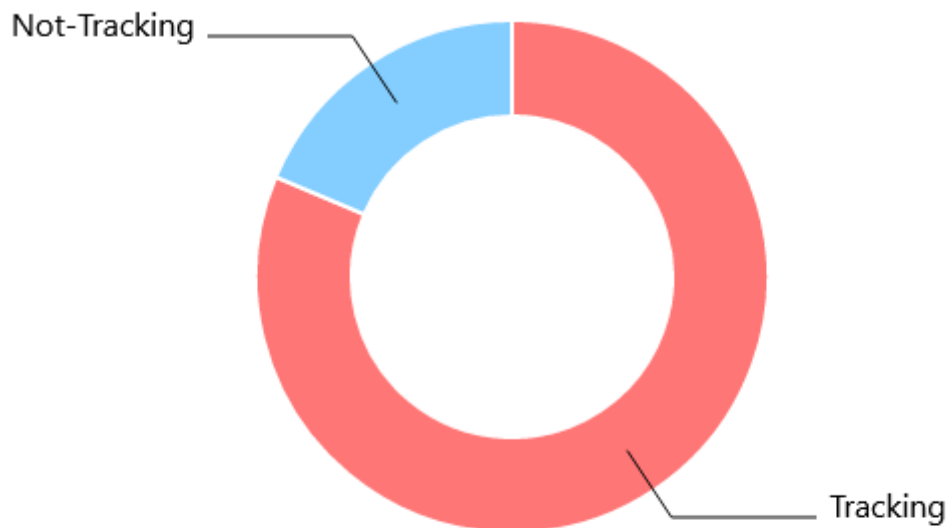


Figura 7: Esbós del disseny de la pàgina web

En el resultat final, que veurem a continuació, es va decidir mostrar primer de tot un carrusel de gràfics amb estadístiques, generades en temps real i que mostren sempre les dades més actualitzades. En el primer gràfic podem veure el percentatge de pàgines webs en les que s'ha trobat almenys un element de *tracking*.



Percentage of Domains using Tracking

81.35% of the 10013 analized Domains use atleast one tracking method.

Figura 8: Estadística de l'ús de tracking.

La segona estadística que es mostra, és la dels tipus de tracking que analitzem, i la seva proporció en aparicions a la web.

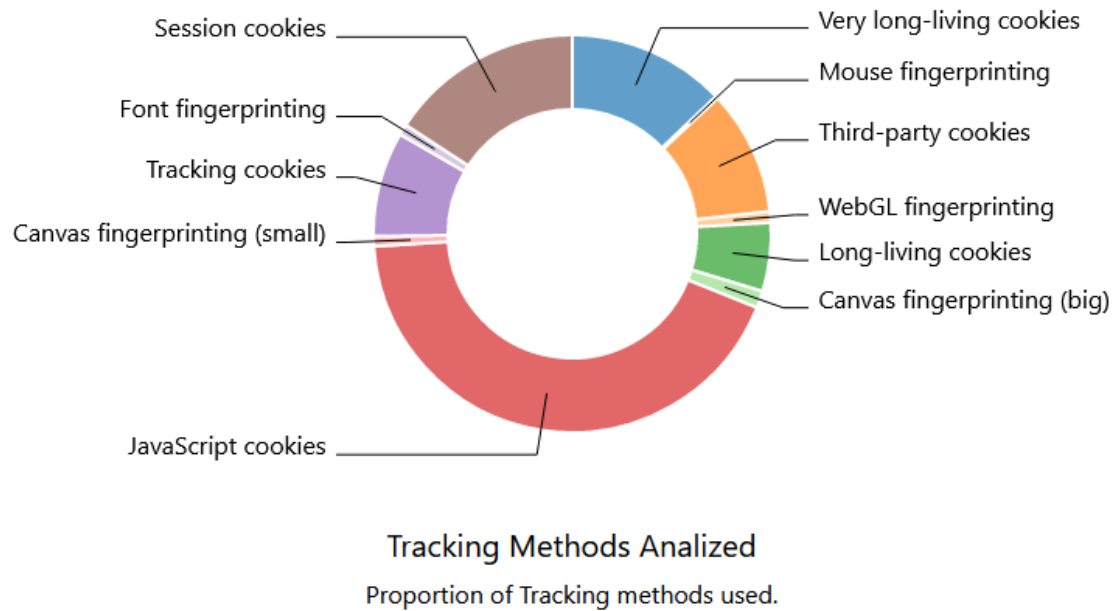


Figura 9: Estadística de l'ús de cada mètode de tracking.

El nom del portal escollit ha sigut ePrivacy Observatory, abreujat ePrivo, i s'espera adquirir els dominis eprivo.eu i eprivo.org en un futur.

El resultat final, que aprofundirem més endavant, un cop acabat el desenvolupament:

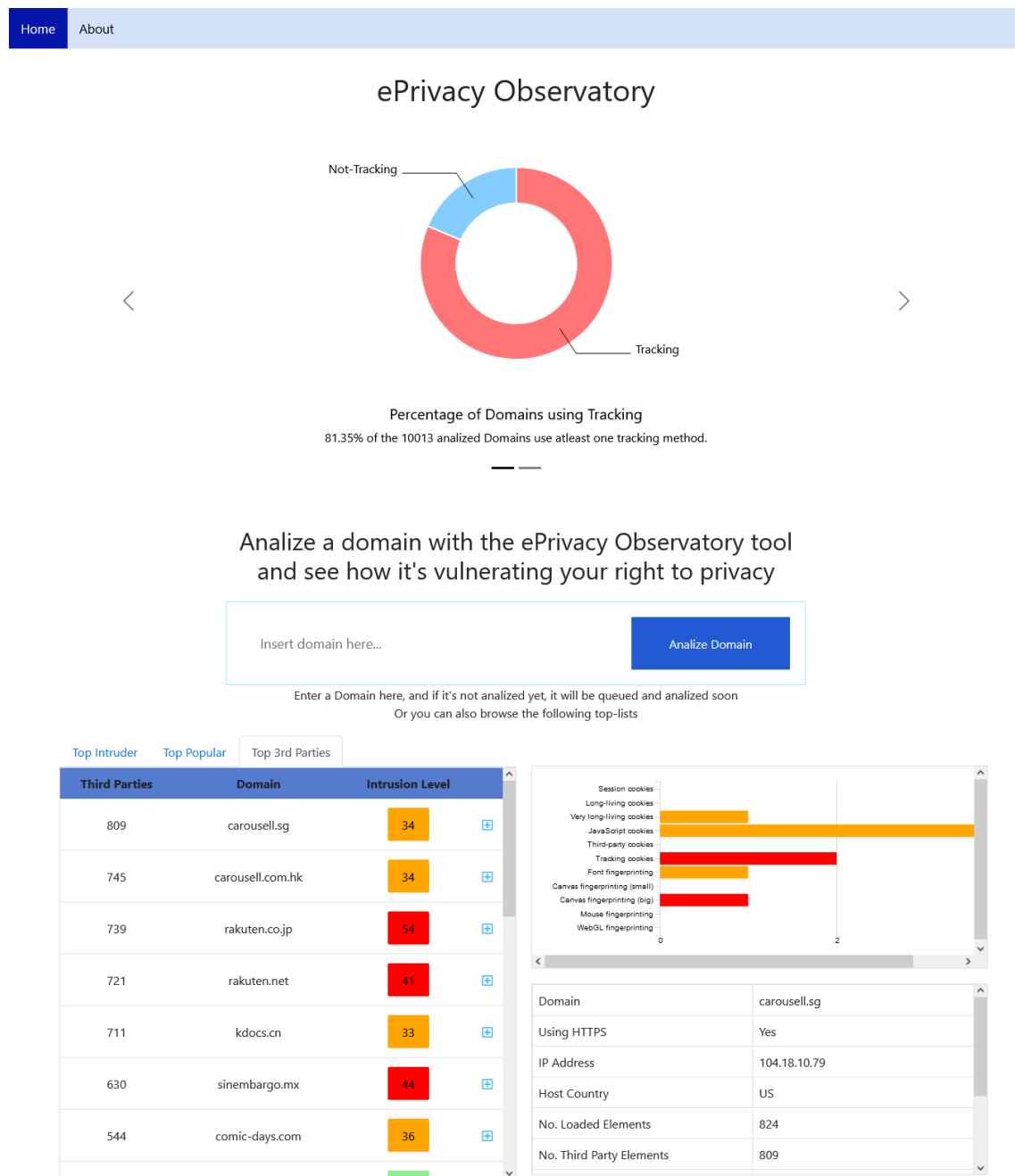


Figura 10: Captura del resultat final de la pàgina principal del servei.

