ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE SISTEMAS

UNIDAD DE TITULACIÓN

PROPUESTA DE UN MÉTODO DE PRESERVACIÓN DE PRIVACIDAD PARA LA APLICACIÓN DE INTELIGENCIA DE FUENTES ABIERTAS (OSINT) DURANTE INVESTIGACIONES DIGITALES FORENSES

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL GRADO DE MAGISTER EN SOFTWARE, MENCIÓN SEGURIDAD

JHONNY JAVIER GUEVARA CAMAS
jhonnyjguevarac@gmail.com

Director: Dr. Denys Alberto Flores Armas
denys.flores@epn.edu.ec

2023

APROBACIÓN DEL DIRECTOR

Como director del trabajo de titulación PROPUESTA DE UN MÉTODO DE PRESERVACIÓN DE PRIVACIDAD PARA LA APLICACIÓN DE INTELIGENCIA DE FUENTES ABIERTAS (OSINT) DURANTE INVESTIGACIONES DIGITALES FORENSES desarrollado por JHONNY JAVIER GUEVARA CAMAS, estudiante de la MAESTRÍA EN SOFTWARE, habiendo supervisado la realización de este trabajo y realizado las correcciones correspondientes, doy por aprobada la redacción final del documento escrito para que prosiga con los trámites correspondientes a la sustentación de la Defensa oral.

Firmado electrónicamente por:
DENYS ALBERTO
FLORES ARMAS

Dr. Denys Alberto Flores Armas
DIRECTOR

DECLARACIÓN DE AUTORÍA

Yo, JHONNY JAVIER GUEVARA CAMAS, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

Firmado electrónicamente por:
JHONNY JAVIER
GUEVARA CAMAS

JHONNY JAVIER GUEVARA CAMAS

ÍNDICE DE CONTENIDO

LISTA DE FIGURAS

LISTA DE TABLAS

LISTA DE ANEXOS

RESUMEN

El presente trabajo propone un método de preservación de privacidad para la aplicación de inteligencia de fuentes abiertas (osint) durante investigaciones digitales forenses. Para el diseño del método se realizó un mapeo de las fases de una investigación forense, principios de privacidad, principios genéricos éticos dados por PRECEPT y las fases del ciclo osint, a partir de esto se definieron requerimientos de privacidad junto a referencias a artículos de la LOPD (Ley orgánica de protección de datos). Se abordan de manera particular las fases de adquisición, recolección y preservación. Posteriormente el método fue implementado de manera practica mediante un crawler web utilizando técnicas de seguridad por diseño, métodos de control de acceso y controles de seguridad. La evaluación del método se llevó a cabo mediante un caso de estudio en donde se evaluó funcionalidad, desempeño y resiliencia. En la evaluación funcional se demostró la efectividad de los lineamientos y requerimientos de privacidad dados por método durante las búsquedas con el crawler versus un navegador convencional. En cuanto a desempeño se evidenció el consumo de recursos mediante tres escenarios con diferente carga en donde la memoria resulto ser el más consumido. Con respecto a resiliencia se realizó un modelado de amenazas basado en el adversario y otro con Microsoft Threat Modelling Tool, sobre estas amenazas se gestionó el riesgo mediante controles del método propuesto y se definió requerimientos críticos de operación. Finalmente se ejecutaron escenarios de ataque donde el sistema mostro resiliencia acorde a los resultados esperados en cada escenario.

Palabras clave: Preservación de Privacidad. Investigación Digital Forense. Osint. Precept

ABSTRACT

This work proposes a privacy preservation method for the application of open-source intelligence (OSINT) during digital forensic investigations. For the design of the method, a mapping of phases of a forensic investigation, privacy principles, generic ethical principles given by PRECEPT and the phases of the osint cycle were carried out, from this, privacy requirements were defined together with references to articles of the LOPD (Organic Law on Data Protection). The phases of acquisition, collection and preservation were particularly considered. Subsequently, the method was put into practice through a web crawler using design security techniques, access control methods and security controls. The evaluation of the method was carried out through a case study where functionality, performance and resilience were evaluated. Functional evaluation, the effectiveness of the guidelines and privacy requirements of the method during searches with the crawler versus a conventional browser was demonstrated. In terms of performance, resource consumption was evidenced by three scenarios with different load where memory was the most consumed. With regard to resilience, a threat modeling based on the adversary and another with Microsoft Threat Modeling Tool was performed, on these threats, risk was managed through controls of the proposed method and critical operating requirements were defined. Finally, attack scenarios were executed where the system showed resilience according to the expected results in each scenario.

Keywords: Privacy Preservation. Digital Forensic Investigation. Osint. Precept

1. INTRODUCCIÓN

En esta sección se introduce al problema que se quiere resolver, se delimita el alcance de la propuesta, se definen los objetivos planteados y se establece la metodología que se utilizó en la investigación.

1.1. Especificación del problema

Es evidente que la tecnología ha cambiado nuestras vidas y está en cada uno de los aspectos que la integran, poniendo a disposición increíbles recursos que ayudan en las actividades cotidianas y profesionales. Sin embargo, es innegable que hay quienes eligen utilizar esta potencia informática para propósitos nefastos, cuando esto sucede, las personas u organizaciones afectadas exigen justicia ante las autoridades y es aquí donde intervienen los expertos forenses informáticos [1]. A este punto, todo tipo de información ligada a estos dispositivos está a disposición de los investigadores como una fuente de posibles evidencias que tras un análisis pueden o no ayudar a probar una hipótesis.

Para llevar a cabo estas tareas de investigación el profesional forense hace uso de múltiples técnicas y herramientas, las cuales permiten acceder, filtrar y extraer información de los dispositivos comprometidos incluso en ocasiones burlando métodos de autenticación y políticas de control de acceso[2]

Otra manera que tiene el investigador de obtener información acerca de un objetivo es a través de la inteligencia de fuentes abiertas (OSINT), que consiste en la recolección y análisis de datos disponibles de manera pública para ser usados en un contexto de inteligencia [3]. OSINT es ampliamente utilizado en investigaciones digitales forenses, pues el hecho de tener a disposición un sin número de fuentes públicas de información genera mucha expectativa acerca de los datos que se pueden obtener, es así que podemos encontrar por ejemplo: bases de datos, números de cédula, números de celular, registros de actividades en redes sociales, números tarjetas de crédito, emails asociados a una persona, registros de ubicación e incluso información recopilada de un proceso OSINT. Toda esta información en primera instancia resulta atractiva y útil para el investigador, ya que facilita la extracción de datos e información que considere a su juicio ser relevante para el contexto de investigación [1]. Sin embargo, también puede ser usada con fines nefastos esto debido a que no existen filtros ni niveles de privacidad y

toda información alojada en estas fuentes es totalmente accesible a cualquier persona y puede pasar a estar en poder de quien sea y para lo que sea.

En los últimos años varios países incluido Ecuador estuvieron involucrados en uno de los escándalos de corrupción más grandes de Latinoamérica relacionados con la empresa ODEBRECHT[4]. Esta problemática se tiene con OSINT cuando la investigación es direccionada a recolectar datos de personas influyentes las cuales son un blanco perfecto para llevar a cabo planes maliciosos como extorsión, tráfico de influencia, lavado de activos, etc.

El uso de OSINT como apoyo a una investigación forense digital como se decía en líneas anteriores, es un hecho que se está dando muy a menudo. Y por otro lado las herramientas que soportan estas tareas también han ido evolucionando, dando las prestaciones para recuperar información más precisa y de una manera más granular gracias a técnicas como machine learning y otras que avanzan en este sentido [3].

Como se puede apreciar conforme lo expuesto, uno de los aspectos vitales que se deben tener en cuenta dentro del marco de una investigación forense digital, es la preservación de la privacidad de la información que se gestiona a lo largo del ciclo forense, y que de ser soportada por prácticas OSINT, estas consideraciones también se deberían extender al su ciclo de implementación, pues ya se ha ejemplificado con escenarios nefastos lo que supondría un mal uso de la información personal o de una organización si cae en las manos equivocadas.

Otro aspecto que debería ser tomado en cuenta e implementado en procesos que implican tratamiento de datos es el legal, actualmente existen leyes que rigen en muchos países del mundo, como por ejemplo en la unión europea con el RGPD [5], y países de nuestra región como Brasil y Chile entre otros que cuentan con leyes que regulan el tratamiento de datos. En Ecuador se aprobó y entro en vigencia hace poco la LOPD [6], hecho que supone un cambio para las empresas que son receptoras y procesadoras de datos en cuanto a gestión de información de terceros.

Finalmente, y considerando lo expuesto, en el presente proyecto se propone diseñar un método de preservación de privacidad para la aplicación de fuentes abiertas (OSINT) durante investigaciones digitales forenses; el cual considere aspectos de interés para el investigador y el propietario de los datos tomando en cuenta principios de privacidad,

principios de capital intelectual, principios éticos, directrices de investigación y aspectos legales referentes a violación de la privacidad en el Ecuador.

1.2. Pregunta de Investigación

¿Es posible desarrollar un método de preservación de privacidad para la aplicación de inteligencia de fuentes abiertas (OSINT) durante investigaciones digitales forenses?

1.3. Objetivo general

Desarrollar un método de preservación de privacidad para la aplicación de inteligencia de fuentes abiertas (OSINT) durante investigaciones digitales forenses.

1.4. Objetivos específicos

- Analizar el estado del arte en relación con la aplicación efectiva de métodos de preservación de privacidad en investigaciones digitales forenses.
- Categorizar los enfoques y técnicas de inteligencia de fuentes abiertas (OSINT) para obtener rastros digitales de personas de interés.
- Utilizar PRECEPT y otros enfoques de preservación de privacidad para la implementación de un prototipo funcional que permita la gestión de la evidencia digital recuperada mediante OSINT.
- Evaluar la aplicabilidad de la propuesta en un caso de estudio.

1.5. Alcance

A lo largo de este proyecto se abordan diferentes aspectos importantes que permiten lograr los objetivos planteados y que junto a la metodología propuesta se considera lo siguiente: El diseño de un método de preservación de privacidad de la información recuperada con OSINT en el contexto de una investigación digital forense mediante el mapeo de sus fases, alineación de principios de privacidad, éticos, capital intelectual, ,actividades de investigación, aspectos legales y requerimientos para preservar la privacidad de la información recolectada. Es de vital importancia mencionar que se hará hincapié y se trabajará sobre las fases de RECOLECCION y PRESERVACIÓN de la información dado que es donde OSINT tiene mayor relevancia y es de interés del presente trabajo, se omiten las fases de IDENTIFICACIÓN, ANÁLISIS, RECONSTRUCCIÓN, INFORMES, REFLEXION Y REVISIÓN. Adicional a esto se implementará un prototipo funcional orientado a la preservación de privacidad de la información recolectada con OSINT, posteriormente se evaluará la aplicabilidad de la

propuesta mediante un caso de estudió evaluando la funcionalidad, el desempeño y resiliencia.

2. MARCO TEÓRICO

En esta sección se tratan a nivel teórico los conceptos relevantes involucrados en la investigación. El presente trabajo se contextualiza dentro de las practicas OSINT y la investigación digital forense por lo cual se describen sus ciclos y fases correspondientes. Además, se contemplan aspectos legales como la ley orgánica de protección de datos (LOPD) dado que rige la gestión de información recolectada. Otro componente abordado y que es eje en la investigación es PRECEPT, del cual se describe su estructura y elementos. En cuanto a enfoques y técnicas para recolectar evidencia de personas de interés se describen varios métodos y herramientas como buscadores, filtros y aplicaciones dedicadas. Por otro lado, se realiza una revisión de literatura en cuanto a trabajos relacionados con el fin de conocer el estado del arte en este contexto, posteriormente se describe la contribución de presente trabajo al campo de investigación.

2.1. Investigación forense digital

El análisis forense informático, en un sentido formal, es definido como un conjunto de principios y técnicas que comprende el proceso de adquisición, conservación, documentación, análisis y presentación de evidencias digitales y que dado el caso puedan ser aceptadas legalmente en un proceso judicial [1].

2.1.1. Ciclo forense

Para llevar a cabo un proceso de investigación forense digital no existe un ciclo o una serie de fases definidas que obligatoriamente se deban cumplir, sino que hay distintos enfoques utilizados por los profesionales de esta disciplina en función del contexto de la investigación. Es por esto que en el presente trabajo se consideran las fases más importantes y que son implementadas en la mayoría de marcos de trabajo como se analiza en [7], incluyendo a PRECEPT que está dentro del alcance de este proyecto.

Dentro del ciclo cada fase responde a las distintas necesidades de una investigación y para esto se deben llevar a cabo las actividades definidas en cada una. A continuación, se detallan brevemente cada una y su implicación en el contexto de una investigación digital forense.

- **Identificación:** En esta fase lo primordial es identificar qué tipo de incidente se ha producido mediante actividades como: analizar informes de la situación, definir relaciones con otras investigaciones (en caso de haberlas), realizar una

evaluación rápida y establecer escenarios del crimen, revisar políticas y legislación vigente.

- **Adquisición:** En esta fase se realiza la adquisición de los dispositivos implicados que contengan información o la información en si misma (posible evidencia). Se pueden realizar actividades como: Incautación física y almacenamiento de dispositivos y datos locales o remotos.
- **Preservación:** En esta fase se debe garantizar la integridad de la información recolectada mediante actividades como: copia y verificación (suma de verificación) de medios, realizar imágenes forenses, imágenes estáticas, Imágenes en vivo, imágenes de memoria, preservación de la información de acceso no autorizado.
- **Búsqueda:** Aquí se realizar todo tipo de búsqueda sobre la información recuperada mediante actividades como: recuperar información de medios físicos incluyendo recuperación de información eliminada, descifrado, fuerza bruta, etc. Recuperar información que responda a preguntas clave de la investigación como: ¿Cómo se configuró el sistema? ¿Qué usuarios accedieron a un recurso? etc.
- **Análisis:** En esta fase se realiza un análisis sobre la evidencia filtrada en la fase anterior y es aquí donde caben actividades como: Realizar Juicios de relevancia, organizar evidencia de hechos suscitados a bajo nivel. Considerar cuatro preocupaciones semi-ortogonales: temporal, espacial, relacional, funcional
- **Reconstrucción:** En esta fase se realizan inferencias basadas en los datos recuperados, filtrados y analizados que ayuden a comprobar hipótesis acerca de lo suscitado.
- **Informes:** En esta fase se realizan informes que permitan evidenciar datos estructurados y resumidos sobre un hecho. Se generan Informes judiciales, comparecencia ante el tribunal, etc.
- **Reflexión y revisión**: En esta fase se realizan retrospectivas acerca del proceso de la investigación en sus fases anteriores, lecciones aprendidas y consideraciones de desempeño y actividades de apoyo como: Generar un documento que detalle qué investigadores forenses estuvieron involucrados y sus actividades específicas, si es posible analizar y desechar los datos irrelevantes que se adquirieron y conservaron, considerar si alguno de los investigadores requiere apoyo o asesoramiento.

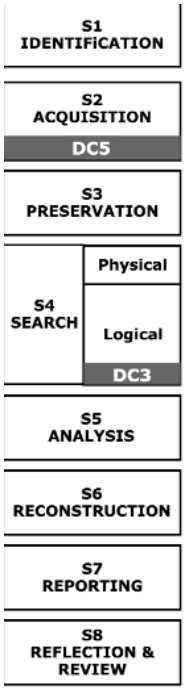


Figura 1- Ciclo forense. [1]

2.2. Inteligencia de fuentes abiertas (OSINT)

Cuando se trata de inteligencia, el término está determinado sobre todo por una controversia de definición, ya que, en su sentido más amplio, se identifica por organización, conocimiento, información y proceso; mientras que, en su sentido más estricto, la idea de inteligencia se refiere a la información analizada como base para la toma de decisiones [1]. Por otro lado, la inteligencia de fuentes abiertas es también una disciplina específica de recopilación de inteligencia, como la inteligencia humana (HUMINT), la inteligencia recopilada por interceptación de señales (SIGINT), la inteligencia imaginaria (IMINT) y el procesamiento científico y técnico de datos recopilados de diferentes fuentes móviles e inmóviles (MASINT) [8]. En otras palabras, es una disciplina cuya característica predominante es la disponibilidad de información no clasificada [9].

Como se puede analizar el proceso OSINT busca obtener inteligencia a partir de información pública y que visto de esta manera los contextos de uso son variados, formales e informales, legales e ilegales, etc. El presente trabajo contempla el uso de esta técnica como apoyo en el contexto de una investigación forense digital para lo cual se considera cada una de las fases que componen el ciclo de vida de OSINT y que en el siguiente apartado se detallan brevemente.

2.2.1. Ciclo OSINT

- **Planificación y Dirección:** El primer paso del ciclo OSINT consiste en planificar las prioridades y requisitos de la misión. Antes de recolectar OSINT, se debe tener una comprensión clara de los tipos de información que necesitan, cómo encontrar esas fuentes y qué esperan lograr con la información adquirida. esta logística precautoria garantizará la productividad y eficiencia de la operación durante las próximas fases del ciclo OSINT.
- **Recolección:** Una vez realizada la planificación adecuada, puede comenzar la recopilación de información, los recursos de OSINT incluyen cualquier material que esté disponible gratuitamente en línea, como artículos de noticias, publicaciones en redes sociales y blogs. Aquí se deben utilizar herramientas y recursos que mayor prestación den para obtener estos datos.
- **Tratamiento y Explotación:** Una vez que se hayan adquirido datos, se puede comenzar a procesar la información para luego compilarlo en un repositorio de evidencia, línea de tiempo o informe común. El objetivo de esta etapa es simplificar el contenido encontrado y hacerlo más legible para los destinatarios de los datos.
- **Análisis y Producción:** Después del procesamiento inicial de los datos recopilados, se realiza un análisis a profundidad de la información. Este es un paso crucial en el ciclo OSINT, ya que permite utilizar los datos adquiridos para interpretar y anticipar eventos. Aquí se debe organizarla información analizada en un documento o presentación detallada, que será leída por una audiencia designada.
- **Difusión e Integración:** El paso final en el ciclo OSINT implica entregar la inteligencia recopilada y analizada a las partes interesadas adecuadas. Luego, los analistas reciben comentarios, que dictaminan si el ciclo OSINT debe comenzar de nuevo.

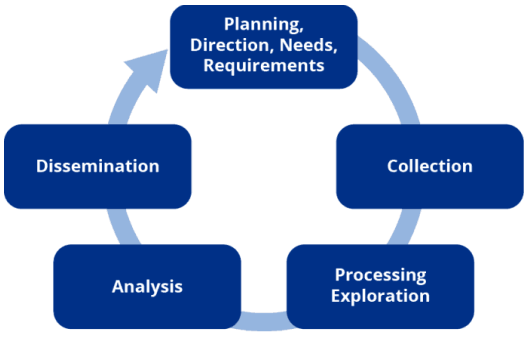


Figura 2- Ciclo OSINT. **[Elaboración Propia]**

2.2.2. Técnicas y enfoques de inteligencia de fuentes abiertas (OSINT)

Como se ha mencionado en líneas anteriores, el uso de OSINT es bastante prometedor y poderoso, pero su implementación también es un desafío. Hoy en día, OSINT es ampliamente adoptado por los gobiernos y los servicios de inteligencia para realizar sus investigaciones y luchar contra el ciberdelito [10]. Sin embargo, no solo se utiliza para asuntos de Estado, sino que se aplica a varios objetivos diferentes, por ejemplo, en [11] se aborda OSINT como herramienta para obtener información de redes sociales para estrategias de mercadeo, análisis de comportamiento, tendencias, etc. En [12] se utiliza OSINT para obtener información acerca de organizaciones terroristas y prevención de fraude utilizando tecnologías web semánticas. Utilizando OSINT, Big Data y el alto poder computacional que actualmente existe, en [13] también se trabaja en este sentido, obteniendo información de datos masivos almacenados históricamente en las empresas. En general los usos de OSINT son muchos y variados ya que pueden apoyan a distintas áreas, sin embargo, de acuerdo con [10] en donde más se ha enfocado su implementación son en tres áreas como se puede ver en la Figura 3.

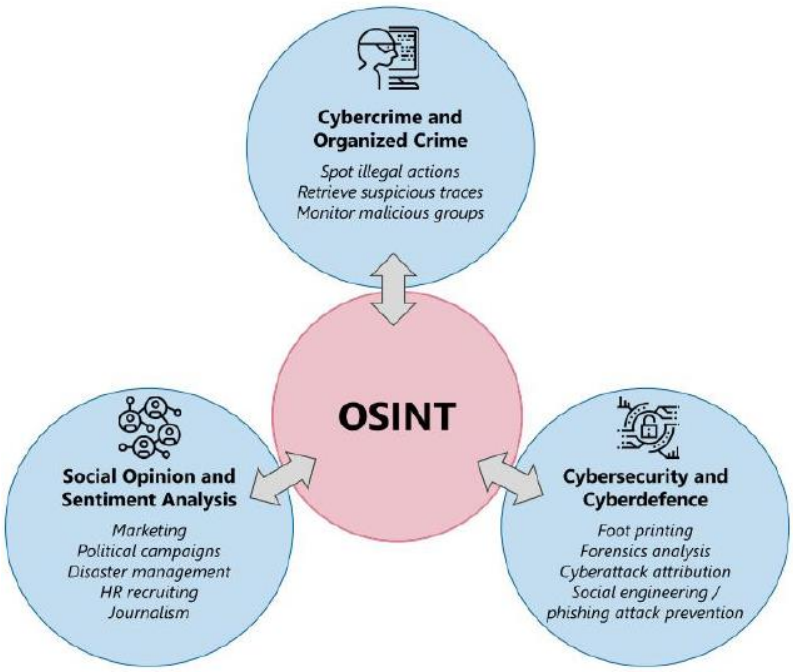


Figura 3 – Enfoques de OSINT. [10]

Analizando los aspectos de la Figura 3 en cuanto a **opinión social y análisis de sentimientos** es evidente que, con el auge de las redes sociales, es posible recopilar las

interacciones de los usuarios, mensajes, intereses y preferencias para extraer conocimiento no explícito. La evidencia acumulada de las redes sociales es de gran alcance y muy ventajosa [10], [14], pudiéndose aplicar para campañas de marketing, políticas, etc.

En cuanto a **Ciberdelincuencia y delincuencia organizada** las practicas OSINT permiten analizar y comparar datos continuamente para detectar intenciones delictivas en una etapa temprana [10], [12], [15]. Teniendo en cuenta los patrones de los adversarios y las relaciones entre delitos graves, OSINT puede brindar a las fuerzas de seguridad la oportunidad de detectar rápidamente acciones ilegales. En este sentido, mediante la explotación de los datos abiertos es posible rastrear la actividad de las organizaciones terroristas sin que estas sospechen[12].

Con respecto a **Ciberseguridad y ciberdefensa** basta con tomar en cuenta los constantes ataques que sufren los sistemas informáticos con el objetivo de interrumpir la disponibilidad de los servicios prestados. La investigación se vuelve, por lo tanto, crucial para defender esos sistemas de los ciber atacantes, concretamente para hacer frente a los desafíos que aún están abiertos en el campo de la ciberseguridad [10], concretamente, las técnicas de minería de datos pueden ayudar a realizar análisis de los ataques diarios, correlacionarlos y respaldar los procesos de toma de decisiones para una defensa eficaz, pero también para una reacción rápida como se describe en [16] en el campo de las redes informáticas.

OSINT como cualquier otro proceso posee ventajas y limitaciones las cuales se deben tener en cuenta ya que estas definen en gran parte las técnicas que se utilizan en el proceso. En la Tabla 1 se describen varios puntos importantes de acuerdo al análisis realizado en [10].

Tabla 1 - Ventajas y desventajas de OSINT. [10]

Pros ✓	Cons ✗
Huge amount of available information	Complexity of data management
High capacity of computing	Unstructured information
Big data and machine learning	Misinformation
Complementary types of data	Data sources reliability
Flexible purpose and wide scope	Strong ethical/legal considerations

Para llevar a cabo el proceso OSINT existen un sin número de técnicas y herramientas que permiten cumplir los objetivos de la investigación, siendo unas más utilizadas que otras según el contexto y los datos iniciales que posee el investigador. A continuación, se describen las más comunes de acuerdo con [10], [17], [18] .

- **Técnica utilizando Motores de Búsqueda**

  Los conocidos Google, Bing o Yahoo!, entre otros, son herramientas ampliamente utilizadas y su so tradicional es la forma más sencilla de hacer OSINT. Estos motores buscan dentro de la Word Wide Web dada una consulta textual tratando de entregar información que coincida con el criterio de búsqueda. Sin embargo, la cantidad de resultados puede ser tan abrumadora que incluso puede ser contraproducente para el usuario. Por esa razón, se debe saber especificar las solicitudes para refinar las búsquedas y recuperar exactamente el tipo de información que se requiere, servicios como Google o Bing admiten este tipo de filtros como se puede ver en la siguiente tabla.

Tabla 2 – Filtros Google y Bing para búsquedas avanzada. [10]

Google/Bing filter	Search operator	Example of use
Force an exact-match search	" "	"University of Murcia"
Exclude a term or phrase	-	university murcia -catholic
Search for X or Y	OR, |	university murcia|cartagena
Search for X and Y (used by default)	AND	university AND of AND murcia
Use of a wildcard	*	university of *
Search for a range of numbers	..	university murcia 2010..2019
Group terms or search operators	()	"university of (murcia|cartagena)"
Search within a given domain	site:	university murcia site:um.es
Search for a certain file type	filetype:	university murcia filetype:pdf
Search in page titles	intitle:	university intitle:umu
Search in URLs	inurl:	university inurl:um
Search in the text of the pages	intext:	university intext:murcia
Search the most recent cached version of a page	cache:	cache:um.es

- **Técnica basada en Redes Sociales**

  Las redes sociales son un blanco perfecto para prácticas OSINT, a continuación, se describen algunas de las técnicas utilizadas en base a estas.

  Facebook es una red social repartida por todo el mundo con millones de usuarios. Podría considerarse un diario de sociedad, donde se puede encontrar información personal muy valiosa para las investigaciones de OSINT. El perfil de nuestro target puede revelar su empleo, educación, edad, ubicación, lugares visitados o grupos de agrado, entre otros. Las fotos y publicaciones también pueden ayudarnos a contextualizar la empresa o persona que estamos investigando, las zonas que frecuenta o el tipo de actividades que realiza. Por otro lado, **YouTube**

la plataforma basada en videos donde las grandes comunidades se conforman en torno a intereses compartidos. No solo es valioso el contenido subido por un usuario específico (temas, imágenes, escenas, lugares y personas que aparecen en los videos), sino también las opiniones y comentarios de los suscriptores. Otra red social muy utilizadas es **Twitter**, que opera sobre una comunicación en vivo donde se encuentran opiniones personales a través de una línea de tiempo ordenado. Aparte de la información que revela cada perfil es valiosa la información como usuarios seguido, seguidores, interacciones, preferencias y las búsquedas basadas en palabras clave, frases exactas, hashtags, idioma, fecha, etc. Con respecto a negocios la más popular es **LinkedIn** que permite buscar por nombre real, empresa, organización, cargo o ubicación. En este caso, los perfiles profesionales pueden revelar datos de contacto completos, números de teléfono celular, direcciones de correo electrónico. Además, también podemos extraer información sobre el empleo, la educación, las habilidades, los idiomas y las relaciones comerciales, como se ve es realmente una fuente de información poderosa. Aparte de las redes sociales mencionadas existen muchas otras de distinta categoría como por ejemplo para citas o encontrar pareja y para esto revela datos personales, gustos y preferencias, también hay redes sociales que se desarrollan únicamente en ciertos puntos geográficos. Como se ve la información de libre acceso se encuentra disponible masivamente en internet y es cuestión de saber dónde buscar para obtener lo deseado, a continuación, en la siguiente tabla se puede ver un resumen entre otras características útiles para realizar OSINT.

Tabla 3- Potencial de Redes Sociales. [10]

Social Network	Type	Scope	Main potential for OSINT
4chan	Online community	Worldwide	Users interested in illicit activities
Badoo	Dating	Worldwide	Intimate and personal details
Cloob	Social connections	Iran	Personal profile, posting and community membership
Draugiem	Social connections	Latvia	Personal profile, publications in blogs, group membership
Facebook	Social connections	Worldwide	Personal profile, preferences and places visited
Facenama	Social connections	Iran	Personal profile, publications, photos and videos
Flickr	Photo-sharing	Worldwide	Activities, hobbies, places and personal relationships
Instagram	Social connections	Worldwide	Habits, locations and personal relationships
LinkedIn	Business	Worldwide	Professional profile, education, skills and languages
Mixi	Social connections	Japan	Personal profile, interests and opinions
Odnoklassniki	Social connections	Mainly Russia	Personal profile of adults, past and present friendships
Qzone	Social connections	Mainly China	Personal profile, preferences, habits
Reddit	Online community	Worldwide	Users trends, behaviors, and publications
Renren	Social connections	Mainly China	Personal profile of students, friendships and discussions
Taringa!	Social connections	Mainly Latin America	Personal profile, publications and community membership
Tinder	Dating	Worldwide	Intimate and personal details
Tumblr	Photo-sharing	Worldwide	Activities, hobbies, places and personal relationships
Twitter	Social connections	Worldwide	Personal profile, opinions and publications
VKontakte (VK)	Social connections	Mainly Russia	Personal profile, preferences and publications
Weibo	Social connections	Mainly China	Personal profile, opinions and publications
YouTube	Video-sharing	Worldwide	Video content, opinions and comments of subscribers

- **Técnica utilizando direcciones de correo electrónico**

  Muchas ocasiones al iniciar una investigación OSINT entre los datos iníciales se encuentra una dirección de correo electrónico a partir de la cual mediante diferentes técnicas y herramientas se puede obtener mucha más información. En la Tabla 4 se listan herramientas las cuales por ejemplo sirven para verificar si la una dirección de correo electrónico es válida o fue comprometida, también se puede obtener información personal asociada a esta dirección.

Tabla 4 – Herramientas OSINT basadas en dirección de correo electrónico. [10]

Email address OSINT service	URL	Main output
Hunter	`hunter.io`	Validity and availability
Have I Been Pwned	`haveibeenpwned.com`	Appearance in public data breaches
Pipl	`pipl.com`	Personal information about the owner

- **Técnica basada en un nombre de usuario**

  De manera similar a la técnica basada en direcciones de correo electrónico, el nombre de usuario también permite obtener más información de otros sitios web y redes sociales, en la siguiente tabla se puede ver una lista de sitios web utilizados para obtener información en base a un nombre de usuario.

Tabla 5- Herramientas OSINT basadas en nombres de usuario. [10]

Username OSINT service	URL	Main output
KnowEm	`knowem.com`	Presence in social networks, domains and online communities
Name Chk	`namechk.com`	
Name Checkr	`namecheckr.com`	
User Search	`usersearch.org`	
NameVine	`namevine.com`	Suggestions of alternative similar usernames
Lullar	`com.lullar.com`	Availability in social networks

- **Técnica de basada en un nombre real**

  Mediante este dato es posible asociar una mayor cantidad de información y más precisa, gracias a sitios como los que se muestra en la Tabla 6 se puede obtener información como direcciones de domicilio, educación, ocupación, etc. Incluso con otras herramientas se puede obtener información genealógica para descubrir familiares de la víctima.

Tabla 6- Herramientas OSINT basadas en nombres reales de usuario. [10]

Real name OSINT service	URL	Main output
Pipl	pipl.com	Personal information
That's Them	thatsthem.com	Personal details, education, professional career, skills, locations, and relatives.
Spokeo	spokeo.com	
Fast People Search	fastpeoplesearch.com	
Nuwber	nuwber.com	
Cubib	cubib.com	
Peek You	peekyou.com	
Yasni	yasni.com	Social networks profiles
Family Search	familysearch.org	Kinship information, relatives
GENi	geni.com	
Family Tree Now	familytreenow.com	
True People Search	truepeoplesearch.com	

- **Técnica basada en la localización de un objetivo**

  Conocer los sitios que frecuenta nuestro target proporciona indicios de sus hábitos, así como también conocer su localización. Mediante sitios como los que se muestra en la Tabla 7 también se puede obtener direcciones a partir de coordenadas y viceversa. También podemos encontrar imágenes históricas de ciertos sitios obtenidos de sitios como Google Maps.

Tabla 7 - Herramientas OSINT basadas en localización. [10]

Location OSINT service	URL	Main output
Google Maps	google.com/maps	Locations from GPS coordinates
Wikimapia	wikimapia.org	
Bing Maps	bing.com/maps	
GPS Coordinates	gps-coordinates.net	GPS coordinates from location
Historic Aerials	historicaerials.com	Historic images of the past
Terra Servers	terraserver.com	
Land Viewer	eos.com	

- **Técnica basada en direcciones IP**

  Las direcciones IP se obtienen a partir de investigaciones de ciberataques, direcciones de correo electrónico o conexiones a través de Internet. También son cruciales para el análisis forense digital con el fin de recopilar la mayor cantidad de información posible de un incidente. El servicio de Localización IP obtiene, a partir de una determinada dirección IP, aspectos de alto nivel como localización (latitud y longitud), país, región, ciudad, nombre de dominio o ISP. En la Tabla 8 se muestran varias herramientas para estos casos.

Tabla 8- Herramientas OSINT basadas en IPs. [10]

IP address OSINT service	URL	Main output
IP Location	iplocation.net	Location, domain and ISP
ViewDNS	viewdns.info	Technical network-based information
That's Them	thatsthem.com/reverse-ip-lookup	Individual or company information
I Know What You Download	iknowwhatyoudownload.com	Torrent files

- **Técnica basada en un nombre de dominio**

  A partir de un dominio se pueden utilizar técnicas similares a las basadas en IPs. Por una parte, se puede obtener información cruzada como datos del propietario del dominio, también se puede conocer subdominios asociados y el tráfico del dominio. Otras herramientas también almacenan copias históricas de los sitios web para posteriores comparaciones y análisis de evolución de las páginas del dominio, varias de estas herramientas se listan en la Tabla 10.

Tabla 9- Herramientas OSINT basadas en dominios. [10]

Domain name OSINT service	URL	Main output
DNS Trails	securitytrails.com/dns-trails	DNS records and related domains
Whoisoly	whoisology.com	Personal or company information
Wayback Machine	web.archive.org/web	Backups of websites
Visual Site Mapper	visualsitemapper.com	Map of subdomains
Threat Crowd	threatcrowd.org	
Whois	who.is	Registration info and DNS records
Alexa	alexa.com	Traffic statics
SimilarWeb	similarweb.com	
FindSubdomains	findsubdomains.com	Subdomains

2.3. Ley de protección de datos

Dentro del contexto del uso de OSINT en una investigación forense digital es necesario tener en cuenta aspectos de legislación en cuanto al manejo y procesamiento de datos ya que son procesos claves dentro de los respectivos marcos de trabajo y que es mandatorio realizar una revisión en lo que concierne a esta investigación.

Como primeros vestigios de la protección del derecho a la vida privada están considerados a mediados del siglo XX en donde el derecho a la vida privada obtiene mayor relevancia dado que La Declaración Universal de Derechos Humanos de 1948 establece las bases del derecho humano a la vida privada. En años posteriores el gran avance tecnológico introduce nuevos paradigmas y con esto retos jurídicos en materia de protección de datos para los cuales los gobiernos y leyes vigentes no están a la par para cubrir estas demandas, conllevando todo esto a la creación de instrumentos jurídicos nacionales y supranacionales.

A nivel mundial muchos países y regiones regulan la protección de datos personales con el objetivo de velar por los derechos de sus ciudadanos y fomentar el desarrollo de empresas de servicios, cuyo objeto de negocios es la información. Además de establecer políticas coherentes de gobierno electrónico con miras a establecer un lenguaje jurídico transnacional.

Uno de los reglamentos más relevantes que se tiene es el RGPD [5] el cual rige a países de la Unión Europea el cual entro en vigencia en mayo del 2018 y actualmente está vigente sirviendo de modelo para generar otras leyes similares en más países. Por otro lado, en Ecuador también se han hecho esfuerzos en este sentido y es así que hace poco tiempo se aprobó la LOPD [6] en nuestro país que entrará en vigencia muy pronto.

En el siguiente apartado se detallan aspectos relevantes en cuanto al RGPD y LOPD los cuales son de interés dentro de este trabajo.

2.3.1. Ley orgánica de protección de datos (LOPD)

La Constitución del Ecuador reconoce y garantiza en el artículo 66 numeral 19 a las personas: “El derecho a la protección de datos carácter personal, que incluye el acceso y la decisión sobre información y datos de este carácter, así como su correspondiente protección”. En este sentido finalmente Ecuador cuenta con la primera ley de protección de datos la cual fue aprobada por la asamblea nacional en julio del 2021 y entrará en vigencia en mayo del presente año. La LOPD [6] es una normativa que permitirá el desarrollo de la innovación y el uso de la tecnología considerando el tratamiento de los datos personales como su eje central de protección. A continuación, una revisión de los aspectos más relevantes.

- **Actores:** En primer lugar, está el **responsable del tratamiento de datos personales** quien puede ser una persona natural o jurídica, pública o privada, que decide sobre la finalidad y propósito del tratamiento de datos personales. Por otro lado, está el **encargado del tratamiento de datos personales** quien puede ser una persona natural o jurídica, pública o privada, que trata los datos personales por cuenta y nombre de un responsable de tratamiento de datos personales y finalmente está el **Destinatario** quien puede ser una persona natural o jurídica que recibe o ha sido comunicada con datos personales.

- **Principios:** a) Juridicidad. b) Lealtad. c) Transparencia. d) Finalidad. e) Pertinencia y minimización de datos personales. f) Proporcionalidad del tratamiento. g) Confidencialidad. h) Calidad y exactitud. i) Conservación. j) Seguridad de datos personales. k) Responsabilidad proactiva y demostrada
  l) Aplicación favorable al titular. m) Independencia del control

2.3.2. Reglamento general de protección de datos (RGPD)

El Reglamento General de Protección de Datos (RGPD) se aplica tanto a organizaciones europeas que procesan datos personales de ciudadanos dentro de la UE como a organizaciones que tienen su sede fuera de la UE y realizan actividades para personas que viven en la UE. A continuación, una revisión de los aspectos más relevantes.

- **Actores:** En primer lugar, se tiene al **sujeto de los datos o el interesado** que se refiere a los individuos que se encuentren en la Unión Europa cuyos datos son tratados, personas naturales que podemos ser distinguidos con derechos sobre nuestros datos personales. Por otro lado, está el **responsable del tratamiento** quien será responsable de los datos como una persona física o jurídica, autoridad pública, agencia u otro organismo que, solo o conjuntamente con otros, determina los fines y los medios del tratamiento de datos personales. Quienes procesan los datos son denominados **procesadores de datos** que pueden ser personas físicas o jurídicas, autoridades públicas u otras agencias y organismos que procesan datos personales en nombre del responsable. Otro actor es el **delegado de protección de datos** que es un garante del cumplimiento de la normativa de la protección de datos en las organizaciones, sin sustituir las funciones que desarrollan las Autoridades de Control. El RGPD dedica una sección entera a esta nueva figura dada la relevancia que tiene, conocida popularmente como DPO (por sus siglas en inglés, Data Protection Officer).
- **Base legal para el tratamiento:** Los datos solo se pueden tratar si existe al menos una base legal para hacerlo y que son las siguientes:
  - El interesado ha dado su consentimiento para el tratamiento de sus datos personales con uno o más propósitos específicos.
  - El tratamiento es necesario para la ejecución de un contrato del que el interesado es parte o para tomar medidas a petición del interesado antes de celebrar un contrato.