Intrusion Level for carousell.sg: 34

This domain was analyzed by ORM on 2021-06-11 21:18:11. To re-analyze it, click here:

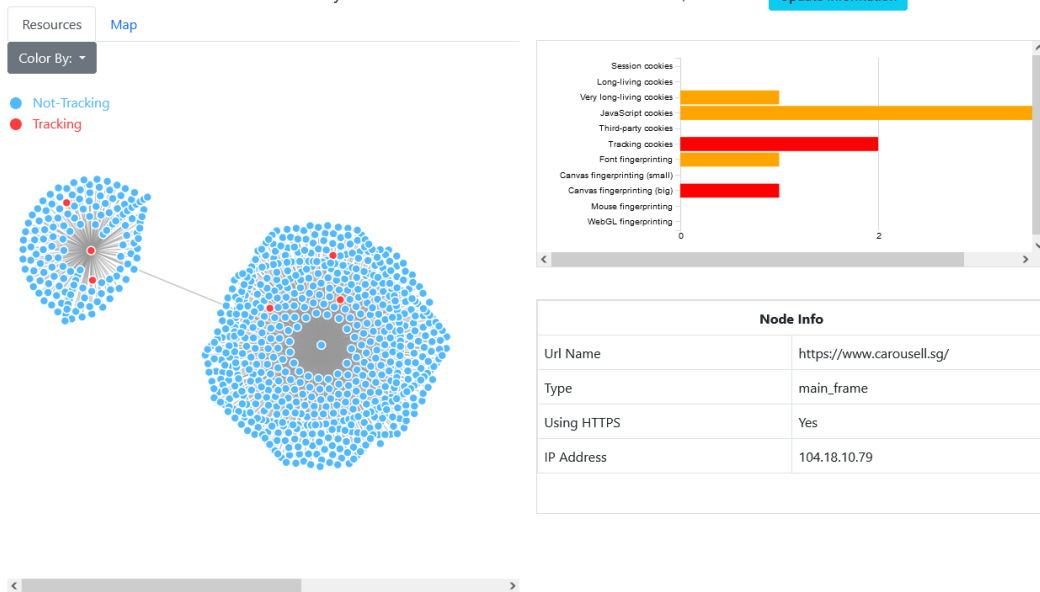
[Update Information](#)

Figura 11: Captura del resultat final de la pàgina d'un domini del servei.

Com molta informació que mostrem a la pàgina és necessari que es vagi actualitzant segons les interaccions dels usuaris, ha sigut necessari l'ús de AJAX (Asynchronous JavaScript and XML). Això és degut a que PHP és l'encarregat d'executar totes les sentències SQL amb les que obtenim les dades que volem mostrar, però PHP s'executa només quan carreguem la pàgina en qüestió, i sempre abans que es mostri qualsevol informació a la web.

AJAX, que no és res més que un conjunt de tècniques de desenvolupament web en *JavaScript*, i no pas un llenguatge de programació ni *framework*, ens soluciona el problema permetent fer crides asíncrones a altres pàgines PHP i mostrant el resultat en la pàgina actual.

```

function showTablePopular() {
    var xhttp = new XMLHttpRequest();
    xhttp.onreadystatechange = function() {
        if (this.readyState == 4 && this.status == 200) {
            document.getElementById("nav-popular").innerHTML =
                this.responseText;
        }
    };
    xhttp.open("GET", "get_top_popular.php", true);
    xhttp.send();
}

```

Figura 12: Exemple d'ús d'Ajax.

Per tant s'han fet la majoria de consultes a la base de dades en fitxers PHP concrets que retornen la informació correctament formatada per a poder-la mostrar directament amb *JavaScript*.

4.3.1.1 Nivell d'intrusió

S'ha decidit mostrar una mètrica associada a cada domini, anomenada *Intrusion Level* (Nivell de Intrusió). Aquesta mètrica pretén poder donar, per a cada domini, un valor numèric que representi en una escala del 0 al 99 quant invasiva és una pàgina web en concret.

Per a calcular-lo, cal saber que per cada tipus de tracking que ORM detecta, se li ha assignat un nivell d'intrusió que va del 1 al 9, ja que no tots els mètodes de rastreig els considerem igual d'invasius:

Tipus de tracking	Nivell d'intrusió
Session Cookies	1
Long-living Cookies	2
Very long-living Cookies	3
JavaScript Cookies	3
Third-party Cookies	4
Font fingerprinting	4
Tracking Cookies	5
Canvas fingerprinting (small)	5
Canvas fingerprinting (big)	9
Mouse fingerprinting	9
WebGL fingerprinting	9

Taula 1: Nivell d'intrusió per a cada tipus de tracking.

Així doncs, per a calcular el nivell d'intrusió, el que es fa és multiplicar el nivell d'intrusió de cada element tracking per el número d'elements tracking en qüestió que la pàgina carrega. S'ha establert un màxim de 9 punts per cada tipus de tracking. Aquesta decisió ha sigut presa perquè si no establíem aquest límit, aquelles webs que carregaven més recursos tracking, pel simple fet de carregar més recursos, ja aconseguien un nivell d'intrusió més elevat. Considerem també que és igual d'invasiu tenir un sol element de Mouse fingerprinting que tenir-ne 5, doncs el correcte seria no tenir-ne cap. També perquè d'aquesta manera, per tal que una web aconsegueixi un nivell alt de tracking és necessari que faci servir molts mètodes de tracking, i no només molts recursos tracking d'un sol tipus, ja que és creuant informació obtinguda a través de diferents mètodes que s'aconsegueix desanonimitzar i rastrejar a un usuari concret.

Finalment es sumen tots els valors i obtenim un número entre el 0 i 99, el nivell d'intrusió de la pàgina web.

Hem decidit acompanyar aquest resultat amb un color:

- Vermell per a nivells d'intrusió superiors o iguals a 40.
- Taronja per a nivells d'intrusió d'entre 21 i 39

- Groc per a nivells d'intrusió entre 1 i 20
- Verd per a nivells d'intrusió 0

En qualsevol cas, la manera de calcular aquesta mètrica pot ser modificada en un futur.

5 Resultat Final

A continuació fem una descripció de totes les funcionalitats del servei web final.

5.1 Pàgina principal

5.1.1 Estadístiques

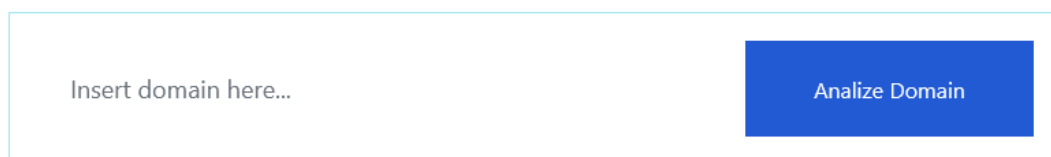
Com s'ha explicat en el punt 4.3.1, la primera informació que mostrem són estadístiques globals, en un format de carrusel. Actualment es mostra una estadística amb el percentatge de pàgines web que fan servir almenys un mètode de tracking, i una altre amb la proporció de mètodes de tracking que hem detectat.

És possible que s'afegeixin més estadístiques, doncs amb la quantitat de dades que obtenim amb ORM, hi ha bastanta informació en cru a la base de dades. Algunes dades que es podrien extreure i processar per a generar estadístiques amb elles podrien ser:

- Països amb més proporció de pàgines web amb elements de tracking.
- Recursos *third-party* més utilitzats per les pàgines web.
- Estadístiques del propi servei: Durant el treball s'han analitzat més de 10.000 dominis, a una mitjana de 30 segons d'anàlisi per domini, s'ha trigat 3 dies i mig en analitzar tots els dominis.

5.1.2 Inserir domini

Analyze a domain with the ePrivacy Observatory tool
and see how it's vulnerating your right to privacy



Enter a Domain here, and if it's not analyzed yet, it will be queued and analyzed soon
Or you can also browse the following top-lists

Figura 13: Formulari de cerca de domini.

Fent servir el formulari de la pàgina principal, podem tant cercar com inserir dominis per a analitzar-los. Si és un domini nou que no estava present a la base de dades i que per

tant no ha estat analitzat per ORM, se'ns redirigeix a una pàgina d'espera mentre ORM l'analitza. Aquesta pàgina només es mostra si el domini inserit té un registre A vàlid. En cas de no tenir-lo, es mostra un error.

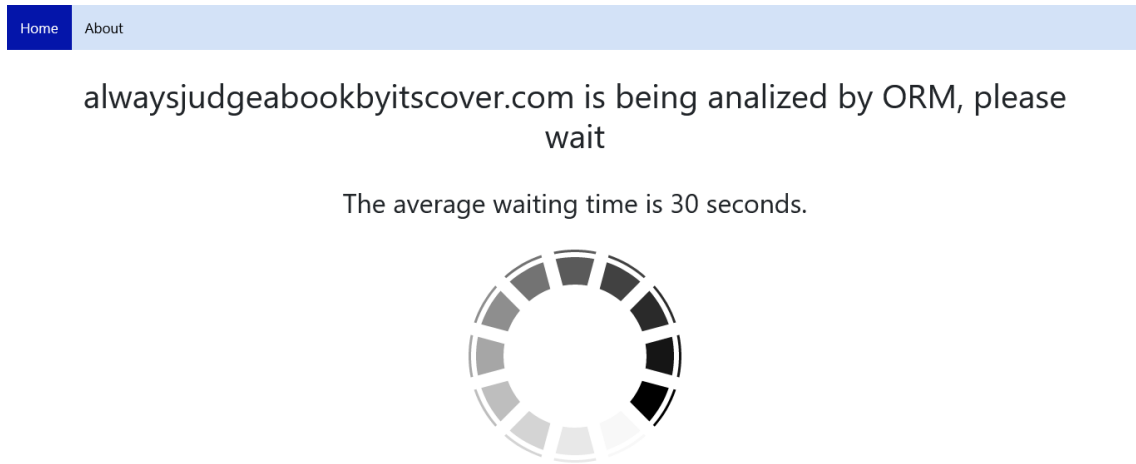


Figura 14: Pàgina d'espera.

Un cop la pàgina ha sigut analitzada, se'ns redirigeix a la pàgina del domini, de la mateixa manera que se'ns hauria redirigit en cas que el domini ja hagués sigut analitzat prèviament.

5.1.3 TOPs

Top Intruder			
Top Popular			
Top 3rd Parties			
Third Parties	Domain	Intrusion Level	
809	carousell.sg	34	+
745	carousell.com.hk	34	+
739	rakuten.co.jp	54	+
721	rakuten.net	41	+
711	kdocs.cn	33	+
630	sinembargo.mx	44	+
544	comic-days.com	36	+

Figura 15: Taula amb diferents pestanyes per a cada TOP.

Es mostra una taula amb diferents pestanyes, que ens permeten ordenar els 20 dominis més populars segons la classificació Alexa, els 20 que més nivell d'intrusió tenen i els 20 que més elements *third-party* carreguen. Cada domini és acompanyat del seu corresponent nivell d'intrusió.