- El tratamiento es necesario para cumplir con una obligación legal a la cual el controlador está sujeto.
- El tratamiento es necesario para proteger los intereses vitales del interesado o de otra persona física.
- El tratamiento es necesario para la realización de una tarea llevada a cabo en interés público o en el ejercicio de la autoridad oficial conferida al controlador.
- El tratamiento es necesario para la realización de una tarea llevada a cabo en interés público o en el ejercicio de la autoridad oficial conferida al controlador.
- El tratamiento es necesario para los fines de los intereses legítimos perseguidos por el responsable o por un tercero, salvo cuando dichos intereses sean anulados por los intereses o los derechos y libertades fundamentales del interesado que requieren protección de datos personales, en particular cuando el interesado es un niño.

- **Consentimiento:** Cuando el consentimiento se utiliza como la base legal para el tratamiento, el consentimiento debe ser explícito para los datos recopilados y los fines para los que se utilizan los datos (Artículo 7, definido en el Artículo 4). Consentimiento para niños debe ser otorgado por el padre o tutor del niño, y verificable (Artículo 8). Los controladores de datos deben poder probar el consentimiento y puede ser retirado.
- **Pseudonimización:** La pseudonimización es un procedimiento de gestión de datos y desidentificación de datos mediante el cual los campos de información personalmente identificables dentro de un registro de datos se sustituyen por uno o más identificadores artificiales. La pseudonimización es una forma de cumplir con las exigencias del RGPD para el almacenamiento seguro de la información personal. Los datos pseudonimizados pueden ser restaurados a su estado original con la adición de información que permita volver a identificar a los individuos.
- **Limitación del fin, datos y almacenamiento:** El interesado tiene derecho a obtener del responsable del tratamiento la rectificación de los datos personales inexactos que le conciernan. Teniendo en cuenta las finalidades del tratamiento, el interesado tiene derecho a que se completen los datos personales que sean incompletas, inclusive mediante una declaración adicional.

- **Derecho de rectificación:** Si una persona considera que sus datos personales son incorrectos, incompletos o inexactos, tiene derecho a rectificarlos o completarlos sin demoras indebidas.
- **Infracción de las normas y sanciones:** El incumplimiento del RGPD puede dar lugar a multas de hasta 20 millones de euros o del 4% del volumen de negocios mundial de la empresa, en determinadas infracciones. La autoridad de protección de datos puede imponer medidas correctivas adicionales, como obligar a poner término al tratamiento de los datos personales.

2.4. Precept

Precept [1] es un framework de soporte para actividades digitales forenses directamente orientado a la privacidad, que basado en sus fases va definiendo lineamientos a considerar en cuanto a principios de privacidad, capital intelectual, y requerimientos de investigación en cada una de ellas. Además, ha sido sometido a rigurosos análisis y afinamientos por expertos de la materia y cuenta con una fuerte base conceptual, la Figura 4 presenta un diagrama general del framework.

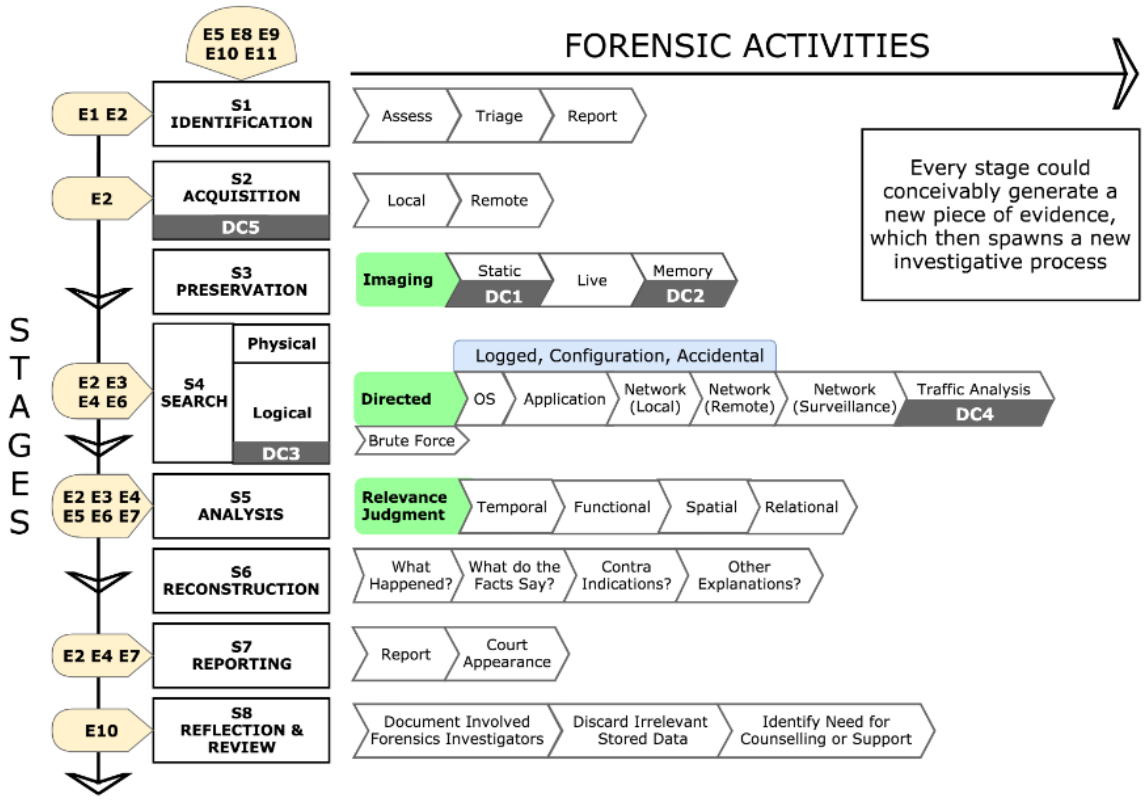


Figura 4- Framework Precept. [1]

2.5. Trabajos relacionados

En este apartado se ha realizado un análisis de los trabajos relacionados con métodos efectivos para la preservación de privacidad para la aplicación de OSINT e investigaciones digitales forenses, esto con el fin de identificar, analizar y considerar aspectos relevantes que se relacionen con el presente trabajo.

Para llevar a cabo el análisis y selección de trabajos relacionados se consideran ciertos elementos de la declaración PRISMA [19], la cual permite documentar y llevar de una manera transparente el proceso de revisión de literatura, detallando términos de búsqueda, buscadores, numero de resultados, trabajos excluidos e incluidos, etc. En este sentido se definen los elementos que se toman en cuenta para el proceso:

- Justificación
- Objetivos
- Criterios de elegibilidad
- Fuentes de información
- Estrategia de búsqueda
- Proceso de selección de los estudios
- Lista de los datos
- Características de los estudios
- Discusión

La justificación de este proceso es dada por la necesidad de conocer el estado del arte actual del campo de estudio que el presente trabajo propone. En cuanto a los **objetivos** que persigue este proceso son, buscar, seleccionar, y analizar trabajos relacionados al campo de estudio, mismos que permitan responder la pregunta de investigación y definir elementos de trabajo.

Previo a realizar las búsquedas es importante definir cuáles serán los **criterios de elegibilidad**, es decir definir ciertas características que deberán cumplir los artículos candidatos a una revisión más minuciosa, además de permitir definir filtros de búsqueda y recuperar artículos que sean de interés y se relacionen con el campo de estudio. A continuación, se definen estos criterios:

- Artículos publicados en los últimos 5 años.
- Artículos publicados en fuentes de información confiables.
- Artículos recuperados con las cadenas de búsqueda definidas.
- Deben ser artículos de investigación dentro de las ciencias de la computación.

Una vez definidos los criterios de elegibilidad para realizar las búsquedas se procede también a definir las **fuentes de información** de donde se extraen los artículos. Actualmente existen una gran lista de fuentes de datos, sin embargo, muchas de estas no son confiables o contiene información duplicada. Por otro lado, dentro de la comunidad de investigadores existen fuentes que son bien conocidas dada su gran reputación, confiabilidad y artículos actualizados que poseen. En este sentido se consideran las siguientes fuentes para las búsquedas: **IEEE**, **Springer Link**, **ACM**, **ScienceDirect.**

Una vez definidos los elementos necesarios se pasa a **la estrategia de búsqueda**, la cual se empieza por la delineación de la cadena de búsqueda inicial y palabras clave, en primera instancia se define: "**PRIVACY PRESERVATION METHOD IN DIGITAL FORENSIC INVESTIGATIONS WITH OSINT",** esta cadena es bastante genérica y permite conocer y analizar el universo de artículos relacionados en las distintas fuentes de datos. Una vez realizado este primer sondeo los resultados fueron variados, por ejemplo, en IEEE 0, ACM 587.837, Springer 120, ScienceDirect 11. En IEEE debido a que la cadena de búsqueda es muy extensa, sin embargo, en ACM se ve una gran cantidad de ítems pero que se relacionan únicamente con una de las palabras que conforman la cadena, pero no tiene relación directa con lo que se está buscando. Tomando en cuenta esto se definen nuevas cadenas de búsqueda y se obtienen menos resultados, pero más acertados con el criterio de búsqueda, en este sentido también se van aplicando los filtros previamente definidos tales como, artículos publicados en un periodo de tiempo, operadores lógicos, también se filtra por tipos de publicación y demás criterios que se pueden ver en la Tabla 10, a su vez se registra el número de resultados obtenidos con cada cadena de búsqueda.

Tabla 10 - Fuentes de información y cadenas de búsqueda. [21]

Fuentes de Información				
IEEE	**ACM**	**Springer**	**ScienceDirect**	**Filtros**
PRIVACY PRESERVATION METHOD IN DIGITAL FORENSIC INVESTIGATIONS WITH OSINT				
0	587,837	120	11	• No se aplica ningún filtro
PRIVACY PRESERVATION **OR** FORENSIC INVESTIGATIONS **OR** OSINT				
4,642	74	73	91.679	• Se aplica **operadores lógicos** entre frases clave
PRIVACY PRESERVATION **AND** FORENSIC INVESTIGATIONS **AND** OSINT				
2.390	58	76	28.536	• Solo publicaciones desde el **2018**
IEEE: PRIVACY **AND** PRESERVING **AND** FORENSIC SPRINGER: PRIVACY **AND** PRESERVATION **AND** FORENSIC AND INVESTIGATIONS **AND** OSINT ACM: "PRIVACY" **AND** "PRESERVING" **AND** "FORENSIC INVESTIGATION" SCIENCEDIRECT: "PRIVACY" **AND** "PRESERVING" **AND** "FORENSIC INVESTIGATION"				
29	60	37	103	• Se aplica operadores lógicos entre frases clave, • Solo publicaciones desde el **2018,** • Research papers, Computer Science(Subject area) (SCIENCEDIRECT), • Book, Conference paper (Springer)

Posterior a la aplicación de filtros se obtiene una lista de artículos reducida de la cual ya es posible aplicar ciertos criterios de inclusión y exclusión de manera manual y obtener una lista mucho más acertada. Este **proceso de selección de estudios** preliminar se realizó basándose en el título de cada artículo logrando reducir la lista a un total de 209 artículos de interés basándose en los criterios que fueron:

Tabla 11 - Criterios de inclusión y exclusión de artículos. **[Elaboración propia]**

Criterios de Inclusión	**Criterios de Exclusión**
1. Artículos que propongan métodos de preservación de privacidad en contextos forenses.	1. Artículos que propongan métodos de preservación de privacidad en contextos no digitales.
2. Artículos que implemente técnicas de preservación de privacidad para data recuperada con OSINT.	2. Artículos que implementan preservación de la privacidad a nivel de hardware.
3. Artículos que implementen algún framework orientado a	3. Artículos de tipo de revisión de literatura sobre

preservación de la privacidad en contextos forenses.	preservación de la privacidad.
4. Artículos que implementen seguridad mediante tecnologías PET.	4. Artículos que aborden la preservación de la privacidad a nivel de herramientas de usuario final y configuración.
5. Artículos referentes a privacidad por diseño en contextos forenses u OSINT.	

Aplicando los criterios descritos se logró refinar la **lista de datos** de 209 artículos a **17 artículos seleccionados** para la revisión. A continuación, se describe las cifras y los criterios que se aplicó para llegar a la selección de la lista de datos.

Tabla 12 - Lista de artículos excluidos. **[Elaboración propia]**

Artículos Excluidos		
Criterio Exclusión	**Fuentes de datos**	**Número excluidos**
Criterio N.-1	IEEE (7), Springer (18), ACM (15), ScienceDirect (31)	71
Criterio N.-2	IEEE (3), Springer (9), ACM (6), ScienceDirect (16)	34
Criterio N.-3	IEEE (6), Springer (5), ACM (2), ScienceDirect (13)	26
Criterio N.-4	IEEE (12), Springer (12), ACM (3), ScienceDirect (20)	47
Duplicados	IEEE (2), Springer (2), ACM (2), ScienceDirect (8)	14
	Total, Excluidos	**192**

Una vez obtenidos los artículos seleccionados para la revisión se pueden destacar **características de cada estudio,** por ejemplo partiendo desde la problemática de preservación de privacidad se tiene trabajos que abordan esto y proponen métodos de solución, como en [20] que aborda la contradicción natural entre los objetivos que persigue la informática forense, la protección de la privacidad y sus limitaciones, de igual

manera en [21] y [22] se propone un método de privacidad basado en dos elementos que son el nivel de sensibilidad de los datos y la relevancia para la investigación, mismos elementos que determinaran el nivel de acceso a la misma y si amerita ser recolectada y analizada, cabe mencionar que depende fuertemente de la clasificación previa que se realice sobre la misma para lo cual sugieren técnicas de machine learning.

En [21] también se analiza el hecho de determinar los niveles de sensibilidad de información para su tratamiento (encriptación) y gestión (configuración de acceso). Por otro lado, en [23] plantea la automatización de tareas para el análisis de información, iniciando por definir palabras clave y la posterior configuración automática de búsquedas que definen parámetros como: permisos efectivos sobre los archivos de búsqueda, filtros de fechas, formatos a buscar, etc. Otro método de privacidad [24] propone una solución basada en tecnologías PET y la aplicación de privacidad por diseño en cuanto a métodos control de acceso, encriptación, patrones de diseño de software, etc. Un aspecto relevante y de interés para el presente trabajo es el análisis que realiza sobre ocho principios de diseño de privacidad (minimizar, separar, agregar, ocultar, informar, controlar, hacer cumplir y demostrar) y las posibles implementaciones que permitirían hacer cumplir estos principios. Por otro lado, en [25] se propone un método en el cual la recolección y análisis de información están condicionados por el usuario de la misma, y de un tercero neutral quien define que información es privada y relevante para la investigación y se recolectará, logrando preservar la privacidad, pero limitando al investigador u ocultando información a su conveniencia.

En métodos como [26], [25], [27] y [28] se considera el cifrado y la anonimización de datos como técnicas eficaces para resolver el problema de privacidad, incluso también se sugiere la utilización de estas prácticas en leyes de protección de datos como RGPD y LOPD. No obstante, como se menciona en [29] las propuestas de este tipo basadas en criptografía, cifran todos los datos del usuario durante la fase de recopilación y posteriormente se descifran cuando encuentran en la fase de análisis. El problema con estos métodos es que el cifrado y descifrado de datos completos del usuario es costoso y una solución ineficiente. Para acelerar un proceso de investigación protegiendo la privacidad, según [29], se debería cifrar únicamente los datos que son relevantes y privados.

Como se puede ver los distintos métodos analizados utilizan distintas técnicas y flujos para preservar la privacidad, a continuación, en la Tabla 13 se presenta un resumen de los artículos revisados y sus características.

Tabla 13 – Métodos de privacidad y sus técnicas. **[Elaboración propia]**

Métodos de privacidad	**Aspectos de interés (Métodos, técnicas, aspectos legales, etc.)**
A Privacy-Preserving Digital Forensics Framework [21]	Encriptación, clasificación de información, IA, control de acceso
Digital Forensics and Privacy-by-Design (2019)[30]	Encriptación, patrones de diseño seguros, control de acceso
Efficient Privacy-Preserving Forensic Method for Camera Model Identification [31]	Detección de imágenes, IA, Deep Learning.
Designing an automated, privacy preserving, and efficient digital forensic framework (2019) [32]	Definición de palabras clave, clasificación de información, automatización de búsquedas, control de acceso.
A SMART Goal-based Framework for Privacy Preserving Embedded Forensic Investigations (2019) [25]	Clasificación de PII, consentimiento del usuario (almacenamiento y cobertura), niveles de privacidad (relevancia y sensibilidad de los datos), control de acceso, participación de un tercero neutral como mediador
Privacy Levels for Computer Forensics: Toward a More Efficient Privacy-preserving Investigation [22]	Clasificación de información, IA, control de acceso
A Comprehensive Analysis of Privacy-Preserving Solutions [23]	Anonimización de datos, protección contra crawling.
A privacy-aware digital forensics investigation in enterprises [18]	Algoritmos de filtrado por niveles basado en entropía, detección de patrones en archivos
Privacy-Preserving social media Forensic Analysis for Preventive Policing of Online Activities [33]	Reconocimiento de voz, detección de rostros, técnicas proactivas contra crawling, RGPD
Gathering Evidence	Control de acceso y autenticación, técnicas

from OSINT Sources [34]	antiforense,
Securing Medical Forensic System Using Hyperledger Based Private Blockchain [35]	Detección de acceso no autorizado, Blockchain, Hyperledger
Privacy-enhanced robust image hashing with bloom filters [36]	Hashing, Criptografía
Privacy and Robust Hashes [37]	Combinación de técnicas de Hash, criptografía, privacidad
A secure searchable encryption scheme for cloud using hash-based indexing [38]	Cifrado homomorfico, optimización de procesos de cálculo de hash, privacidad en la nube
Log pseudonymization: Privacy maintenance in practice [39]	Pseudonimización, Elasticsearch, Logstash y Kibana
A traceable and revocable multi-authority access control scheme with privacy preserving for mHealth [40]	Revocación de permisos, multi autorización de recursos, preservación de privacidad
Secure and efficient privacy protection system for medical records [41]	Criptografia, Biohashing

Una vez expuestos y analizados los artículos y sus características, se realiza una **discusión**. Tras concluir la revisión de trabajos relacionados se evidencia que la mayoría de métodos de preservación de privacidad se encuentran basados en técnicas criptográficas y métodos de control de acceso, en segundo lugar y no tan lejano esta la utilización de técnicas de inteligencia artificial como machine learning, aprendizaje supervisado, etc. Todas estas técnicas están aplicadas en su mayoría a data recolectada en un proceso judicial o de negocio, es decir data que fue recolectada con un fin distinto al análisis forense, y que recibe un tratamiento distinto a data que proviene de fuentes abiertas en donde la naturaleza de la información es impredecible y masiva, proponiendo retos tecnológicos y denotando la falta de estudios en este campo. Sin embargo, las soluciones genéricas que propone la privacidad por diseño con el uso de tecnologías son bastante útiles y efectivas ya que parten de necesidades comunes como la gestión de control de acceso y cifrado de información y que al parecer serían soluciones bastante factibles para satisfacer los requerimientos de seguridad en un contexto dado.

2.6. Contribución

El presente proyecto es una contribución académica importante dentro del campo de la preservación de la privacidad para OSINT en investigaciones digitales forenses, ya que realizando una revisión de literatura se sabe que es el primero en llevar la estructura conceptual de **PRECEPT** hacia un método practico y extender sus directrices hasta un proceso de inteligencia de fuentes abiertas, dejando como referencia y aporte un método base que identifica necesidades y propone soluciones de manera practica basadas en los conceptos del framework y tecnologías de la información. La propuesta también contribuye y se diferencia de otras alternativas añadiendo un elemento que considera aspectos legales basadas en la LOPD la cual debe ser tomada en cuenta en los procesos de gestión de la información recolectada.

3. PROPUESTA DE UN MÉTODO DE PRESERVACIÓN DE PRIVACIDAD

En esta sección se describe la metodología, diseño, técnicas y herramientas utilizadas para el diseño e implementación del método de preservación de privacidad. Dentro de la metodología se describen sus etapas e implementación mediante sus siete líneas guía. En el diseño del método se definen sus componentes y relaciones, mediante mapeos de fases y actividades, también se describen las herramientas y técnicas a utilizar en la implementación del prototipo funcional de software.

3.1. Metodología

Design Science Research fue la metodología con la que se decidió trabajar debido a que permite medir de forma cuantitativa los resultados y busca generar nuevo conocimiento, en el caso particular de este proyecto, se ha genera conocimiento que puede ser utilizado en investigaciones forenses digitales soportadas por OSINT. Además de la metodología mencionada también se trabajó con Scrum, que es una metodología ágil mayormente utilizada para el desarrollo de software que define ciertos aspectos de interés que se han tomado como soporte para la metodología principal.

Designó Science Research se puede aplicar a varias ramas de la ciencia y de manera general, parte de un problema, luego, define los objetivos de la solución al problema, a continuación, se diseña y desarrolla la solución, se realiza una demostración de esta, se lleva a cabo una evaluación y, finalmente, se debe realizar una socialización de la solución encontrada. En la presente investigación, estas etapas se detallan a continuación:

- **Problema**: Preservación de la privacidad en la aplicación de OSINT durante investigaciones forenses digitales soportadas por OSINT.
- **Objetivos de la solución:** Implementación de un método de preservación de la privacidad en la aplicación de OSINT durante investigaciones forenses digitales soportadas por OSINT.
- **Diseño, desarrollo, demostración de la solución:** Para estas etapas, se utilizó el marco de trabajo PRECEPT, ciclo forense, ciclo OSINT y la LOPD que se detallan en el capítulo 4.
- **Evaluación:** Para esta etapa se utilizó la herramienta JVISUALVM herramienta propia de java que permite evaluar los recursos consumidos por una aplicación,

así como también un modelado de amenazas basado en STRIDE que se detalla en el capítulo 5.

- **Socialización:** Esta etapa se cumple con la elaboración de este documento y su posterior divulgación en medios especializados.

Para llevar a cabo estas etapas, la metodología “Design Science Research” considera 7 líneas guía, las cuales se presentan en la Tabla 14.

Tabla 14 - Líneas guía de la metodología Design Science Research. **[Elaboración propia]**

#	**Línea guía**	**Descripción**
1	Diseño de un artefacto	Esta metodología debe producir un artefacto, en este caso, será un método de preservación de privacidad para OSINT durante investigaciones forenses digitales el cual se describe en el apartado 3.2 de este capítulo. Para el diseño de este artefacto se utilizará SCRUM ya que define procesos y actividades las cuales soportan varias actividades dentro de esta fase de diseño como se puede observar en la Figura 5.
2	Problema relevante	El objetivo de esta metodología es desarrollar una solución basada en tecnología para solventar un problema empresarial de importancia, en este caso, se enfoca en el problema de preservación de privacidad para OSINT durante investigaciones forenses digitales el cual se describe en el capítulo 1. Además, para el desarrollo de la solución tecnológica se utilizará SCRUM, ya que soporta la gestión de todo el ciclo de desarrollo de software como se puede ver en la Figura 5.
3	Diseño de la evaluación	En Design Science Research es necesario que sea posible evaluar el artefacto diseñado, en la presente investigación, la evaluación se enfocará en la aplicabilidad de la propuesta en un caso de estudio de interés público que se describe en el capítulo 5. Para llevar a cabo la evaluación se realizará un modelado

		de amenazas con STRIDE y un análisis y control de riesgos basado en los requerimientos de privacidad dado por el método propuesto
4	Contribuciones de la investigación	En la metodología seleccionada, se debe contribuir al área que se investiga, en el caso de esta investigación, se aportará en el ámbito de la preservación de la privacidad para OSINT durante investigaciones digitales forenses que se detalla en el capítulo 2.
5	Rigor de la investigación:	Tanto para el marco de privacidad, como para su evaluación, se utilizarán métodos rigurosos que deberán ser definidos durante el proceso de investigación. Para esto se utilizará el ciclo de vida para investigaciones digitales forenses el cual está contemplado en PRECEPT y se trabajará a lo largo de sus fases haciendo hincapié en las fases definidas en el alcance.
6	Diseño como proceso de búsqueda	La metodología indica que será necesario utilizar los medios disponibles para alcanzar los fines deseados, siempre que se satisfagan los requerimientos de los problemas planteados. Para esto se ha realizado una revisión sistemática de literatura la cual se explica de mejor manera en el apartado 2.5 del presente documento que corresponde a trabajos relacionados.
7	Comunicación de la investigación	En Design Science Research, se debe presentar la investigación tanto a personas con conocimientos en tecnología, como a personas orientadas a la gestión. En este caso particular, se lo hará mediante una presentación a un grupo de profesionales tecnológicos y su posterior publicación de un artículo científico.

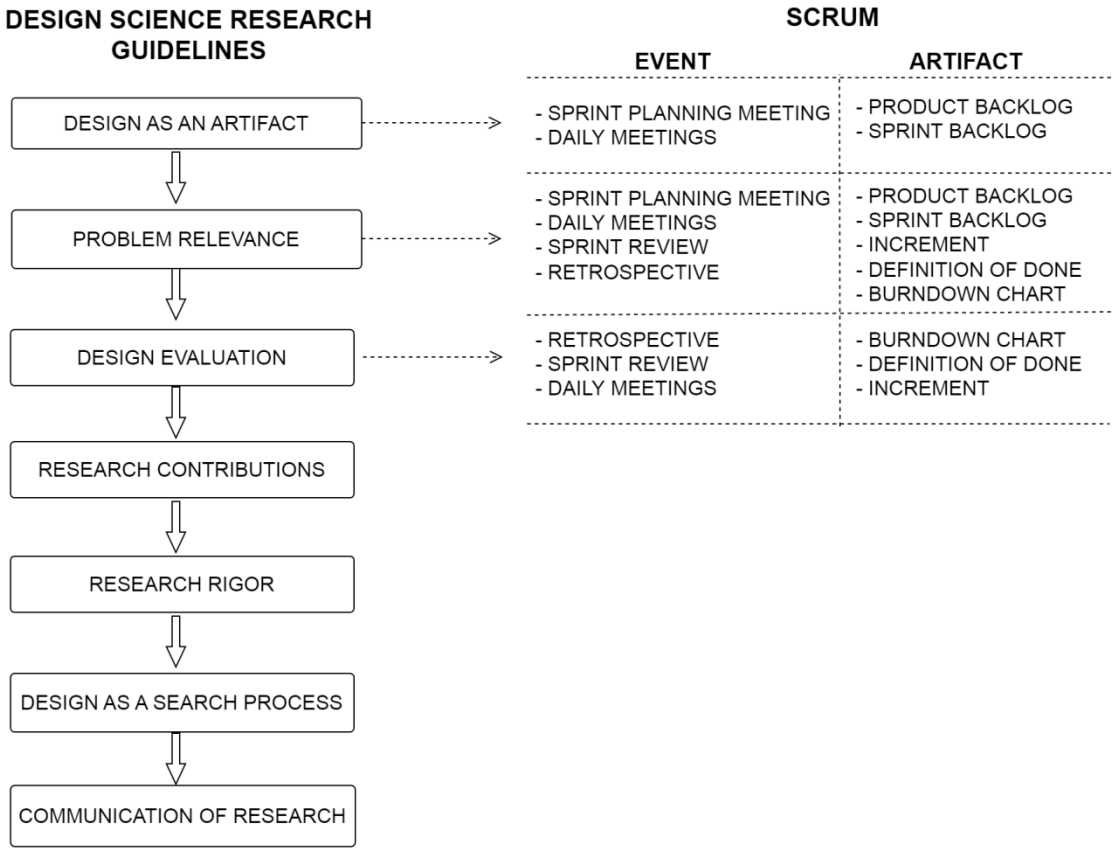


Figura 5- Mapeo de procesos de metodologías DSC y SCRUM. **[Elaboración propia]**

Para cumplir con las líneas guía planteadas por la metodología que se utilizó, en primer lugar, se realizó una búsqueda sistemática de material bibliográfico respecto a marcos de trabajo y guías relacionadas con protección de la privacidad en prácticas OSINT e investigaciones digitales forenses. De la misma manera, se identificaron principios de privacidad, principios éticos, aspectos legales y directrices para prácticas de OSINT en investigaciones forenses digitales respetuosos de la privacidad, los cuales fueron insumos necesarios para mapearlos en las fases de una investigación forense y el ciclo OSINT.

3.2. Diseño del método de preservación de privacidad

En este apartado se presentan y describen los elementos, relaciones y aspectos relevantes para el diseño del método de preservación de privacidad.

Describiendo de manera general, las actividades que se han definido parten desde el ciclo de vida de una investigación digital forense junto con el ciclo OSINT para posteriormente mapear principios de privacidad definidos en PRECEPT junto con

consideraciones legales y finalmente definir requerimientos de privacidad y lineamientos técnicos. Por otro lado, se tendrá los requerimientos generales del prototipo OSINT los cuales serán caracterizados por los requerimientos de privacidad obtenidos previamente logrando definir una lista de requerimientos funcionales enriquecidos en aspectos de privacidad y que definirán la arquitectura y diseño del aplicativo. En la Figura 6 se puede ver una estructura inicial de lo mencionado.

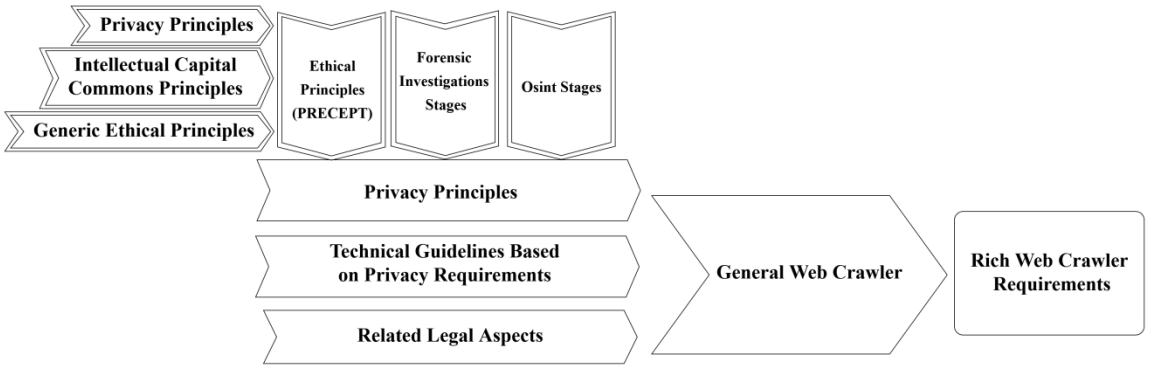


Figura 6 – Esquema inicial de método de preservación de privacidad. **[Elaboración propia]**

3.2.1. Mapeo de fases

Para el diseño del método de preservación propuesto se parte de dos pilares que son el ciclo de vida de una investigación digital forense y el ciclo de vida de un proceso OSINT, esto debido a que el contexto del problema que se plantea resolver esta principalmente dentro de estos dos procesos.

En primera instancia se realizó un mapeo de sus fases basado en las actividades que cada una dictamina ya que al tener actividades en común es primordial conocer los vínculos que tiene estos dos procesos poseen para definir un diseño óptimo y trazable como se puede ver en la Figura 7. Una versión detallada a nivel de actividades se puede encontrar en las columnas CICLO OSINT y CICLO FORENSE del mapeo general de fases y requerimientos del Anexo II .

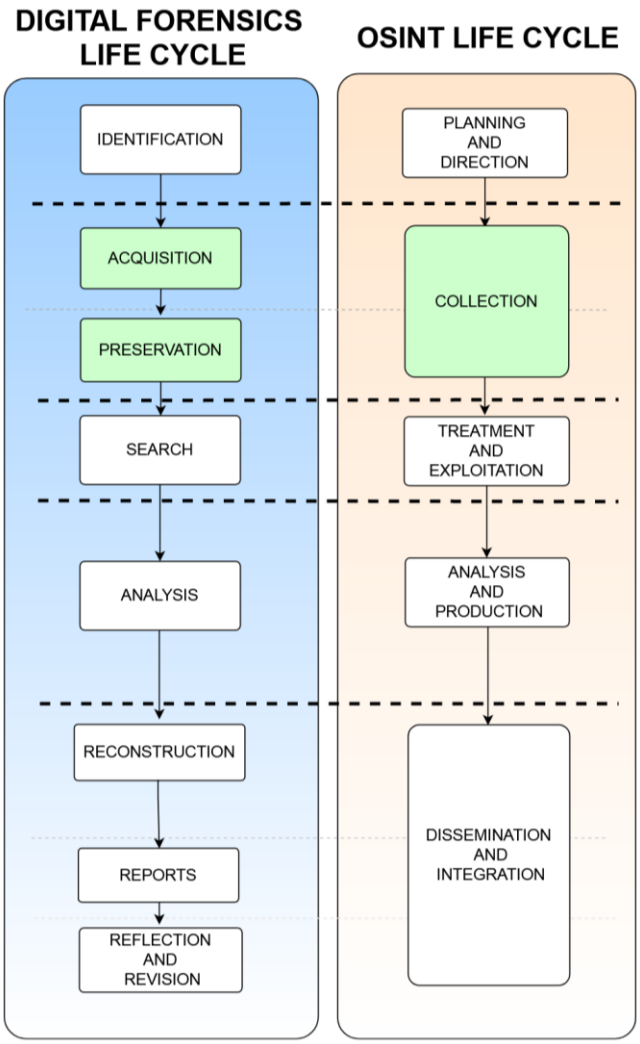


Figura 7 - Mapeo de fases ciclo forense y OSINT. **[Elaboración propia]**

Además, en la Figura 8 se puede observar el aporte de este mapeo de fases en el diseño del método.

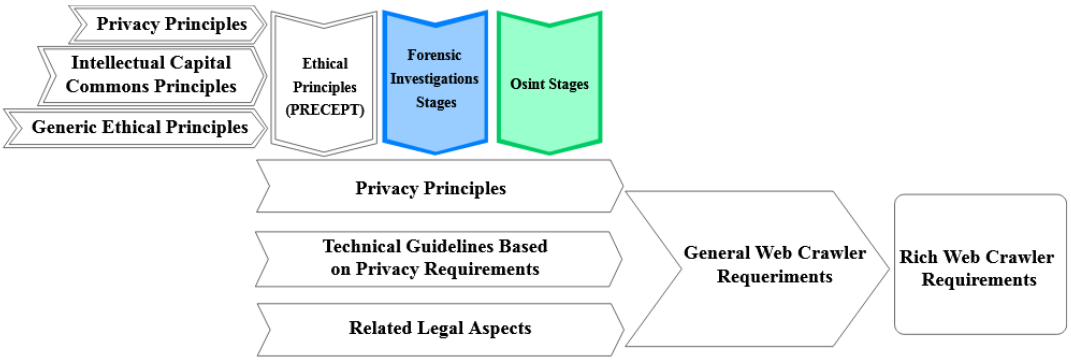


Figura 8 – Esquema del método, mapeo de fases. **[Elaboración propia]**

3.2.2. Mapeo de principios éticos y de privacidad

Una vez establecida la relación entre las fases de los ciclos OSINT y forense es necesario analizar dichas actividades para darles un enfoque orientado a preservación de privacidad. Para esto se utilizan los principios éticos y de privacidad que PRECEPT ya define para las fases del ciclo forense de acuerdo a la Tabla 15 y Tabla 17, estos principios también aplican para las fases de OSINT gracias al mapeo realizado previamente.

Una vez mapeados estos principios éticos también se logra una referencia hacia otros principios como son: principios de privacidad y principios para una comunicación efectiva de capital intelectual que también son dados por PRECEPT tal como se observa en la Tabla 15, Tabla 16 y Tabla 17. De esta manera finalmente se incluye un tercer elemento (**ETHICAL PRINCIPLE)** importante en cuestiones de privacidad tal como se puede observar en la Figura 9.