5.1.4 Diagrama de barres i taula de domini

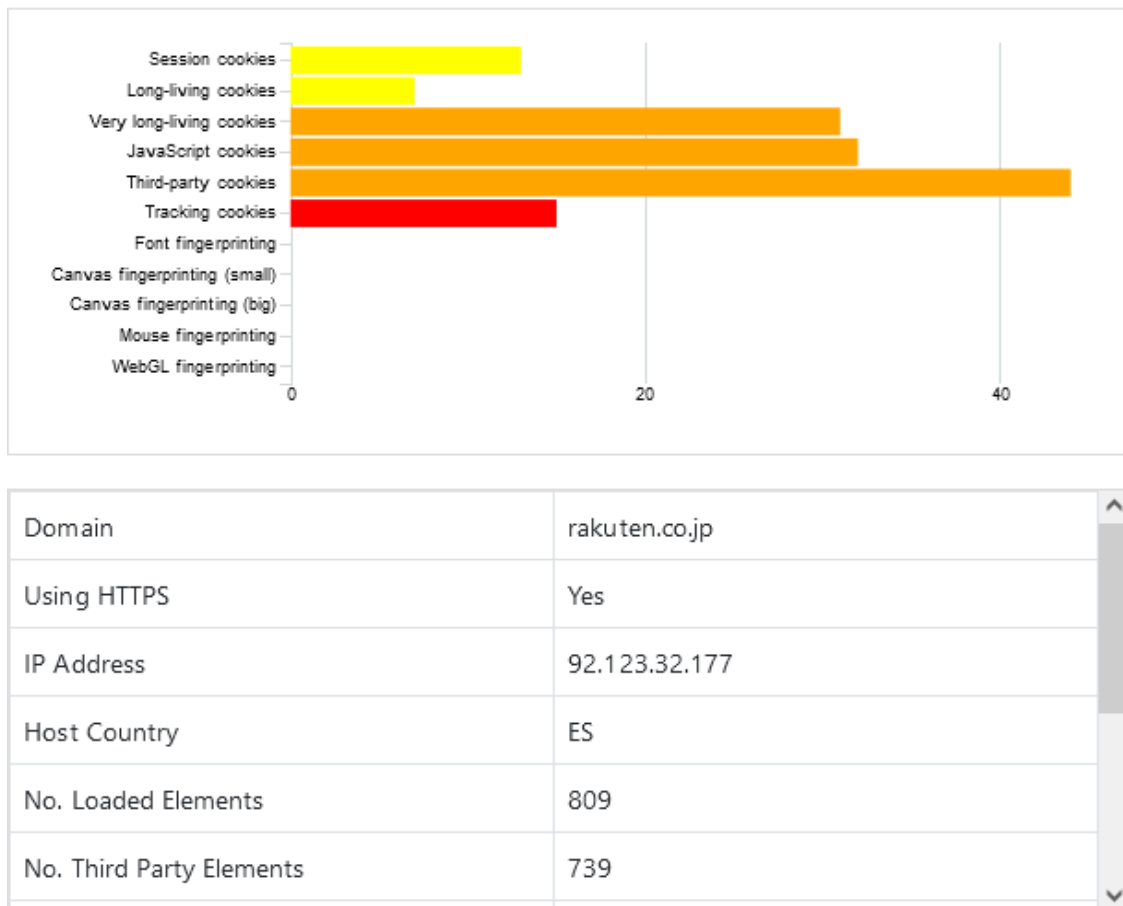


Figura 16: Diagrama de barres i taula de domini.

Per a cada domini que cliquem de la taula s'actualitza la informació associada a aquest, a la dreta de la pantalla. En primer lloc veiem un gràfic de barres que mostra el nombre d'elements de tracking que conté, i després una taula que ens informa de diferent informació relativa a la pàgina web.

5.2 Pàgina de domini

5.2.1 Actualitzar domini i nivell d'intrusió

Intrusion Level for trypap.com: 6

This domain was analyzed by ORM on 2021-06-16 12:05:25. To re-analyze it, click here: [Update Information](#)

Figura 17: Nivell d'intrusió i botó d'actualitzar domini.

En la mateixa pàgina del domini, se'ns indica el nivell de intrusivitat i la data en que va ser analitzat per última vegada. Podem actualitzar la informació mostrada tornant-lo a analitzar.

5.2.2 Graf de recursos

Es mostra un graf on cada node és un recurs carregat per la pàgina web. Les relacions entre nodes, representades amb arestes, mostren les dependències entre recursos. És a dir, un recurs pot carregar un o més recursos.

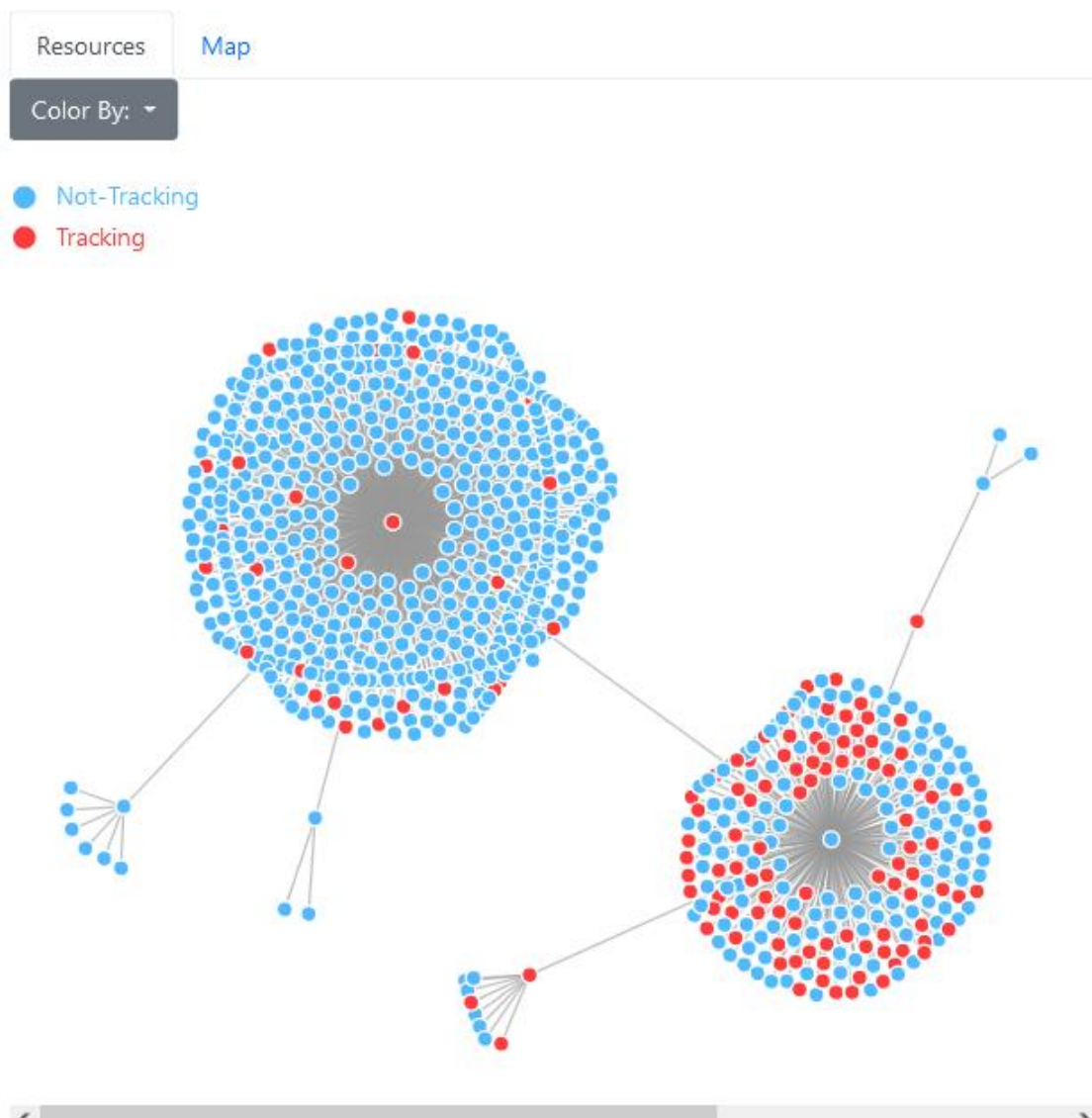


Figura 18: Graf de domini.

Podem dibuixar el graf de nodes amb diferents colors, ja sigui mostrant quins recursos són elements de tracking, quins recursos són elements *third-parthy* o simplement mostrant el tipus de recurs que és cada node.

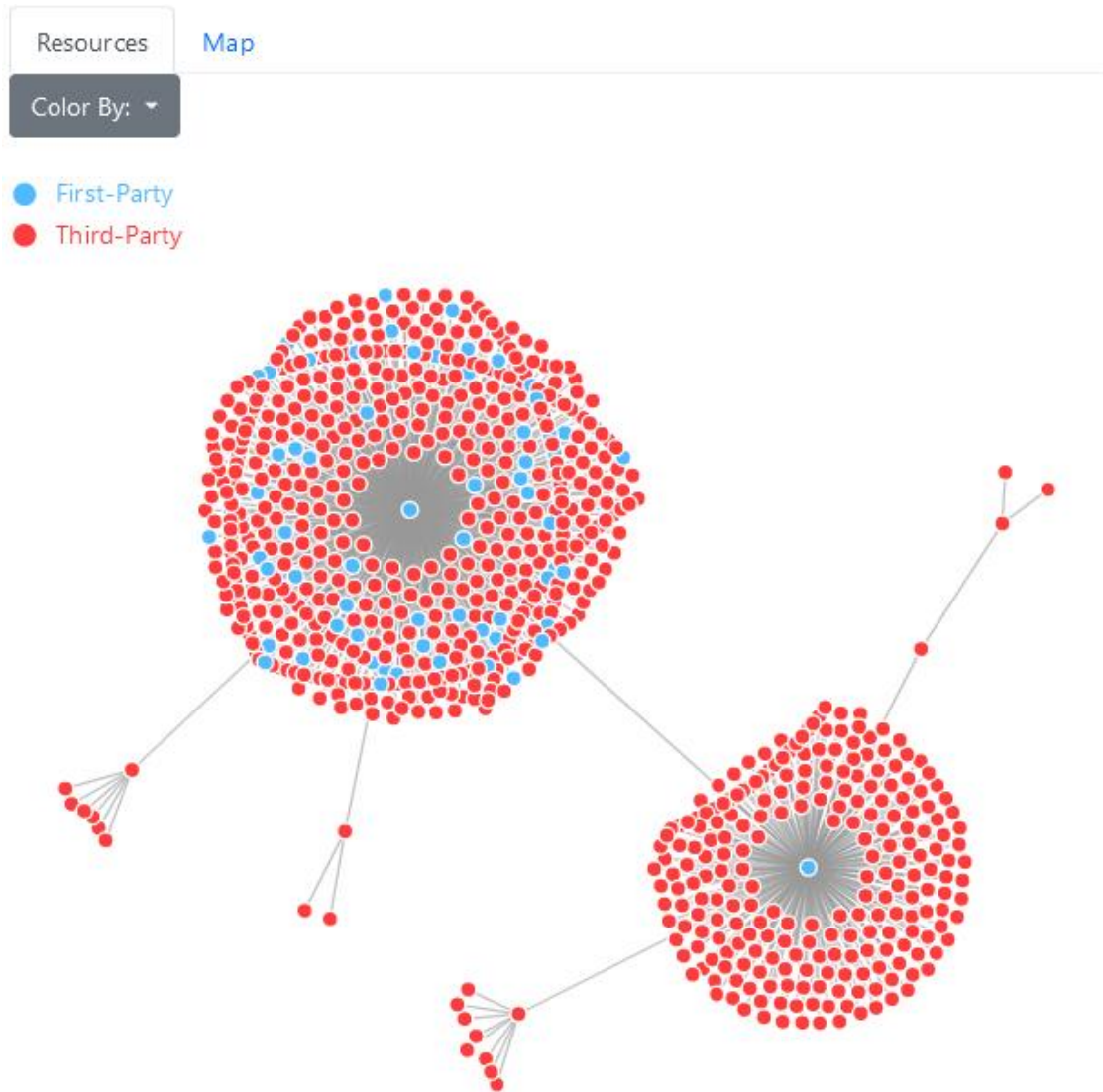


Figura 19: Graf amb colors segons si els recursos són *third-party*.



Figura 20: Graf amb colors segons el tipus d'element que és cada node.

5.2.3 Mapa de recursos

Si fem clic sobre la pestanya de Mapa, podem observar un mapamundi, en el que cada país té un color segons el nombre de recursos que carrega per a aquella pàgina web. Passant el ratolí per sobre del país, ens apareix un requadre amb el nombre de recursos carregats.

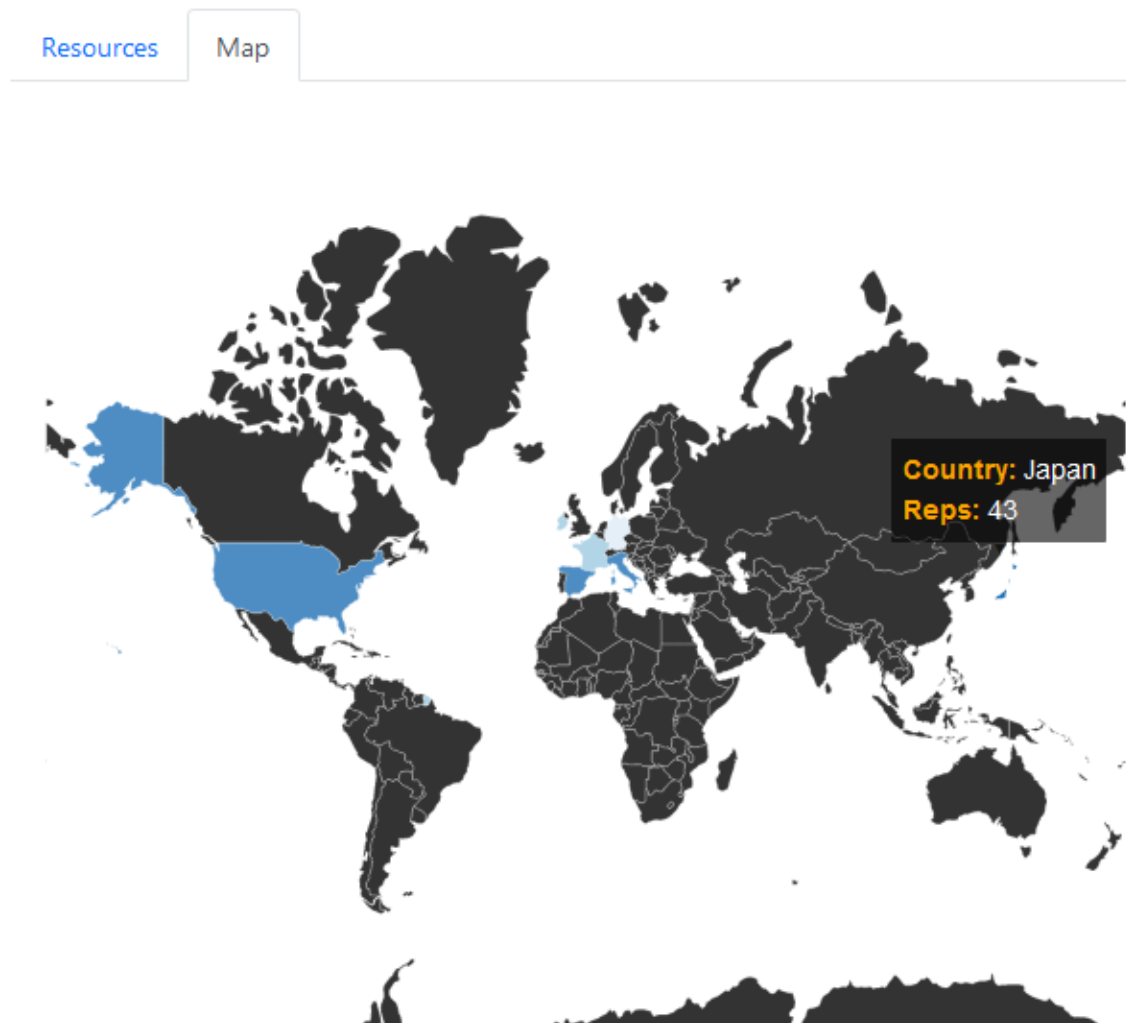
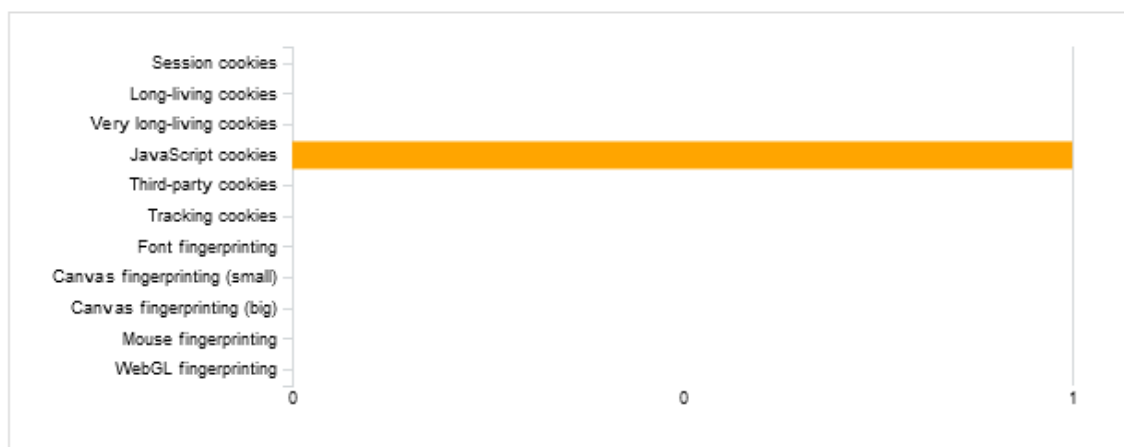


Figura 21: Mapamundi de recursos carregats.

5.2.4 Diagrama de barres i taula de recurs



Node Info	
Url Name	https://r.r10s.jp/com/inc/home/20080930/spt/common/js/pc_ichiba_top_async-1.4.6.js
Type	script
Using HTTPS	Yes
IP Address	151.101.242.63

Figura 22: Diagrama de barres i taula amb informació sobre un node.

Finalment, a la part dreta de la pàgina de domini, tenim de nou un diagrama de barres i una taula amb informació, però aquest cop s'anirà modificant a mesura que cliquem sobre els nodes del graf.

6 Gestió del Projecte

6.1 Metodologia

La metodologia empleada per a dur a terme el projecte ha sigut *Agile*, i concretament Scrum (22). Es tracta d'un model iteratiu i incremental, en el qual s'han definit *sprints* de dues setmanes. Aquesta metodologia està pensada per a treballs en equip, on cadascú té un rol. En cada *sprint* s'assignen aquestes tasques, que haurien d'estar acabades al final de l'*sprint*.