Tabla 15– Principios de privacidad. [1]

P1	Consent & choice	P7	Openness, transparency & notice
P2	Purpose legitimacy and specification	P8	Individual participation and access
P3	Collection limitation	P9	Accountability
P4	Data minimization	P10	Information security controls
P5	Use, retention and disclosure limitation	P11	Compliance
P6	Accuracy and quality		

Tabla 16- Principios de comunicación efectiva del capital intelectual. [1]

IC1	Clear link to future value creation	IC6	Alignment of interests between company and investors
IC2	Transparency of methodology	IC7	Prevention of information overflow
IC3	Standardization	IC8	Reliability and responsibility
IC4	Consistency over time	IC9	Risk assessment
IC5	Balanced trade-off between disclosure and privacy	IC10	Effective disclosure placement and timing

Tabla 17 – Mapeo principios éticos. [1]

Ethical Principle	Detail	Privacy Principle	IC Comms Principle	Generic Ethical Principle
E1	**Delineate Remit:** Commence by carefully delineating the remit of the investigation (Nikkel, 2014, Srinivasan, 2007).	P2, P3		Respect Social Responsibility
E2	**Respect the privacy of the subject:** The privacy of the subject should be protected by only investigating topics identified as being of interest to the investigation (Law *et al.*, 2011; Dehghantanha and Franke, 2014). In particular, examination scope should be identified *before* the investigation proceeds.	P3	IC5, IC7	Justice Beneficence
E3	**Only investigate other parties if there is evidence of their involvement:** The privacy of third parties should be protected by only investigating them if there is evidence that they have been implicated in the topic of the investigation (Van Staden, 2013).	P3	IC10	Justice Beneficence
E4	**Exclude private information:** During investigation, bookmark private information that is irrelevant to the investigation so that it is not included in any report. Examples are personal credit card numbers, personal passport numbers, and national insurance numbers (John, 2012; Dehghantanha and Franke, 2014).	P4		Respect
E5	**Document all actions:** Document all data that was examined, judged private and irrelevant, and relevant to the investigation (Saleem *et al.*, 2014; Srinivasan, 2007).	P6, P7	IC8	Integrity
E6	**Facilitate audits:** Facilitate post-investigation scrutiny (Gay, 2012).	P9		Integrity
E7	**Report all investigative activities:** When the investigation is concluded, the report should include details of exactly what was examined, who was included in the investigation, which devices were examined (and who they belonged to) (Losavio *et al.*, 2015), how data was classified as relevant (to be reported), confidential (only to be reported if the court so orders), irrelevant (not to be divulged) and how the data was preserved to prevent any alteration (Saleem *et al.*, 2014; Roux and Falgoust, 2012).	P5, P10		Integrity
E8	**Be transparent about the extent of the investigation, and the gathered information:** Subjects, and their counsel, have to be given the right to know what data was processed and how it was processed (Saleem *et al.*, 2014).	P9		Respect Justice
E9	**Investigators should undergo regular training:** Investigators should undergo frequent proficiency training and testing (Saleem *et al.*, 2014).	P11		Social Responsibility
E10	**Information's Integrity and Confidentiality should be maintained:** Investigators should carry out investigations lawfully and with integrity, and confidentiality (ACPO, 2012; Srinivasan, 2007).	P11	IC10	Integrity
E11	**Consideration for the well being of investigators** as highlighted by Burruss, *et al.* (2018).			Social Responsibility

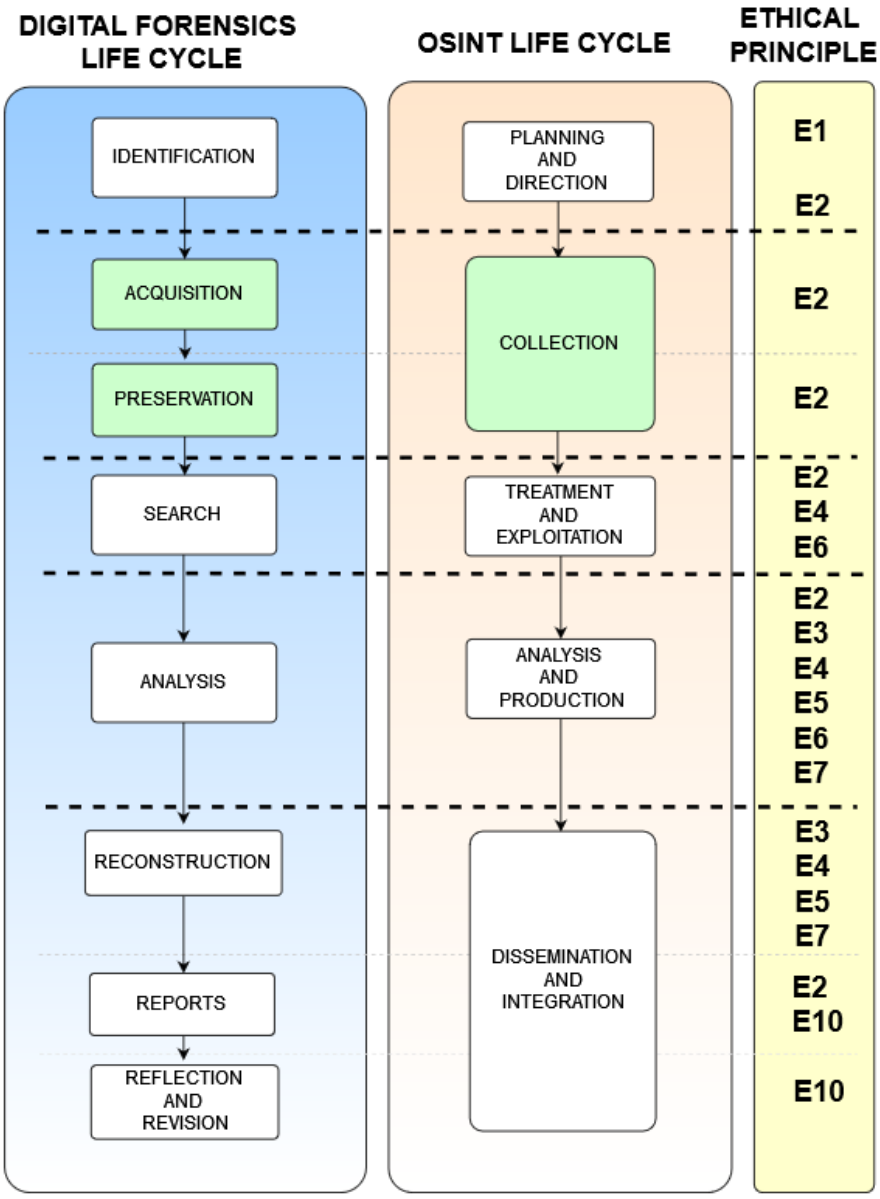


Figura 9 – Mapeo ciclos forense, OSINT y principios éticos. **[Elaboración propia]**

Una versión detallada que incluye y mapea principios éticos, principios de privacidad y principios de capital intelectual se puede encontrar en el mapeo general de fases y requerimientos del Anexo II. En la Figura 10 se resaltar la inclusión del nuevo elemento (**Privacy Principles**) al mapeo de principios de privacidad del método propuesto.

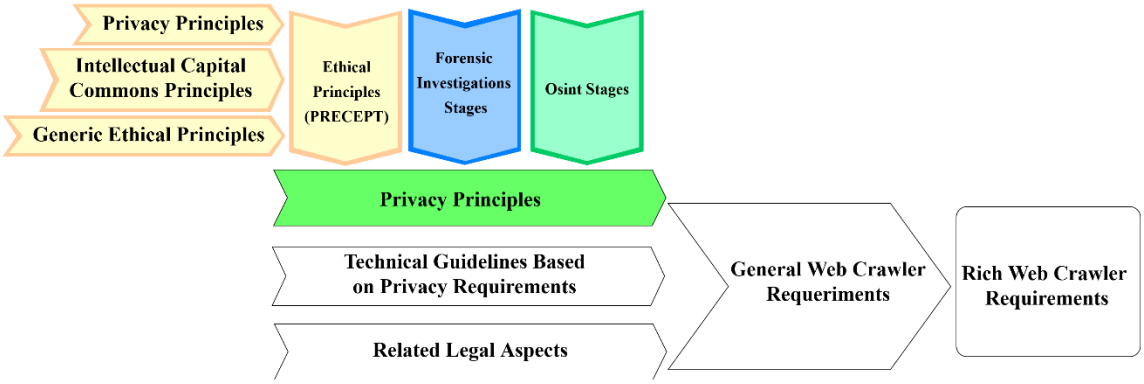


Figura 10 - Esquema del método, principios éticos y de privacidad. **[Elaboración propia]**

3.2.3. Requerimientos de privacidad

Una vez realizados los mapeos de fases, actividades y principios, es necesario definir un cuarto elemento que de sentido y conjugue estos elementos de manera práctica, para lo cual se han definido requerimientos que permitan hacer respetar los principios de privacidad que rigen en las actividades de cada fase de los procesos en cuestión, tal como se puede ver en la figura Figura 11 y Tabla 18. Además, para cada requerimiento de privacidad se definen lineamientos técnicos para su implementación los cuales más adelante serán tomados en cuenta para la implementación del prototipo funcional de software. Estos requerimientos de privacidad (**PRIVACY REQUERIMIENT** en la Figura 11) se pueden revisar a detalle en la columna LINEAMIENTOS PRESERVACIÓN DE PRIVACIDAD del mapeo general de fases y requerimientos del Anexo II**.**

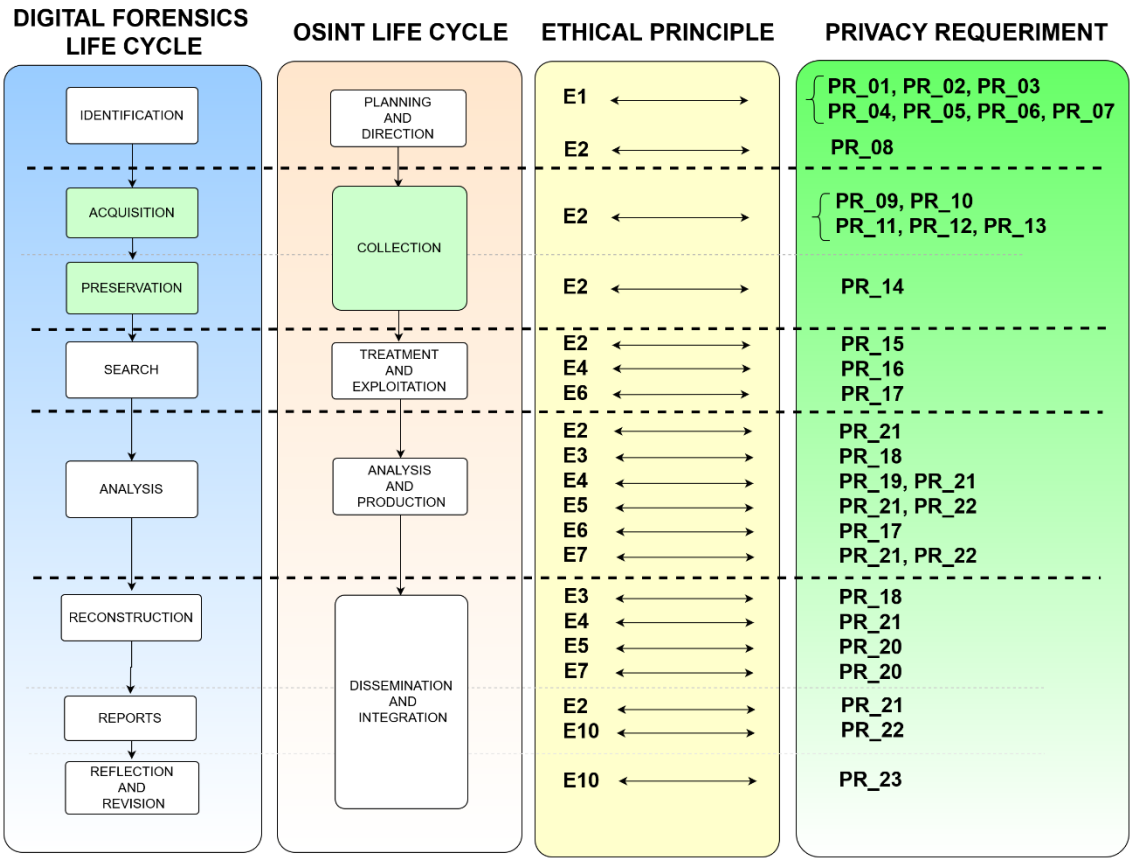


Figura 11 – Mapeo Requerimientos de privacidad

Tabla 18 – Requerimientos de Privacidad y lineamientos técnicos

PR_01	Definir las fuentes y dominios de información de acuerdo al ámbito establecido para la investigación.
Technical Guidelines	• **1.1** Definir plan de búsquedas de acuerdo al ámbito de investigación • **1.2** Implementar mecanismos que permitan definir los orígenes de búsqueda (Dominios, subdominios, sitios, white list)
PR_02	Definir el volumen de información a ser almacenado.
Technical Guidelines	• **2.1** Definir el volumen de información a ser almacenado para las instancias de búsqueda delimitado de acuerdo al ámbito de investigación. • **2.2** Implementar mecanismos que permitan configurar el volumen máximo de información a ser almacenado.
PR_03	Definir los formatos (png, pdf, xtml, etc.) de archivos que serán considerados

	y que se almacenarán.
Technical Guidelines	• **3.2** Definir una lista de formatos de información a ser almacenada delimitados de acuerdo al ámbito de investigación. • **3.1** Implementar mecanismos que permitan configurar los formatos de información que se almacenaran (White list).
PR_04	Definir un espacio de tiempo (años: 2012, 2020, 2023, etc.) que serán considerados para realizar las búsquedas.
Technical Guidelines	• **4.1** Definir los espacios de tiempo considerados para el almacenamiento de recursos delimitados de acuerdo al ámbito de investigación. • **4.2** Implementar mecanismos que permitan definir listas de años (2012, 2013, 2018, etc.) que serán considerados para realizar búsquedas.
PR_05	Definir una lista de idiomas (inglés, español, etc.) que serán considerados para realizar las búsquedas.
Technical Guidelines	• **5.1** Definir los idiomas de los recursos a ser almacenados delimitados de acuerdo al ámbito de investigación. • **5.2** Implementar funcionalidades que permitan definir listas de idiomas (inglés, español, etc.) que serán considerados para realizar las búsquedas en sitios con información en estos idiomas.
PR_06	Definir un límite de búsquedas.
Technical Guidelines	• **6.1** Definir un número máximo de búsquedas a realizar delimitado de acuerdo al ámbito de investigación. • **6.2** Implementar funcionalidades que permitan definir parámetros para el control del número máximo de búsquedas que se permitirán realizar.
PR_07	Definir información que se almacenará como auditoria respecto a las búsquedas realizadas.
Technical Guidelines	• **7.1** Definir las instancias dentro del proceso de investigación que requieran el registro de una justificación para su ejecución. • **7.2** Implementar funcionalidades que permitan registrar información de causa en relacionada con la actividad realizada.
PR_08	Definir criterios de búsqueda (palabras, frases, patrones, etc.).
Technical Guidelines	• **8.1** Definir cadenas de búsqueda para las instancias de búsqueda delimitadas de acuerdo al ámbito de investigación.

	• **8.2** Implementar funcionalidades que permitan definir palabras, frases o patrones de búsqueda.
PR_09	Verificar que la información que se almacenará provenga únicamente de fuentes del ámbito(dominios) establecido.
Technical Guidelines	• **9.1** Implementar controles para filtrar y realizar peticiones únicamente a sitios del ámbito(dominios) establecido.
PR_10	Verificar que la información que se almacenará no exceda el volumen máximo establecido.
Technical Guidelines	• **10.1** Implementar controles para verificar que la información que se almacene no exceda el volumen máximo permitido. • **10.2** Implementar mecanismos de alerta al supera el volumen de almacenamiento máximo configurado.
PR_11	Verificar que la información que se almacenará sea únicamente de los formatos establecidos.
Technical Guidelines	• **11.1** Implementar controles para verificar que la información que se almacene sea de los formatos establecidos.
PR_12	Verificar que la información que se almacene no supere lo configurado para cada instancia de búsqueda.
Technical Guidelines	• **12.1** Implementar mecanismos para verificar el volumen de almacenamiento para una instancia de búsqueda configurada. • **12.2** Implementar mecanismos de alerta al supera el volumen de almacenamiento máximo configurado.
PR_13	Registrar datos acerca del origen de la información.
Technical Guidelines	• **13.1** Implementar políticas y mecanismos por defecto para registrar información de los orígenes de la información recolectada.
PR_14	Establecer mecanismos para el aseguramiento de la seguridad de la información.
Technical Guidelines	• **14.1** Definir políticas de control de acceso. • **14.2** Implementar métodos de control de acceso que se ajusten al modelo de negocio. • **14.3** Encriptación y anonimización de datos. • **14.4** Gestión de usuarios y contraseñas. • **14.5** Seguridad de redes. • **14.6** Política de desarrollo seguro

	• **14.7** Pruebas de seguridad de sistemas
PR_15	Recuperar metadatos
Technical Guidelines	• **15.1** Implementar mecanismos que permitan recuperar metadatos de la información almacenada tal como: tamaño, extensión, formato, fecha y hora de procesamiento, fuentes de origen, etc.
PR_16	Definir un nivel de sensibilidad de información para cada unidad de dato procesada (clasificar la información).
Technical Guidelines	• **16.1** Implementar técnicas de etiquetado de información (publica, confidencial, etc.) • **16.2** Definir políticas de gestión de la información de acuerdo al nivel de sensibilidad.
PR_17	Definir momentos y acciones relevantes dentro del proceso, a fin de registrar información para relevante para auditorias y trazabilidad.
Technical Guidelines	• **17.1** Implementar métodos proactivos para registros de log, firmas digitales, estampas de tiempo, registros de auditoría, etc. • **17.2** Implementar métodos que permitan trazabilidad en las actividades y procesos dentro de la investigación.
PR_18	Previo a incluir a terceros en hipótesis considerar la evidencia encontrada.
Technical Guidelines	• **18.1** Definir criterios de inclusión en la investigación. • **18.2** Definir perfiles de atacantes.
PR_19	Definir mecanismos de filtrado de información.
Technical Guidelines	• **19.1** Definir nivel de sensibilidad de la información. • **19.2** Definir flujos para el tratamiento de la información. • **19.3** Definir tipo de filtrado (automático o manual) de la información.
PR_20	Definir tipos de informes a presentar en función de: Tipo de información que contiene (Nivel de sensibilidad), Perfil del lector, Relevancia en cuanto a la investigación
Technical Guidelines	• **20.1** Definir perfiles y dominios de uso de información. • **20.2** Definir formatos de presentación de información. • **20.3** Definir métodos de amonificación de información presentada.
PR_21	Definir tipos y formatos de información a ser presentada para análisis.
Technical Guidelines	• **21.1** Implementar funcionalidades que permitan generar reportes en base al nivel de acceso y sensibilidad de información. • **22.1** Implementar técnicas de anonimización y enmascaramiento de

	datos.
PR_22	Analizar la revocación de privilegios a ciertos perfiles en cuanto a acceso de información.
Technical Guidelines	• **22.1** Definir políticas para revocación de privilegios. • **22.2** Implementar funcionalidades que permitan revocación de privilegios de acceso a información (Configuración dinámicas de permisos)
PR_23	Analizar la conservación la información recolectada en la investigación.
Technical Guidelines	• **23.1** Definir métodos para eliminación de datos segura. • **23.2** Implementar funcionalidades que permitan eliminar información recolectada en la investigación de manera segura.

En la Figura 12 se resalta el nuevo elemento (**Technical Guidelines Based on Privacy Requeriments**) dentro del método propuesto.

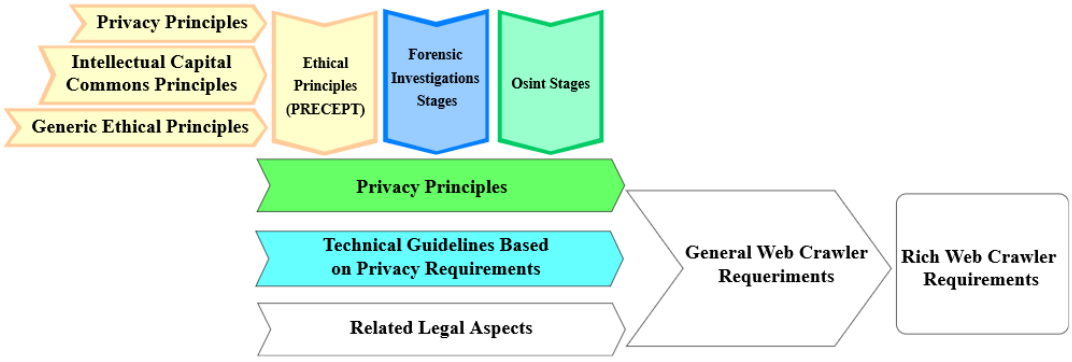


Figura 12 - Esquema del método, requerimientos de privacidad y lineamientos técnicos. **[Elaboración propia]**

3.2.4. Consideraciones legales

Dentro del alcance del proyecto está el aspecto legal y para cubrir este punto se considera dentro del diseño a la ley orgánica de protección de datos LOPD, de la cual se extraen artículos y aspectos que tiene implicaciones con las actividades de los ciclos forense, OSINT y principios de privacidad previamente mapeados, esto con el fin de contextualizar aspectos importantes de la ley al momento de su implementación tales como:

- **Ámbito de aplicación integral:** Tratamiento legítimo de datos personas, interés legítimo, etc.
- **Principios: Finalidad:** Pertinencia y minimización de datos personales, seguridad de datos personales, conservación, etc.
- **Categorías especiales de datos:** Datos personales, tratamiento de datos sensibles, derechos de los Titulares de Datos Crediticios.
- **Transferencia o comunicación y acceso a datos personales por terceros:** Acceso a datos personales por parte del encargado, Acceso a datos personales por parte de terceros, comunicación de datos personales.
- **Seguridad de datos personales:** Protección de datos personales desde el diseño y por defecto, análisis de riesgo, amenazas y vulnerabilidades.

Este mapeo y los artículos relacionados se pueden revisar a detalle en la columna **CONSIDERACIONES LEGALES** del mapeo general de fases y requerimientos del Anexo II. Además, en la Figura 13 se resalta el nuevo elemento incluido (Related Legal Aspects) en el método propuesto.

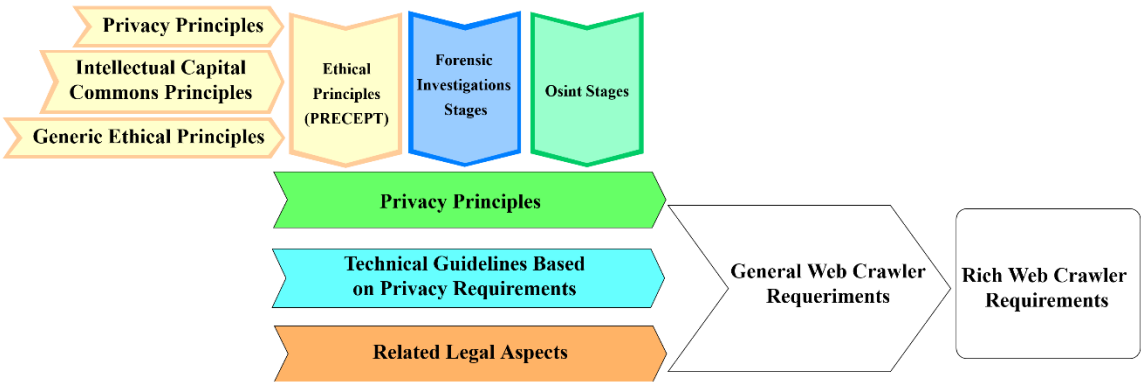


Figura 13 - Esquema del método, aspectos legales relacionados. **[Elaboración propia]**

3.2.5. Requerimientos Web Crawler

Ya definidos los requerimientos de privacidad para el método propuesto nos enfocaremos en el prototipo funcional de software, para ello se parte de la idea de construir un aplicativo básico que permita realizar OSINT mediante búsquedas que coincidan con parámetros dados por un usuario y que almacene los resultados obtenidos. La idealización de este aplicativo podría ser definida por los requerimientos descritos en la

Tabla 19 y en términos funcionales cumpliría su objetivo. Sin embargo, este aplicativo básico no define ninguna característica de seguridad.

Tabla 19 – Requerimientos generales herramienta OSINT. **[Elaboración propia]**

Código	Descripción
RG_01	El sistema debe permitir registrar usuarios en el sistema.
RG_02	El sistema debe permitir autenticar usuarios mediante id y contraseña.
RG_03	El sistema debe permitir la navegación en el sistema mediante un menú.
RG_04	El sistema debe permitir realizar búsquedas según un criterio de búsqueda ingresado por el usuario.
RG_05	El sistema debe presentar al usuario los resultados de las búsquedas.
RG_06	El sistema debe almacenar las coincidencias las búsquedas.

La definición y selección de requerimientos del aplicativo está dada en función de soportar las actividades de los ciclos OSINT y Forense para lo cual se consideran los requerimientos de privacidad de la Tabla 18 los cuales tienen referencia a las actividades mediante sus fases. A continuación, en la Tabla 20 se alinean los requerimientos funcionales y no funcionales del prototipo (ver Tabla 23) con los requerimientos de privacidad involucrados.

Tabla 20 - Mapeo requerimientos de privacidad y requerimientos web crawler**.**
[Elaboración propia]

Requerimiento de privacidad	Requerimiento Funcional Web Crawler
PR_14	RF_01
PR_14	RF_02
PR_14	RF_03
PR_14	RF_04
PR_14	RF_05
PR_14	RF_06
PR_14	RF_07
PR_01	RF_08
PR_14	RF_09
PR_01	RF_10
PR_02	RF_11
PR_03	RF_12
PR_04	RF_13
PR_05	RF_14
PR_06	RF_15

PR_07	RF_16
PR_14	RF_17
PR_14	RF_18
PR_08	RF_19
PR_08	RF_20
PR_09	RF_21
PR_10	RF_22
PR_11	RF_23
PR_04	RF_24
PR_05	RF_25
PR_13, PR_15	RF_26
PR_06	RF_27
PR_14	RF_28
PR_17	RF_29
PR_20, PR_18	RF_30
PR_16	RF_31
PR_22	RF_32
PR_23	RF_33
PR_19, PR_20, PR_21	RF_34
PR_14	RNF_01 (**Requerimiento No Funcional**)
PR_17	RNF_02 (**Requerimiento No Funcional**)
PR_14	RNF_04 (**Requerimiento No Funcional**)
PR_17	RNF_05 (**Requerimiento No Funcional**)

En la Figura 14 se resalta la inclusión de estos dos últimos elementos (**General Web Crawler Requeriments**, **Rich Web Crawler Requeriments**) método propuesto.

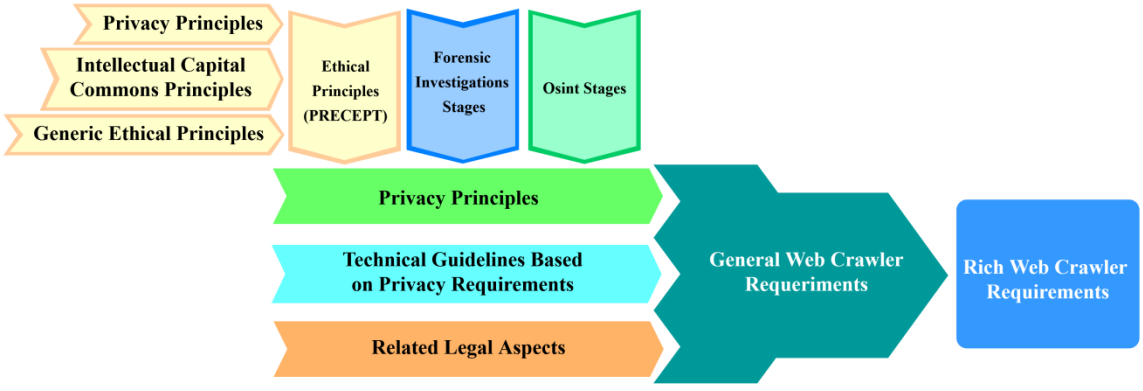


Figura 14 - Esquema del método, requerimientos web crawler generales y enriquecidos. **[Elaboración propia]**

3.3. Selección de técnicas y herramientas a utilizar

Las técnicas y herramientas utilizadas para llevar a cabo el presente proyecto han sido seleccionadas después de realizar un análisis sobre las características y prestaciones que requiere el proyecto y que estas pueden satisfacer.

3.3.1. Control de acceso

Uno de los atributos de seguridad que es primordial en este trabajo, es la privacidad, la cual consiste en garantizar que solo aquellas personas que están autorizadas para acceder a un recurso puedan hacerlo [42], en este caso el recurso custodiado es la información recolectada producto de un proceso OSINT.

Una de las técnicas más efectivas para cubrir este requerimiento de seguridad a nivel empresarial es la implementación de controles de acceso, de los cuales existen varios tipos que pueden ser caracterizados por la granularidad de su configuración o por los niveles jerárquicos de roles que permiten, etc. Por ejemplo, en [2], [42], [43] se analizan las ventajas y desventajas de varios de los modelos más utilizados.

Tabla 21 – Comparación de modelos de control de acceso. [43]

Access Control Models	Advantages	Disadvantages
DAC	• Data integrity	• Does not support dynamic alteration • Requires a high system management
MAC	• Flexibility	• Global policy • Malicious software/programs • Information flow
RBAC	• Authorization management • Hierarchical roles • Separation of duties • Least privileges	• Role explosion • Not preferred in dynamic environment • Not possible to change access rights without changing the roles
ABAC	• Supports dynamic environments	• Hierarchical ABAC • Separation of Duties • Low Expressiveness

Como se puede ver en la Tabla 21 vel modelo RBAC posee mayores ventajas y características que son consideradas vitales tal como lo describe la NIST en su manual NISTR7874 [44] de "Directrices para las métricas de evaluación del sistema de control de acceso" en el que describe como propiedades de cumplimiento aspectos como: Policy

combination, composition, and constraint, Bypass, Least privilege principle support, Separation of Duty (SoD) entre otras.

Por otro lado, también se tiene el método ABAC es un modelo basado en roles el cual de acuerdo con [43], [45] desde 2018 está en tendencia y está siendo utilizado en un porcentaje del 78% a 100% mientras que RBAC es más antiguo y se utiliza entre un 50% -100%. En este sentido en [46] analizan estos dos modelos y se concluyen de acuerdo con la Tabla 21 que RBCA es el modelo más confiable.

Tabla 22- Análisis modelos RBAC y ABAC. [46]

Issues	RBAC	ABAC
Trend in 2018	Medium	High
Global Agreement	No	Yes
Flexibility	No	Yes
Easiness	Yes	No
Dynamicity	No	Yes
Authorization Decision	Locally	Globally
Granularity	Low	High
Manageability	Simple	Complex
Conviction	Locally	Globally
Confusing deputy	No	Yes
Changing privileges	Complex	Simple
Role explosion problem	Yes	No

3.3.2. Web crawler

Uno de los objetivos a cubrir en esta investigación es la implementación de un prototipo funcional que permita la gestión de la evidencia digital recuperada mediante OSINT, para lo cual se analizó varias opciones.

Actualmente las tareas de OSINT están soportadas por decenas de plataformas gratuitas y de pago las cuales en el fondo son web crawler orientados a extraer información de sitios web populares. Por otro lado, también exponen sus APIS para ser consumidas por ingenieros de software que deseen generen sus propias herramientas OSINT especializadas, el problema de estas soluciones es que, pese a obtener información de las búsquedas que realicemos y almacenemos en nuestra base de datos, en primera instancia esta información ya queda registrada en las bases de datos de las empresas que prestan estos servicios, lo cual no es para nada deseable tomando en cuenta el objetivo de este trabajo.

Otra alternativa es implementar un web crawler propio y dotarle de características de privacidad desde su diseño lo cual no es factible en tiempo y esfuerzo dado el contexto en el que se desarrolla el actual proyecto.

En base a lo expuesto se ha optado por la utilización de un Core Web Crawler de código abierto que cumpla con requisitos técnicos básicos para implementar sobre él una herramienta OSINT que posea las características de privacidad deseadas.

En este sentido se ha analizado varios artefactos de software disponibles en plataformas como GitHub de los cuales se optó por crawler4j esto debido a su simplicidad, escalabilidad, configuración y que además posee ciertos lineamientos a modo de documentación, lo que permitió fluidez en el desarrollo de las demás características y funcionalidades de la herramienta OSINT.

3.3.3. Generación de expresiones regulares

Una de las funcionalidades principales del prototipo funcional es realizar búsquedas, en este sentido con el fin de mejorar este proceso se implementó la funcionalidad de búsquedas avanzadas que permitan al investigador ser más específico en sus criterios de búsqueda. Para llevar a cabo esta tarea a nivel técnico se requiere la utilización de expresiones regulares para lo cual la mayoría de lenguajes de programación cuenta con métodos que ayudan a esta tarea. Sin embargo, en este caso se requiere construir expresiones regulares de manera dinámica generadas por el usuario y ocultando el nivel de implementación a bajo nivel, por lo que se analizó varias librerías con las características descritas y finalmente se seleccionó VerbalExpressions que cumple lo requerido y además tiene una nomenclatura de sus métodos semántica y su curva de aprendizaje es baja.

4. IMPLEMENTACIÓN DEL PROTOTIPO ORIENTADO A LA PRESERVACIÓN DE LA PRIVACIDAD

Esta sección está dedicada a la implementación del prototipo, para la cual se parte del modelo del sistema que define entre otras cosas los requerimientos funcionales, product backlog, y el modelo arquitectónico elementos que son la base para la implementación. Posteriormente se definen a mayor detalle diagramas de secuencia, algoritmos y demás artefactos necesarios que describen la funcionalidad del prototipo y que son vitales para una correcta implementación de solución de software.

4.1. Modelo del sistema

En esta sección se hace referencia a los aspectos que conforman el modelo del prototipo orientado a la preservación de la privacidad. Aquí se describen los requerimientos y componentes que posee el sistema, así como su funcionalidad e interacción.

4.1.1. Definición de requerimientos funcionales

Los requerimientos que componen el prototipo funcional se han definido en primer lugar identificando requerimientos generales básicos para una herramienta OSINT, posteriormente sobre estos se definió otros enriquecidos basados en los requerimientos de privacidad del método propuesto. La Tabla 23 a continuación describe los requerimientos funcionales obtenidos.

Tabla 23 – Requerimientos funcionales del sistema. **[Elaboración propia].**

CÓDIGO	DESCRIPCIÓN
RF_01	El sistema debe permitir autenticación usuarios mediante usuario y contraseña.
RF_02	El sistema debe permitir autorizar recursos a usuarios.
RF_03	El sistema debe permitir resetear la clave de un usuario.
RF_04	El sistema debe permitir gestionar usuarios.
RF_05	El sistema debe permitir gestionar perfiles.
RF_06	El sistema debe permitir gestionar permisos.
RF_07	El sistema debe permitir la navegación en el sistema mediante un menú generado en base a sus privilegios.
RF_08	El sistema debe permitir configurar en un perfil de búsqueda.
RF_09	El sistema debe permitir el ingreso de información sensible mediante teclado en pantalla.

RF_10	El sistema debe permitir configurar en un perfil de búsqueda una lista blanca de dominios y subdominios sobre los que se realizará búsquedas.
RF_11	El sistema debe permitir configurar en un perfil de búsqueda parámetros para el control del volumen máximo de información que se almacenará.
RF_12	El sistema debe permitir configurar en un perfil de búsqueda una lista blanca de extensiones de archivos (JPG, PDF, TXT, etc.) sobre los que se realizara búsquedas.
RF_13	El sistema debe permitir configurar en un perfil de búsqueda una lista blanca de años (2015, 2017, 2019, etc.) según los cuales el sistema realizara búsquedas.
RF_14	El sistema debe permitir configurar en un perfil de búsqueda una lista blanca de idiomas según los cuales el sistema realizará búsquedas.
RF_15	El sistema debe permitir configurar en un perfil de búsqueda un parámetro para el control del número máximo de búsquedas se realizará.
RF_16	El sistema debe permitir configurar en un perfil de búsqueda un rango de tiempo(fechas) de activación.
RF_17	El sistema debe permitir configurar en un perfil de búsqueda un código de activación.
RF_18	El sistema debe permitir o denegar iniciar el proceso de búsqueda a un usuario de acuerdo a las condiciones de activación del perfil de búsqueda.
RF_19	El sistema debe permitir realizar búsquedas mediante un parámetro dado por el usuario.
RF_20	El sistema debe permitir realizar búsquedas avanzadas mediante uno o varios parámetros dados por el usuario.
RF_21	El sistema debe permitir o restringir el análisis de sitios web basándose en su url y a una lista blanca de dominios y subdominios configurada en el perfil de búsqueda.
RF_22	El sistema debe permitir o restringir el almacenamiento de información de acuerdo al volumen máximo de información configurado, configurado en el perfil de búsqueda.
RF_23	El sistema debe permitir o restringir el almacenamiento de archivos de acuerdo a una lista blanca de extensiones de archivo configurada en el perfil de búsqueda.
RF_24	El sistema debe permitir o restringir el análisis y almacenamiento de sitios web con información publicada en años distintos o iguales a los configurados

	en el perfil de búsqueda.
RF_25	El sistema debe permitir o restringir el análisis y almacenamiento de sitios web con información publicada en idiomas distintos o iguales a los configurados en el perfil de búsqueda.
RF_26	El sistema debe almacenar datos y metadatos acerca de las búsquedas realizadas
RF_27	El sistema debe permitir o restringir a un usuario realizar búsquedas de acuerdo a un número máximo de búsquedas configurado en el perfil de búsqueda.
RF_28	El sistema implementar encriptación sobre los datos recolectados en las búsquedas.
RF_29	El sistema debe implementar registro de información de eventos relevantes para auditorias.
RF_30	El sistema debe permitir generar reportes en base al nivel de acceso de usuario y sensibilidad de información.
RF_31	El sistema debe identificar y etiquetar la información almacenada de acuerdo a un nivel de sensibilidad (publica, confidencial, etc.)
RF_32	El sistema debe permitir revocar o restringir acceso a la información haciendo efectivos los periodos de tiempo asignados.
RF_33	El sistema debe permitir eliminar información recolectada asociada a un proceso de investigación
RF_34	El sistema debe generar reportes con información que refleje el desempeño y responsabilidad de los actores del proceso de investigación.

Cabe mencionar que, de acuerdo al alcance definido, el prototipo de software implementará únicamente los requerimientos para las fases de ADQUISICIÓN, RECOLECCIÓN Y ALMACENAMIENTO definidos dentro de los Sprints 1,2,3 y 4 de acuerdo a la Tabla 25.

4.1.2. Definición de requerimientos no funcionales

Entre las varias categorías de requerimientos no funcionales con las que podría contar la herramienta OSINT se ha priorizado la seguridad sobre el resto, esto debido a la caracterización marcada y el objetivo que persigue la investigación con esta propuesta, la Tabla 24 a continuación describe los requerimientos funcionales obtenidos.

Tabla 24 – Requerimientos no funcionales. **[Elaboración propia].**

CÓDIGO	DESCRIPCIÓN
RNF_01	El sistema debe implementar seguridad en sus procesos.
RNF_02	El sistema debe mostrar un mensaje de error que muestra la descripción del evento en caso de presentarse alguna excepción.
RNF_03	El sistema debe implementar ToolTips de ayuda para evitar que el usuario ingrese datos errados.
RNF_04	El sistema debe pedir confirmación al usuario antes de guardar los cambios en el sistema.
RNF_05	El sistema registrará un log de los cambios realizados, detallando el módulo, el tipo de movimiento, el usuario que ejecutó la operación, hora y fecha.

4.1.3. Product backlog

Una vez listados los requerimientos se define la prioridad en que estos deben ser implementados, de acuerdo con SCRUM se genera el product backlog y sus correspondientes sprints que se muestran en la Tabla 25. Cabe mencionar que estos requerimientos corresponden a todas las fases del mapeo OSINT forense, sin embargo, como ya se ha mencionado anteriormente se implementa únicamente las etapas de ADQUISICIÓN, RECOLECCIÓN Y ALMACENAMIENTO las cuales corresponden a los sprints 1,2,3 y 4.