6.1.1 Eines de Treball

- **Meet:** Google Meet és una aplicació de videoconferències de Google per a navegadors web i dispositius mòbils. És la principal eina de comunicació entre el desenvolupador, el director del TFG i altres estudiants. El desenvolupador és conscient que Google i les seves diferents eines de productivitat no són les més indicades per a realitzar un treball sobre privacitat, però és l'entorn que la Facultat d'Informàtica de Barcelona ens proporciona als estudiants per comunicar-nos entre nosaltres i amb els professors.
- **Slack:** *Slack* és una plataforma de missatgeria basada en canals. És l'eina de comunicació emprada quan no hi ha reunions concertades.
- **Git:** És un sistema de control de versions gratuït i de codi obert.
- **VScode:** Editor de codi amb diverses funcionalitats que ajuden al desenvolupador a programar de manera ràpida, eficient i amb menys errors.

6.1.2 Seguiment

Els integrants del projecte, és a dir el director i altres estudiants, ens reunim els divendres cada dues setmanes de manera telemàtica per videoconferència. Igualment mantenim comunicació a través de *slack*.

Durant aquestes reunions es fa un repàs a les tasques assolides durant les dues setmanes, problemes que han sorgit i s'estudia com solucionar-los, així com definir la feina per les dues setmanes següents. També aprofitem per valorar les últimes notícies

sobre el món tecnològic i la privacitat de les dades, ja que qualsevol canvi important (noves lleis en matèria de protecció de dades, nous mètodes de *tracking*...) ha d'estar contemplat en el nostre projecte.

6.2 Planificació Temporal

El Treball de Fi de Grau té una durada aproximada de 4 mesos, des de finals de febrer amb la gestió del projecte fins finals de juny amb el torn de lectura. La dedicació setmanal al Treball és d'unes 30 hores, que pot variar en funció de les tasques a realitzar i dels problemes que puguin sorgir durant el desenvolupament d'aquestes.

Així doncs tenim 89 dies lectius i 35 no laborables, entre els que haurem de repartir les 540 hores aproximades que s'han de dedicar a realitzar completament el Treball de Fi de Grau. Les primeres 80-90 hores de treball es van dedicar a realitzar el GEP, que va començar el 23 de febrer i es va allargar fins el 22 de març. En el transcurs d'aquestes hores es fa la gestió del projecte, analitzant el context, definint l'abast i la planificació temporal, l'anàlisi de riscos i costos i finalment l'informe de sostenibilitat.

Paral·lelament es dediquen dues hores de manera bisetmanal a reunions de seguiment del Treball de Fi de Grau, així com les hores necessàries a realitzar les tasques inicials que el director del treball troba pertinents.

Un cop acabat el GEP, al 22 de Març es va començar a dedicar la totalitat de les hores al desenvolupament seguint la metodologia Scrum, fins al 14 de Juny que és la data estimada de finalització del treball. La última setmana, del 14 al 21 de Juny s'ha dedicat a la documentació i revisió del treball. Així doncs s'espera invertir unes 390 hores en el desenvolupament del treball.

6.2.1 Descripció de les Tasques

6.2.1.1 Curs GEP

- **GP1 - Contextualització i abast:** Definir el context i l'abast del projecte.
 - Duració: 20h
 - Dependències: -

- Recursos Humans: Desenvolupador
- Recursos Materials: Ordinador amb accés a Internet, Google Drive, Atenea.
- **GP2 - Planificació temporal:** Realitzar la definició de tasques i la planificació temporal.
 - Duració: 20h
 - Dependències: GP1
 - Recursos Humans: Desenvolupador
 - Recursos Materials: Ordinador amb accés a Internet, Google Drive, Atenea, Ganttproject
- **GP3 - Gestió econòmica i sostenibilitat:** Realitzar un pla econòmic i un informe de sostenibilitat
 - Duració: 20h
 - Dependències: GP2
 - Recursos Humans: Desenvolupador
 - Recursos Materials: Ordinador amb accés a Internet, Google Drive, Atenea.
- **GP4 - Definició final del projecte:** Agrupar els documents anteriors corregint els possibles errors.
 - Duració: 20h
 - Dependències: GP1, GP2, GP3
 - Recursos Humans: Desenvolupador
 - Recursos Materials: Ordinador amb accés a Internet, Google Drive, Atenea.

6.2.1.2 *Desenvolupament*

- **DV1 - Instal·lació ORM en local:** Preparar l'entorn del treball, descarregar els programes, llenguatges i llibreries necessàries per al funcionament d'ORM.
 - Duració: 15h
 - Dependències: -
 - Recursos Humans: Desenvolupador

- Recursos Materials: Ordinador amb accés a Internet, Git, editors de text, *slack*.
- **SG1 - Reunió comprovació ORM:** Comprovar que els resultats de l'ORM són els correctes.
 - Duració: 2h
 - Dependències: DV1
 - Recursos Humans: Desenvolupador, Director
 - Recursos Materials: Ordinador amb accés a Internet, meet.
- **DV2 - Instal·lació ORM en servidor:** Preparar l'entorn del treball, descarregar els programes, llenguatges i llibreries necessàries per al funcionament d'ORM en un servidor.
 - Duració: 20h
 - Dependències: DV1, SG1
 - Recursos Humans: Desenvolupador
 - Recursos Materials: Ordinador amb accés a Internet, Git, editors de text, servidor.
- **DV3 - Dockerització ORM**
 - Duració 40h
 - Dependències: DV2
 - Recursos Humans: Desenvolupador
 - Recursos Materials: Ordinador amb accés a Internet, Git, editors de text, servidor, docker.
- **SG2 - Seguiment Dockerització**
 - Duració 2h
 - Dependències: DV3
 - Recursos Humans: Desenvolupador, director
 - Recursos Materials: Ordinador amb accés a Internet, meet.
- **DC1 - Recerca tecnologies Front-End**
 - Duració: 20h
 - Dependències: -
 - Recursos Humans: Desenvolupador
 - Recursos Materials: Ordinador amb accés a Internet, editor de text.

- **DC2 - Documentar i justificar elecció software de Front-End pel servei.**
 - Duració: 10h
 - Dependències: DC1
 - Recursos Humans: Desenvolupador
 - Recursos Materials: Ordinador amb accés a Internet, editor de text.
- **SG3 - Reunió justificació tecnologies frontend.**
 - Duració: 2h
 - Dependències: DC2
 - Recursos Humans: Desenvolupador, Director
 - Recursos Materials: Ordinador amb accés a Internet, meet.
- **DV4 - Configurar servidor web per al frontend del treball.**
 - Duració: 30h
 - Dependències: SG3
 - Recursos Humans: Desenvolupador
 - Recursos Materials: Ordinador amb accés a Internet, editor de text, git, servidor.
- **DV5 - Primer Prototip frontend: Mostrar les dades recopilades per ORM.**
 - Duració: 40h
 - Dependències: DV2, DV4,
 - Recursos Humans: Desenvolupador
 - Recursos Materials: Ordinador amb accés a Internet, editor de text, git, servidor.
- **SG4 - Seguiment Prototip frontend**
 - Duració 2h
 - Dependències: DV5
 - Recursos Humans: Desenvolupador, Director
 - Recursos Materials: Ordinador amb accés a Internet, meet.
- **DC3 - Documentar procés d'instal·lació d'ORM i del software necessari per al servidor.**
 - Duració: 20h
 - Dependències: DV2, DV3, DV4
 - Recursos Humans: Desenvolupador

- Recursos Materials: Ordinador amb accés a Internet, editor de text.
- **DV6 - Automatització execució ORM.**
 - Duració: 50h
 - Dependències: DV2
 - Recursos Humans: Desenvolupador
 - Recursos Materials: Ordinador amb accés a Internet, editor de text, servidor.
- **SG5 - Seguiment Funcionament ORM en servidor**
 - Duració 2h
 - Dependències: DV2, DV3, DV6
 - Recursos Humans: Desenvolupador, Director
 - Recursos Materials: Ordinador amb accés a Internet, meet.
- **DV7 - Implementació mòduls per al ORM.**
 - Duració: 75h - Hi ha uns 5 mòduls, 15h per mòdul.
 - Dependències: DV2
 - Recursos Humans: Desenvolupador, desenvolupadors dels mòduls.
 - Recursos Materials: Ordinador amb accés a Internet, editor de text, servidor, meet.
- **DC4 - Documentar modulació ORM**
 - Duració: 20h
 - Dependències: DV7
 - Recursos Humans: Desenvolupador, desenvolupadors dels mòduls.
 - Recursos Materials: Ordinador amb accés a Internet, editor de text, servidor, meet.
- **DV8 - Disseny Final Front-End ORM modular:**
 - Duració: 10h
 - Dependències: DV4,DV5, DV7
 - Recursos Humans: Desenvolupador
 - Recursos Materials: Ordinador amb accés a Internet, editor de text, git, servidor.
- **DV9 - Correcció de Bugs i altres millores**
 - Duració: 50h

- Dependències: DV6, DV7, DV8
- Recursos Humans: Desenvolupador.
- Recursos Materials: Ordinador amb accés a Internet, editor de text, servidor, meet.
- **SG6 - Seguiment Final**
 - Duració 2h
 - Dependències: DV9
 - Recursos Humans: Desenvolupador, Director
 - Recursos Materials: Ordinador amb accés a Internet, meet.
- **DC5 - Crear la memòria**
 - Duració: 30h
 - Dependències: DC1, DC2, DC3, DC4, DC4, SG6
 - Recursos Humans: Desenvolupador
 - Recursos Materials: Ordinador amb accés a Internet, drive.
- **DC6 - Preparació Presentació**
 - Duració 20h
 - Dependències: DC5
 - Recursos Humans: Desenvolupador
 - Recursos Materials: Ordinador amb accés a Internet, drive.