Tabla 25 - Product backlog. **[Elaboración propia].**

CÓDIGO	DESCRIPCIÓN
SPRINT 1	
RF_01	El sistema debe permitir autenticación usuarios mediante usuario y contraseña.
RF_02	El sistema debe permitir autorizar recursos a usuarios.
RF_04	El sistema debe permitir gestionar usuarios.
RF_05	El sistema debe permitir gestionar perfiles.
RF_06	El sistema debe permitir gestionar permisos.
RF_07	El sistema debe permitir la navegación en el sistema mediante un menú generado en base a sus privilegios.
RNF_01	El sistema debe implementar seguridad en sus procesos.
SPRINT 2	
RF_08	El sistema debe permitir configurar en un perfil de búsqueda.
RF_09	El sistema debe permitir el ingreso de información sensible mediante teclado en pantalla.
RF_10	El sistema debe permitir configurar en un perfil de búsqueda una

	lista blanca de dominios y subdominios sobre los que se realizará búsquedas.
RF_11	El sistema debe permitir configurar en un perfil de búsqueda parámetros para el control del volumen máximo de información que se almacenará.
RF_12	El sistema debe permitir configurar en un perfil de búsqueda una lista blanca de extensiones de archivos (JPG, PDF, TXT, etc.) sobre los que se realizara búsquedas.
RF_13	El sistema debe permitir configurar en un perfil de búsqueda una lista blanca de años (2015, 2017, 2019, etc.) según los cuales el sistema realizara búsquedas.
RF_14	El sistema debe permitir configurar en un perfil de búsqueda una lista blanca de idiomas según los cuales el sistema realizará búsquedas.
RF_15	El sistema debe permitir configurar en un perfil de búsqueda un parámetro para el control del número máximo de búsquedas se realizará.
RF_16	El sistema debe permitir configurar en un perfil de búsqueda un rango de tiempo(fechas) de activación.
RF_17	El sistema debe permitir configurar en un perfil de búsqueda un código de activación.
SPRINT 3	
RF_08	El sistema debe permitir configurar en un perfil de búsqueda.
RF_09	El sistema debe permitir el ingreso de información sensible mediante teclado en pantalla.
RF_10	El sistema debe permitir configurar en un perfil de búsqueda una lista blanca de dominios y subdominios sobre los que se realizará búsquedas.
RF_11	El sistema debe permitir configurar en un perfil de búsqueda parámetros para el control del volumen máximo de información que se almacenará.
RF_12	El sistema debe permitir configurar en un perfil de búsqueda una lista blanca de extensiones de archivos (JPG, PDF, TXT, etc.) sobre los que se realizara búsquedas.
RF_13	El sistema debe permitir configurar en un perfil de búsqueda una lista blanca de años (2015, 2017, 2019, etc.) según los cuales el sistema realizara búsquedas.
RF_14	El sistema debe permitir configurar en un perfil de búsqueda una lista blanca de idiomas según los cuales el sistema realizará búsquedas.
RF_15	El sistema debe permitir configurar en un perfil de búsqueda un parámetro para el control del número máximo de búsquedas se realizará.
RF_16	El sistema debe permitir configurar en un perfil de búsqueda un rango de tiempo(fechas) de activación.
RF_17	El sistema debe permitir configurar en un perfil de búsqueda un código de activación.
PRINT 4	
RF_29	El sistema debe implementar registro de información de eventos relevantes para auditorias.
RNF_02	El sistema debe mostrar un mensaje de error que muestra la descripción del evento en caso de presentarse alguna excepción.

RNF_03	El sistema debe implementar ToolTips de ayuda para evitar que el usuario ingrese datos errados.
RNF_04	El sistema debe pedir confirmación al usuario antes de guardar los cambios en el sistema.
RNF_05	El sistema registrará un log de los cambios realizados, detallando el módulo, el tipo de movimiento, el usuario que ejecutó la transacción, hora y fecha.
SPRINT 5	
RF_03	El sistema debe permitir resetear la clave de un usuario.
RF_26	El sistema debe almacenar datos y metadatos acerca de las búsquedas realizadas
RF_28	El sistema implementar encriptación sobre los datos recolectados en las búsquedas.
RF_30	El sistema debe permitir generar reportes en base al nivel de acceso de usuario y sensibilidad de información.
RF_31	El sistema debe identificar y etiquetar la información almacenada de acuerdo a un nivel de sensibilidad (publica, confidencial, etc.)
RF_32	El sistema debe permitir revocar o restringir acceso a la información haciendo efectivos los periodos de tiempo asignados.
RF_33	El sistema debe permitir eliminar información recolectada asociada a un proceso de investigación-búsqueda.
RF_34	El sistema debe generar reportes con información que refleje el desempeño y responsabilidad de los actores del proceso de investigación.

4.1.4. Modelo arquitectónico

La arquitectura seleccionada para el prototipo de preservación de la privacidad está definida como cliente-servidor. Ya que se concibe como una aplicación web que es desplegada sobre un servidor de aplicaciones el cual sirve peticiones generadas por los clientes desde un navegador web. A continuación, mediante la Figura 15, se presenta una arquitectura cliente servidor típica.

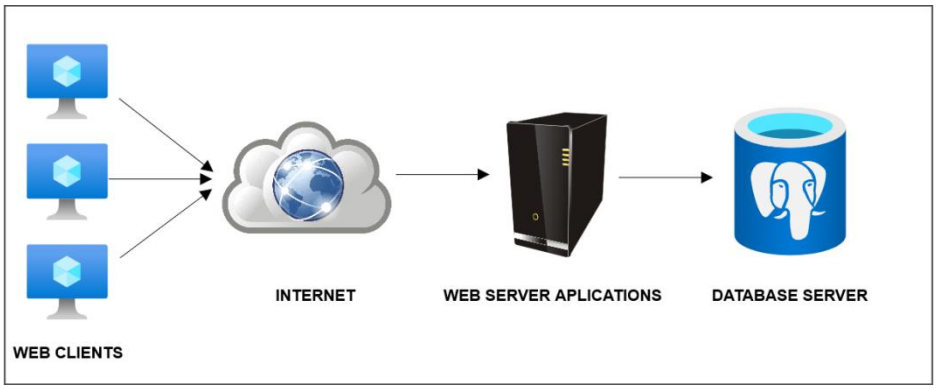


Figura 15 – Modelo arquitectónico. **[Elaboración propia]**

- **Cliente:** En primera instancia está representado por el usuario final quien demanda los servicios del sistema a través de un navegador web el cual realiza peticiones al servidor de aplicaciones.
- **Internet:** Es la red mundial global conformada por un sin número de clientes, servidores y bases de datos que permite la interconexión entre sistemas. Es a través de esta red y los protocolos necesarios que un cliente puede comunicarse con un servidor.
- **Web Server Aplications:** Es el encargado de atender las peticiones de los clientes sobre los recursos que administra. Tiene la capacidad de recibir procesar y enviar los recursos que el cliente necesite y puede estar conformado por un equipo de computación específico o un recurso físico.
- **Data Base Server:** Es el encargado de gestionar los datos que almacenan las aplicaciones, proporcionando toda la data que estas requieran para llevar a cabo sus procesos y proporcionar servicios.

4.2. Desarrollo del prototipo

Para el desarrollo del prototipo es necesario definir sus componentes e interacción mediante ciertas notaciones que permitan explicar sus roles y funcionalidades dentro del prototipo funcional que se plantea. Por tal motivo a continuación se utilizan varias notaciones y diagramas que permiten la comprensión del prototipo y sus componentes.

4.2.1. Diagrama de componentes

Como se puede ver en la Figura 16 se halla un diagrama el cual refleja las relaciones entre los componentes individuales del sistema mediante una vista de diseño estática.

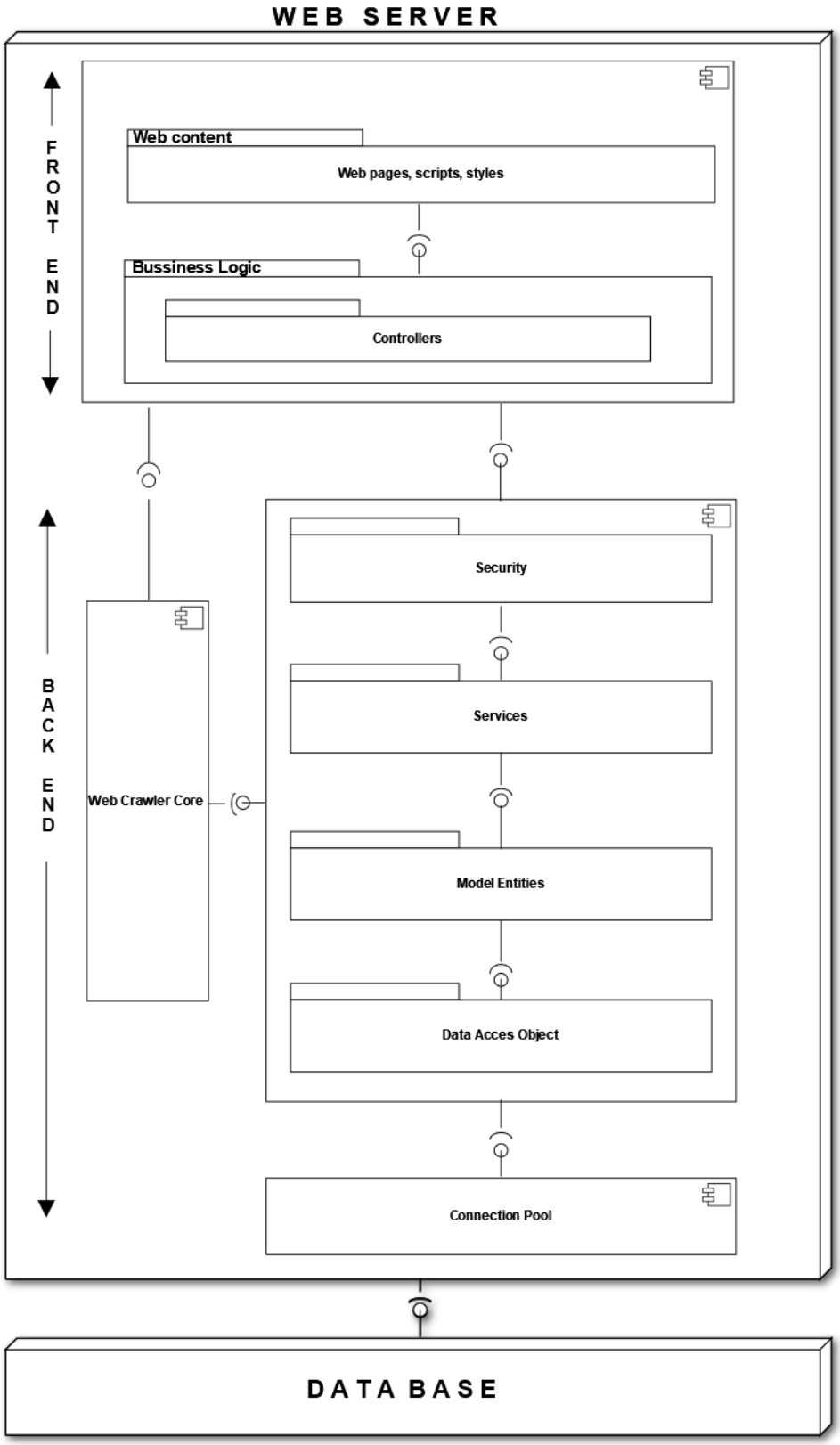


Figura 16- Diagrama de componentes del sistema. **[Elaboración propia]**

A continuación, se describe cada uno de los componentes del diagrama.

- **Web Content:** Representa las páginas HTML, estilos y scripts del lado del cliente (JavaScript) dentro del modelo MVC representa las vistas.
- **Bussiness Logic:** Representa la lógica de negocio del lado del servidor que soporta a las peticiones de las vistas, dentro del modelo MVC representa los controladores.
- **Security:** Representa las clases y filtros necesarios para implementar los servicios de seguridad para control de acceso.
- **Services:** Representa los servicios disponibles que implementa la aplicación que sirven de soporte a la lógica de negocio disponibles como una API.
- **Data Access Object:** Representa el conjunto de interfaces y clases genéricas encargadas de gestionar las consultas realizadas a la base de datos.
- **Model:** Representa las tablas de la base de datos y sus relaciones mapeadas como clases.
- **Connection Pool:** Representa un conjunto limitado de conexiones hacia una base de datos que es manejado por un servidor de aplicaciones.
- **Web Crawler Core:** Representa el core de un WebCrawler como una dependencia que implementa una API mediante la cual se lo puede integrar en otras aplicaciones.
- **DataBase:** Representa la base de datos que utiliza la aplicación para almacenar y recuperar información necesaria.

4.2.2. Funcionalidades, usuarios y requerimientos

Los usuarios y funcionalidades definidas para el web crawler se describen en las tablas Tabla 26 y Tabla 27 respectivamente.

Tabla 26 – Usuarios del sistema. **[Elaboración propia]**

Tipo de usuario	Descripción
Administrador	Usuario encargado de gestionar el módulo de control de acceso que consiste en gestión de usuarios, roles, recursos y perfiles.
Facilitador	Usuario encargado de configurar perfiles de búsqueda y asociar usuarios a estos.
Investigador	Usuario que realiza las búsquedas dentro de un perfil de búsqueda asignado.

Tabla 27 – Módulos del sistema. **[Elaboración propia]**

Módulo	Descripción	Usuario
Autenticación de usuarios	Módulo que realiza la autenticación de usuarios mediante id y contraseña.	Administrador Facilitador Investigador
Gestión de control de acceso	Módulo que permite a un usuario administrador gestionar usuarios, roles, recursos y perfiles basado en un modelo RBCA.	Administrador
Configuración y asignación de perfil de búsqueda	Módulo que permite a un usuario facilitador configurar un perfil de búsqueda, asignarlo a un usuario investigador y definir políticas que se aplicarán a sus búsquedas.	Facilitador
Búsquedas en el Web Crawler	Módulo que permite a un usuario investigador realizar búsquedas dentro de un perfil de búsqueda asignado aplicando las políticas configuradas.	Investigador

4.2.3. Diagramas de secuencia y algoritmos

Dentro de la funcionalidad del prototipo existen procesos que requieren mayor atención en cuanto a diseño e implementación ya que representan requerimientos enfocados a la preservación de la privacidad, lo cual es característica principal de este trabajo y que en este apartado se describen mediante diagramas de secuencia y algoritmos escritos en pseudocódigo.

4.2.3.1. Solicitud genérica de un recurso

La dinámica de una aplicación web está basada en solicitudes de recursos desde un navegador hacia a un servidor, el mismo que atiende estas solicitudes entregando el recurso solicitado. Aquí cabe el análisis de una validación de seguridad la cual la mayoría de aplicaciones implementa y es la autenticación de usuario y posterior inicio de sesión, la misma que le permitirá al usuario realizar peticiones válidas hacia el servidor. Es decir, tras iniciar sesión el servidor es capaz de entregar cualquier recurso que el usuario solicite y este pueda resolver ya que es su trabajo. Sin embargo, aquí claramente puede darse un escenario de acceso no autorizado hacia un recurso generado con información sensible violando así la privacidad de los titulares de los datos entregados.

En base a lo expuesto y considerando las características en cuanto a preservación de la privacidad que el prototipo debe implementar, se ha diseñado un flujo para el manejo de solicitudes por parte de la aplicación ya que en última instancia es quien recupera y procesa los datos de respuesta.

Para esto se implementó un filtro de seguridad lógico que verifica los privilegios que un usuario posee sobre un recurso solicitado mediante una consulta hacia un componente de control de acceso. Esta verificación se realizará en todas las peticiones recibidas, sin excepción, logrando así mantener el control sobre un recurso solicitado todo el tiempo.

El diagrama de secuencia ilustrado en la Figura 17 define el flujo y los objetos que intervienen en el proceso.

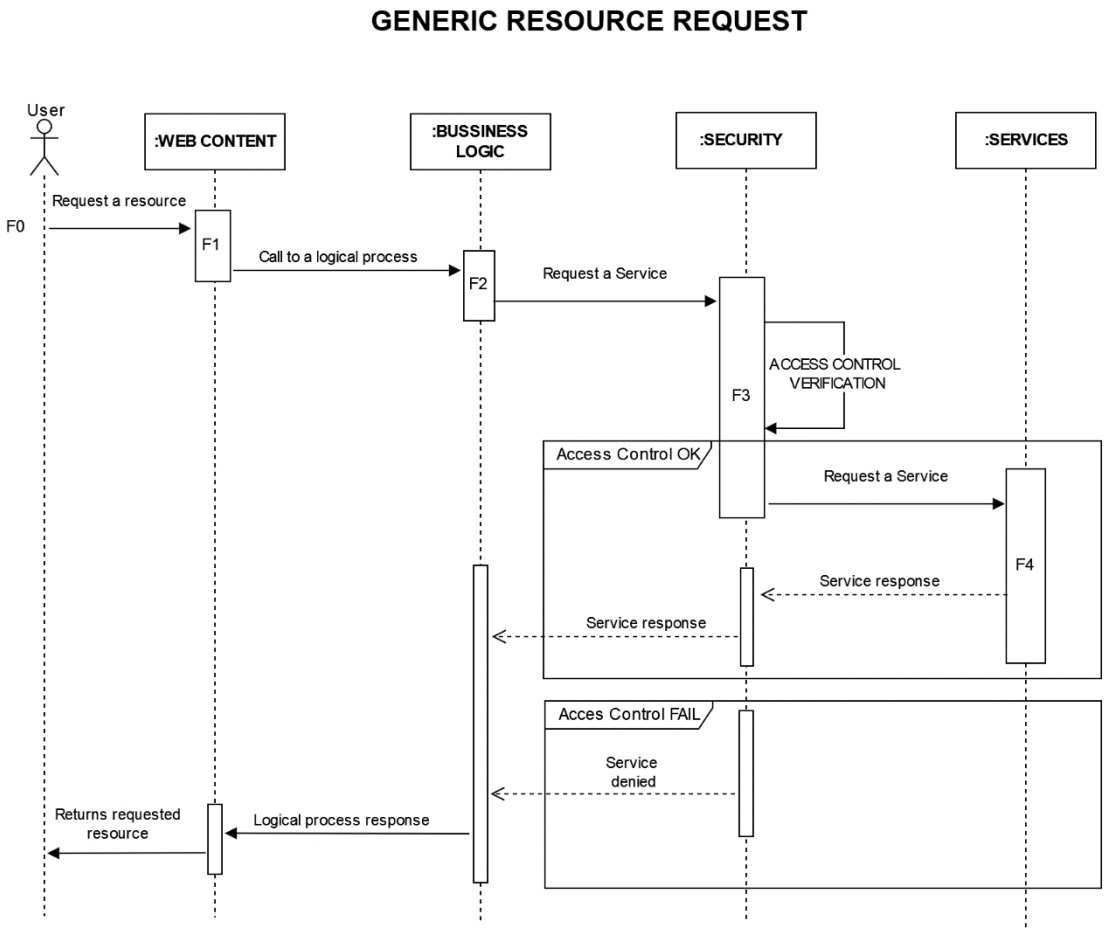


Figura 17- Solicitud genérica de un recurso. **[Elaboración propia]**

Los algoritmos implementados para este proceso se describen a continuación:

Tabla 28- Algoritmos de solicitud genérica de un recurso. **[Elaboración propia]**

USER	
F0	function mainCall() parameters[user] <- userIdValue parameters[rol] <- rolValue parameters[codResource] <- codResourceValue result <- WEB_CONTENT.getSource(parameters) PRINT "ACCESS FOR THE SOURCES IS :" + result endfunction
WEB CONTENT	
F1	function getSource(array parameters) result <- BUSSINESS_LOGIC.proccesAndGetSource(parameters) returnresult endfunction
BUSSINES LOGIC	
F2	function proccesAndGetSource(array parameters) result <-null doSomeLogicProccess(parameters) autorization_result<- SECURITY.validateResourceAutorization(parameters); IF autorization_result = OK: result <- SERVICES.getService(parameters) returnresult endfunction
SECURITY	
F3	function validateResourceAutorization(array parameters) result <- null flag <- ACCES_CONTROL.checkAutorization(parameters) returnflag endfunction
SERVICES	
F4	function getService(array parameters) doSomeLogicProccess(parameters) result <- DATA_ACCESS_JPA.executeQuery(JPQL,parameters) returnresult

	endfunction

4.2.3.2. Control de acceso

Una de las implementaciones más efectivas usadas para gestionar la autorización de recursos son los sistemas de control de acceso, ya que en base a un usuario previamente autentificado se puede autorizar en función de los privilegios que posea si tiene o no acceso a un recurso solicitado.

Continuando con la caracterización que el prototipo debe tener, la implementación de un módulo para el control de acceso también está considerado dentro del diseño, siendo un componente que centraliza la autorización de un recurso solicitado, toma un papel fundamental en el sentido de preservación de la privacidad que se busca con este trabajo. Con esta implementación las necesidades que tienen los distintos componentes de verificar los privilegios de un usuario quedan cubiertos. El diagrama de secuencia ilustrado en la Figura 18 define el flujo y los objetos que intervienen en el proceso.

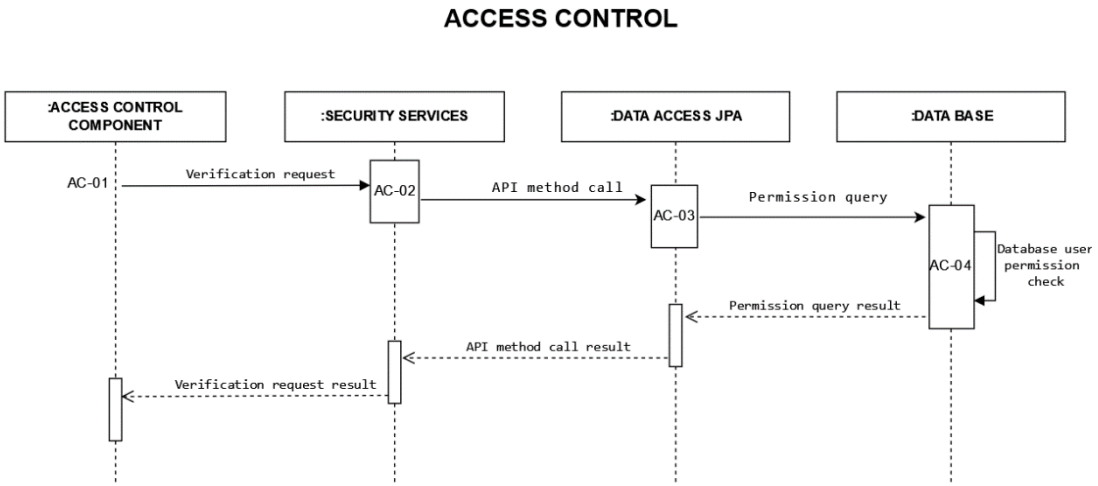


Figura 18 - Control de acceso. **[Elaboración propia]**

Los algoritmos implementados para este proceso se describen a continuación.

Tabla 29 - Algoritmos de control de acceso. **[Elaboración propia]**

ACCESS CONTROL COMPONENT	
AC-01	function checkAutorization() parameters[usuario] <- userIdValue parameters[rol] <- rolValue parameters[codResource] <- codResourceValue result <- SECURITY_SERVICES.autorizationService(parameters) PRINT "ACCESS FOR THE SOURCES IS :" + result endfunction
SECURITY SERVICES	
AC-02	function autorizationService(array parameters) result <- DATA_ACCESS_JPA.executeQuery(JPQL,parameters) returnresult endfunction
DATA_ACCESS_JPA	
AC-03	function executeQuery(array parameters) result <- CONECTION.executeNativeQuery(SQL,parameters) returnresult endfunction
DATABASE	
AC-04	function validationDBMS(array parameters) result <- internalMethodUserValidation() returnresult endfunction

4.2.3.3. Solicitud de búsqueda al web crawler

Como ya se ha expuesto en el primer capítulo acerca de la problemática que busca resolver este trabajo y la propuesta de implementar un web crawler con características de seguridad, en este apartado se describe su proceso.

Una vez autorizada la acción mediante el control de acceso se hace la invocación al crawler quien recibe una cadena de búsqueda y una lista de restricciones definidas por el administrador, las mismas que configuran el motor de búsqueda y restringirán el almacenamiento y análisis de ciertos sitios web, cabe mencionar que estas restricciones son específicas para cada perfil de búsqueda esto a discreción de administrador.

Una vez configurada la búsqueda, ésta se ejecuta, y de la lista de sitios web que va obteniendo decide cuales debería visitar en función de restricciones tales como dominios permitidos, tamaño de archivos, formatos de archivos, etc.

Posteriormente recupera los sitios web seleccionados para analizar a detalle su contenido y coincidencias con la cadena de búsqueda o expresión regular proporcionada por el usuario, si existen coincidencias pasa a cifrar y almacenar la información del sitio web caso contrario ignora el sitio. Todo este proceso se ejecuta hasta alcanzar el número de búsquedas o cantidad de almacenamiento máximo configurado para la búsqueda en cuestión. Posteriormente retorna un resumen de la búsqueda y más tarde realiza acciones de almacenamiento masivo en segundo plano de acuerdo a lo configurado. El diagrama de secuencia ilustrado en la Figura 19 define el flujo y los objetos que intervienen en el proceso.

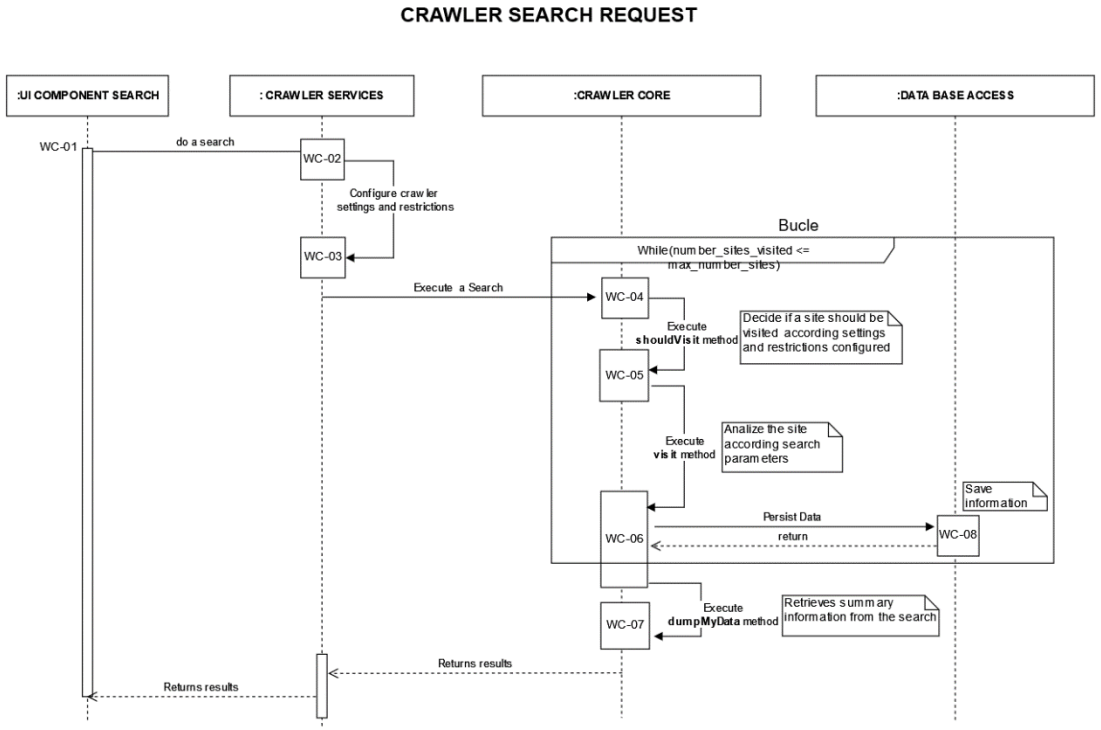


Figura 19- Solicitud de búsqueda al crawler. **[Elaboración propia]**

Los algoritmos implementados para este proceso se describen a continuación

Tabla 30- Algoritmos de solicitud de búsqueda al crawler. **[Elaboración propia]**

UI COMPONET SEARCH	
WC-01	function searchRequest() parameters[searhParams] <- searchParamsValue parameters[usuario] <- userIdValue parameters[rol] <- rolValue parameters[codResource] <- codResourceValue result <- CRAWLER_SERVICES.doSearch(parameters) PRINT "SHOW RESULT:" + result end function
CRAWLER SERVICES	
WC-02	function doSearch(array parameters) CRAWLER_SERVICES.configCrawlerSettingsRestrictions(parameters) CRAWLER_CORE.executeSearch() end function
CRAWLER SERVICES	
WC-03	function configCrawlerSettingsRestrictions(JPQL,parameters) GlobalCrawlerInstance<- newCrawler() GlobalCrawlerInstance<- setConfiguration(parameters) GlobalCrawlerInstance<- setRestriction(parameters) endfunction
CRAWLER CORE	
WC-04	function executeSearch(array parameters) numSitesVisited<- 0 shouldVisit<- false WHILE numSitesVisited<= crawlerConfig.MAX_SITES shouldVisit<- shouldVisit() if(shouldVisit = true): visit() numSitesVisited<- numSitesVisited + 1 return dumpMyData() endfunction
WC-05	function shouldVisit(url site) IF(crawlerConfig.domainsAllowed IN site.url AND crawlerConfig.extensionsAllowed IN site.extension):

	return TRUE ELSE: return FALSE end function
WC-06	function visit(url site) IF(stringSearch IN site.content OR regularExpression IN site.content): DATA_BASE_ACCESS.saveData(content,url,etc) end function
WC-07	function dumpMyData() globalSummaryInformation<- getSummary() return globalSummaryInformation endfunction
DATA BASE ACCESS	
WC-08	function saveData(array parameters) result<-CRUD_SERVICES.persistData(parameters) returnresult endfunction

4.2.3.4. Generación de expresiones regulares

Como se explica en el tercer capítulo, se implementa la construcción dinámica de expresiones regulares para permitir al investigador ser más específico en sus criterios de búsqueda y así lograr mejores resultados en sus búsquedas, trabajando en este sentido se implementó esta funcionalidad generada principalmente en interacciones del usuario con componentes front end que a bajo nivel representan intenciones y parámetros de búsqueda que van construyendo la expresión regular a buscar y que es enviada al crawler. El diagrama de secuencia ilustrado en la Figura 20 define el flujo y los objetos que intervienen en el proceso.

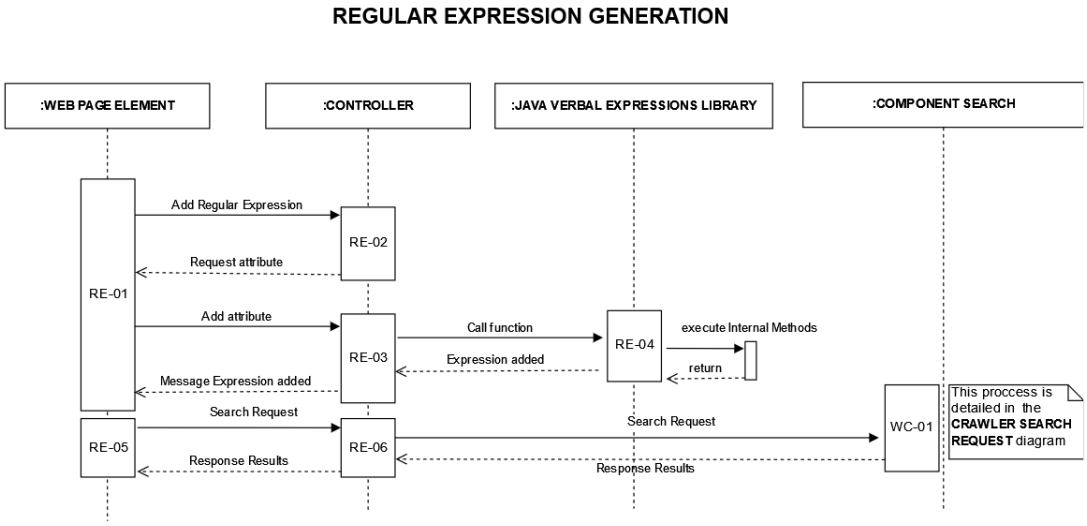


Figura 20- Generación de expresiones regulares. **[Elaboración propia]**

Los algoritmos implementados para este proceso se describen a continuación.

Tabla 31 – Algoritmos de generación de expresiones regulares. **[Elaboración propia]**

WEB PAGE ELEMENT	
RE-01	function configAdvancedSearchUI(Event ev) requiredAttribute<- CONTROLLER.addExpression(ev) attribute <- readAttributeFromUser(ev) messageExpression<- CONTROLLER.addAttribute(attribute) endfunction
RE-05	function requestAdvancedSearchUI(Event ev) responseResult<- CONTROLLER.searchRequest(ev) showResultUI(responseResult) endfunction
CONTROLLER	
RE-02	function addExpression(Event ev) globalExpression.add(ev.nameExpression) return globalExpression.attributeRequired end function
RE-03	function addAttribute(Object attribute) globalExpression<- JAVA_VERBAL_EXPRESSION_LIBRARY.

	getExpression(globalExpression(attribute)) return "EXPRESSION ADDED:" + globalExpression endfunction
RE-06	function requestAdvancedSearch(array parameters) responseResult<- COMPONENT_SEARCH.executeSearch(parameters) returnresponseResult endfunction
JAVA VERBAL EXPRESSIONS LIBRARY	
RE-04	function getExpression(Object expression) executeInternalMethods(expression) return "EXPRESSION ADDED:" + globalExpression end function

5. EVALUACIÓN Y RECOLECCIÓN DE RESULTADOS

Esta sección está dedicada a la evaluación del método de preservación de privacidad, para lo cual en primera instancia se plantea llevar a cabo un caso de estudio realizando búsquedas a través del aplicativo y recolectando evidencia existente. Otra evaluación se enfoca en el desempeño del aplicativo en cuanto al consumo y utilización de recursos computacionales sobre las búsquedas que realiza el aplicativo. La resiliencia es otro nivel de evaluación que se lleva a cabo, para esto se desarrolla un modelo de ataque y amenazas basado en un adversario potencial del aplicativo, con el fin de definir las mayores amenazas, defensas más óptimas y requerimientos críticos del aplicativo. Posteriormente se pone a prueba la resiliencia del aplicativo ejecutando escenarios de ataque sobre las funcionalidades críticas. Finalmente, se discuten los resultados obtenidos de las evaluaciones.

5.1. Propuesta de caso de estudio

El caso de estudio que se ha tomado para la evaluación de la aplicabilidad se describe a continuación.

Contexto

Ha ocurrido un robo en una propiedad privada de la cual se han sustraído varios artículos (discos de Vinyl, Cds, cassettes de audio, etc.) los cuales tienen un alto valor comercial dado que son piezas de colección. El propietario de los artículos ha puesto en conocimiento de la policía lo sucedido quienes según las versiones tomadas e interrogatorios realizados han logrado obtener información relevante que podría servir para definir líneas de investigación.

¿En dónde sucedió?

Sucedió en la parroquia Itchimbia, ciudad de Quito provincia de Pichincha.

¿Cuándo sucedió, a qué hora?

Sucedió el sábado 18 de septiembre del 2022 entre la 01:00 HH y 02:30 HH aproximadamente.

¿Se conoce algún detalle de los artículos sustraídos?

La víctima del suceso cuenta con una gran cantidad de piezas musicales de colección sin embargo se sustrajeron únicamente las siguientes:

- LP KISS - TheBestof Vinyl 7' Julio Jaramillo - Fatalidad (Primera edición, autografiado con dedicatoria)
- LP Pink Floyd, The Dark Side of The Moon (Primera edición, sello Warners, incluye figura coleccionable)
- LP Radiohead, Kid A (Prensaje de 1988, sello SONY)
- LP Michael Jackson, Thriller (Primera edición, autografiado)
- Cassette Bob Dylan, Blood on The Tracks (Primera edicion, cinta de cromo)
- Cassette The Beatles, Sgt. Pepper's Lonely Hearts Club Band (Disco Picture, autografiado)
- CD Metallica, ...And JusticeForAll (Primera edición, autografiado, incluye poster A3)
- CD The Rolling Stones, StickyFingers (Primera edición, sello ICARVS, incluye figura colecionable)

¿La víctima tiene indicios de algún posible sospechoso?

Según sus versiones comenta que no muchas personas conocían de todas sus piezas de colección a excepción de las siguientes:

Tabla 32 - Lista de personas sospechosos. [**Elaboración propia]**

Nombres	**NickName**	**Celular**	**Correo electrónico**
Juan Guerra	Collector 1975	0983016685	collecto34uio@hotmail.com
Lucia López	Vintage Soul		luciavintage1988@gmail.com
Mario Pineda		0954115774	

¿Existe alguna información adicional relevante?

La víctima supone que en varias tiendas físicas de segunda mano y sitios web de compra y venta del país se podrían estar ofertando estos objetos.

Tabla 33- Posibles sitios físicos y virtuales. [**Elaboración propia]**

Tiendas físicas	**Sitios Web**
- Tienda de discos La Tola (Quito) - Vintage Valle (Quito) - Gabys Rock (Cuenca) - El SotanoRecs.(Ambato)	- Only Analog is real - Coleccionistas vinyl Ecuador - Amor vinyl Quito - Discogs - Mercado libre

	- OLX - Amazon - Ebay

5.2. Evaluación funcional

En esta sección se realizan las evaluaciones al modelo propuesto dentro de las fases de adquisición, almacenamiento y preservación de la información recolectada en un proceso OSINT basado en el caso de estudio descrito anteriormente.

5.2.1. Fase de Adquisición

De acuerdo con el ciclo forense dado por PRECEPT y el ciclo OSINT esta fase está dedicada a la adquisición de información y dentro del método propuesto corresponde a la fase de **ADQUISICIÓN** del mapeo de fases de la Figura 11.

Previo a iniciar el proceso se requiere un insumo inicial que es el plan de búsquedas el cual es producto de la fase previa (**identificación - planificación**) la cual no está dentro del alcance del presente trabajo pero que se ha generado debido a la necesidad. A continuación, se presenta el formato de las búsquedas y un registro de ejemplo, en el Anexo V se puede visualizar la tabla de búsquedas completa que se ha generado para el caso de estudio.

Tabla 34 – Formato de búsquedas. [**Elaboración propia]**

#	**Cadena de Búsqueda**	**Dominio de búsqueda**	**Filtros**
S01	Julio Jaramillo Ecuador	https://ec.ebay.com/	**Formato:** html,png,jpg; **Idioma:** Español **Vol.Max:** 0.5 GB **Max.Inst:** 500 **Max.Pag:** 1500

Una vez definidos los criterios para las búsquedas se inicia el proceso de evaluación basado en los lineamientos de privacidad definidos por PRECEPT y OSINT.

Como se pudo ver en la revisión de técnicas y enfoques OSINT en el apartado 2.2 del presente documento estas búsquedas se las realiza en internet mediante buscadores

más o menos especializados en alguna temática especifica aplicando filtros y cadenas de búsquedas, sin embargo, adquirir información de esta manera tiene ciertos problemas en cuanto a privacidad como, por ejemplo:

- Fuentes de información no controlados.
- Formatos de información.
- Historial de búsqueda.
- Volumen de información.

El método propuesto aborda estos problemas y propone soluciones basadas en principios de privacidad que se incorporan en cada fase tal como se ve en la Figura 11 y su descripción en la Tabla 18 .

Tabla 35 – Lineamientos de privacidad fase de adquisición. [**Elaboración propia]**

LINEAMIENTOS DE PRIVACIDAD – FASE DE ADQUISICIÓN
Definir las fuentes de información de acuerdo al ámbito establecido para la investigación
Definir el volumen de Información total máximo a ser almacenado.
Definir los formatos (jpg ,mp4, pdf,etc.) de archivos que serán considerados y que se almacenarán.
Definir los años (2012, 2013, 2018,etc.) que serán considerados para realizar las búsquedas.
Definir una lista de idiomas (Inglés, Español, etc) que serán considerados para realizar las búsquedas.
Definir un parámetro que indique le número máximo de búsquedas que serán permitidas realizar.
Definir criterios en cuanto historiales de las búsquedas y quien las realizó.