6.2.2 Estimacions

6.2.2.1 Taula d'estimacions

ID	Tasca	Hores	Dependències
GP1	Contextualització i abast	20h	-
GP2	Planificació temporal	20h	GP1
GP3	Gestió econòmica i sostenibilitat	20h	GP2

GP4	Definició final del projecte	20h	GP1, GP2, GP3
DV1	Instal·lació ORM en local	15h	-
SG1	Reunió comprovació ORM:	2h	DV1
DV2	Instal·lació ORM en servidor	20h	DV1, SG1
DV3	<i>Dockerització ORM</i>	40h	DV2
SG2	Seguiment <i>Dockeritzacio</i>	2h	DV3
DC1	Recerca tecnologies Front-End	20h	-
DC2	Documentar i justificar elecció software de Front-End pel servei	10h	DC1
SG3	Reunió justificació tecnologies <i>frontend</i> .	2h	DC2
DV4	Configurar servidor web per al <i>frontend</i> del treball.	30h	SG3
DV5	Primer Prototip Front-End: Mostrar les dades recopilades per ORM.	40h	DV2, DV4
SG4	Seguiment Prototip <i>frontend</i>	2h	DV5
DC3	Documentar procés d'instal·lació d'ORM i del software necessari per al servidor.	20h	DV2, DV3, DV4
DV6	Automatització execució ORM.	50h	DV2
SG5	Seguiment Funcionament ORM en servidor	2h	DV2, DV3, DV6
DV7	Implementació mòduls per al ORM.	5*15h = 70h	DV2

DC4	Documentar modulació ORM	20h	DV7
DV8	Disseny Final Front-End ORM modular:	10h	DV4,DV5, DV7
DV9	Correcció de Bugs i altres millores	50h	DV6, DV7, DV8
SG6	Seguiment Final	2h	DV9
DC5	Crear la memòria	30h	DC1, DC2, DC3, DC4, DC4, SG6
DC6	Preparació Presentació	20h	DC5
TOTAL		537h	

Taula 2: Estimacions.

6.2.2.2 Gantt

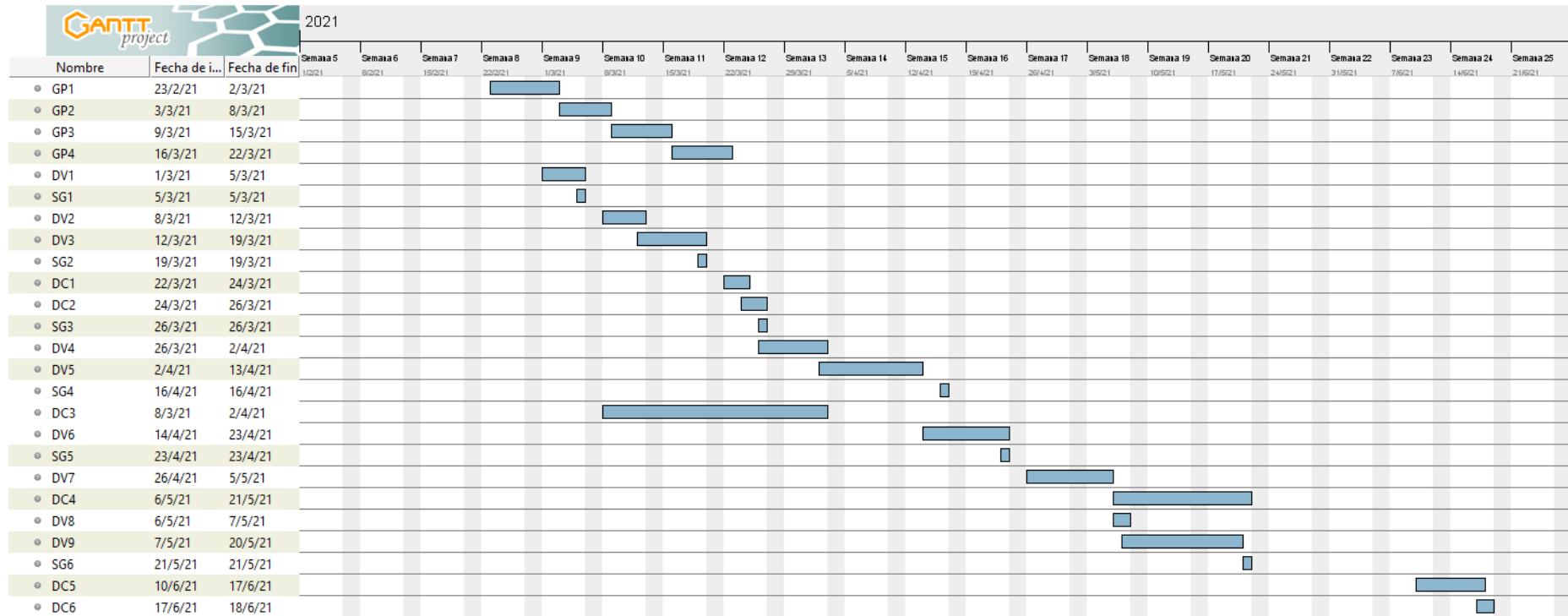


Figura 23: Diagrama de Gantt.

6.2.3 Desviacions

Finalment no hi ha hagut temps per realitzar la *dockerització* d'ORM. Un cop acabat el projecte tampoc considerem que sigui una cosa imprescindible. Els usuaris no s'aprofitarien d'aquesta *dockerització*, ja que l'objectiu principal és que facin servir la pàgina web, doncs mostra tota la informació necessària de forma gràfica, a més de que tenint-ho tot centralitzat ampliem la mostra de dominis analitzats.

La part de gestionar els mòduls d'ORM des del servei no s'ha aprofundit durant el treball, i finalment s'ha optat per a analitzar completament cada pàgina web, no essent possible fer només un anàlisi d'un cert tipus de tracking. Cal dir que el temps que triga ORM en analitzar completament una web no és significatiu comparant-lo en quant trigaria en analitzar només un tipus de tracking, per tant no es justifica els canvis que serien necessaris per realitzar una execució que només analitzés uns mètodes de tracking concrets.

Estava previst gestionar un *crawler* o aranya que revisés internet en cerca de nous dominis per a afegir a la base de dades, però no ha sigut possible realitzar-lo dins el termini. Igualment la base de dades conté actualment més 10.000 dominis, obtinguts de la classificació Alexa, amb la qual cosa ja és una dada bastant significativa del que ve a ser les webs més utilitzades, i que ens permet donar unes estadístiques generals sobre l'estat d'Internet bastant acurades.

Per altra banda, s'han dedicat moltes més hores de les previstes en la instal·lació d'ORM en el servidor de la primera arquitectura, així com en el *frontend* de la pàgina final, les dues coses degut a la falta d'experiència del desenvolupador.

6.3 Gestió del Risc

6.3.1 Obstacles i riscos

Els diferents obstacles i riscos que podrien afectar el desenvolupament del projecte o reduir la seva qualitat, s'avaluen a continuació.

- **Inexperiència en algunes tecnologies:** Cal fer el *frontend* del servei, i el desenvolupador té molt poca experiència en tecnologies *frontend* així com un baix nivell en el disseny d'aquest per tal que sigui amigable.
- **Bugs:** Com qualsevol projecte, cal estar pendent constantment dels bugs que van sorgint i anar-los arreglant, amb el contratemps que comporta.
- **Calendari tancat:** El fet de tenir una data de presentació tancada fa que potencialment no es puguin implementar totes les funcions volgudes o en la qualitat desitjada, per tal de no passar-se de terminis.
- **Implementació de mòduls de detecció de *tracking*:** Al ser un projecte modular, i que aquests mòduls estan programats per terceres persones, hi ha la possibilitat de que aquests no compleixin amb la qualitat desitjada.

Com ja s'ha mencionat anteriorment, hi ha una sèrie de riscos que poden obstaculitzar la correcta realització del treball i inclús dilatar les dates d'entrega.

És per això que a continuació s'ha calculat la possibilitat de que aquests obstacles apareguin així com una possible solució per a contenir-los.

- **Inexperiència en algunes tecnologies**
 - Probabilitat: Mitja
 - Aprendre noves tecnologies comporta un temps d'aprenentatge i recerca, que ja s'ha tingut en compte. Igualment s'ha fet una sobreestimació de les hores dedicades a desenvolupament en tecnologies en les que no hi ha un gran experiència.
- **Bugs:**
 - Probabilitat: Alta
 - Com qualsevol projecte, la fase de *debugging* i *testing* és molt important. No es pot ignorar que apareixeran *bugs* i que s'hauran de realitzar millores. Per això, es pot dedicar hores i tasques a exclusivament solucionar aquesta mena de problemes.
- **Calendari tancat:**
 - Probabilitat: Baixa

- Aquest és un escenari poc probable, doncs les dates són conegudes des del principi i s'ha fet una planificació del treball. En cas de no arribar a les fites, s'hauria de dedicar més hores de les planificades a la setmana.
- **Implementació de mòduls de detecció de tracking:**
 - Probabilitat: Mitja
 - Que apareguin errors en la implementació de mòduls degut a incompatibilitats està assumit, però com cada mòdul de detecció de *tracking* és independent, en cas d'error es pot simplement eliminar i fer els anàlisis sense aquell mòdul en concret. Això pot provocar que el resultat final no sigui tant acurat.

7 Pressupostos

7.1 Identificació i estimació de costos

Per a desenvolupar aquest projecte seran necessaris diferents tipus de recursos humans i materials. Els recursos humans són les persones que treballen en el projecte i els materials el hardware i software necessaris per al desenvolupament i execució d'aquests.

S'ha de tenir en compte també els costos indirectes, com les instal·lacions des d'on treballa el personal.

Finalment també s'ha calculat un cost estimat per a fer front a contingències i imprevistos.

Com a nota, tots els costos s'han arrodonit a l'alça per a fer un sobreestimament d'aquests.