- **Definir las fuentes de información de acuerdo al ámbito establecido para la investigación:** En un buscador convencional los datos adquiridos provienen de cualquier fuente, esto no es deseable desde el punto de vista de privacidad ya se puede estar adquiriendo información sensible o personal de fuentes que no se desea.

  Para solucionar esto el método propuesto define un perfil de búsqueda en el cual se configura los orígenes de datos o dominios de búsqueda permitidos para un

determinado Investigador. Es decir, las búsquedas que un Investigador realice con el crawler serán únicamente de los dominios de búsquedas configurados previamente por el usuario Configurador de Búsquedas, de esta manera se intenta preservar la privacidad restringiendo las búsquedas en dominios establecidos y restringiendo los que estén fuera de la lista.

En el contexto del caso de estudio esto se ve reflejado mediante la configuración de un perfil de búsqueda en el cual se definen los dominios de búsqueda para el campo **DOMINIOS DE BUSQUEDA** de acuerdo con la búsqueda **S01**, la Tabla 34 y Tabla 35 .

- **Definir el volumen de Información y máximo de búsquedas permitidas:** Los buscadores convencionales generan un volumen de información abrumadora, y esto no es bueno ni malo ya que puede ser interpretado de muchas formas. Sin embargo, en lo que concierne a prácticas OSINT y privacidad, a mayor data mayor la probabilidad de incurrir en la adquisición de datos personales o privados innecesarios para la investigación, ya que las actuales herramientas de análisis masivo de datos generan mayor información de la que inicialmente se pretendía conocer, siendo así el volumen de datos algo contraproducente en aspectos de privacidad y capacidad computacional.

  Para solucionar esto el método propuesto define un perfil de búsqueda en el cual se configura el volumen máximo de almacenamiento y el número máximo de búsquedas para cada perfil de búsqueda. Es decir, las búsquedas que un Investigador realice con el crawler estarán limitadas en cuanto volumen de almacenamiento y número de búsquedas, esto es configurado por el usuario Configurador de Búsquedas, de esta manera se intenta preservar la privacidad restringiendo el volumen de data disponible para posteriores análisis.

  En el contexto del caso de estudio esto se ve reflejado mediante la configuración de un perfil de búsqueda en el cual se define el volumen máximo de almacenamiento y el número máximo de búsquedas para los campos **NUMN.MAX.INSTANCIAS**, **NUM.MAX.BUSQ.INSTANCIA** y **TAMAÑO MAX ALMACENAMIENTO** de acuerdo a la búsqueda **S01,** la Tabla 34 y Tabla 35 .

- **Definir los formatos (jpg, mp4, pdf, etc.) de archivos que serán considerados y que se almacenarán:** Mientras más formatos de información sea capaz de adquirir un buscador de mejor utilidad se vuelve para el investigador y esto es relativo, ya que, desde el punto de vista de privacidad no se requiere que el investigador sea capaz de acceder a todos los recursos (fotos, imágenes, videos, etc.) relacionados con la búsqueda cuando solo se buscaba una simple entrada en un blog.

  Para solucionar esto el método propuesto define un perfil de búsqueda en el cual se configura una lista de formatos de información permitidos para cada perfil de búsqueda. Es decir, las búsquedas que un Investigador realice con el crawler estarán limitadas a adquirir información únicamente en los formatos establecidos, esto es configurado por el usuario Configurador de Búsquedas, de esta manera se busca preservar la privacidad restringiendo el acceso a información en formatos no establecidos y reduciendo el volumen de información.

  En el contexto del caso de estudio esto se ve reflejado mediante la configuración de un perfil de búsqueda en el cual se define el formato de información mediante el campo **EXTENSION DE ARCHIVO** de acuerdo a la búsqueda **S01**, la Tabla 35 y la Figura 21.

- **Definir criterios en cuanto a historiales de las búsquedas y quien las realizó:** La adquisición de información en internet es libre, en un proceso OSINT funciona igual y el investigador decide qué, donde, y como buscar en este proceso no se lleva un historial de búsquedas que registre estos eventos y a posteriori permita conocer que búsquedas realizo, donde las realizo, a qué hora, etc. De esta manera la privacidad puede verse comprometida y no existiría un registro que sirva de evidencia.

  Para solucionar esto el método propuesto define un registro de auditoría para cada búsqueda. Es decir, las búsquedas que un Investigador realice con el crawler generarán internamente un registro de auditoría con información necesaria para posteriores informes y análisis permitiendo la trazabilidad de un incidente o auditoría.

- **Definir una lista de idiomas (inglés, español, etc.) que serán considerados para realizar las búsquedas:** Como se mencionó anteriormente limitar el dominio de búsqueda es importante por las razones expuestas y no todos los buscadores permiten filtrar resultados por idioma.

  Para solucionar esto el método propuesto define un perfil de búsqueda en el cual se configura una lista de idiomas permitidos para cada perfil de búsqueda. Es decir, las búsquedas que un Investigador realice con el crawler estarán limitadas a adquirir información únicamente en los idiomas establecidos, esto es configurado por el usuario Facilitador de Búsquedas, de esta manera se busca preservar la privacidad restringiendo el acceso a información en idiomas no establecidos y reduciendo el volumen de información.

  En el contexto del caso de estudio esto se ve reflejado mediante la configuración de un perfil de búsqueda en el cual se define el formato de información mediante el campo **IDIOMA** de acuerdo a la búsqueda **S01** la Tabla 35 y la Figura 21

Figura 21 – Formulario de configuración de Perfil de Búsqueda. **[Elaboración propia]**

5.2.2. Fase de Recolección

De acuerdo con el ciclo forense dado por PRECEPT y el ciclo OSINT esta fase está dedicada al almacenamiento de información y dentro del método propuesto corresponde a la fase de **RECOLECCIÓN** del mapeo de fases de la Figura 11. En la Tabla 36 a continuación se pueden ver los lineamientos definidos.

Tabla 36 - Lineamientos de privacidad fase de Recolección. [**Elaboración propia]**

LINEAMIENTOS DE PRIVACIDAD – FASE DE RECOLECCIÓN
Verificar que la información que se almacenará provenga únicamente de fuentes del ámbito(dominios) establecido.
Verificar que la información que se almacenará no exceda el volumen máximo establecido.
Verificar que la información que se almacenará sea únicamente de los formatos establecidos.
Verificar que la información que se almacenará sea únicamente de los años publicados establecidos.
Verificar que la información que se almacenará sea únicamente de los formatos (jpg, mp4, pdf, etc.) de archivos establecidos.
Verificar que el número máximo de búsquedas no exceda el máximo establecido previo a almacenar un nuevo resultado.
Registrar datos (y metadatos) acerca de las fuentes de donde se obtiene la información.

Como se ve en la Tabla 36 los lineamientos de esta fase se complementan con los lineamientos de la fase de **RECOLECCIÓN,** pero en un sentido de cumplimiento de los filtros configurados que permiten almacenar los resultados que cumplan con estos.

En el capítulo 4 mediante la Figura 19 se ha definido el diagrama de secuencia y el pseudocódigo para las búsquedas que realiza el crawler, sin embargo, para analizar a detalle la aplicación de las restricciones para esta fase de recolección se presenta las siguientes ilustraciones.

Tabla 37 – Fases de recolección de resultados de búsqueda. [**Elaboración propia]**

#	**Fases de validación**
1	Recuperación de restricciones configuradas
2	Configuración del crawler

3	Inicio de ejecución
3.1	Validación de sitio
3.2	Visita del sitio
3.3	Almacenamiento del sitio
4	Fin de ejecución

Como se puede ver en la Tabla 37 el crawler implementa ciertas fases que permiten recolectar la información correcta de acuerdo a lo configurado para cada perfil de búsqueda. A continuación, se describe cada una de ellas.

- **Recuperación de restricciones configuradas:** En esta fase se recuperan las restricciones configuradas para el perfil de búsqueda tales como formatos de archivo, idiomas, volumen máximo de información etc. Para esto únicamente se verifica que este dentro de las fechas habilitadas.
- **Configuración del crawler:** Al crear una instancia del crawler este posee configuraciones iniciales que definen su comportamiento básico de las cuales interesan y se configuran las siguientes:

Tabla 38 – Parámetros crawler [Elaboración propia]

PARAMETRO CRAWLER	**PARAMETRO CONFIGURADO EN EL PERFIL DE BÚSQUEDA**
Nivel de profundidad de búsqueda	3
Páginas web semilla (página de la cual inicia recolectando URLs)	Lista de dominios de búsqueda
Permitir archivos binarios(multimedia)	Formatos de archivos

- **Inicio de ejecución:** En esta fase el crawler se inicia y empiezan a ejecutar sus dos fases internas de manera recursiva.
- **Validación de sitio:** Todas las url que el crawler recolecte entra en esta validación en la cual se verifica si:
    - Proviene de alguno de los dominios configurados para el perfil de búsqueda.
    - Se agrega a una lista de sitios **POR VISITAR** caso contrario se ignora.

- **Visita del sitio:** Todas las Url que cumplieron con la validación de sitio, es decir que provienen de un dominio permitido entran en esta fase y se realiza lo siguiente:
    - Se realiza la petición del contenido del sitio para analizar su contenido.
    - Se verifican otras restricciones tales como: Coincidencia con la cadena de búsqueda, Idioma, Formato de información, etc.
    - En caso de cumplir con las restricciones configuradas, se invoca asíncronamente al proceso de almacenamiento de información.
- **Almacenamiento del sitio:** En esta fase se persiste la información que ya fue validada y se realiza lo siguiente:
    - Se recibe el contenido del sitio.
    - Se encripta su contenido.
    - Se genera un registro de auditoría.
- **Fin de ejecución:** Una vez que se llega al límite de búsquedas permitidas o el volumen de información almacenada el proceso finaliza y en la interface del usuario se puede ver un resumen de la búsqueda.

De acuerdo al proceso general de búsquedas del crawler, a continuación, se muestran estas fases y su correspondencia con los lineamientos de privacidad para la fase de **RECOLECCION.** Con esto se puede evidenciar como los lineamientos de privacidad del método están siendo implementados en la funcionalidad del crawler.

Tabla 39 – Fases de validación y lineamientos de privacidad [Elaboración propia]

#	**Fases de validación**	**Lineamiento de privacidad**
1	Recuperación de restricciones configuradas	NA
2	Configuración del crawler	Verificar dominios de búsqueda
3	Inicio de ejecución	
3.1	Validación de sitio (Se obtiene solo la URL)	Verificar dominios de búsqueda Verificar formatos de información
3.2	Visita del sitio (Se obtiene el contenido del sitio)	Verificar cadena de búsqueda Verificar idioma del sitio Verificar máximo volumen de almacenamiento Verificar máximo de páginas visitadas Verificar fecha de publicación

3.3	Almacenamiento del sitio	Registro de sitio o recurso Registro de auditoría
4	Fin de ejecución	NA

En el contexto del caso de estudio esto se ve reflejado mediante la verificación de la información recolectada, los parámetros definidos para la búsqueda y el comportamiento del sistema. En este sentido a continuación, se muestran los resultados obtenidos en cuanto a los dominios visitados, número de páginas visitadas, idiomas, formatos de información y el volumen de data recolectada. Esto permite validar el cumplimiento de los requerimientos definidos para esta fase.

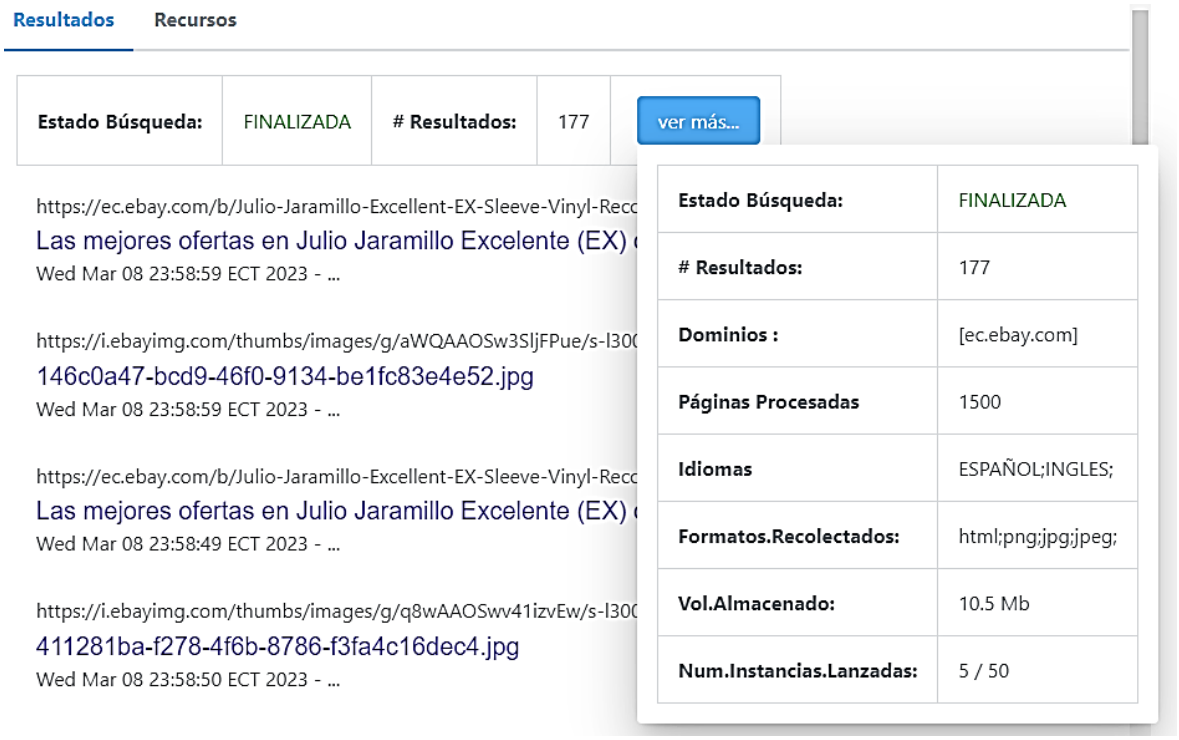


Figura 22 - Dominios de búsqueda recolectados. [Elaboración propia].

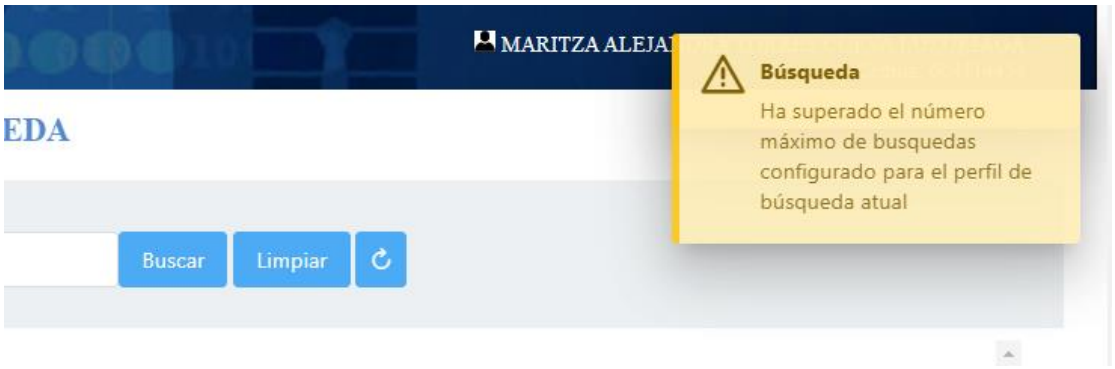


Figura 23 – Mensaje máximo de búsquedas alcanzado. [Elaboración propia]

Tabla 40 – Resultado de búsquedas fase de recolección [Elaboración propia]

PARAMETROS	**VALORES**		**ESTADO**
Dominios de búsquedas	https://ec.ebay.com	https://ec.ebay.com	**OK**
Max.volumen Almacenamiento	20MB	[10.5 MB]	**OK**
Max.Busqueda x Instancia	50	[50]	**OK**
Max.Pag.Busqueda	1500	[1500]	**OK**
Formatos Información	html, jpg, png,jpeg	[html, jpg, png,jpeg]	**OK**
Idiomas	Español, Ingles	[Español, Ingles]	**OK**
Años de publicación	2022, 2023	[2023]	**OK**

Por otro lado, la búsqueda con un navegador convencional los resultados que arroja son abrumadores por ejemplo se obtuvieron con 4.020.000 resultados de varios sitios web y recurso de todo tipo como imágenes, videos, música como se puede ver en la siguiente figura.

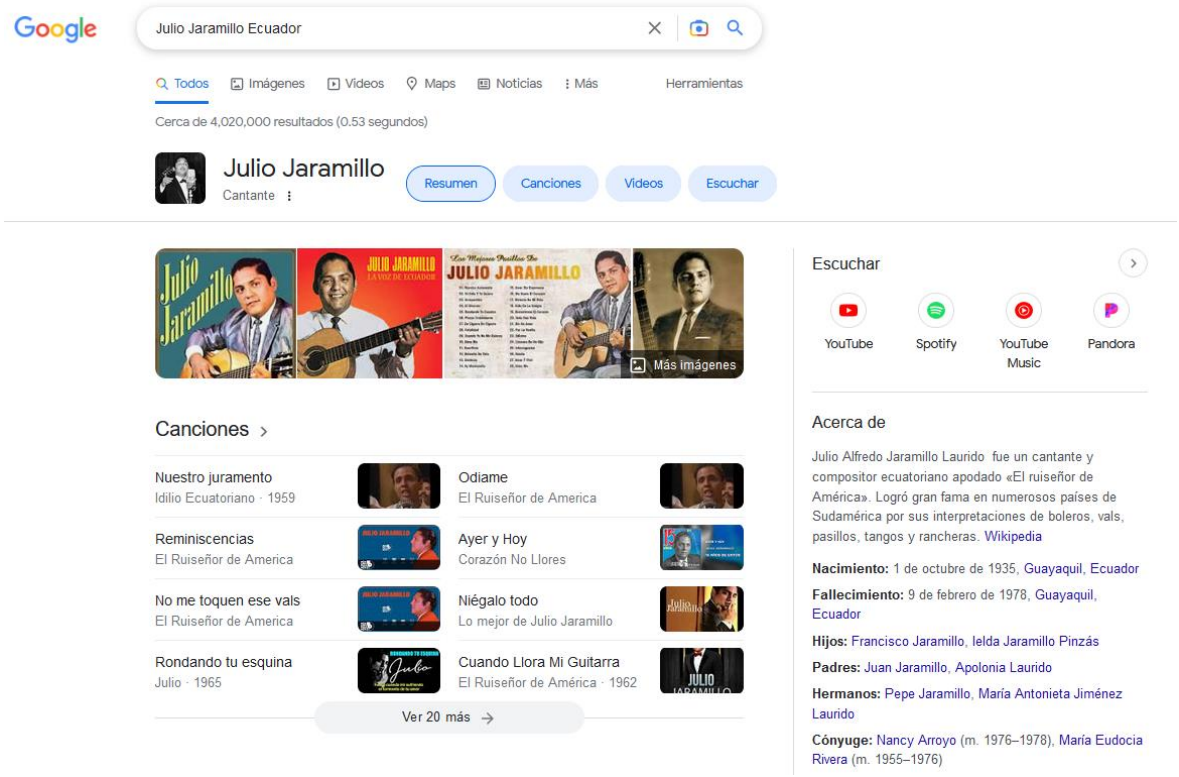


Figura 24 – Resultados búsqueda buscador convencional

5.2.3. Fase de Preservación

En un buscador convencional luego de realizar una búsqueda y ver los resultados del proceso finaliza ya que no es de mayor relevancia almacenar la información resultante, es más, en la mayoría de buscadores estos registros de búsquedas no se almacenan pues son de tipo volátil y una vez el investigador cierra el navegador finaliza todo el proceso.

Sin embargo, en un contexto forense es distinto ya que este requiere que la información obtenida se preserve correctamente ya que posteriormente podría ser presentada como evidencia en un proceso judicial. Es por este motivo que la fase de PRESERVACIÓN dentro del método propuesto toma relevancia y hace necesario implementar controles que permitan salvaguardar la privacidad de la información recolectada ante posibles escenarios como:

- Accesos no autorizados a través de la aplicación (Crawler).
- Accesos no autorizados a nivel de base de datos.

Para esto el método propuesto define los lineamientos de privacidad de la Tabla 41 los mismos que serán validados.

Tabla 41 - Lineamientos de privacidad fase de preservación. [**Elaboración propia]**

LINEAMIENTOS DE PRIVACIDAD - FASE DE PRESERVACIÓN
Implementar métodos de control de acceso
Implementar encriptación de datos sobre la data recolectada

- **Control de acceso a nivel de aplicación:** Dado que el sistema posee funcionalidades que gestionan la información recolectada de las búsquedas esta podría verse comprometida por otro usuario de la aplicación ya que estos podrían alterar o eliminar la información lo cual dentro de un proceso forense se considera como evidencia judicial y está ya no sería válida. Para resolver este problema a el método propuesto define la implementación de un método de control de acceso el cual permite autorizar las funcionalidades requeridas para cada usuario de acuerdo a sus funciones, con esto se logra:

    - Protección contra accesos no autorizados.
    - Segregación de roles y funciones.
    - Autorización centralizada de recursos del sistema.
    - Administración dinámica de usuarios y privilegios.

    Para esto el sistema implementa un módulo dedicado para control de acceso basado en roles **(RBAC)** que asigna privilegios específicos a cada rol, de manera que la estrategia para asignar permisos es definir grupos de usuarios de acuerdo a sus privilegios y posteriormente definir roles con estos privilegios e ir asignando a cada usuario un rol de acuerdo a sus responsabilidades con el sistema.

- **Control de acceso a nivel de base de datos:** La información recolectada con OSINT reposa en una base de datos a la cual también es vulnerable ya que es el camino más directo para un atacante y no es necesario vulnerar la aplicación ya que accede directamente a los datos. Es por esto que los motores de bases de datos proporcionan nativamente un sistema de control de acceso basado en roles usuario y grupos con los cuales se puede gestionar los accesos y operaciones permitidas para cada objeto de base de datos con esto se logra:

- Restringir operaciones de **SELECT, INSERT, DELETE, UPDATE** sobre cada tabla de la base de datos.
- Definir usuarios específicos para cada módulo de la aplicación
- Aplicar regla de privilegio mínimo como contramedida en caso de robo de credenciales.

- **ENCRIPTACIÓN DE DATOS:** Una vez implementadas las medidas a nivel de control de acceso con el fin de proteger la privacidad esta aún no ha sido cubierta en su totalidad, ya que, si bien es cierto solo usuarios específicos pueden acceder a ella por otro lado aún hay otras consideraciones que se deben tomar en cuenta como:
  - La información es legible para el usuario autorizado.
  - Si la base de datos es robada la información sigue siendo legible.

  Debido a esto, una segunda medida de seguridad con el fin de preservar la privacidad de la data es encriptarla. Así, quien tenga acceso directamente a la base de datos o robe la base de datos no le servirá de mucho ya que al estar encriptada los datos son ilegibles. De modo que si alguien intenta alterar la data sería necesario desencriptarla modificarla y volverla a encriptar, operación que resulta demasiado costosa si se desconoce la clave.

  En este sentido el sistema adquiere la data mediante el crawler, la encripta y almacena en la base de datos, para esta implementación se utiliza un algoritmo de clave simétrica (AES) debido a que:
  - El volumen a encriptar es alto y un algoritmo de clave simétrica no requiere un alto nivel computacional.
  - Se maneja una misma clave para encriptar y desencriptar y en el sistema esto lo llevan a cabo distintos usuarios (Configurador de Búsqueda, Investigador)

Cabe mencionar que en apartado 5.4 que corresponde a Evaluación de Resiliencia se realiza un análisis de riesgos del cual se identifican amenazas y controles de los cuales se generan escenarios de ataque que permiten evaluarlos.

5.3. Evaluación de Desempeño

Para la evaluación de desempeño se tomaron en cuenta factores que están directamente relacionados con el funcionamiento del aplicativo, y que su monitoreo permite conocer la demanda que realiza sobre estos y la medida en la que requerirían para un escenario dado. Factores como tiempo de respuesta, uso de CPU y uso de memoria se evaluarán mediante 3 escenario basados en el volumen de información que procesará el aplicativo. En la Tabla 42 se puede visualizar un resumen de lo mencionado.

Tabla 42 - Escenarios y recursos para evaluación de desempeño. **[Elaboración propia]**

Escenario	**Volumen de carga**	**Recurso a evaluar**
A (Carga Alta)	Número máximo de páginas procesadas: 2500	• Uso de memoria • Uso de CPU • Tiempo de Respuesta
B (Carga Media)	Número máximo de páginas procesadas: 1500	
C (Carga Baja)	Número máximo de páginas procesadas: 750	

Para definir los escenarios se partió de una prueba inicial en la cual se configuró al aplicativo para recibir un número elevado de páginas a procesar con el fin de conocer la configuración del sistema antes de su colapso. Tras realizar estas pruebas se conoció que el sistema colapsa por falta de recursos, específicamente de memoria, y que con los recursos que se describen en la Tabla 43 el número máximo de páginas a procesar es de 3.000.

En base a esto se definió lo que sería una carga alta para el sistema que corre sobre un servidor de las características de la Tabla 43 y una profundidad de búsqueda de nivel 2 (Figura 25).

Tabla 43 – Características servidor de aplicaciones. **[Elaboración propia]**

Recurso	**Detalle**
Procesador	Intel® Core™ i3-7020U 2.3GHz
Memoria	4 GB RAM

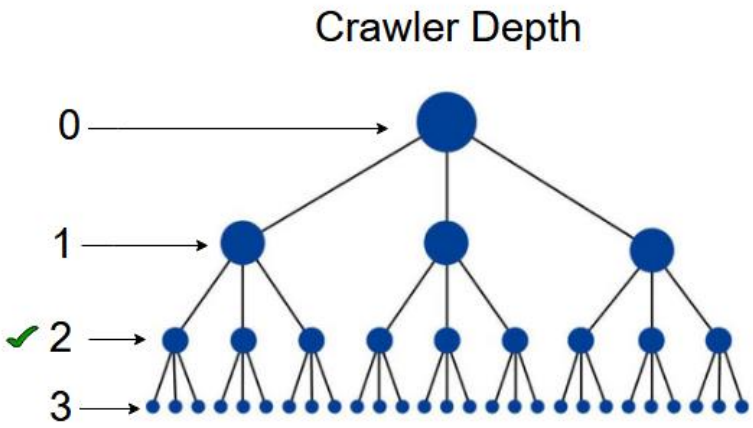


Figura 25 - Profundidad de búsqueda del crawler [Elaboración propia]

5.3.1. Ejecución del escenario A

El escenario A corresponde a una carga alta por lo cual se procede a configurar el perfil de búsqueda que define un número máximo de páginas procesar igual a 2500 como se ve en la Figura 26, y posteriormente se ejecuta la búsqueda desde el perfil de investigador.

Figura 26 – Configuración del Perfil de Búsqueda para Escenario A. **[Elaboración propia]**

Paralelamente a esto se inicia la herramienta VisualVM la cual permite monitorizar los recursos consumidos por las aplicaciones que se están ejecutando en una Máquina Virtual Java (JVM), de la cual se obtiene los gráficos de la Figura 27, Figura 28 y Figura 29 que reflejan el consumo de recursos durante el tiempo que dura el proceso de búsqueda

.

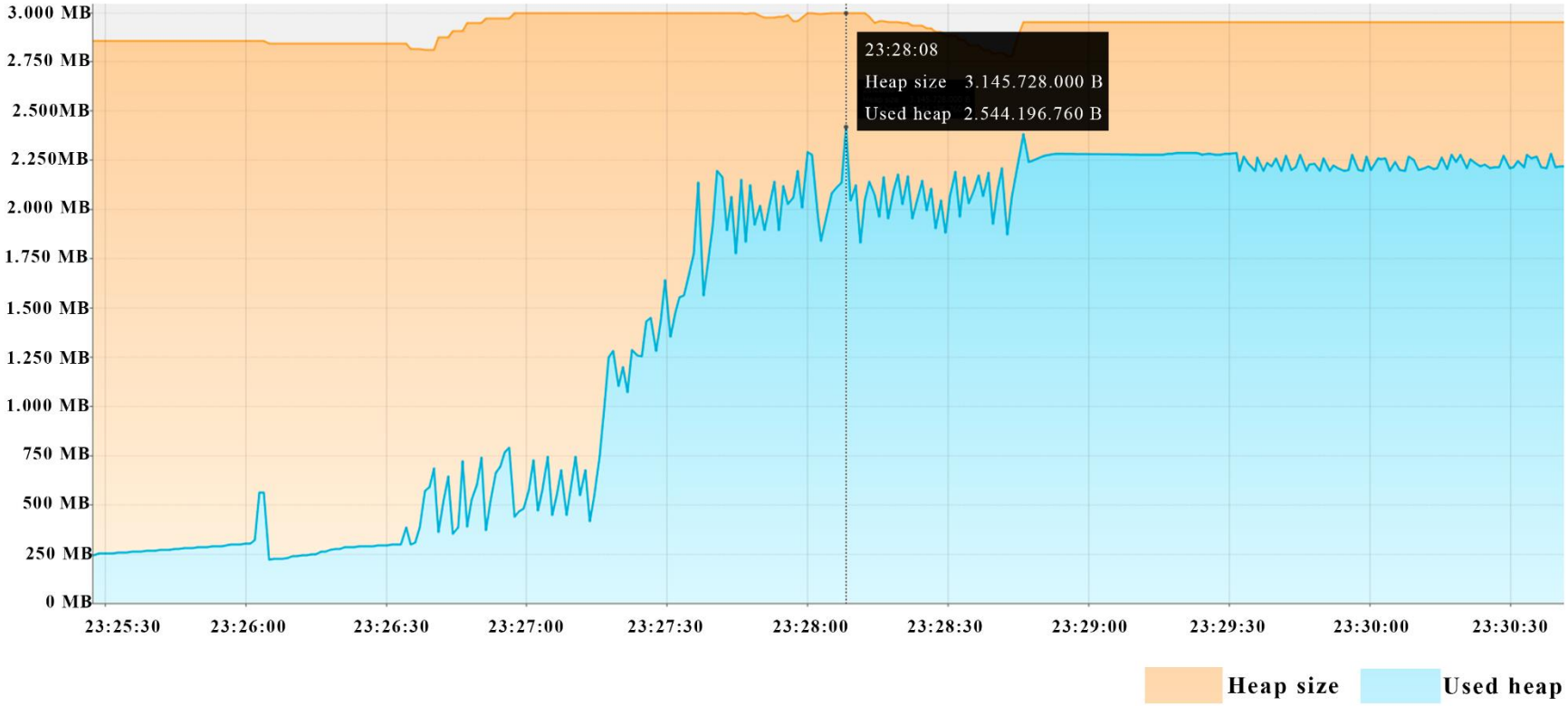


Figura 27 – Consumo de memoria Escenario A. **[Elaboración propia]**

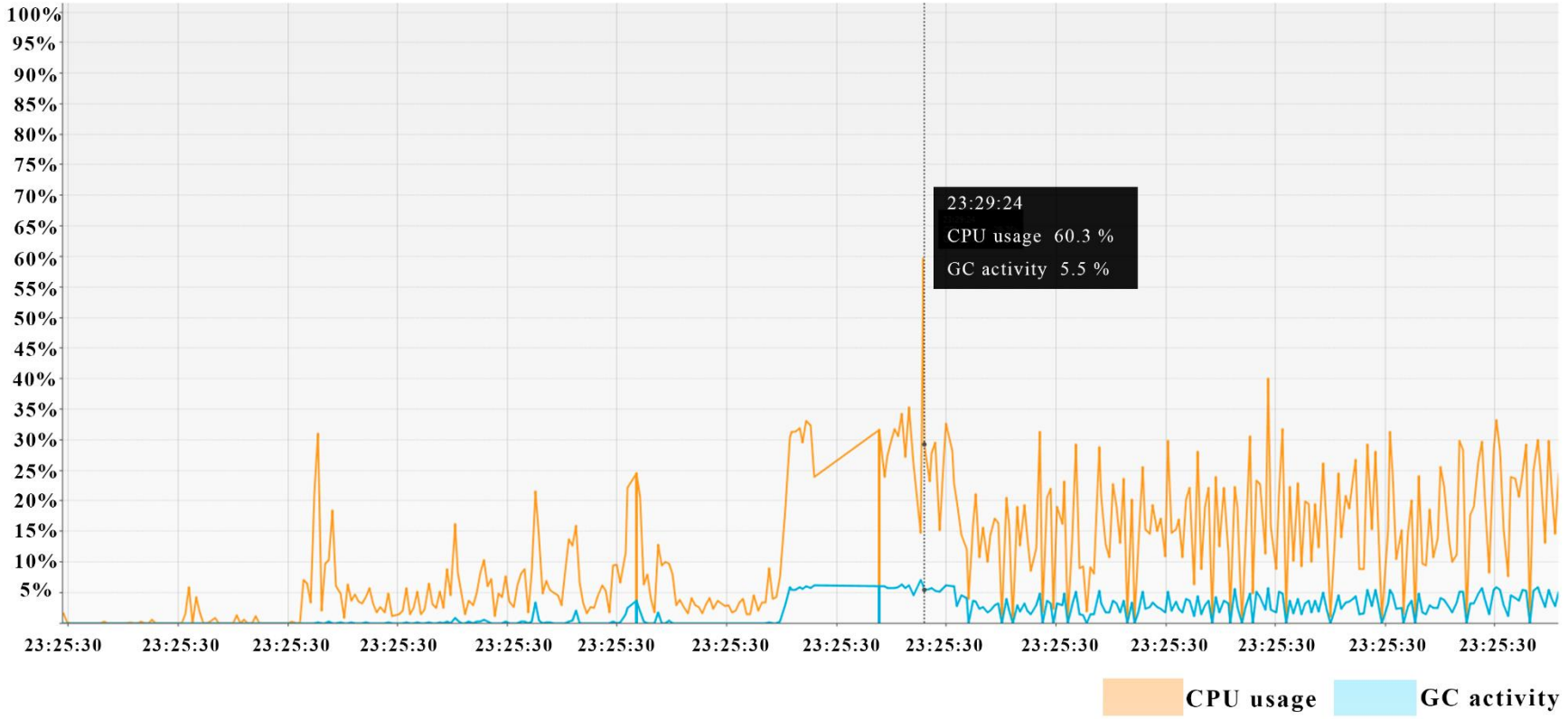


Figura 28 – Consumo de CPU Escenario A. **[Elaboración propia]**

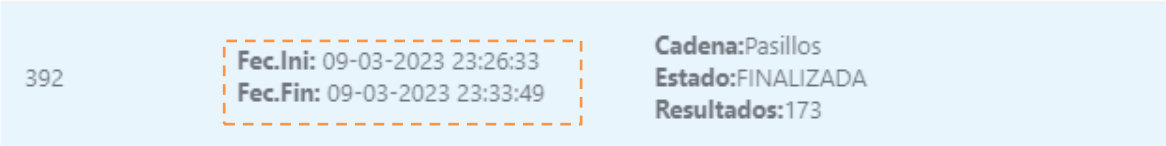


Figura 29 – Tiempo de respuesta Escenario A. **[Elaboración propia]**

Con una carga alta, de acuerdo al gráfico se obtienen los siguientes resultados.

Tabla 44 – Resultados escenario A. **[Elaboración propia]**

Escenario	**Volumen de carga**
A (Carga Alta)	Número máximo de páginas procesadas: 2500
• Uso de memoria: 2.3GB máximo. • Uso de CPU 60.3% máximo. • Tiempo de Respuesta: 00:07:16	

5.3.2. Ejecución del escenario B

El escenario B corresponde a una carga alta por lo cual se procede a configurar el perfil de búsqueda que define un número máximo de páginas procesar igual a 1500 como se ve en la Figura 30, y posteriormente se ejecuta la búsqueda desde el perfil de investigado.

Figura 30 - Configuración del Perfil de Búsqueda para Escenario B. **[Elaboración propia]**

Paralelamente a esto se inicia la herramienta **VisualVM** la cual permite monitorizar los recursos consumidos por las aplicaciones que se están ejecutando en una Máquina Virtual Java (JVM), de la cual se obtiene los gráficos de la Figura 31, Figura 32 y Figura 33 que reflejan el consumo de recursos durante el tiempo que dura el proceso de búsqueda.

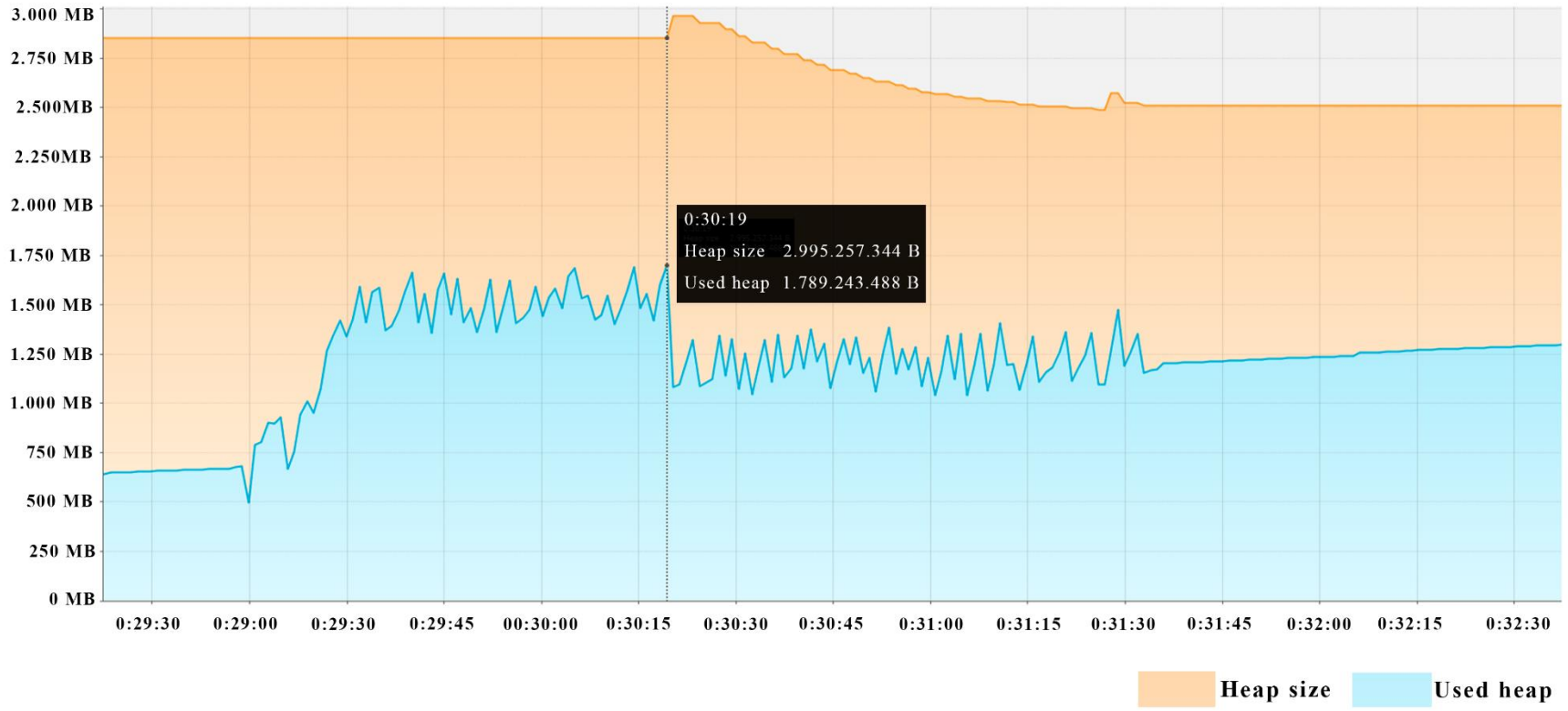


Figura 31 – Consumo de memoria Escenario B. **[Elaboración propia]**

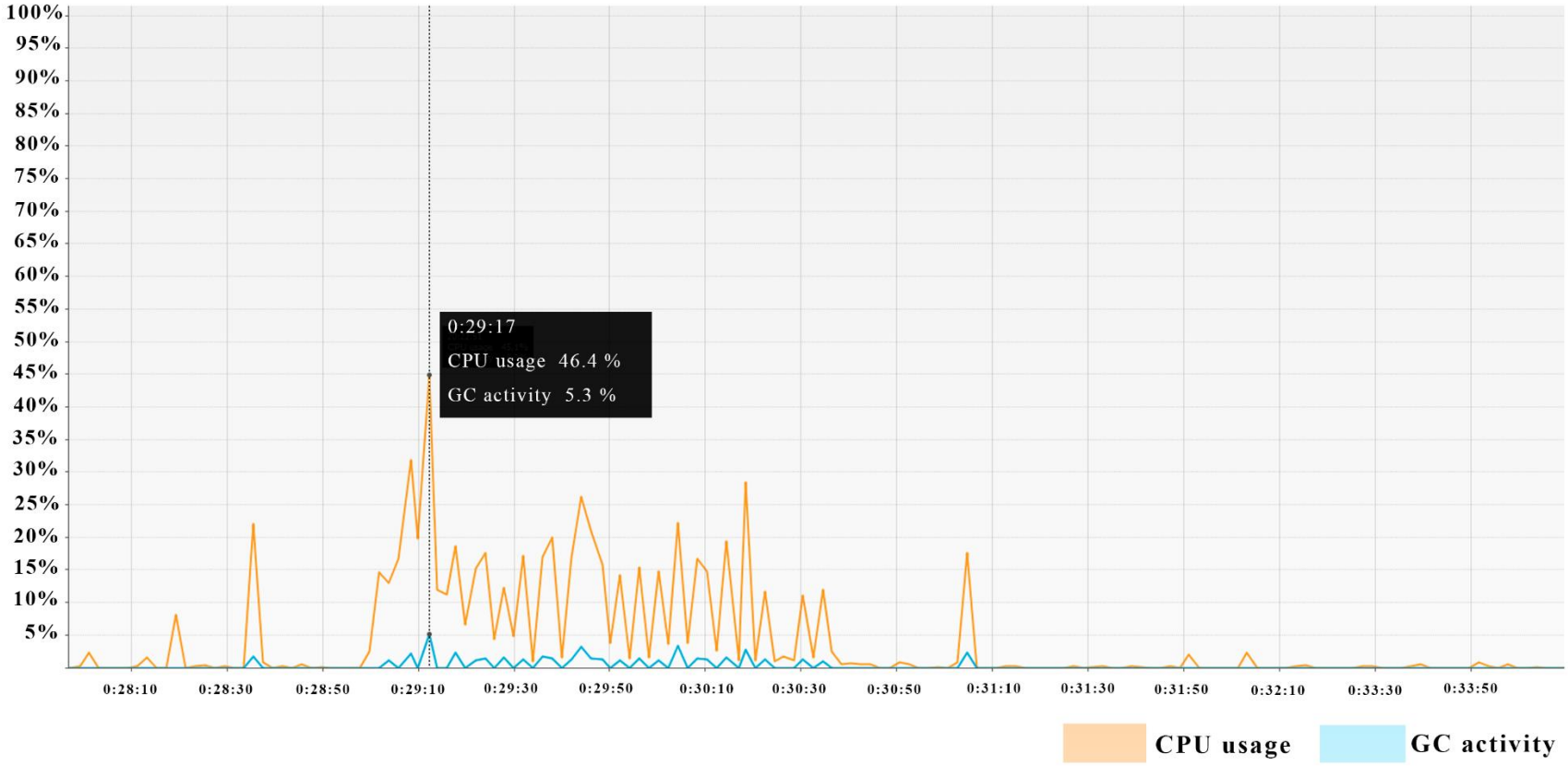


Figura 32 – Consumo de CPU Escenario B. **[Elaboración propia]**

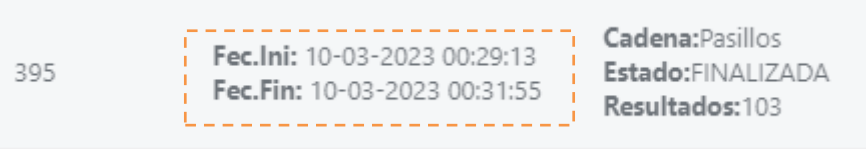


Figura 33 – Tiempo de respuesta Escenario B. **[Elaboración propia]**

Con una carga media, de acuerdo al gráfico se obtienen los siguientes resultados:

Tabla 45 - Resultados escenario B. **[Elaboración propia]**

Escenario	**Volumen de carga**
B (Carga Media)	Número máximo de páginas procesadas: 1500
• Uso de memoria: 1.7 GB máximo. • Uso de CPU 46.4% máximo. • Tiempo de Respuesta: 00:02:42	

5.3.3. Ejecución del escenario C

El escenario C corresponde a una carga alta por lo cual se procede a configurar el perfil de búsqueda que define un número máximo de páginas procesar igual a 750 como se ve en la Figura 34 y posteriormente se ejecuta la búsqueda desde el perfil de Investigador.

Figura 34 - Configuración del Perfil de Búsqueda para Escenario C. **[Elaboración propia]**

Paralelamente a esto se inicia la herramienta **VisualVM** la cual permite monitorizar los recursos consumidos por las aplicaciones que se están ejecutando en una Máquina Virtual Java (JVM), de la cual se obtiene los gráficos de la Figura 35, Figura 36 y Figura 37 que reflejan el consumo de recursos durante el tiempo que dura el proceso de búsqueda.

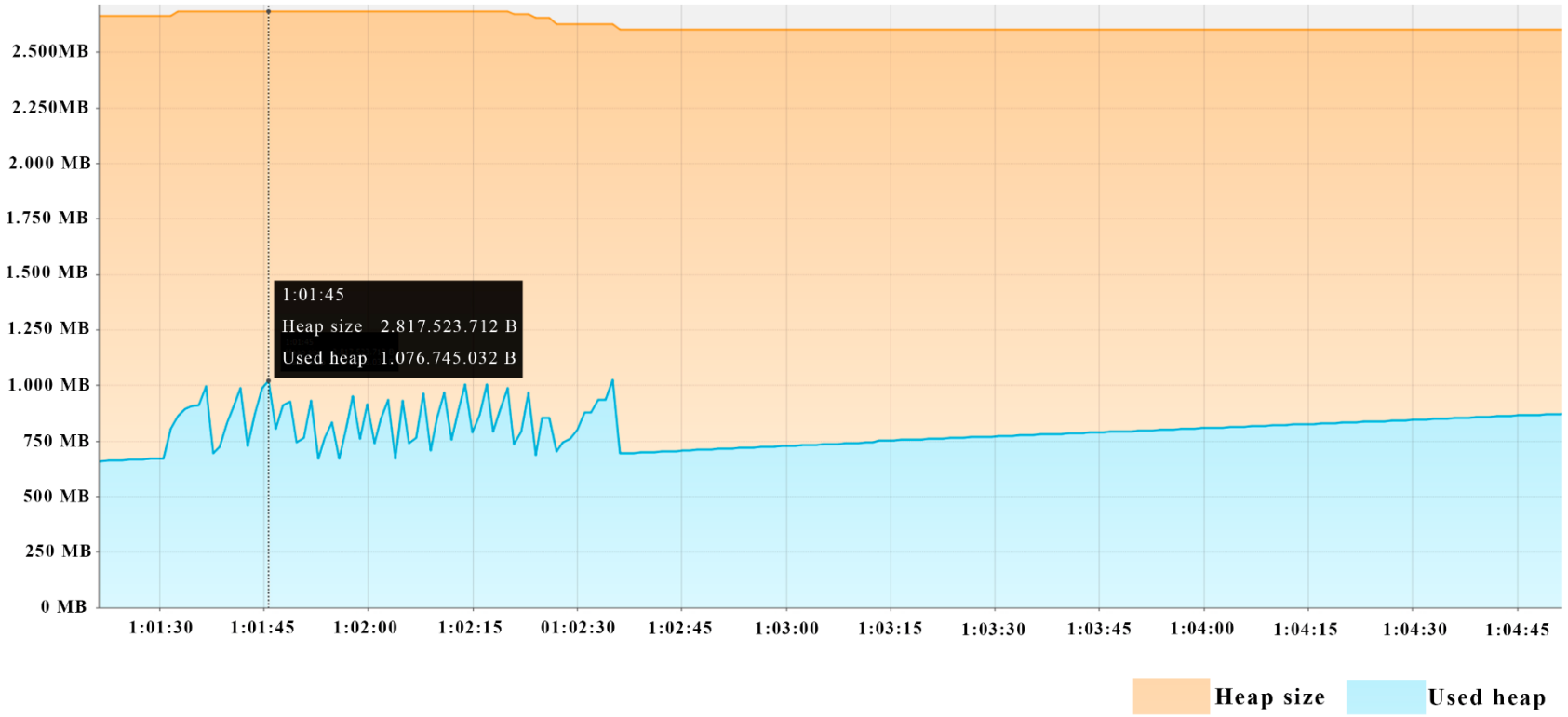


Figura 35 – Consumo de memoria Escenario C. **[Elaboración propia]**

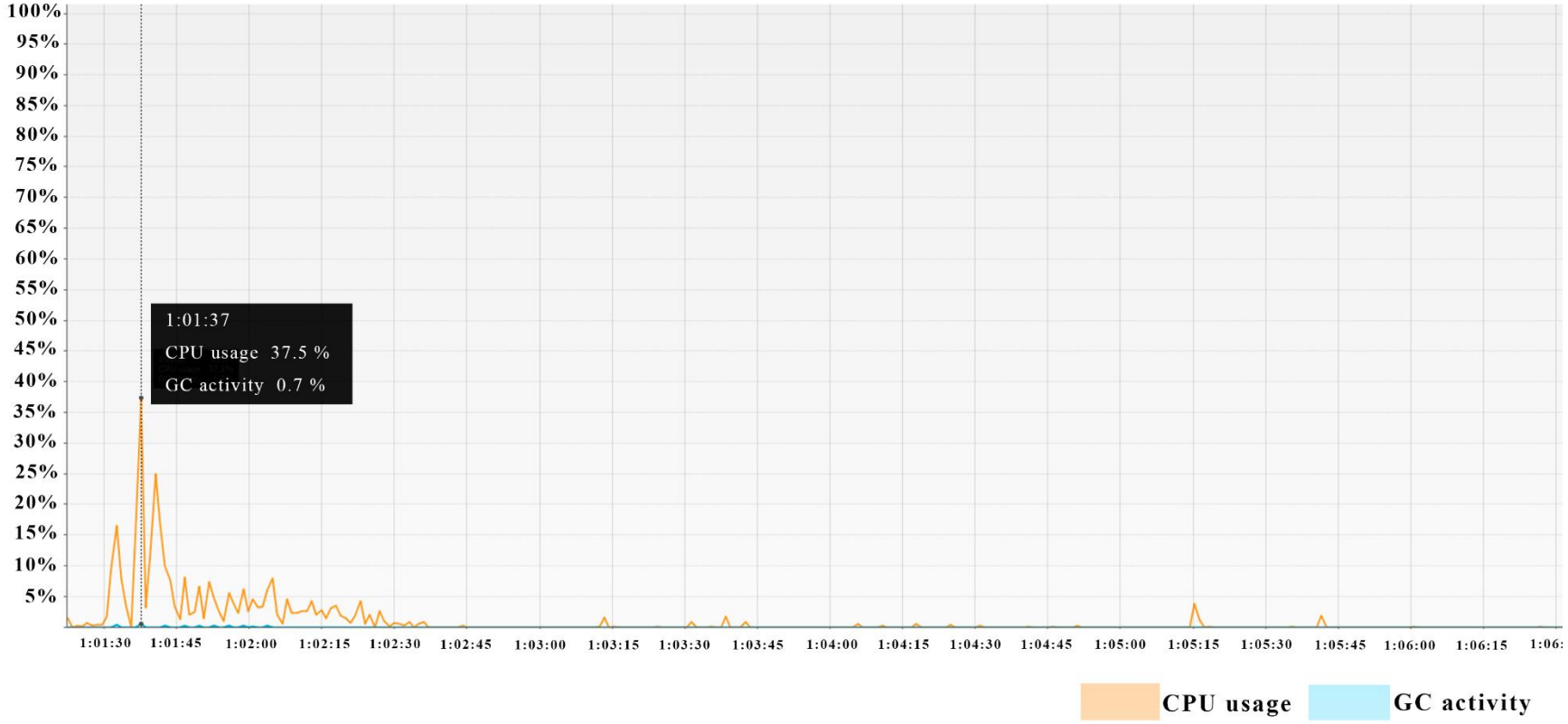


Figura 36 – Consumo de CPU Escenario C. **[Elaboración propia]**

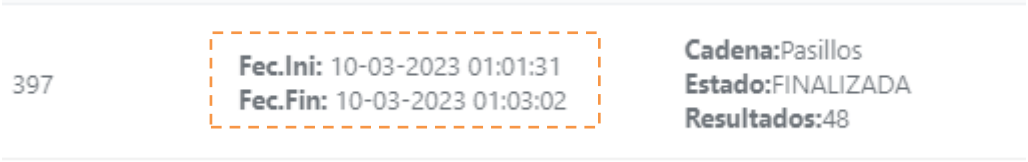


Figura 37 – Tiempo de respuesta Escenario C. **[Elaboración propia]**

Con una carga alta, de acuerdo al gráfico se obtienen los siguientes resultados:

Tabla 46 - Resultados escenario B. **[Elaboración propia]**

Escenario	**Volumen de carga**
C (Carga Baja)	Número máximo de páginas procesadas: 750
• Uso de memoria: 1.08 GB máximo. • Uso de CPU 37.5% máximo. • Tiempo de Respuesta: 00:01:02	

Tabla 47 – Resumen resultados evaluación de desempeño. **[Elaboración propia]**

Escenario	**Volumen de Carga**	**Memoria**	**CPU**	**Tiempo Respuesta**
A(Alta)	2500	2.3 GB	60.30%	00:07:16
B(Media)	1500	1.7 GB	46.4%	00:02:42
C(Baja)	750	1.08 GB	37.5%	00:01:02

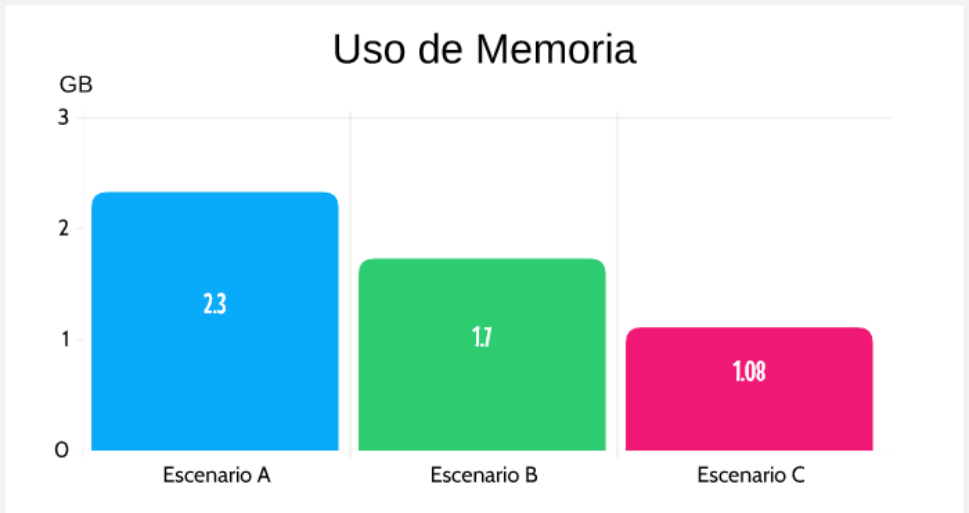


Figura 38 – Comparativa de escenarios en uso de memoria. **[Elaboración propia]**

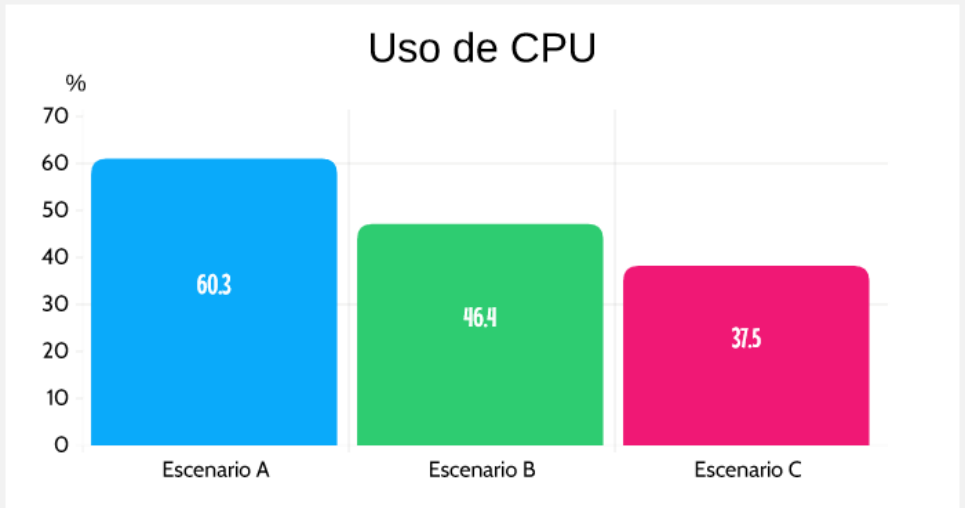


Figura 39 - Comparativa de escenarios en uso de CPU. **[Elaboración propia]**

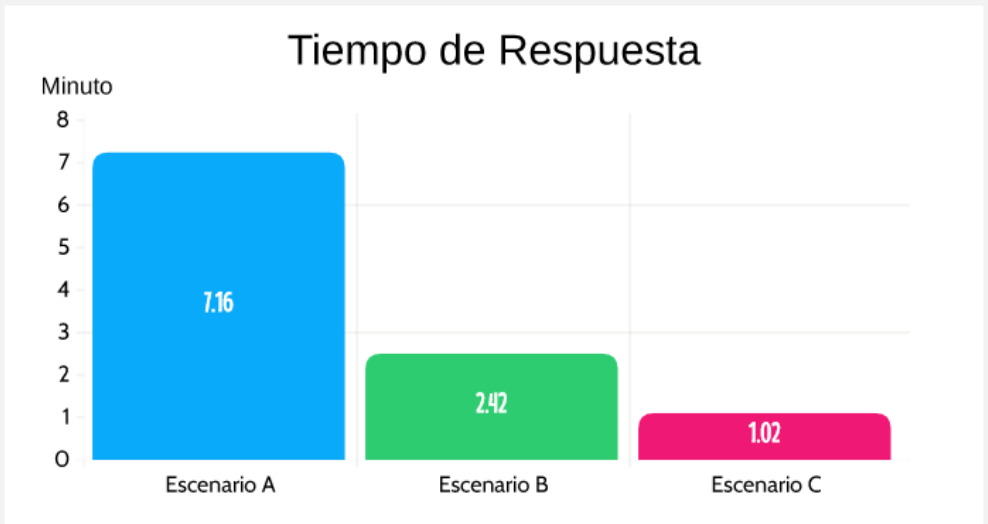


Figura 40 - Comparativa de escenarios en tiempo de respuesta. **[Elaboración propia].**

5.4. Evaluación de Resiliencia

Dado que el método propuesto se implementa en el prototipo software es necesario evaluar la resiliencia que esta muestra ante escenarios adversos los cuales intentarán burlar los controles de seguridad implementados para preservar la privacidad de la información recolectada en un proceso OSINT.

5.4.1. Modelo del Adversario

Para definir los escenarios adversos es necesario definir una perspectiva o modelo de ataque y generalmente para esto se utilizan modelos de amenazas que si bien cubren diversos tipos de amenazas a su vez también requieren numerosas y costosas implementaciones en controles de seguridad, de los cuales muy pocos son eficientes contra los ataques utilizados por los adversarios.

Tomando en cuenta esto se opta por utilizar **ADVERSARIAL THREAT MODEL** que como su nombre indica es un modelo de amenazas basado en el adversario, el cual hace énfasis en el análisis del adversario y sus métodos de ataque para posteriormente definir los controles más adecuados. De esta manera se logra definir escenarios de ataque más realistas y controles más eficientes.

5.4.1.1. Determinar adversario

Identificar nuestro adversario es fundamental para cualquier estrategia de defensa, en este caso se identificó como potencial adversario a un usuario interno tomando en cuenta que tiene interacción directa y conoce el funcionamiento del sistema.

Atacante interno: Usuario de la aplicación y miembro del equipo de investigación.

5.4.1.2. Comprensión del adversario

Ya identificado nuestro adversario es necesario comprender las ventajas y desventajas que posee frente al objetivo de ataque que queremos defender, así como sus posibles técnicas. En la Tabla 48 se presenta un análisis de lo comentado.

Tabla 48 – Conocimiento del adversario. **[Elaboración propia]**

Atacante interno
Comprensión del adversario: • Tiene asignado un perfil dentro de la aplicación con ciertos privilegios. • No requiere burlar seguridad a nivel de infraestructura. • No requiere burlar métodos de autenticación. • Conoce perfectamente las funcionalidades de la aplicación. • Conoce al personal comprometido con la seguridad de la organización. • Acceso a la superficie de la aplicación para test. • Facilidad para ejecutar ataques basados en formularios como Sql injection, XSS, CSRF, Brute Force Attacks (Dictionary), etc. • Explotar vulnerabilidades de Control de Acceso.
Ataques: • Social engineering • Broken Access Control • SQL injection attacks • Weak or stolen credentials • Input Capture • CSRF (cross site request forgery) • Key logger • Insufficient Encryption • Weak credentials • Brute Force Attacks (Dictionary)

5.4.1.3. Ataques del adversario.

Luego de realizar el análisis sobre el adversario un primer insumo es definir una lista de ataques que podría utilizar, esto para tener una visión más clara y saber cuándo podría ejecutarlos. Es así que se los clasifica de acuerdo a las fases de la cadena de muerte **"KILL CHAIN"** [47], que es un procedimiento típico que siguen los ciberdelincuentes para completar un ataque cibernético con éxito. En la siguiente tabla se describen las amenazas del adversario.

Phase	
Reconnaissance	**Web Application Mapping**
	Port Scanners
	Adversary-in-the-Middle
	Collecting email from compromised hosts
	with publicly available tools
Weaponization	**SQL injection attacks**
	Broken Access Control
	Weak or stolen credentials
	Phishing (social engineering)
	Input Capture
	Network mapping (nmap)
Delivery	**Phishing attack**
	Brute Force Attacks(Dictionary)
	Denial of Service Attack
Exploitation	**CSRF(cross site request forgery)**
	Key logger
	Executables
Installation	**Insuficient Encryption**
	Hydra tools, Custom Batch scripts
Call Back	**Insuficient Encryption**
	RDP, ComRAT, Custom CS (BISCUIT)

Tabla 49– Amenazas del adversario**. [Elaboración propia]**

5.4.1.4. Diseño de defensas para ataques del adversario.

Ya definidas los posibles ataques de nuestro adversario se definen las defensas para estos ataques clasificándolas en acciones para detectar, denegar, interrumpir, degradar y engañar. Lo comentado se muestra a continuación.

Phase	Threat	Detect	Deny	Disrupt	Degrade	Deceive
Reconnaissance	**Web Application Mapping**	IDS , HoneyPot Vulnerability Information and Analysis	Access Control, IPS, FW	HoneyNet, connections limits data limits, IPS	Timeout connections	HoneyPot
	Collecting email from compromised hosts with publicly available tools	Dark Web Scan	Delete Email			Fake Employe
Weaponization	**SQL injection attacks**	SQL sanitization,Threats information,NIDS Vulnerability Information and Analysis,	Threats information, Pentests,System and application patching, S.O and Aplplications Version hidden, NIPS	Hardening, Version hidden	Unused services disabling	
	Broken Access Control	Log review	Access Control	Locked user		
	Weak or stolen credentials	Threats information, Captacha code, Two-Factor Authentication, Vulnerability Information and Analysis	Threats information Strong Password Policy	Password expiration policy		
	Phishing (social engineering)	Threats information, NIDS Vulnerability Information and Analysis,	NIPS, Threats information, Pentests	Antivirus		
	Input Capture	Threats information, Antivirus Vulnerability Information and Analysis	User Permission on System	Antivirus		
Delivery	**Phishing attack**	Email Security, Endpoint Protection (EPP, EDR), IDS, Firewall	Proxy Filter, DNS security, Network IPS Host Intrusion Prevention System (HIPS), Firewall, Change fabric settings	Hardening In-line AV	Fake Employe	HoneyPot
	Brute Force Attacks(Dictionary)	Accounts locked IDS, FW	Force captchas after multiple failed logins Change default password	Locked acccounts		
Exploitation	**CSRF(cross site request forgery)**	Vulnerability Information and Analysis	Token validation	Token validation		
	Key logger	Antivirus Scan	Screen Keyboard for sentitive data input	Antivirus policies		
	Security Misconfiguration	Change fabric settings	Application configuration			
Installation	**Insuficient Encryption**		Sensitive Data Encryption			
Call Back	**Insuficient Encryption**		Sensitive Data Encryption, Audit Records and Reports			
	Remote Desktop	NIDS, SIEM, TI Feed, NIDS, HIDS	Whitelisting FW, ACL, DNS Filtering Firewall Access Control Lists,	NIPS, HIPS	QoS, Tarpit	HoneyPot, DNS Redirect

Tabla 50 – Defensas para ataques del adversario. **[Elaboración propia]**

5.4.1.5. Defensas de alto costo para el adversario

Ya definidas las defensas estas se priorizan en función de su eficacia, tomando en cuenta las fases y defensas que más le costarán superar al adversario. Esto se muestra en la siguiente tabla a continuación.

Phase	Threat	Detect	Deny	Disrupt	Degrade	Deceive
Reconnaissance	**Web Application Mapping**	IDS , HoneyPot Vulnerability Information and Analysis	Access Control, IPS, FW	HoneyNet, connections limits data limits, IPS	Timeout connections	HoneyPot
	Collecting email from compromised hosts with publicly available tools	Dark Web Scan	Delete Email			Fake Employe
Weaponization	**SQL injection attacks**	SQL sanitization,Threats information,NIDS Vulnerability Information and Analysis,	Threats information, Pentests,System and application patching, S.O and Aplplications Version hidden, NIPS	Hardening, Version hidden	Unused services disabling	
	Broken Access Control	Log review	Access Control	Locked user		
	Weak or stolen credentials	Threats information, Captacha code, Two-Factor Authentication, Vulnerability Information and Analysis	Threats information Strong Password Policy	Password expiration policy		
	Phishing (social engineering)	Threats information, NIDS Vulnerability Information and Analysis,	NIPS, Threats information, Pentests	Antivirus		
	Input Capture	Threats information, Antivirus Vulnerability Information and Analysis	User Permission on System	Antivirus		
Delivery	**Phishing attack**	Email Security, Endpoint Protection (EPP, EDR), IDS, Firewall	Proxy Filter, DNS security, Network IPS Host Intrusion Prevention System (HIPS), Firewall, Change fabric settings	Hardening In-line AV	Fake Employe	HoneyPot
	Brute Force Attacks(Dictionary)	Accounts locked IDS, FW	Force captchas after multiple failed logins Change default password	Locked acccounts		
Exploitation	**CSRF(cross site request forgery)**	Vulnerability Information and Analysis	Token validation	Token validation		
	Key logger	Antivirus Scan	Screen Keyboard for sentitive data input	Antivirus policies		
	Security Misconfiguration	Change fabric settings	Application configuration			
Installation	**Insuficient Encryption**		Sensitive Data Encryption			
Call Back	**Insuficient Encryption**		Sensitive Data Encryption, Audit Records and Reports			
	Remote Desktop	NIDS, SIEM, TI Feed, NIDS, HIDS	Whitelisting FW, ACL, DNS Filtering Firewall Access Control Lists,	NIPS, HIPS	QoS, Tarpit	HoneyPot, DNS Redirect

Tabla 51– Defensas de alto costo para el adversario. **[Elaboración propia]**

5.4.1.6. Defensas de bajo costo para el defensor

Ahora interesa determina las fases y defensas que implican menor costo de implementación para prevenir, detectar, responder ataques del adversario. Esto se muestra en la siguiente tabla continuación.

Phase	Threat	Detect	Deny	Disrupt	Degrade	Deceive
Reconnaissance	**Web Application Mapping**	IDS , HoneyPot Vulnerability Information and Analysis	Access Control, IPS, FW	HoneyNet, connections limits data limits, IPS	Timeout connections	HoneyPot
	Collecting email from compromised hosts with publicly available tools	Dark Web Scan	Delete Email			Fake Employe
Weaponization	**SQL injection attacks**	SQL sanitization,Threats information,NIDS Vulnerability Information and Analysis,	Threats information, Pentests,System and application patching, S.O and Aplplications Version hidden, NIPS	Hardening, Version hidden	Unused services disabling	
	Broken Access Control	Log review	Access Control	Locked user		
	Weak or stolen credentials	Threats information, Captacha code, Two-Factor Authentication, Vulnerability Information and Analysis	Threats information Strong Password Policy	Password expiration policy		
	Phishing (social engineering)	Threats information, NIDS Vulnerability Information and Analysis,	NIPS, Threats information, Pentests	Antivirus		
	Input Capture	Threats information, Antivirus Vulnerability Information and Analysis	User Permission on System	Antivirus		
Delivery	**Phishing attack**	Email Security, Endpoint Protection (EPP, EDR), IDS, Firewall	Proxy Filter, DNS security, Network IPS Host Intrusion Prevention System (HIPS), Firewall, Change fabric settings	Hardening In-line AV	Fake Employe	HoneyPot
	Brute Force Attacks(Dictionary)	Accounts locked IDS, FW	Force captchas after multiple failed logins Change default password	Locked acccounts		
Exploitation	**CSRF(cross site request forgery)**	Vulnerability Information and Analysis	Token validation	Token validation		
	Key logger	Antivirus Scan	Screen Keyboard for sentitive data input	Antivirus policies		
	Security Misconfiguration	Change fabric settings	Application configuration			
Installation	**Insuficient Encryption**		Sensitive Data Encryption			
Call Back	**Insuficient Encryption**		Sensitive Data Encryption, Audit Records and Reports			
	Remote Desktop	NIDS, SIEM, TI Feed, NIDS, HIDS	Whitelisting FW, ACL, DNS Filtering Firewall Access Control Lists	NIPS, HIPS	QoS, Tarpit	HoneyPot, DNS Redirect

Tabla 52- Defensas de bajo costo para el defensor. **[Elaboración propia]**

5.4.1.7. Defensas priorizadas

Finalmente se obtiene una lista priorizada de defensas efectivas para los ataques del adversario, las cuales se caracterizan por tener un bajo costo de implementación y un alto costo para los ataques del adversario. Lo comentado se muestra en la Tabla 53 a continuación

Tabla 53 – Defensas priorizadas. **[Elaboración propia]**

Prioridad	**Defensas**	**Threat**
1	Access Control	Broken Access Control, Elevation of privilege
2	Changefabric settings	Penetration Tests
3	Locked acccounts	Denial of Service
4	Strong PasswordPolicy	Spoofing Identity
5	Screen Keyboard for sentitive data inputs	Spoofing Identity
6	Force captchas after multiple failed logins	Denial of Service
7	Sql Sanitization	Tampering, Elevation of privilege
8	Audit Records and Reports	Repudiation, Insufficient Auditing
9	Sensitive Data Encryption	Information Disclosure
10	Token validation	Spoofing Identity
11	Two-Factor Authentication	Spoofing Identity

5.4.2. Modelado de amenazas y control de riesgos

Llevar a cabo un modelado de este tipo nos permite identificar y entender potenciales amenazas dentro de un sistema para establecer controles de seguridad que permitan mitigarlos. Es por esto que mediante la herramienta MICROSOFT THREAT MODELLING TOOL se realiza un modelado de amenazas del sistema el cual nos arroja una lista de amenazas para las cuales se realiza un análisis y control de riesgo. Aquí es importante mencionar que para definir los controles aplicados para gestionar el riesgo se toman los requerimientos de privacidad del método propuesto de la Tabla 18 del tercer capítulo. De manera resumida en la Tabla 54 se puede ver como la gestión de riesgos mediante los controles mencionados reduce estos a niveles aceptables. En el Anexo I se puede encontrar más detalles de este proceso.

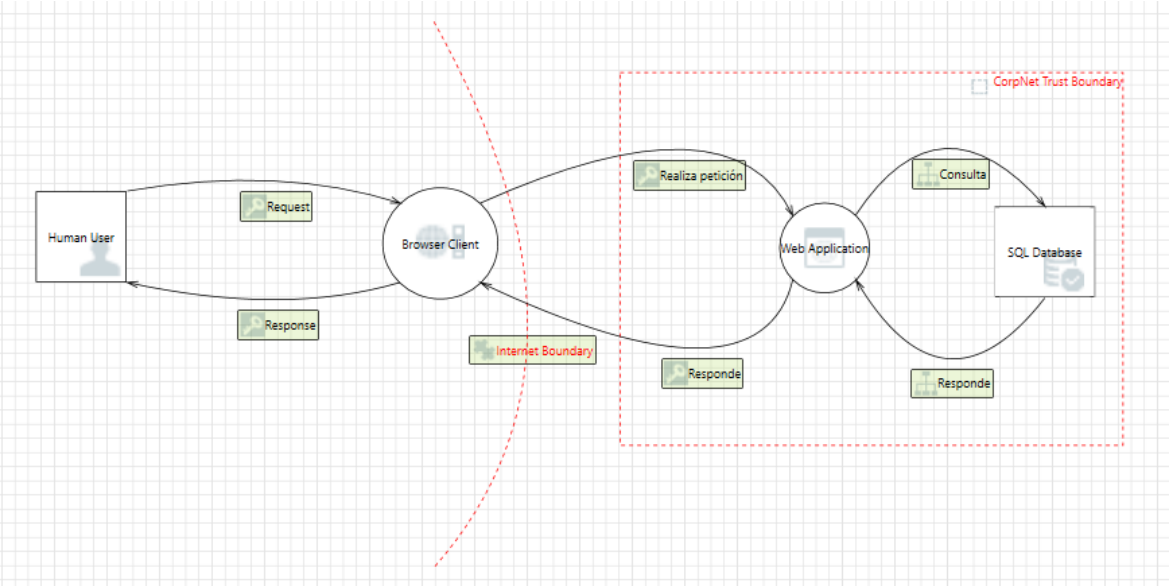


Figura 41 – Diagrama de la solución en Microsoft Threat Modelling

Tabla 54 – Análisis y control de riesgos del prototipo software [Elaboración propia]

#	Amenaza	Categoría	Riesgo Inherente			Tratamiento	Riesgo Residual		
			Impacto	Probabilidad	Riesgo	Control	Impacto	Probabilidad	Riesgo
A_02	Elevation Using Impersonation	E	2	2	4	PR_14.1, PR_14.2, PR_14.4 PR_14.6	1	1	2
A_05	Cross Site Scripting	T	2	2	3	PR_14.6 PR_14.7	0	1	1
A_06	Potential Data Repudiation by Web Application	R	1	2	3	PR_17.1 PR_17.2 PR_13.1	1	1	2
A_07	Potential Process Crash or Stop for Web Application	D	1	2	3	PR_02.1 PR_03.1 PR_04.1 PR_06.1 PR_10.1 PR_10.2	1	1	2
A_21	Risks from Logging	T	2	2	4	PR_14	1	1	2

A_24	Weak Access Control for a Resource	I	2	2	4	PR_14.1, PR_14.2, PR_14.4 PR_14.7	1	0	1
A_27	Potential SQL Injection Vulnerability for SQL Database	T	2	2	4	PR_14.6 PR_14.7	1	1	2
A_30	Insufficient Auditing	R	1	2	3	PR_15.1 PR_17.1	0	1	1
A_32	Weak Credential Storage		2	2	4	PR_14.3 PR_14.4 PR_14.7	0	1	1

5.4.3. Definición de requerimientos críticos de operación

Para definir los requerimientos críticos de operación se toma en cuenta 2 factores:

- **Modelo de amenazas del adversario e implicaciones en módulos del sistema:** Permite identificar amenazas y ataques de adversario y sus defensas más eficientes.
- **Análisis y control de riesgos:** permite identifica amenazas, nivel de riesgo y sus controles.

En primer lugar, se analizan los componentes del sistema responsables de la preservación de privacidad de la información recolectada, así como las amenazas del modelo adversarial que comprometen su correcto funcionamiento. La Figura 42 , Tabla 55 y Tabla 56 muestran cómo se relacionan estos factores y la incidencia en cada módulo del sistema.