7.1.1 Recursos Humans

Per poder saber quant costarà el cost humà total del projecte, cal definir el sou de tots els rols que en formen part.

En el meu projecte participen un desenvolupador i un director, però per a la part modular del projecte participen altres 3 estudiants de grau i 2 de màster.

Així doncs consideraré que els 3 estudiants de grau cobren 9 euros/hora i els de màster 10 euros/hora, tal i com estipula la normativa de la FIB (23).

Per al rol de director, que el duu a terme un professor associat a la universitat, he fet servir el portal GlassDoor (24), que és un directori que permet consultar els salaris de diferents posicions de treball.

Rol	Sou/Hora	Cost/Hora (S.S incl)
Desenvolupador Grau	9 euros	11,7 euros
Desenvolupador Màster	10 euros	13 euros
Director	24,80 euros	32,24 euros

Taula 3: Sous i costos de cadascun dels rols del projecte.

A continuació calculem el cost total del desenvolupament del projecte.

Per a les tasques DV7 i DC4 assumim que els desenvolupadors dels mòduls dedicaran menys hores que el desenvolupador del projecte.

Id Tasca	Hores	Participants	Cost (euros)
GP1	20h	Desenvolupador	234
GP2	20h	Desenvolupador	234
GP3	20h	Desenvolupador	234
GP4	20h	Desenvolupador	234
DV1	15h	Desenvolupador	175,5
SG1	2h	Desenvolupador, Director, Desenvolupadors dels mòduls (3g, 2m)	210,08
DV2	20h	Desenvolupador	234
DV3	40h	Desenvolupador	468
SG2	2h	Desenvolupador, Director, Desenvolupadors dels mòduls (3g, 2m)	210,08
DC1	20h	Desenvolupador	234
DC2	10h	Desenvolupador	117
SG3	2h	Desenvolupador, Director, Desenvolupadors dels mòduls (3g, 2m)	210,08
DV4	30h	Desenvolupador	351

DV5	40h	Desenvolupador	468
SG4	2h	Desenvolupador, Director, Desenvolupadors dels mòduls (3g, 2m)	210,08
DC3	20h	Desenvolupador	234
DV6	50h	Desenvolupador	585
SG5	2h	Desenvolupador, Director, Desenvolupadors dels mòduls (3g, 2m)	210,08
DV7	5*15h = 70h	Desenvolupador	819
	15h	Desenvolupadors dels mòduls (3g, 2mr)	916,5
DC4	20h	Desenvolupador	234
	3h	Desenvolupadors dels mòduls (3g, 2m)	183,3
DV8	10h	Desenvolupador	117
DV9	50h	Desenvolupador	585
SG6	2h	Desenvolupador, Director, Desenvolupadors dels mòduls (3g, 2mr)	210,08
DC5	30h	Desenvolupador	351
DC6	20h	Desenvolupador	234
Total			8502,78

Taula 4: Costos de cadascuna de les tasques del projecte.

7.1.2 Recursos Materials

El projecte es durà a terme en un portàtil Asus VivoBook de 700 euros. La vida útil esperada d'un portàtil és de 5 anys aproximadament. Per tant cada any del portàtil costa

140 euros, és a dir 11,67 euros al mes. Així doncs, pels 4 mesos de durada ens costarà 46,68 euros.

Un altre cost associat és el del servidor on s'executarà el servei.

El servei es podria executar en un VPS de Contabo (25), un hosting econòmic alemany.

El preu mensual per un VPS amb 4 vCPU Cores, 8 GB RAM, 200 GB SSD i connexió a 200 Mbit/s és de 4,99 euros/mes amb un cost d'instal·lació de 3,99 euros.

Considerarem 3 mesos de funcionament.

Recurs	Cost total (euros)	Cos/mes prorratejat	Cost pels 4 mesos de projecte (euros)
Asus VivoBook	700	11,67	
Servidor	18,96	4,74	
Total	718,96	16,41	65,64

Taula 5: Costos materials.

7.1.3 Recursos Indirectes

En els recursos indirectes cal considerar el cost de les instal·lacions des d'on treballen els desenvolupadors i director. Per a facilitar els càlculs i no haver de calcular el cost individual del serveis d'internet, electricitat, aigua etc. assumirem que es treballa des de un *coworking* a la ciutat de Barcelona.

S'ha escollit el *coworking* MyMonday (26), que per 299 euros + iva permet tenir una taula fixa de treball a la Barceloneta amb tots els serveis bàsics.

Recurs	Preu/Mes	Total (4 mesos)
Coworking MyMonday	361,79 iva incl.	1.447,16

Taula 6: Costos indirecte.

7.1.4 Contingències

S'assumeix un cost fix del 10% per a contingències. La taula següent mostra els costos anteriorment calculats amb aquest percentatge aplicat.

Recurs	Cost (euros)	Cost +10%
Recursos Humans	8502,78	9.353,06
Recursos Materials	65,64	72,21
Recursos Indirectes	1.447,16	1.591,88

Taula 7: Costos amb contingència aplicada.

7.1.5 Imprevistos

Els imprevistos que afecten a la planificació temporal ja s'han considerat prèviament en el GEP.

Pel que fa a riscos econòmics, podem suposar que l'ordinador personal del desenvolupador falla i que per tant l'ha de reparar. Assumeixo un cost de reparació de 200 euros, que seria l'equivalent a que falla 1 peça més el preu típic per hora de reparació. Com el desenvolupador té molta cura del portàtil i li fa manteniment, assumirem que la probabilitat de que falli és només del 5%

En cas de que el *coworking* tanqués, es podria ràpidament trobar un altre de característiques i preus similars a Barcelona o rodalies, per tant no el considerem en aquest apartat.

Recurs	Reparació	Probabilitat	Cost
Asus VivoBook	200 euros	5%	10 euros

Taula 8: Costos d'imprevistos.

7.1.6 Pressupost Final

Així doncs, en la següent taula podem veure el cost total del projecte:

Recurs	Cost amb contingències (euros)
Cost Recursos Humans	9.353,06
Cost Recursos Materials	72,21
Cost Recursos Indirectes	1.591,88
Partida per imprevistos	13
Total	11.030,15

Taula 9: Costos totals del projecte.

7.2 Control de Gestió

Cada cop que una tasca s'hagi realitzat, s'ha de comprovar que la planificació inicial s'ha complert i corregir les possibles desviacions en el pressupost. Per a fer-ho es farien servir les següents fórmules:

- **Desviació d'hores consumides per tasca**
 - $(\text{Hores estimades} - \text{Hores reals}) * \text{Cost estimat}$
- **Desviació dels costos segons les hores consumides per tasca**
 - $(\text{Hores estimades} - \text{Hores reals}) * \text{Cost real}$
- **Desviació dels costos en recursos humans per tasca**
 - $(\text{Cost estimat} - \text{Cost real}) * \text{Hores reals}$
- **Desviació total dels costos materials**
 - $\text{Cost material estimat} - \text{Cost material real}$
- **Desviació total dels costos indirectes**
 - $\text{Cost indirecte estimat} - \text{Cost indirecte real}$
- **Desviació total dels imprevistos**
 - $\text{Cost imprevist estimat} - \text{Cost imprevist real}$
- **Desviació total dels costos de personal**
 - $\text{Cost personal estimat} - \text{Cost personal real}$
- **Desviació total d'hores**

- Hores estimades – Hores reals
- **Desviació total dels costos**
 - Cost total estimat – Cost total real

8 Informe de sostenibilitat

8.1 Projecte posat en producció

8.1.1 Dimensió ambiental

El cost ambiental del projecte ha sigut el corresponent a tenir un ordinador portàtil, amb un consum aproximat de 0.05 kWh encès durant el desenvolupament del projecte, al qual s'han dedicat finalment unes 550 hores.

El consum total ha sigut, per tant, de 27,5 kWh.

8.1.2 Dimensió econòmica

Durant el desenvolupament s'han fet servir servidors VPS, els ja comentats de Kimsufi i Hetzner, tal i com estava pressupostat.

8.1.3 Dimensió social

A nivell personal aquest projecte m'ha permès aprendre noves tecnologies i aprofundir en llenguatges de programació. Realitzant aquest treball he pogut reafirmar el meu interès per a la privacitat de les dades i ha sigut gratificant poder participar en un projecte que espero que segueixi en desenvolupament més enllà de la realització d'aquest projecte.

8.2 Vida útil

8.2.1 Dimensió ambiental

El projecte posat en producció consta de 3 servidors amb les característiques descrites a l'aparat 3.2. Cal tenir en compte que els servidors s'han muntat fent servir peces antigues ja utilitzades i reutilitzades. En total tenim 2 Intel Xeon E5450, que consumeixen 0,08 kWh i 4 Intel Xeon L5640 amb un consum de 0,06 kWh. Per tant en total la nostre solució consumeix 0,4 kWh. Una manera de reduir aquest consum seria amb processadors més eficients.

8.2.2 Dimensió econòmica

El cost econòmic final del projecte ha sigut menor de l'esperat, doncs els servidors no es troben allotjats per un proveïdor de servidors, si no que es troben a l'UPC i han sigut construïts amb components reutilitzats.

8.2.3 Dimensió social

La dimensió social del servei durant la seva vida útil probablement sigui la més important de totes.

S'espera que el servei esdevingui un lloc popular entre els usuaris interessats en la privacitat de les dades, on consultin per a cada web que fan servir com estan sent rastrejats.

Idealment ePrivacy Observatory podria esdevindre un referent en l'anàlisi de pàgines webs en el que a la privacitat respecta, tal i com *virustotal* ho és per a l'anàlisi de malware, facilitant a tots els usuaris prendre decisions sobre si decideixen navegar en una web o no en base a la informació que mostrem.