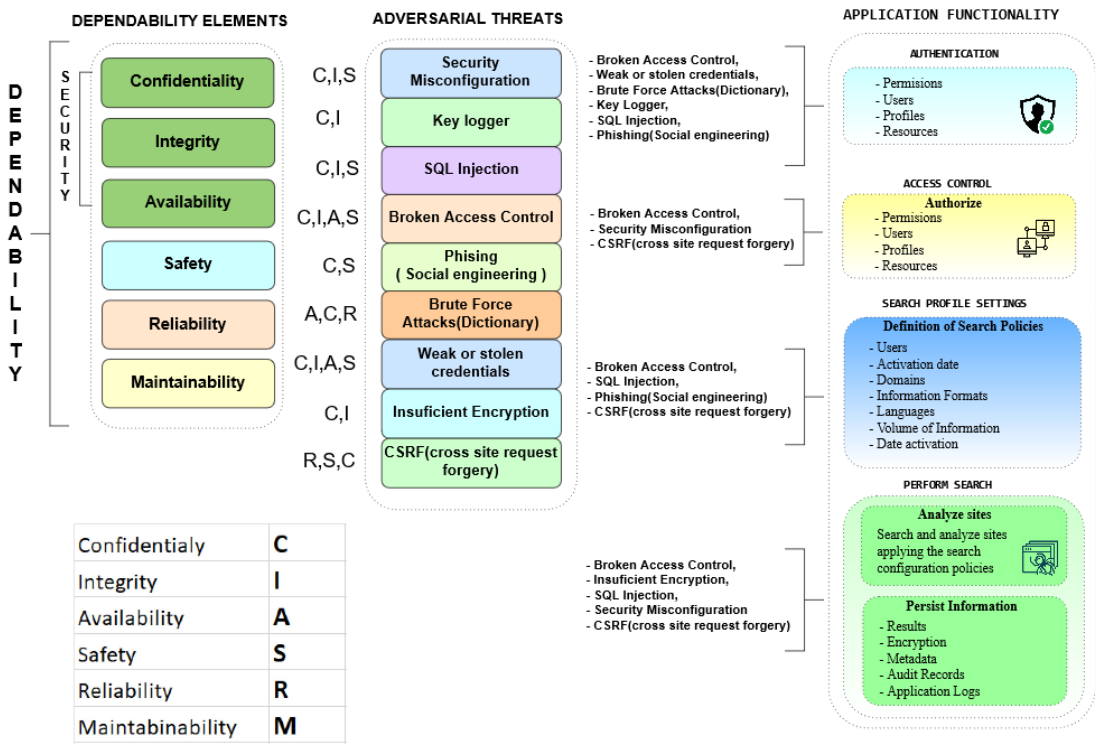


Figura 42 – Amenazas y funcionalidades del sistema. **[Elaboración propia]**

Tabla 55 – Matriz de amenazas y módulos del sistema**. [Elaboración propia]**

		APPLICATION MODULES				
		AUTHENTICATION	**ACCESS CONTROL**	**SEARCH PROFILE SETTINGS**	**PERFORM SEARCH**	**TOTAL**
THREATS	**Security Misconfiguration**		✓		✓	**2**
	Key Logger	✓				**1**
	SQL Inyección	✓		✓	✓	**3**
	Broken Access Control	✓	✓	✓	✓	**4**
	Phishing	✓				**1**
	Brute Torce Attacks	✓				**1**
	Weak or stolen credentials	✓				**1**
	Insuficient Encryption				✓	**1**
	CSRF		✓	✓	✓	**3**

Luego de analizar los módulos del sistema y las amenazas del adversario estas se ranquean de acuerdo al número de módulos que comprometería un ataque de este tipo.

Tabla 56 – Amenazas y criticidad. **[Elaboración propia]**

Cod	**Threat**
ADV_01	Broken Access Control
ADV_02	SQL Injection
ADV_03	CSRF
ADV_04	Security Misconfiguration
ADV_05	Weak or stolen credentials
ADV_06	Phishing
ADV_07	Insuficient Encryption
ADV_08	Key logger
ADV_09	Brute Force Attacks

A continuación, se resumen las amenazas acuerdo al modelado de amenazas con **Microsoft Threat Modeling y** el **Modelo del Adversario** lo cual permite tener una visión más amplia de sus implicaciones en el sistema y con esto definir los requerimientos críticos y su evaluación.

Tabla 57 – Mapeo de amenazas [Elaboración propia]

Mapeo de amenazas				
Microsoft Threat Modeling			**Modelo Adversarial**	
Cod	**Amenaza**	**Riesgo**	**Cod**	**Amenaza**
A_02	Elevation Using Impersonation	4	ADV_01	Broken Access Control, Elevation of privilege
A_05	Cross Site Scripting	4	ADV_03	Tampering
A_06	Potential Data Repudiation by Web Application	3	ADV_08	Repudiation, Insufficient Auditing
A_07	Potential Process Crash or Stop for Web Application	3	ADV_03	Denial of Service
A_21	Risks from Logging	4	ADV_01, ADV_08	Broken Access Control, Repudiation
A_24	Weak Access Control for a Resource	4	ADV_01	Broken Access Control
A_27	Potential SQL Injection Vulnerability for SQL Database	4	ADV_02	Tampering, Elevation of privilege
A_30	Insufficient Auditing	3	ADV_08	Repudiation
A_32	Weak Credential Storage	4	ADV_09	Information Disclosure

Tomando en cuenta las amenazas que mayor nivel de riesgo representan para el sistema, los controles definidos y las defensas del adversario se define los siguientes requerimientos críticos de la Tabla 58 , RC_01, RC_02, RC_03, RC_04 los cuales corresponden a control de acceso y serán evaluados mediante escenarios de ataque que permitan demostrar la resiliencia del sistema, mientras que los requerimientos RC_06, y RC_07 se evalúan mediante un análisis de vulnerabilidades estático y dinámico lo cual se puede ver en el Anexo VI.

Tabla 58 – Requerimientos críticos [Elaboración propia]

REQUERIMIENTOS CRÍTICOS			
Código	**Descripción**	**Amenazas**	**Evaluación**
RC_01	El módulo de control de acceso debe permitir y hacer cumplir la política de privilegio mínimo.	ADV_01, ADV_05, A_24, A_01	✓
RC_02	El módulo de control de acceso debe permitir configurar reglas dinámicas para separación de roles.	ADV_01, ADV_05, A_24, A_01	✓
RC_03	El módulo de control de acceso debe autorizar recursos de acuerdo a las reglas configuradas para cada rol.	ADV_01, ADV_05, A_24, A_01	✓
RC_04	El módulo de control de acceso debe validar el acceso a un recurso solicitado en cada petición realizada.	ADV_01, ADV_05, A_24, A_01	✓
RC_06	El sistema debe ser resiliente ante ataques de tipo Cross Site Scripting	ADV_03, A_05	X
RC_07	El sistema debe ser resiliente ante ataques de tipo SQL Injection	A_27, ADV_02	X

5.5. Escenarios de ataque

Una vez definidos los requerimientos críticos de operación se pasa a describir los escenarios de ataque que permitirán validarlos.

5.5.1. Escenario de ataque 1

Tabla 59- Escenario de ataque 1. **[Elaboración propia]**

Requerimientos críticos involucrados: • **RC_01:** El módulo de control de acceso debe permitir configurar y hacer cumplir la política de privilegio mínimo. • **RC_02:** El módulo de control de acceso debe hacer cumplir la política de privilegio mínimo.
Perfil del atacante: Usuario con perfil de Investigador
Objetivo del atacante: Acceder a información privilegiada mediante usuario ficticios.
Descripción del escenario:

El Administrador del sistema configura el perfil de usuario Facilitador con políticas de privilegio mínimo. Posteriormente el atacante ingresa al sistema mediante credenciales robadas con el fin de crear usuarios ficticios con privilegios elevados. Tras un login exitoso el atacante ingresa al sistema, sin embargo, solo se le despliega un menú con las opciones mínimas configuradas para el perfil autenticado las cuales no poseen privilegios avanzados para gestión de usuarios.
Resultado esperado: • El sistema permite la configuración de privilegio mínimo. • El sistema despliega un menú dinámico únicamente con las opciones permitidas para el perfil del usuario y rol autenticado.

Control de Acceso a nivel de aplicación

Dentro de las responsabilidades que tiene un módulo de control de acceso está la de autorizar recursos en base a una serie de políticas definidas las cuales deben ser configuradas previamente. Si bien es en cierto con esto se estaría autorizando recursos adecuadamente, pero existe un problema que surge implícitamente con la configuración de estos accesos, ya que, al delegar responsabilidades operacionales sobre un perfil del sistema estas deberían ser únicamente las mínimas necesarias para realizar la función asignada, esto para:

- Evitar accesos innecesarios a recursos del sistema.
- Reducir el área de ataque ante posibles robos de credenciales.

Con el fin de mitigar esto, el sistema implementa un módulo de control de acceso que permite la **configuración dinámica** de accesos a recursos del sistema haciendo posible implementar una **política de privilegios mínimos** para cada usuario los mismos que son **validados en cada petición recibida**. A continuación, se describe los usuarios y sus privilegios mínimos dentro del sistema

- **Administrador:** Usuario encargado de gestionar el módulo de control de acceso que consiste en gestión de usuarios, roles, recursos y perfiles.
- **Facilitador de Búsquedas:** Usuario encargado de configurar perfiles de búsqueda y asociarlos a usuarios.
- **Investigador:** Usuario que realiza las búsquedas dentro de un perfil de búsqueda asignado.

Tabla 60 – Configuración de privilegios mínimos por usuario [Elaboración propia]

RECURSOS DEL SISTEMA		**USUARIOS**		
NOMBRE	**URL**	**ADMINISTRADOR**	**CONFIGURADOR**	**FACILITADOR**
USUARIOS	/Administrador/Gestion/gestionUsuario.owc	✓		
PROCESOS	/Administrador/Gestion/gestionProceso.owc	✓		
PERFILES	Administrador/Gestion/gestionPerfil.owc	✓		
PERFIL PROCESO	/Administrador/Gestion/gestionPerfilProceso.owc	✓		
CONFIGURAR PERFIL DE BÚSQUEDA	/Administrador/Gestion/gestionPerfilBusqueda.owc		✓	
BUSQUEDAS ASIGNADAS	/Busqueda/perfilBusquedaInvestigador.owc			✓
BUSQUEDA GENERAL	/Busqueda/busqueda.owc			✓

Ejecución del escenario de ataque

De acuerdo al escenario se procede a configurar los privilegios mínimos para el usuario Facilitador, como se puede ver en la Figura 43 y Figura 44, primeramente, se configura los permisos para el perfil mencionado y posteriormente se asigna este perfil al usuario. De esta manera se puede evidenciar que el módulo de control de acceso **es capaz de permitir configurar las políticas de privilegio mínimo** ya que inicialmente un usuario tiene denegados los accesos y dinámicamente el administrador va otorgándole los privilegios necesarios.

Figura 43 – Perfil Facilitador. **[Elaboración propia]**

Figura 44 – Privilegios y restricciones del perfil Facilitador. **[Elaboración propia]**

Posteriormente el atacante logra ingresa al sistema con las credenciales robadas del usuario Facilitador con el fin de crear usuarios con privilegios elevados. Sin embargo, dado que el usuario únicamente fue configurado con los privilegios mínimos para cumplir con sus responsabilidades no cuenta con la opción de crear usuarios como se puede ver

en la Figura 46 y el objetivo principal del atacante se ve frustrado. De esta manera se puede evidenciar que el módulo de control de acceso **es capaz de hacer cumplir la política de privilegio mínimo**.

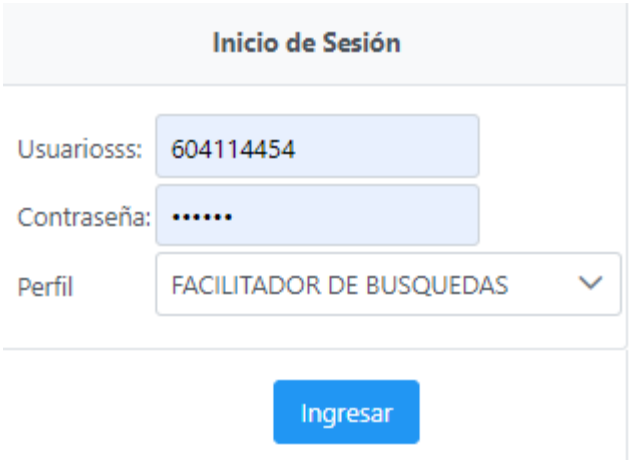


Figura 45 – Ingreso al sistema con credenciales robadas. **[Elaboración propia]**

Figura 46 – Acciones permitidas para el perfil Facilitador. **[Elaboración propia]**

Control de acceso a nivel de base de datos

De la misma manera que se autoriza recursos a nivel de aplicación, la base de datos también requiere una configuración de privilegios a nivel de tablas por usuario. En este sentido se aplica una política de control de acceso a nivel de usuario de base de datos en función de lo requerido por la aplicación y el módulo de control de acceso que implementa. Con esto se busca asegurar que las conexiones que requiera la aplicación hacia la base de datos sean gestionadas por el servidor de aplicaciones en el cual existe

una configuración establecida de acuerdo a las necesidades de cada módulo evitando así cualquier implementación directa hacia la base de datos. La configuración de los pools se puede ver en el Anexo IV.

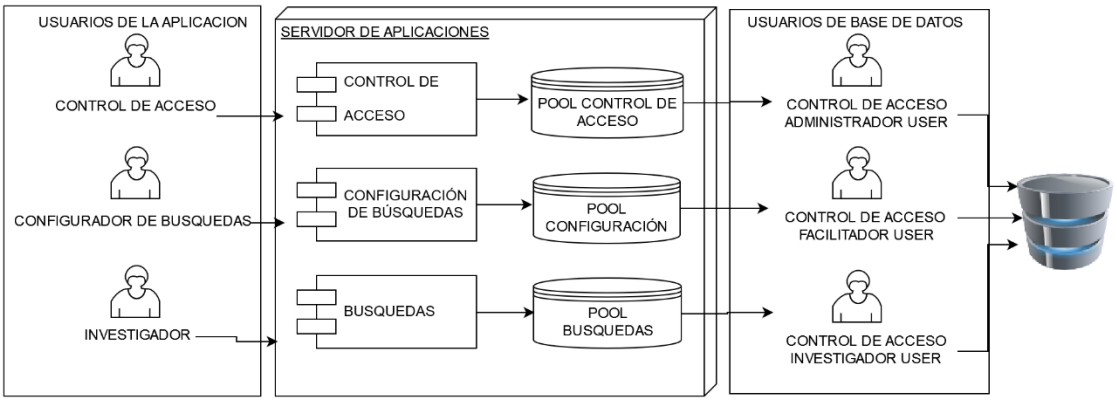


Figura 47 – Esquema de acceso a datos mediante pools y usuarios [Elaboración propia]

Con la implementación de este esquema se logra mantener un mayor control en cuanto a las operaciones que puede realizar cada módulo de la aplicación ya que si desde la aplicación se intenta recuperar o modificar data de una tabla para la cual no tiene privilegios recibirá una excepción siendo así resiliente ante este tipo de ataques.

Una vez la información se halla en la base de datos el sistema de seguridad del DBMS se encarga de gestionar los accesos de acuerdo a los privilegios configurados. Es por esto que se define una política de a nivel de usuario, tabla y operación que garantice el acceso necesario para cada módulo configurado en el pool de conexiones del servidor de aplicaciones. El modelo físico de la base de datos y la asignación de permisos de usuarios se puede ver en el Anexo III.

Tabla 61 – Política de acceso a recursos de la base de datos [Elaboración propia]

TABLAS	**USUARIOS**											
	Administrador				**Facilitador**				**Investigador**			
	S	**I**	**U**	**D**	**S**	**I**	**U**	**D**	**S**	**I**	**U**	**D**
owc_filtro_busqueda						✓	✓		✓			
owc_instancia_busqueda									✓	✓		
owc_perfil	✓	✓	✓									
owc_usuario	✓	✓	✓									

owc_perfil_proceso	✓	✓	✓									
owc_perfil_usuario	✓	✓	✓									
owc_proceso	✓	✓	✓									
owc_perfil_busqueda					✓	✓	✓		✓		✓	
owc_resultado_busqueda									✓	✓		
owc_secuencia	✓	✓	✓		✓	✓	✓		✓	✓	✓	
owc_trazabilidad	✓	✓				✓				✓		
S: Select; **I**: Insert; **U**: Update; **D**: Delete												

A continuación, se intenta alterar un registro de auditoría mediante un usuario de base de datos. Sin embargo, el sistema de seguridad del motor de base de datos recupera y valida los privilegios del usuario atacante los cuales no son suficientes para llevar a cabo este objetivo, ya que, dentro de sus funciones este no debería ser capaz de alterar registros de esta tabla mostrando así resiliencia ante un ataque de este tipo.

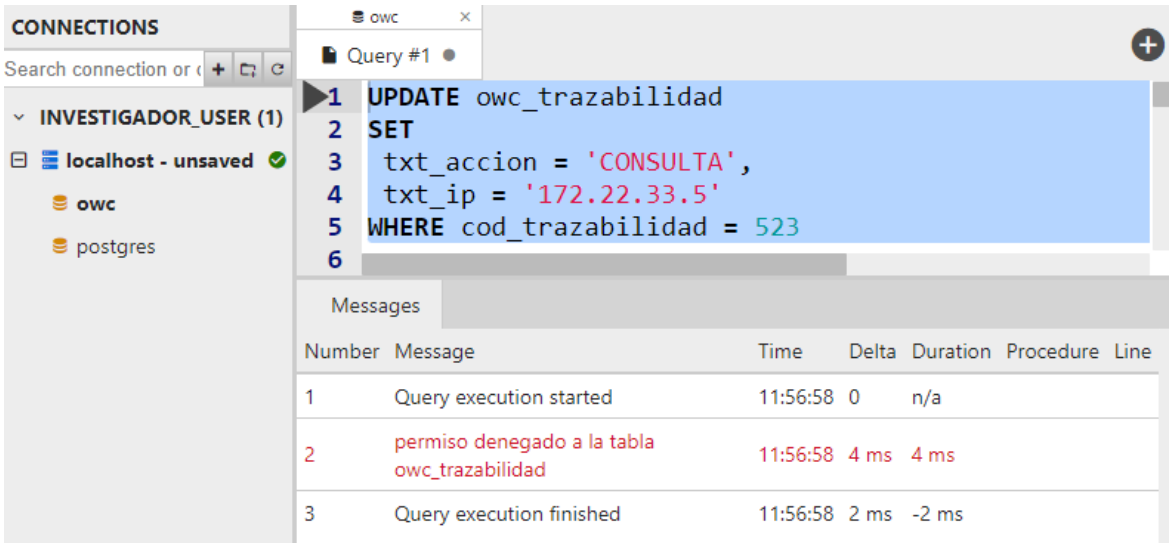


Figura 48 – Intento de acceso a objetos de base de datos no autorizados [Elaboración propia]

5.5.2. Escenario de ataque 2

Tabla 62- Escenario de ataque 2. **[Elaboración propia]**

Requerimientos involucrados: • **RC_03:** El módulo de control de acceso debe permitir configurar reglas dinámicas para separación de roles. • **RC_04:** El módulo de control de acceso debe autorizar recursos a un usuario de acuerdo al rol del usuario autenticado.
Perfil del atacante: Usuario con perfiles de Investigador y Revisor
Objetivo del atacante: Obtener acceso a la información recolectada en búsquedas OSINT.
Descripción del escenario: Una vez el usuario ha sido autenticado con rol de Investigador intenta acceder a funcionalidades de su rol de Revisor aprovechando la sesión activa.
Resultado esperado: • El sistema permite la configuración de reglas dinámicas para separación de roles. • El sistema despliega un menú únicamente con las opciones permitidas para el perfil del usuario autenticado. • El sistema permite o deniega el acceso a los recursos solicitados de acuerdo a los privilegios que posee el perfil del usuario autenticado.

Segregación de roles y funciones

Identificar roles y definir sus funciones es vital dentro de la gestión de procesos ya que permite llevar mejores controles organizacionales y tomar medidas en cualquiera de sus instancias, como, por ejemplo:

- Mal uso de la información
- Fraude, robo y otros riesgos relacionados con la seguridad.

Dentro del proceso de investigación digital forense se tiene entre los activos de información críticos a los recursos digitales recolectados por búsquedas OSINT, por lo cual implementar políticas de segregación de roles y funciones es necesario para evitar cualquiera de los problemas mencionados.

Tomando en cuenta lo dicho, el módulo de control de acceso que se propone provee de funcionalidades para gestionar usuarios y recursos involucrados dentro del proceso de investigación, lo que a su vez hace posible implementar las políticas requeridas para

preservar la privacidad de los recursos obtenidos en búsquedas OSINT. La definición de roles y funcionalidades se definen mediante las Tabla 60 y Figura 49.

Administración dinámica de usuarios y privilegios.

Dentro de la administración de sistemas informáticos en general, la configuración de parámetros que definen su comportamiento es indispensable ya que permite adaptabilidad dentro del proceso automatizado. En este caso concreto se refiere a la gestión de privilegios los cuales no pueden ser estáticos ya que dentro de un proceso legal por orden superior cualquier persona puede quedar excluida de un proceso de investigación lo cual requiere revocación inmediata de privilegios, de la misma manera puede ocurrir lo contrario y por esta razón la administración dinámica de usuarios y privilegios es un requerimiento mandatorio.

Ejecución del escenario de ataque

De acuerdo al escenario el atacante tiene varias responsabilidades y funciones dentro del sistema por lo cual tiene configurado dos perfiles, a continuación, se procede a configurar las reglas dinámicas para cada perfil y se asignar al usuario en cuestión.

Figura 49 – Perfiles del usuario atacante. **[Elaboración propia]**

Figura 50 – Configuración dinámica de permisos para perfil Investigador. **[Elaboración propia]**

Figura 51 – Configuración dinámica de permisos para perfil Revisor. **[Elaboración propia]**

Figura 52 – Asignación de perfiles al usuario. **[Elaboración propia]**

Para verificar que el sistema autoriza los recursos de acuerdo a lo configurado previamente, se inicia sesión con cada uno de los roles donde se puede evidenciar que efectivamente el sistema genera un menú dinámico de acuerdo a los privilegios asignados al rol del usuario autenticado, garantizando así la flexibilidad y eficacia en la aplicación de las reglas configuradas.

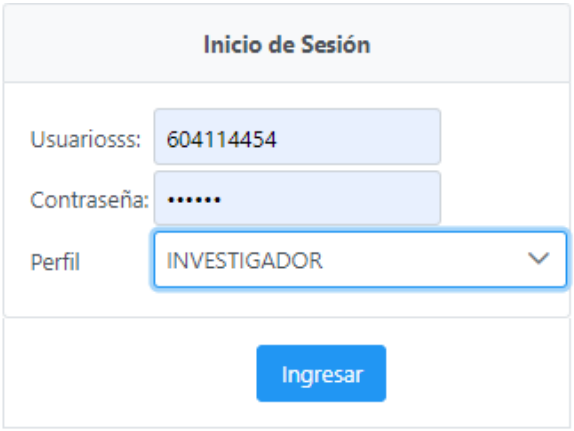


Figura 53 – Inicio de sesión con perfil Investigador. **[Elaboración propia]**

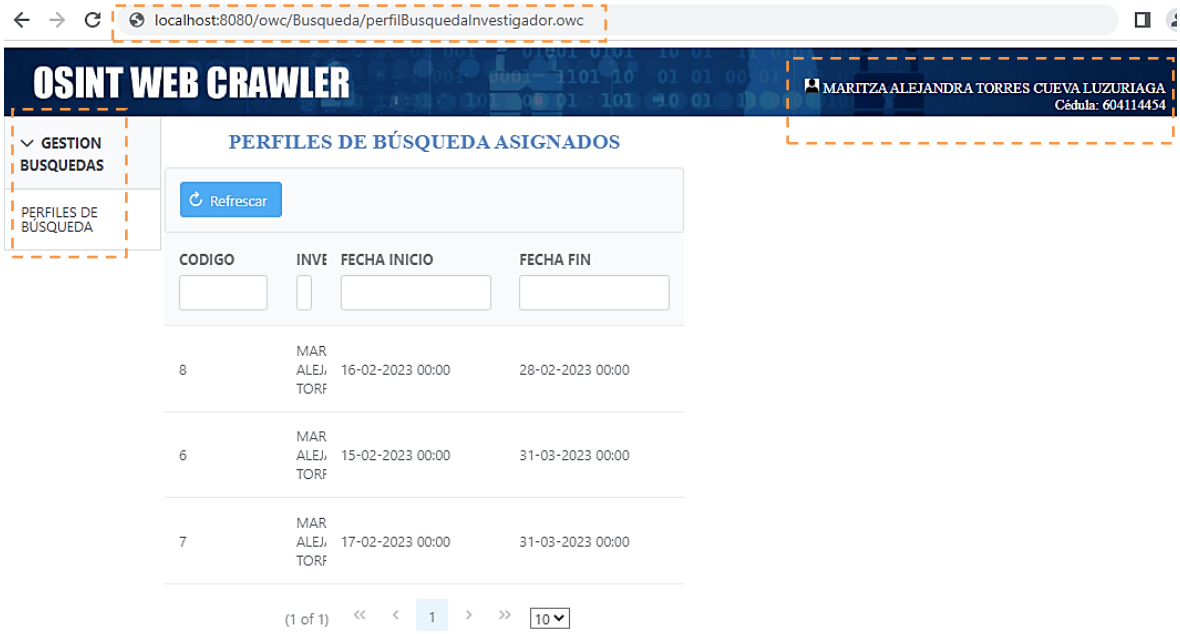


Figura 54 – Menú dinámico de acuerdo al perfil autenticado (Investigador). **[Elaboración propia]**

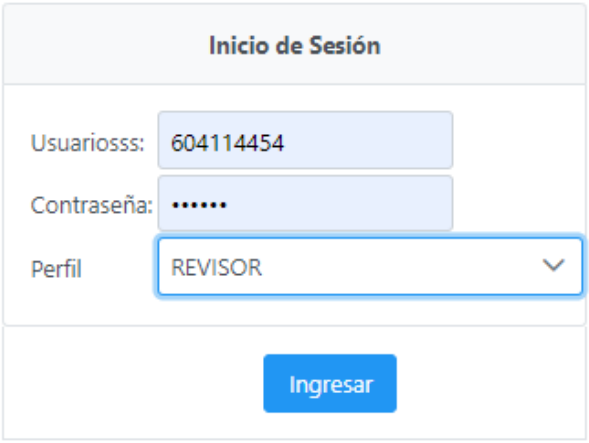


Figura 55 – Inicio de sesión con perfil Revisor. **[Elaboración propia]**

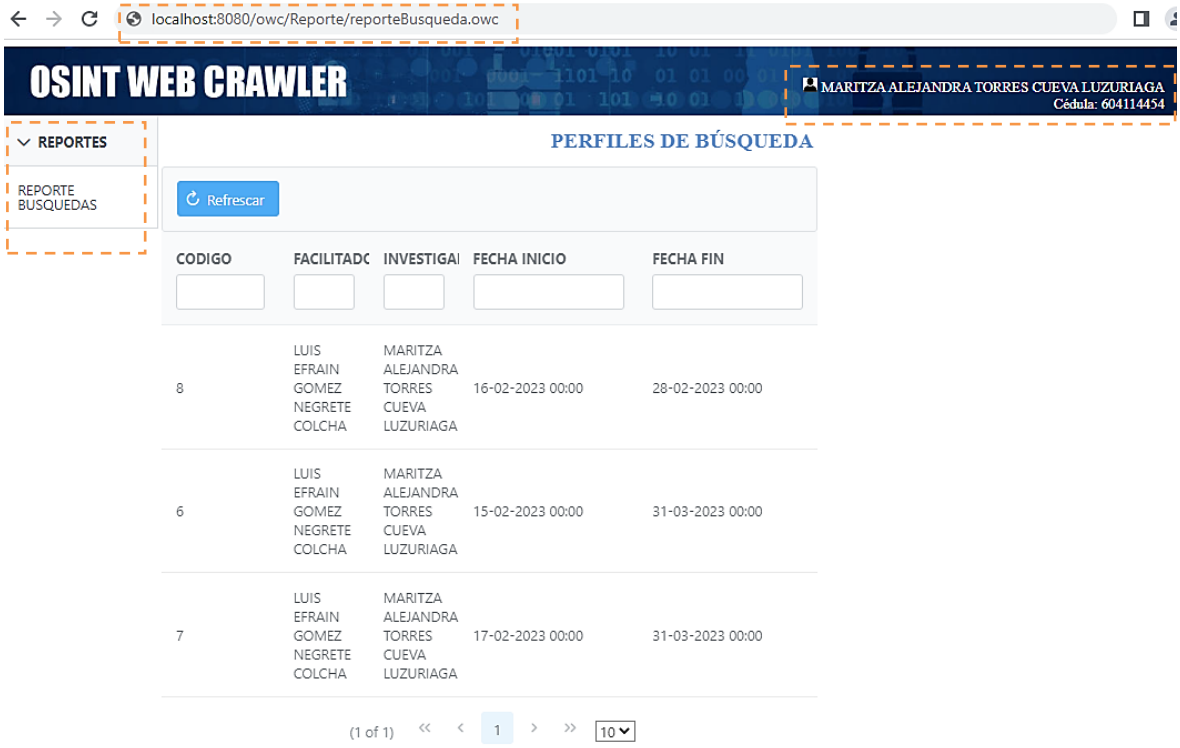


Figura 56 – Menú dinámico de acuerdo al perfil autenticado (Revisor). **[Elaboración propia]**

Como se pudo observar en primera instancia despliego un menú dinámico con opciones de acuerdo a cada uno de los roles configurados con los que se inició sesión, esto se observa en la Figura 54 y Figura 56. Con esto se puede evidenciar que el módulo de control de acceso es capaz de:

- Permitir configurar reglas dinámicas para separación de roles.
- Autorizar recursos a un usuario de acuerdo al rol del usuario autenticado.

Con esto se verifica el cumplimiento de los requerimientos críticos para este escenario de ataque.

5.5.3. Escenario de ataque 3

Tabla 63- Escenario de ataque 3. **[Elaboración propia]**

Requerimientos críticos a evaluar: • **RC_05:** El módulo de control de acceso debe validar el acceso a un recurso solicitado en cada petición realizada.
Perfil del atacante: Usuario con perfil de Investigador con permisos limitados pretende acceder funcionalidades del perfil de Facilitador.
Objetivo del atacante: Modificar los parámetros del perfil de búsqueda al que fue asignado.
Descripción del escenario: Una vez el usuario ha sido autenticado es direccionado a la página de inicio donde aprovechando de una sesión activa el atacante intentará acceder a un recurso no autorizado ingresando manualmente en la barra de navegación una URL correspondiente a una pantalla del usuario Facilitador.
Resultado esperado: • El sistema verifica los permisos para el recurso solicitado de acuerdo al usuario y rol autenticado • El sistema deniega su acceso e informa al usuario con una pantalla de acceso denegado. • El evento se registra en el log de la aplicación.

Forzar todas las solicitudes a validaciones de control de acceso

La manera general en la que un usuario navega dentro de un sistema es un menú el cual va redirigiendo a las páginas solicitadas, cada ítem del menú genera una solicitud al servidor y esta entrega el recurso solicitado al navegador. En este modelo existe un problema ya que la restricción de navegabilidad está dada por el menú entregado inicialmente y se asume que el usuario accederá únicamente a las páginas de este menú ya que desconoce las url de otras funcionalidades, a esto se le conoce como **seguridad por oscuridad.** Sin embargo, no es nada recomendable este tipo de prácticas ya que

descubrir más urls podría ser sencillo debido al historial de navegación del navegador y la utilización de técnicas de ingeniería social.

Para mitigar este problema es necesario definir los tipos de recursos que posee el sistema.

- **Privado:** Todo recurso que requiere una sesión activa.
- **Público:** Todo recurso que no requiere una sesión activa.

De acuerdo a esta clasificación se aplica una política de filtrado únicamente para los **recursos privados** la cual consiste en validar si el usuario que realiza la petición tiene privilegios sobre el recurso solicitado para el rol con el que inició sesión. Con este control se garantiza que toda petición sobre un recurso que no sea publico sea autorizada de manera correcta acorde los privilegios definidos a ese momento, de esta manera cualquier intento de acceso a un recurso ya sea desde el menú de la aplicación o generado manualmente será sometido a este control.

Registrar todos los eventos de control de acceso

Como se ha mencionado en líneas anteriores la implementación de un proceso de autorización es vital para garantizar el correcto acceso a los recursos de un sistema, sin embargo, dado que los recursos gestionados permiten manipular data considerada activo de información crítica es necesario registrar información respecto a las autorizaciones que realiza el sistema como control proactivo de auditoría y que a posteriori permitan responder preguntas como, por ejemplo:

- ¿Qué usuarios solicitaron acceso a un recurso determinado?
- ¿A que recursos accedió un usuario en una fecha dada?
- ¿A que recursos intento acceder un usuario en una fecha dada?

El registro del evento de autorización en el archivo de log de la aplicación incluye información como:

- **Usuario:** Usuario autenticado que solicita un recurso.
- **Rol:** Rol con el que el usuario inicio sesión.
- **Recurso:** Recurso solicitado en la petición.
- **Fecha y Hora:** Marca de tiempo en la que se recibe la petición.

- **Estado autorización:** Autorizado/Denegado

Por otro lado, también se implementa un registro de auditoría de mayor detalle para cada operación que realiza el sistema que se detalla a continuación.

- **Nombre Pantalla:** Usuario autenticado que realiza la operación.
- **Acción:** Recurso solicitado en la petición.
- **Observación/Motivo:** Marca de tiempo en la que se recibe la petición.
- **Fecha y Hora:** Marca de tiempo en la que se realiza la acción.
- **IP:** Dirección IP desde donde se realiza la petición.
- **Usuario:** Usuario que realiza la acción.

cod_ [PK]	txt_nom_pantalla character varying (50)	txt_accion character (100)	txt_observacion character varying (200)	fec_registro timestamp without time zone	txt_ip character varyi	txt_usuario_login character varying (
1	Gestión Roles	REGISTRAR ...	Registro de nuevo Rol	2023-01-29 18:57:25.673	192.168.1.5	1785542569
2	Gestión Perfil de Búsqueda	REGISTRAR	Asignación de perfil de búsque...	2023-01-29 19:01:21.677	192.168.1.12	06504092498

Figura 57 – Registro de auditoria [Elaboración propia]

Este tipo de registro además de permitir trazabilidad dentro de un proceso podría ayudar a garantizar el no repudio en un caso dado.

Ejecución del escenario de ataque

De acuerdo al escenario el atacante intenta acceder a funcionalidades del sistema para las cuales no posee permisos. En primera instancia el atacante inicia sesión mediante sus credenciales y rol asignado (Investigador) y como se puede ver en la Figura 59 el atacante únicamente puede realizar búsquedas y acceder a esta funcionalidad mediante la URL correspondiente que es generada por la interacción con el menú.

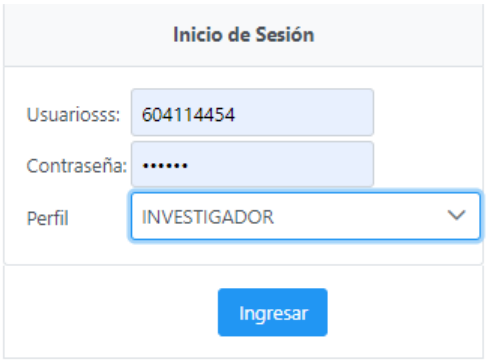


Figura 58 – Inicio de sesión con perfil Investigador. **[Elaboración propia]**

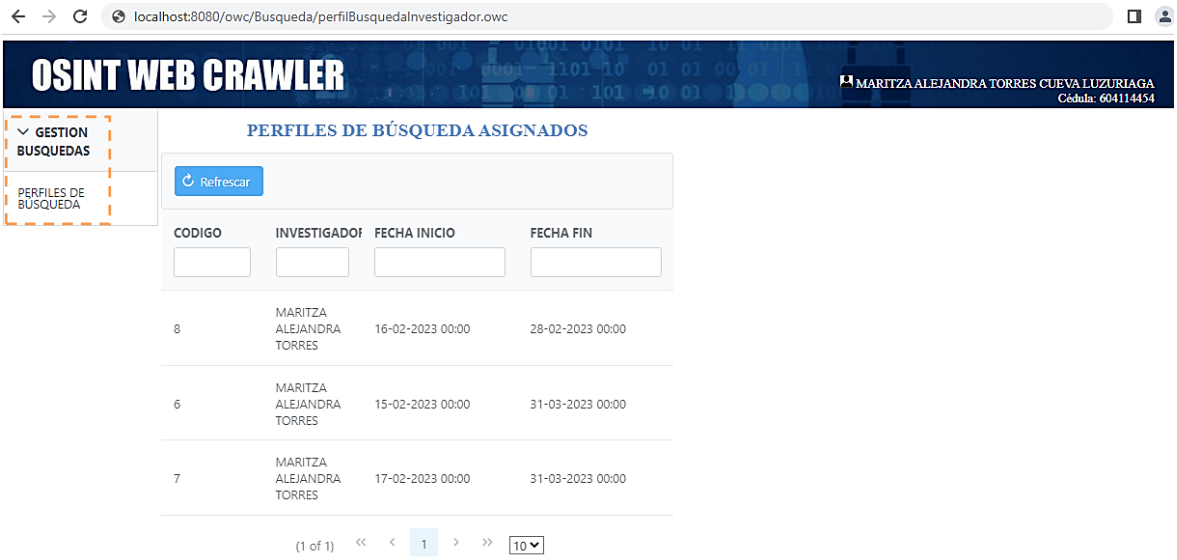


Figura 59 – Acceso a recursos configurados para el rol Investigador. **[Elaboración propia]**

Sin embargo, este ha logrado hacerse de las URL que corresponden a funcionalidades de rol de Facilitador e intenta ingresar a ellas aprovechando la sesión activa que pose.

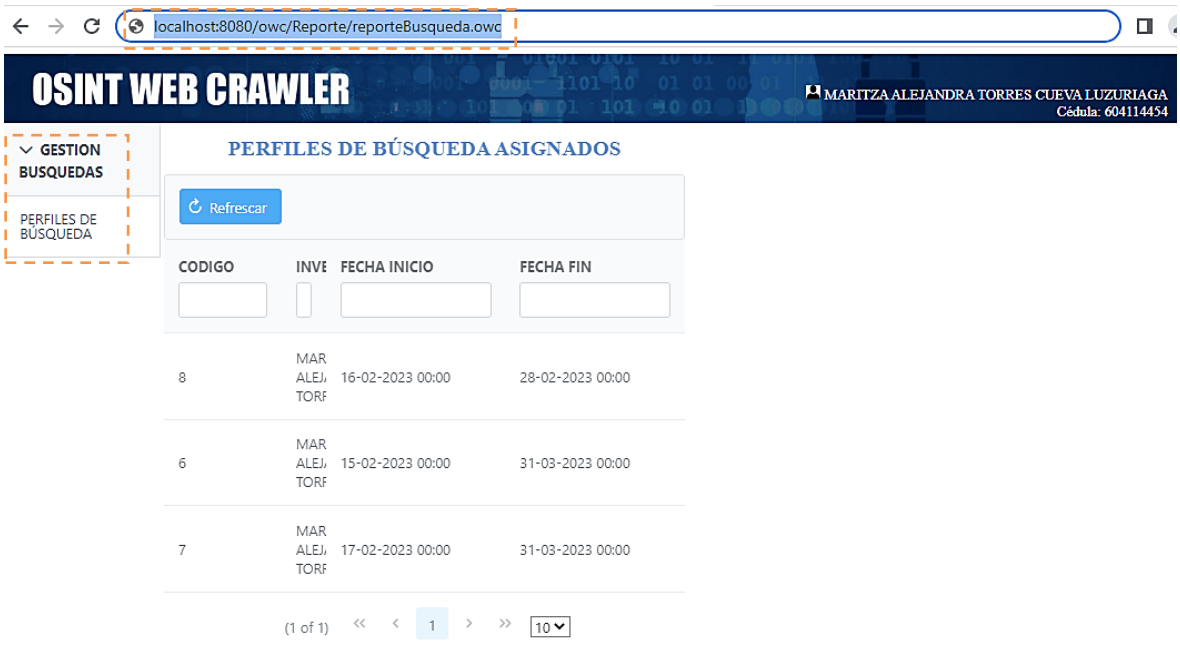


Figura 60 – Intento de acceso a un recurso de otro perfil. **[Elaboración propia]**

Figura 61 – Denegación de acceso a un recurso no autorizado. **[Elaboración propia]**

Como se puede ver en la Figura 60 el atacante ingresa una URL que pertenecea a una pantalla del usuario Facilitador ante lo cual el sistema hace efectiva la autorización de recursos para cada petición mostrando una pantalla de acceso restringido, de lo cual se demuestra la resiliencia del sistema ante este tipo de ataques .