Si s'aconsegueix que suficients usuaris siguin conscients dels seus drets a la privacitat, aquests podrien decidir deixar de navegar per aquelles webs que no els respecten, o fer pressió per a que siguin respectats.

8.3 Riscos

8.3.1 Dimensió ambiental

Si el servei tingués molt d'èxit, potser s'hauria d'ampliar el nombre de servidors per tal de fer front a la demanda. Tant mateix, els CPU són del 2010 aproximadament, amb la qual cosa és probable que s'hagin de reemplaçar en no massa temps, amb la petjada ambiental que suposa qualsevol tipus de residu.

8.3.2 Dimensió econòmica

El projecte, al no tenir fins de lucre, no hi ha escenaris que impedeixin la seva viabilitat econòmica en el futur. Igualment, com s'ha comentat en cas de reemplaçar els servidors

o traslladar-ho tot a un proveïdor extern seria assumible, ja que no és un cost massa elevat.

8.3.3 Dimensió social

Respecte a la dimensió social, no sembla que hi hagi cap possibilitat de que es produeixi un escenari en que el nostre servei sigui perjudicial per a cap segment de la població. Al ser un lloc de consulta i de codi obert, tampoc crearia cap mena de dependència deixant als usuaris en una posició de debilitat.

9 Conclusions i Competències tècniques.

9.1 Conclusions

Considero el treball realitzat amb èxit, doncs s'ha complert l'objectiu principal de dissenyar i crear el servei ePrivacy Observatory, amb el qual es poden analitzar pàgines web sota demanda i veure el resultat dels anàlisis a través de diferents gràfics.

Durant el desenvolupament del treball s'han anat superant satisfactòriament els reptes que han anat apareixent, podent mostrar finalment una web operativa i funcional.

S'ha pogut recopilar informació suficient com per demostrar que més del 80% de les pàgines web analitzades contenen almenys un element de tracking, una xifra molt elevada. La gran majoria de webs més populars fan servir multitud de sistemes de *tracking*, moltes vegades de manera injustificada.

Finalment, creiem que el servei és realment útil per als usuaris, que poden consultar per a cada pàgina web el seu nivell d'intrusió, saber quins mètodes de tracking fa servir i conèixer si les seves dades s'envien a tercers.

9.2 Treball futur

De cara al futur es farà la web més auto-explicativa, ja que per exemple el graf de nodes no sembla intuïtiu per a tothom, doncs alguns usuaris diuen no entendre què vol dir la informació mostrada. Això clarament és una prioritat a millorar, ja que l'objectiu és que la web sigui accessible per a usuaris de qualsevol nivell.

Seria molt interessant mantenir instàncies d'ORM en servidors a diferents països del món, per veure si les pàgines webs s'adapten a diferents normatives segons el país al que pertany l'usuari. Per exemple, la GDPR diu que si una web està disponible per a un ciutadà europeu, la pàgina ha de complir aquesta llei, encara que el servidor no es trobi a la Unió Europea (27). En aquest cas si seria interessant realitzar una *dockerització* del servei, per tal de facilitar l'execució d'instàncies d'ORM des de diferents servidors.

També, actualment la web es troba disponible només en anglès. Estaria bé fer una feina de traducció per a poder-la consultar en català, castellà i altres idiomes. Aquesta feina

podria ser realitzada per nosaltres mateixos, per simpatitzants del projecte a través d'aportacions al codi, ja que és obert, o per traductors professionals. Igualment, s'hauria de realitzar una tasca publicitària del servei, per a popularitzar-lo. Òbviament queda descartada qualsevol acció amb els serveis principals de publicitat a internet com Google AdSense o Facebook Ads. El més lògic seria fer entrades en blogs i fòrums relacionats amb la privacitat, o inclús aconseguir que mitjans digitals d'articles i notícies en fessin menció.

Finalment s'espera poder millorar ORM per tal d'integrar-li un *crawler* de manera que els dominis no només s'insereixin per via dels usuaris o per via del Top Alexa, si no que es vagi revisant quins dominis de la base de dades ja no són operatius i anar-ne afegint de nous de manera automàtica per a que es vagin analitzant.

9.3 Competències Tècniques

En l'apartat 3 d'aquesta memòria s'ha explicat l'arquitectura del nostre servei, i en el punt 4.1 s'ha explicat com ha sigut la instal·lació de requisits per a posar-lo en marxa.

Per tant considero les competències següents assolides:

CTI2.2: Administrar i mantenir aplicacions, sistemes informàtics i xarxes de computadors (els nivells de coneixement i de comprensió són a les competències tècniques comunes).

CTI3.3: Dissenyar, implantar i configurar xarxes i serveis.

Pel que fa a la competència **CTI3.1: Concebre sistemes, aplicacions i serveis basats en tecnologies de xarxa, tenint en compte Internet, web, comerç electrònic, multimèdia, serveis interactius i computació ubiqüa**, s'ha assolit durant el desenvolupament i disseny del servei final, que està detallat en els punts 4.2 en endavant i tot el punt 5.

Finalment la competència **CTI1.3: Seleccionar, desplegar, integrar i gestionar sistemes d'informació que satisfacin les necessitats de l'organització amb els criteris de cost i**

qualitat identificats, es pot considerar assolida considerant ORM com un sistema d'informació que administra, recol·lecta, recupera, processa i emmagatzema la informació rellevant per a poder fer funcionar el servei web final.

10 Referències

1. *Dobleo*. [En línia] [Data: 15 / 06 / 2021.] <https://www.dobleo.com/blog/google-2/definicion-y-tipos-de-cookies-informatica>.
2. *Github CBA-UPC*. [En línia] [Data: 18 / 06 / 2021.] <https://github.com/CBA-UPC/ORM>.
3. *GDPR.eu*. [En línia] <https://gdpr.eu/cookies/>.
4. *Webbrowstertools*. [En línia] [Data: 18 / 6 / 2021.] 2021. <https://webbrowstertools.com/webgl-fingerprint/>.
5. *Vice*. [En línia] [Data: 18 / 06 / 2021.] <https://www.vice.com/en/article/78k8pz/how-you-move-your-mouse-could-help-others-track-you-even-on-tor>.
6. *Ionos*. [En línia] [Data: 18 / 06 / 2021.] <https://www.ionos.es/digitalguide/online-marketing/analisis-web/canvas-fingerprinting-el-web-tracking-sin-cookies/>.
7. *Browserleaks*. [En línia] [Data: 18 / 06 / 2021.] <https://browserleaks.com/fonts>.
8. *Europa.eu*. [En línia] [Data: 18 / 06 / 2021.] https://ec.europa.eu/info/law/law-topic/data-protection_es.
9. *Boe*. [En línia] [Data: 18 / 06 / 2021.] <https://www.boe.es/buscar/pdf/2018/BOE-A-2018-16673-consolidado.pdf>.
10. *Virustotal*. [En línia] [Data: 18 / 06 / 2021.] <https://www.virustotal.com/gui/>.
11. *Hetzner*. [En línia] [Data: 18 / 06 / 2021.] <https://www.hetzner.com/>.
12. *PyPi*. [En línia] [Data: 18 / 06 / 2021.] <https://pypi.org/project/pip/>.
13. *Mimecast*. [En línia] [Data: 18 / 06 / 2018.] <https://community.mimecast.com/s/article/Content-Examination-Definitions-Using-Fuzzy-Hashing-809049828>.
14. *Selenium*. [En línia] [Data: 18 / 06 / 2021.] <https://www.selenium.dev/documentation/en/webdriver/>.
15. *Xorg*. [En línia] [Data: 18 / 06 / 2021.] <https://www.x.org/releases/X11R7.6/doc/man/man1/Xvfb.1.xhtml>.
16. *Ubuntu*. [En línia] [Data: 18 / 06 / 2021.] <https://help.ubuntu.com/community/UFW>.
17. *Fail2Ban*. [En línia] [Data: 18 / 06 / 2021.] https://www.fail2ban.org/wiki/index.php/Main_Page.
18. *W3Shools*. [En línia] [Data: 18 / 06 / 2021.] https://www.w3schools.com/sql/sql_injection.asp.
19. *Bootstrap*. [En línia] [Data: 18 / 06 / 2021.] <https://getbootstrap.com/>.
20. *d3js*. [En línia] [Data: 18 / 06 / 2021.] <https://d3js.org/>.
21. *Alexa*. [En línia] [Data: 18 / 06 / 2021.] <https://www.alexa.com/topsites>.

22. *Proyectosagiles*. [En línia] [Data: 18 / 06 / 2021.] <https://proyectosagiles.org/que-es-scrum/>.
23. *Fib*. [En línia] [Data: 18 / 06 / 2021.] <https://www.fib.upc.edu/ca/empresa/practiques-en-empresa>.
24. *Glassdoor*. [En línia] [Data: 18 / 06 / 2021.] https://www.glassdoor.es/Sueldos/profesor-adjunto-sueldo-SRCH_KO0,16.htm.
25. *Contabo*. [En línia] [Data: 18 / 06 / 2021.] <https://contabo.com/en/>.
26. *MyMonday*. [En línia] [Data: 18 / 06 / 2021.] <https://mymonday.es/edificios/barcelona/barceloneta/>.
27. *Bigwing*. [En línia] [Data: 21 / 06 / 2021.] <https://bigwing.com/blog/what-is-the-gdpr-and-does-it-apply-to-my-website/>.
28. *F-secure*. [En línia] [Data: 18 / 06 / 2021.] https://www.f-secure.com/sw-desc/tracking_cookie.shtml.
29. *techtarget.com*. [En línia] [Data: 18 / 06 / 2021.] <https://whatis.techtarget.com/definition/third-party-cookie>.
30. *Allaboutcookies*. [En línia] [Data: 18 / 06 / 2021.] <https://www.allaboutcookies.org/cookies/session-cookies-used-for.html>.