Al mismo tiempo que el atacante intenta vulnerar al sistema este va registrando información referente a los recurso que autoriza o deniega tal como se puede ver en la Figura 62 que existen 2 recursos autorizados y uno denegado, este último corresponde al intento de acceso a la página de Facilitador el cual era su objetivo.

server.log

```
[2023-03-16T21:51:18.211-0500] [Payara 5.2021.8] [INFORMACIÓN] [] [com.owc.seguridad.FiltroSeguridad] [tid: _ThreadID=119
  Usuario: 604114454; Perfil: INVESTIGADOR; Recurso: /owc/Busqueda/perfilBusquedaInvestigador.owc- AUTORIZADO]]

[2023-03-16T21:51:18.239-0500] [Payara 5.2021.8] [INFORMACIÓN] [] [com.owc.seguridad.FiltroSeguridad] [tid: _ThreadID=632
  Usuario: 604114454; Perfil: INVESTIGADOR; Recurso: /owc/Busqueda/busqueda.owc- AUTORIZADO]]

[2023-03-16T21:51:18.355-0500] [Payara 5.2021.8] [INFORMACIÓN] [] [] [tid: _ThreadID=632 _ThreadName=http-thread-pool::htt
  constructor]

[2023-03-16T21:51:18.888-0500] [Payara 5.2021.8] [INFORMACIÓN] [] [com.owc.seguridad.FiltroSeguridad] [tid: _ThreadID=118
  Usuario: 604114454; Perfil: INVESTIGADOR; Recurso: /owc/Busqueda/perfilBusquedaInvestigador.owc- AUTORIZADO]]

[2023-03-16T21:51:18.933-0500] [Payara 5.2021.8] [INFORMACIÓN] [] [com.owc.seguridad.FiltroSeguridad] [tid: _ThreadID=630
  Usuario: 604114454; Perfil: INVESTIGADOR; Recurso: /owc/Busqueda/busqueda.owc- AUTORIZADO]]

[2023-03-16T21:51:19.080-0500] [Payara 5.2021.8] [INFORMACIÓN] [] [] [tid: _ThreadID=630 _ThreadName=http-thread-pool::htt
  constructor]]
[2023-03-16T21:51:20.029-0500] [Payara 5.2021.8] [INFORMACIÓN] [] [com.owc.seguridad.FiltroSeguridad] [tid: _ThreadID=631
  Usuario: 604114454; Perfil: INVESTIGADOR; Recurso: /owc/Busqueda/perfilBusquedaInvestigador.owc- AUTORIZADO]]

[2023-03-16T21:51:20.054-0500] [Payara 5.2021.8] [INFORMACIÓN] [] [com.owc.seguridad.FiltroSeguridad] [tid: _ThreadID=119
  Usuario: 604114454; Perfil: INVESTIGADOR; Recurso: /owc/Busqueda/busqueda.owc- AUTORIZADO]]

[2023-03-16T21:51:20.087-0500] [Payara 5.2021.8] [INFORMACIÓN] [] [] [tid: _ThreadID=119 _ThreadName=http-thread-pool::htt
  constructor]]

[2023-03-16T21:51:21.251-0500] [Payara 5.2021.8] [INFORMACIÓN] [] [com.owc.seguridad.FiltroSeguridad] [tid: _ThreadID=631
  Usuario: 604114454; Perfil: INVESTIGADOR; Recurso: /owc/Busqueda/perfilBusquedaInvestigador.owc- AUTORIZADO]]

[2023-03-16T21:51:23.631-0500] [Payara 5.2021.8] [INFORMACIÓN] [] [com.owc.seguridad.FiltroSeguridad] [tid: _ThreadID=630
  Usuario: 604114454; Perfil: INVESTIGADOR; Recurso: /owc/Busqueda/perfilBusquedaInvestigador.owc- AUTORIZADO]]

[2023-03-16T21:51:29.652-0500] [Payara 5.2021.8] [INFORMACIÓN] [] [com.owc.seguridad.FiltroSeguridad] [tid: _ThreadID=632
  Usuario: 604114454; Perfil: INVESTIGADOR; Recurso: /owc/Busqueda/busqueda.owc- AUTORIZADO]]

[2023-03-16T21:51:39.298-0500] [Payara 5.2021.8] [INFORMACIÓN] [] [com.owc.seguridad.FiltroSeguridad] [tid: _ThreadID=119
  Usuario: 604114454; Perfil: INVESTIGADOR; Recurso: /owc/Administrador/Gestion/gestionPerfil.owc- DENEGADO]]
```

Figura 62 – Registro de eventos de autorización. **[Elaboración propia]**

Como se pudo observar el sistema en primera instancia autorizó el acceso a un recurso sobre el cual el usuario tenía privilegios y posteriormente se intentó acceder a funcionalidades de un rol distinto mediante el ingreso manual de la URL ante lo cual es sistema fue resiliente y denegó el acceso, además todos estos eventos de autorización fueron registrados en logs de auditoría. Con esto se puede evidenciar que el módulo de control de acceso es capaz de:

- Validar el acceso a un recurso solicitado en cada petición realizada de acuerdo a las restricciones configuradas.
- Llevar un registro de auditoría de los recursos solicitados.

Con esto se verifica el cumplimiento de los requerimientos críticos para este escenario de ataque.

5.6. Discusión de resultados

En esta sección se realiza un análisis de los resultados obtenidos de las evaluaciones realizadas en el capítulo anterior.

5.6.1. Evaluación funcional

La evaluación funcional basada en un caso de estudio permitió llevar a la práctica los lineamientos y requerimientos planteados por el método propuesto en cada fase en donde se pudo evidenciar ciertos aspectos que se discuten a continuación.

Adquisición

En esta fase se partió de la premisas e hipótesis del caso de estudio plasmadas en una tabla de búsquedas la cual sirvió para la configuración de perfiles de búsqueda. En este punto se evidenció que los controles dados por el método soportan totalmente las necesidades para realizar una búsqueda mientras se preserva la privacidad en esta instancia de la investigación.

Las limitaciones identificadas en la implementación de ciertos controles en el crawler fueron el determinar en qué año en fue publicado un sitio ya que al no tener un histórico indexado de sitios como Google, aplicar este filtro se vuelve más complejo y se limita a metadatos que defina el sitio. Otra limitación concierne a la restricción de dominio de búsqueda ya que al partir de un dominio establecido en ciertos casos los recursos multimedia embebidos de un sitio proceden de un subdominio o de sitios ajenos al dominio original los cuales en primera instancia deben ser restringidos para ser ignorados

en la fase de recolección, provocando que el sitio padre no se almacene con la totalidad de sus recursos. Sin embargo, se restringieron únicamente los recursos de dominios ajenos.

Recolección

La recolección de información se realizó de acuerdo a las configuraciones definidas para cada perfil de búsqueda, en donde se pudo observar que la implementación de estos controles permite filtrar y almacenar únicamente la data deseada como es origen, formato y volumen de la data, así como formatos y demás restricciones como se puede ver en la Tabla 40 que resume los resultados de búsqueda obtenidos con el prototipo y los obtenidos con un navegador convencional donde el volumen es mucho menor y la información se halla más filtrada lo cual cumple con lo requerido.

Una limitación hallada fue la precisión de las búsquedas ya que al buscar por coincidencias y expresiones regulares se pueden omitir ciertos sitios importantes para la investigación que con técnicas especializadas de inteligencia artificial podrían ser identificados, evidentemente superar esta limitación sobrepasa el alcance del presente trabajo sin embargo deja brecha para un trabajo futuro en este sentido.

Otro aspecto a tomar en cuenta en materia de seguridad es el análisis de los recursos binarios recuperados de las fuentes de información ya que bien se podría estar almacenando cualquier tipo de malware camuflado de un recurso inofensivo el cual estallaría el ejecutarlo.

Preservación

Preservar la privacidad de la información almacenada está directamente ligada a la eficacia de los métodos de seguridad definidos para lograr este objetivo. El nivel de granularidad de la política de base de datos implementada es de tabla, usuario y operación como se puede ver en la Tabla 61 la cual se pudo validar frente a las funcionalidades y necesidades de la aplicación en el caso de estudio durante la cual no existieron conflictos en los privilegios definidos. A nivel de aplicación el esquema de recursos de conexión de la Figura 47 definido para cada módulo finalmente permitió la implementación de las funcionalidades del prototipo asegurando el nivel de acceso adecuado.

El nivel de granularidad que da el método de control de acceso RBAC Implementado en la solución satisface las necesidades de configuración para especificar reglas de la política definida. Sin embargo, de requerirse controles más minuciosos se puede escalar a nivel de roles con el fin de no afectar los ya existentes y entrar en un conflicto de reglas.

5.6.2. Evaluación de desempeño

Tras ejecutar las pruebas correspondientes y registrar los resultados en la Tabla 47 se puede analizar que el nivel de consumo del recurso es directamente proporcional al volumen de carga soportando en cada escenario y de igual manera el tiempo de respuesta mantiene esta relación. Cabe mencionar que el nivel de profundidad es de 2 de acuerdo a la Figura 25 y bajo los recursos de hardware definidos previo a la ejecución de las pruebas.

Los aspectos relevantes a tomar en cuenta y que influyen en los resultados son:

- Técnica de búsqueda de coincidencias basada en expresiones regulares la cual afecta el tiempo de respuesta y consumo de CPU.
- Tamaño de archivos analizados, a mayor tamaño mayor consumo de memoria.

De acuerdo a lo expuesto se infiere que para búsquedas de mayor profundidad y numero paginas procesadas se requieren mayores prestaciones de hardware principalmente en cuanto a memoria ya que es el recurso que mayor demanda la solución implementada.

5.6.3. Evaluación de resiliencia

De la evaluación realizada se pueden discutir varios aspectos entre ellos el modelado del adversario, el cual resulto de mucha utilidad para definir las defensas más eficientes y de bajo costo de implementación, además definió el contexto y perfil del adversario. Algo importante es tener claros los supuestos (infraestructura, despliegue de la solución, tipos de usuario, etc.), ya que esto permite descartar controles innecesarios que se da por hecho que se tiene cubiertos y enfocarse en lo que realmente importa.

Por otro lado, el modelado de amenazas realizado con Microsoft Threat Modeling Tool permitió dar un universo de posibilidades más amplio lo cual fue positivo ya que incremento la lista de posibles ataques que posteriormente se complementó con el modelo del adversario permitiendo definir los módulos más críticos del sistema.

Tras definir los requerimientos críticos y ejecutar los escenarios de ataque se pudo evidenciar que las implementaciones de los controles definidos en el método propuesto pueden reducir el riesgo de ciertas amenazas a niveles aceptables como por ejemplo amenazas relacionadas con elevación de privilegios y manipulación de datos lo cual afecta directamente al cumplimiento de uno de los objetivos del método que es preservar la privacidad de la data almacenada.

Por otro lado, se pudo verificar que la implementación del método de control de acceso implementado de manera transversal desde la base de datos hasta el módulo de autorización permite al sistema ser resiliente en distintos escenarios de ataque.

Además de las amenazas evaluadas en los escenarios de ataque existen otras de menor probabilidad y criticidad que fueron mitigadas mediante técnicas de desarrollo seguro y herramientas de análisis de vulnerabilidades de código estático a lo largo de la implementación del prototipo como se puede ver en el Anexo VI.

6. CONCLUSIONES Y RECOMENDACIONES

6.1. Conclusiones

- Se llevó a cabo una revisión de literatura que permitió conocer el estado del arte en cuanto a la aplicación efectiva de métodos preservación de privacidad en investigaciones digitales forenses, para esto se utilizaron elementos de la declaración PRISMA que permitieron realizar una revisión más transparente y objetiva lo cual se encuentra documentado mediante la Tabla 11 y Tabla 12. En cuanto a los métodos analizados se concluye lo siguiente:
    a. La mayoría de métodos recurren a la utilización de sistemas de control de acceso como primera medida para preservar la privacidad de la data almacenada, esto se puede ver en la Tabla 13 donde 8 de los 15 revisados implementan métodos control de acceso.
    b. Las técnicas más utilizadas para preservar la privacidad son métodos de control de acceso, criptografía e inteligencia artificial.
    c. Las técnicas de inteligencia artificial son implementadas para preservar la privacidad en las fases de recolección y análisis mediante patrones de reconocimiento de datos sensibles y la explotación de data recolectada.
    d. Las técnicas criptográficas son ampliamente recomendadas a lo largo del ciclo de vida de la información recolectada en una investigación, ya sea para garantizar la privacidad e integridad. Sin embargo, el alto costo computacional limita su uso dado su costo beneficio.
- Se realizó un análisis de las técnicas y enfoques de fuentes abiertas para obtener rastros digitales de personas de interés de lo cual se identificaron 3 enfoques que son: Opinión social y análisis de sentimientos, Ciberdelincuencia y delincuencia organizada y Ciberseguridad y ciberdefensa como se puede ver en la Figura 3. Estos enfoques permitieron abordar el caso de estudio de manera más acertada y definir un plan de búsquedas más completo que permita realizar una evaluación funcional acorde a los enfoques más utilizado actualmente en prácticas OSINT. El plan de búsquedas completo se lo puede ver en el Anexo V.

- Se identificó que las técnicas más utilizadas en prácticas OSINT son las basadas en direcciones de correo electrónico, nombre de usuario, nombre real, localización geográfica, direcciones IP y nombre de dominio, esta categorización sirvió de insumo para definir funcionalidades clave como: búsqueda avanzada,

implementación de búsqueda por expresiones regulares, definición de dominios entre otras definidas en los requerimientos funcionales.

- Se utilizó PRECEPT y otros enfoques de preservación de privacidad como privacidad por diseño y controles proactivos para la implementación del prototipo funcional que permitieron una adecuada gestión de la evidencia digital recolectada. En esta implementación se llevó a cabo lo siguiente:

    a) Se realizó un mapeo de fases de los ciclos forense y OSINT junto con principios de privacidad de los cuales se generaron requerimientos funcionales que se implementaron en el prototipo de software.
    b) Mediante privacidad por diseño de implemento un método de control de acceso transversal desde el nivel de almacenamiento de datos hasta la autorización de recurso en la aplicación, mismo que fue evaluado mediante el cumplimiento requerimientos críticos frente a escenarios de ataque.
    c) El enfoque legal en cuanto a preservación de la privacidad de la información dado por la LOPD también contribuyo al método mediante los artículos mapeados en el Anexo II.

- La aplicabilidad del método propuesto se evaluó desde 3 puntos de vista que son el aspecto funcional, el desempeño y la resiliencia frente escenarios de ataque. En lo que concierne a la parte funcional se pudo evidenciar que en la fase de Adquisición los controles definidos para realizar las búsquedas brindaron el soporte suficiente para delimitar elementos de privacidad como origen de información, volumen de información, formatos requeridos, entre otros que se resumen en la Figura 21. De la misma manera en la fase de Recolección se evidenció que la efectividad de las restricciones configuradas logra delimitar ´la información a únicamente lo necesario, estos se pueden corroborar cuantitativamente contra los datos recolectados con un navegador convencional en la Tabla 40. Por otro lado, en la fase de Preservación el método de control de acceso RBAC implementado junto a principios como segregación de roles, validación en cada petición y autorización dinámica permitieron cumplir con los resultados esperado definidos en cada escenario de ataque.
- En cuanto a evaluación de desempeño se evidenció de acuerdo al resumen de la Tabla 47 que el recurso provoca el colapso del sistema es la memoria, esto se debe al gran número de páginas cargadas en memoria para las validaciones correspondientes.

- Para la evaluación de resiliencia se realizó un modelo basado en el adversario del cual se identificaron y priorizaron 11 amenazas y defensas como se puede ver en la Tabla 53 las cuales fueron consideradas para el definir los requerimientos críticos de la Tabla 58 validados en escenarios de ataque.
- Se realizó un modelado de amenazas con Microsoft Threat Modelling Tool de la solución implementada donde se obtuvieron 16 amenazas de las cuales se realizó un análisis y control de riesgo mediante los requerimientos de privacidad del método propuesto y se logró controlar el riesgo reduciéndolo a niveles aceptables en 10 amenazas tal como se puede ver en la Tabla 54 y el Anexo I.

6.2. Recomendaciones

- Un aspecto sobre el cual se recomienda un estudio a profundidad es la identificación de información sensible mediante patrones durante la búsqueda, ya que la detección y etiquetado de la información tempranamente hace posible un tratamiento más adecuado durante la investigación.
- Se recomienda que para la implementación de la presente solución y similares se cuente con altas prestaciones de hardware tales como memoria y procesamiento tomando como referencia los recursos utilizados en la evaluación de desempeño realizada.
- En lo referente a recuperación de información de fuentes desconocidas se recomienda implementar controles de seguridad adicionales para detectar posible malware camuflado de recursos disfrazados de imágenes, archivos, etc.
- En cuanto a lo legal se recomienda siempre tener en cuenta la legislación vigente para la adecuada gestión de la información recolectada con prácticas OSINT ya que constantemente se incorporan nuevas leyes que se pudiesen estar violando.

REFERENCIAS BIBLIOGRÁFICAS

[1] R. I. Ferguson, K. Renaud, S. Wilford, and A. Irons, "PRECEPT: a framework for ethical digital forensics investigations," *Journal of Intellectual Capital*, vol. 21, no. 2, pp. 257–290, May 2020, doi: 10.1108/JIC-05-2019-0097.

[2] P. Langmead, "Comparative Evaluation of Access Control Models," 2022.

[3] N. A. Hassan, "Gathering Evidence from OSINT Sources," in *Digital Forensics Basics*, Apress, 2019, pp. 311–322. doi: 10.1007/978-1-4842-3838-7_10.

[4] M. B. Arroyo, "Las conexiones secretas de Odebrecht en Ecuador," *#PerDebate*, vol. 3, Oct. 2019, doi: 10.18272/pd.v3i1.1555.

[5] el Parlamento Europeo, "rgpd," 2016.

[6] Ley, "LEY ORGÁNICA DE PROTECCIÓN DE DATOS PERSONALES," 2021. [Online]. Available: www.lexis.com.ec

[7] Robert André Furuhaug, "Open Source Intelligence Methodology," 2019.

[8] I. Dobák and T. Tóth, "Social Engineering," *Belügyi Szemle*, vol. 69, no. 2, pp. 195–212, Feb. 2021, doi: 10.38146/bsz.2021.2.2.

[9] H. Gibson, "Acquisition and preparation of data for OSINT investigations," in *Advanced Sciences and Technologies for Security Applications*, Springer, 2016, pp. 69–93. doi: 10.1007/978-3-319-47671-1_6.

[10] J. Pastor-Galindo, P. Nespoli, F. Gomez Marmol, and G. Martinez Perez, "The not yet exploited goldmine of OSINT: Opportunities, open challenges and future trends," *IEEE Access*, vol. 8, pp. 10282–10304, 2020, doi: 10.1109/ACCESS.2020.2965257.

[11] F. Buccafurri, V. de Angelis, and M. F. Idone, "Implementing Multiple-Social-Network Meta-APIs to Support OSINT Programming," in *2020 12th International Conference on Advanced Infocomm Technology (ICAIT)*, IEEE, Nov. 2020, pp. 124–128. doi: 10.1109/ICAIT51223.2020.9315457.

[12] D. chalicheemala and D. chalicheemala, "What is Open-Source Intelligence and How it Can Prevent Frauds," *SSRN Electronic Journal*, 2022, doi: 10.2139/ssrn.4170882.

[13] B. Senekal and E. Kotzé, "Open source intelligence (OSINT) for conflict monitoring in contemporary South Africa: Challenges and opportunities in a big data context," *African Security Review*, vol. 28, no. 1, pp. 19–37, Jan. 2019, doi: 10.1080/10246029.2019.1644357.

[14] A. Majeed, S. Khan, and S. O. Hwang, “A Comprehensive Analysis of Privacy-Preserving Solutions Developed for Online Social Networks,” *Electronics (Switzerland)*, vol. 11, no. 13. MDPI, Jul. 01, 2022. doi: 10.3390/electronics11131931.

[15] M. M. Toro-Alvarez, M. Leonardo, and B. Duitama, “Investigación-del-Cibercrimen y Delitos-Informáticos Utilizando OSINT Cybercrime Project View project Applied cybercriminology for cyberpolicing improvement and counteracting cybersecurity threats. View project,” 2018, doi: 10.13140/RG.2.2.21594.59849.

[16] I. Vacas, I. Medeiros, and N. Neves, “Detecting Network Threats using OSINT Knowledge-Based IDS,” in *2018 14th European Dependable Computing Conference (EDCC)*, IEEE, Sep. 2018, pp. 128–135. doi: 10.1109/EDCC.2018.00031.

[17] Y. al Mahmeed, W. Elmedany, and M. S. Sharif, “Eagle-Eye: Open-Source Intelligence Tool for IoT Devices Detection,” in *2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, IEEE, Nov. 2022, pp. 526–530. doi: 10.1109/3ICT56508.2022.9990658.

[18] K. Uehara *et al.*, “Basic Study on Targeted E-mail Attack Method Using OSINT,” 2020, pp. 1329–1341. doi: 10.1007/978-3-030-15032-7_111.

[19] M. J. Page *et al.*, “Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas,” *Revista Española de Cardiología (English Edition)*, vol. 74, no. 9, pp. 790–799, Sep. 2021, doi: 10.1016/j.rec.2021.07.010.

[20] R. Verma and G. Gupta, “Perception of Data Privacy in Digital Forensic Investigation,” 2018.

[21] M. Abulaish, N. A. H. Haldar, and J. Jahiruddin, “P2DF: A Privacy-Preserving Digital Forensics Framework,” *International Journal of Digital Crime and Forensics*, vol. 13, no. 6, pp. 1–15, Nov. 2021, doi: 10.4018/ijdcf.288547.

[22] W. Halboob, R. Mahmod, N. I. Udzir, and M. D. T. Abdullah, “Privacy levels for computer forensics: Toward a more efficient privacy-preserving investigation,” in *Procedia Computer Science*, Elsevier B.V., 2015, pp. 370–375. doi: 10.1016/j.procs.2015.07.222.

[23] R. Verma, J. Govindaraj, and G. Gupta, “DF 2.0: DESIGNING AN AUTOMATED, PRIVACY PRESERVING, AND EFFICIENT DIGITAL FORENSIC FRAMEWORK,” 2018. [Online]. Available: https://commons.erau.edu/adfsl

[24] B. J. Koops, J. H. Hoepman, and R. Leenes, “Open-source intelligence and privacy by design,” *Computer Law and Security Review*, vol. 29, no. 6, pp. 676–688, Dec. 2013, doi: 10.1016/j.clsr.2013.09.005.

[25] J. Pathak, S. Sankaran, and K. Achuthan, “A SMART Goal-based Framework for Privacy Preserving Embedded Forensic Investigations,” in *Proceedings of the 2019 International Symposium on Embedded Computing and System Design, ISED 2019*, Institute of Electrical and Electronics Engineers Inc., Dec. 2019, pp. 6–10. doi: 10.1109/ISED48680.2019.9096232.

[26] Sri Shakthi Institute of Engineering and Technology, Institute of Electrical and Electronics Engineers. Madras Section, All-India Council for Technical Education, and Institute of Electrical and Electronics Engineers, *Investigation on Privacy Preserving using K-Anonymity Techniques*. 2020.

[27] Cadre privé, “Information technology-Security techniques-Privacy framework,” 2011.

[28] I. Jayaraman and A. Stanislaus Panneerselvam, “A novel privacy preserving digital forensic readiness provable data possession technique for health care data in cloud,” *J Ambient Intell Humaniz Comput*, vol. 12, no. 5, pp. 4911–4924, May 2021, doi: 10.1007/s12652-020-01931-1.

[29] L. Englbrecht and G. Pernul, “A privacy-aware digital forensics investigation in enterprises,” in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Aug. 2020. doi: 10.1145/3407023.3407064.

[30] D. Billard and B. Bartolomei, “Digital Forensics and Privacy-by-Design: Example in a Blockchain-Based Dynamic Navigation System,” 2019, pp. 151–160. doi: 10.1007/978-3-030-21752-5_10.

[31] Y. Chen, T. Qiao, F. Retraint, and G. Hu, “Efficient Privacy-Preserving Forensic Method for Camera Model Identification,” *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 2378–2393, 2022, doi: 10.1109/TIFS.2022.3185769.

[32] R. Verma, “Digital Forensics 2.0: an automated, efficient, and privacy preserving digital forensic investigation framework,” 2018.

[33] S. Naqvi *et al.*, “Privacy-Preserving Social Media Forensic Analysis for Preventive Policing of Online Activities,” in *2019 10th IFIP International Conference on New*

Technologies, Mobility and Security (NTMS), IEEE, Jun. 2019, pp. 1–6. doi: 10.1109/NTMS.2019.8763830.

[34] N. A. Hassan, "Gathering Evidence from OSINT Sources," in *Digital Forensics Basics*, Berkeley, CA: Apress, 2019, pp. 311–322. doi: 10.1007/978-1-4842-3838-7_10.

[35] M. Ahmed, S. Reno, N. Akter, and F. Haque, "Securing Medical Forensic System Using Hyperledger Based Private Blockchain," in *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, IEEE, Dec. 2020, pp. 1–6. doi: 10.1109/ICCIT51783.2020.9392686.

[36] U. Breidenbach, M. Steinebach, and H. Liu, "Privacy-enhanced robust image hashing with bloom filters," in *Proceedings of the 15th International Conference on Availability, Reliability and Security*, New York, NY, USA: ACM, Aug. 2020, pp. 1–10. doi: 10.1145/3407023.3409212.

[37] M. Steinebach, S. Lutz, and H. Liu, "Privacy and Robust Hashes," in *Proceedings of the 14th International Conference on Availability, Reliability and Security*, New York, NY, USA: ACM, Aug. 2019, pp. 1–8. doi: 10.1145/3339252.3340105.

[38] N. Andola, S. Prakash, V. K. Yadav, Raghav, S. Venkatesan, and S. Verma, "A secure searchable encryption scheme for cloud using hash-based indexing," *J Comput Syst Sci*, vol. 126, pp. 119–137, Jun. 2022, doi: 10.1016/j.jcss.2021.12.004.

[39] A. Varanda, L. Santos, R. L. de C. Costa, A. Oliveira, and C. Rabadão, "Log pseudonymization: Privacy maintenance in practice," *Journal of Information Security and Applications*, vol. 63, p. 103021, Dec. 2021, doi: 10.1016/j.jisa.2021.103021.

[40] L. Zhang, C. Zhao, Q. Wu, Y. Mu, and F. Rezaeibagha, "A traceable and revocable multi-authority access control scheme with privacy preserving for mHealth," *Journal of Systems Architecture*, vol. 130, p. 102654, Sep. 2022, doi: 10.1016/j.sysarc.2022.102654.

[41] M. Ramzan, M. Habib, and S. A. Khan, "Secure and efficient privacy protection system for medical records," *Sustainable Computing: Informatics and Systems*, vol. 35, p. 100717, Sep. 2022, doi: 10.1016/j.suscom.2022.100717.

[42] K. K. Mojtaba Mohamamdi, "Analysis of Common Access Control Models and Their Limitations in Cloud Computing Environment," 2015.

[43] M. Mulimani and R. Rachh, “Analysis of Access Control Methods in Cloud Computing,” 2016, doi: 10.20944/preprints201607.0012.v1.

[44] V. C. Hu and K. Scarfone, “Guidelines for Access Control System Evaluation Metrics,” Gaithersburg, MD, Sep. 2012. doi: 10.6028/NIST.IR.7874.

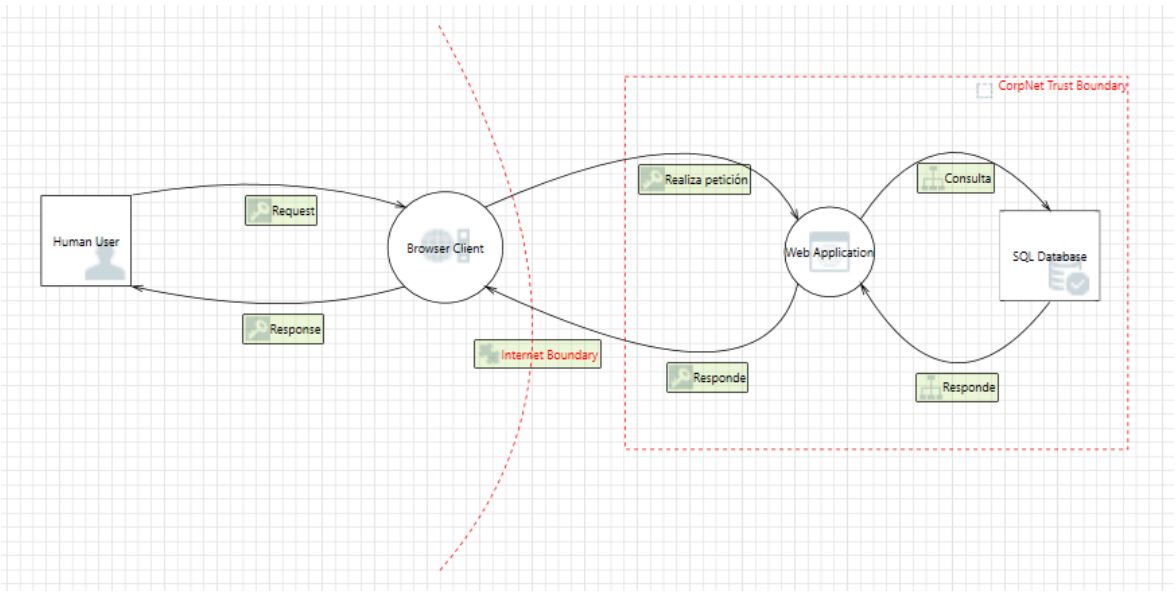
[45] M. Liu, C. Yang, H. Li, and Y. Zhang, “An Efficient Attribute-Based Access Control (ABAC) Policy Retrieval Method Based on Attribute and Value Levels in Multimedia Networks,” *Sensors*, vol. 20, no. 6, p. 1741, Mar. 2020, doi: 10.3390/s20061741.

[46] M. umar Aftab, Z. Qin, Zakria, S. Ali, Pirah, and J. Khan, “The Evaluation and Comparative Analysis of Role Based Access Control and Attribute Based Access Control Model,” in *2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, IEEE, Dec. 2018, pp. 35–39. doi: 10.1109/ICCWAMTIP.2018.8632578.

[47] M. M. Yamin, M. Ullah, H. Ullah, B. Katt, M. Hijji, and K. Muhammad, “Mapping Tools for Open Source Intelligence with Cyber Kill Chain for Adversarial Aware Security,” *Mathematics*, vol. 10, no. 12, p. 2054, Jun. 2022, doi: 10.3390/math10122054.

ANEXOS

Anexo I - Modelado de amenazas con Microsoft Threat Modeling Tool



Threat List

ID	Title	Priority
83	Spoofing the Human User External Entity	High
87	Elevation Using Impersonation	High
198	Spoofing the Browser Client Process	High
199	Browser Client Process Memory Tampered	High
200	Cross Site Scripting	High
201	Potential Data Repudiation by Web Application	High
202	Potential Process Crash or Stop for Web Application	High
203	Data Flow Realiza petición Is Potentially Interrupted	High
204	Elevation Using Impersonation	High
205	Web Application May be Subject to Elevation of Privilege Using Remote Code Execution	High
206	Elevation by Changing the Execution Flow in Web Application	High
207	Spoofing the Web Application Process	High
208	Web Application Process Memory Tampered	High
209	Potential Data Repudiation by Browser Client	High
210	Potential Process Crash or Stop for Browser Client	High
211	Data Flow Responde Is Potentially Interrupted	High
212	Elevation Using Impersonation	High
213	Browser Client May be Subject to Elevation of Privilege Using Remote Code Execution	High
214	Elevation by Changing the Execution Flow in Browser Client	High
220	Spoofing of Source Data Store SQL Database	High
222	Cross Site Scripting	High
223	Persistent Cross Site Scripting	High
224	Weak Access Control for a Resource	High
225	Spoofing of Destination Data Store SQL Database	High
227	Potential SQL Injection Vulnerability for SQL Database	High
233	Potential Excessive Resource Consumption for Web Application or SQL Database	High
241	Risks from Logging	High
242	Risks from Logging	High
243	Lower Trusted Subject Updates Logs	High
244	Data Logs from an Unknown Source	High
245	Insufficient Auditing	High
246	Potential Weak Protections for Audit Data	High
247	Weak Credential Storage	High

Nro	Amenaza	Categoría	Descripción			Antes de Tratamiento (Riesgo Inherente)			Tratamiento		Después del Tratamiento (Riesgo Residual)		
				Activo	Responsable	Impacto	Probabilidad	Riesgo	Opción (1,2,3,4)	Control (si opcion = 1)	Impacto	Probabilidad	Riesgo
A_02	Elevation Using Impersonation	Elevation Of Privilege	Browser Client may be able to impersonate the context of Human User in order to gain additional privilege.	Application	Web Master	2	2	4	1	PR_14 (Autenticación)	1	1	2
A_05	Cross Site Scripting	Tampering	The web server 'Web Application' could be a subject to a cross-site scripting attack because it does not sanitize untrusted input.	Application	Web Master	2	2	4	1	PR_14 (Token de sesión)	0	1	1
A_06	Potential Data Repudiation by Web Application	Repudiation	Web Application claims that it did not receive data from a source outside the trust boundary. Consider using logging or auditing to record the source, time, and summary of the received data.	Application	Web Master	1	2	3	1	PR_14 (Firma digital, Reg uditoria, Logs) PR_15 PR_17	1	1	2
A_07	Potential Process Crash or Stop for Web Application	Denial Of Service	Web Application crashes, halts, stops or runs slowly; in all cases violating an availability metric.	Application	Web Master	1	2	3	1	PR_02 PR_03 PR_04	1	1	2
A_12	Spoofing the Web Application Process	Spoofing	Web Application may be spoofed by an attacker and this may lead to unauthorized access to Browser Client. Consider using a standard authentication mechanism to identify the source process.	Application	TI Department	1	2	3	2		1	1	2
A_21	Risks from Logging	Tampering	Log readers can come under attack via log files. Consider ways to canonicalize data in all logs. Implement a single reader for the logs, if possible, in order to reduce attack surface area. Be sure to understand and document log file elements which come from untrusted sources.	Data Base	* DB Administrator * Infraestructure Administrator	2	2	4	1	PR_14 (Controlde acceso)	1	1	2
A_22	Cross Site Scripting	Tampering	The web server 'Web Application' could be a subject to a cross-site scripting attack because it does not sanitize untrusted input.	Application	Web Master	1	2	3	1	PR_14 (Filtros de entrada)	0	1	1
A_24	Weak Access Control for a Resource	Information Disclosure	Improper data protection of SQL Database can allow an attacker to read information not intended for disclosure. Review authorization settings.	Data Base	DB Administrator	2	2	4	1	PR_14 (Control de acceso, encripatción)	1	1	0
A_25	Spoofing of Destination Data Store SQL Database	Spoofing	SQL Database may be spoofed by an attacker and this may lead to data being written to the attacker's target instead of SQL Database. Consider using a standard authentication mechanism to identify the destination data store.	Application	TI Department	1	1	2	1	PR_14 (4eg, auditoria, Logs)	1	1	2
A_27	Potential SQL Injection Vulnerability for SQL Database	Tampering	SQL injection is an attack in which malicious code is inserted into strings that are later passed to an instance of SQL Server for parsing and execution. Any procedure that constructs SQL statements should be reviewed for injection vulnerabilities because SQL Server will execute all syntactically valid queries that it receives. Even parameterized data can be manipulated by a skilled and determined attacker.	Application	Web Master	2	2	4	1	PR_14 (Filtro de entrada)	1	1	2

A_28	Lower Trusted Subject Updates Logs	Repudiation	If you have trust levels, is anyone other outside of the highest trust level allowed to log? Letting everyone write to your logs can lead to repudiation problems. Only allow trusted code to log.	Data Base		1	3	4	1	PR_14 (Control de acceso)	1	0	1
A_29	Data Logs from an Unknown Source	Repudiation	Do you accept logs from unknown or weakly authenticated users or systems? Identify and authenticate the source of the logs before accepting them.	* Data Base * Server	* DB Administrator * Infraestructure Administrator	1	1	2	1	PR_14 (Control de acceso)	0	0	0
A_30	Insufficient Auditing	Repudiation	Does the log capture enough data to understand what happened in the past? Do your logs capture enough data to understand an incident after the fact? Is such capture lightweight enough to be left on all the time? Do you have enough data to deal with repudiation claims? Make sure you log sufficient and appropriate data to handle a repudiation claims. You might want to talk to an audit expert as well as a privacy expert about your choice of data.	* Data Base *	* DB Administrator * Infraestructure Administrator	1	2	3	1	PR_14 (Reg auditoria, Logs) PR_15 PR_17	0	1	1
A_31	Potential Weak Protections for Audit Data	Repudiation	Consider what happens when the audit mechanism comes under attack, including attempts to destroy the logs, or attack log analysis programs. Ensure access to the log is through a reference monitor, which controls read and write separately. Document what filters, if any, readers can rely on, or writers should expect	* Data Base * Server	* DB Administrator * Infraestructure Administrator	1	1	2	2		1	0	1
A_32	Weak Credential Storage	Information Disclosure	Credentials held at the server are often disclosed or tampered with and credentials stored on the client are often stolen. For server side, consider storing a salted hash of the credentials instead of storing the credentials themselves. If this is not possible due to business requirements, be sure to encrypt the credentials before storage, using an SDL-approved mechanism. For client side, if storing credentials is required, encrypt them and protect the data store in which they're stored	Data Base	DB Administrator	2	2	4	1	PR_14 (Encriptación)	0	1	1
A_33	Potential Excessive Resource Consumption for Web Application or SQL Database	Denial Of Service	Does Web Application or SQL Database take explicit steps to control resource consumption? Resource consumption attacks can be hard to deal with, and there are times that it makes sense to let the OS do the job. Be careful that your resource requests don't deadlock, and that they do timeout.	Data Base	DB Administrator	1	0	1	1	PR_02 PR_03 PR_04 PR_06	0	0	0

Anexo II - Mapeo de fases Precept Osint, investigación digital forense

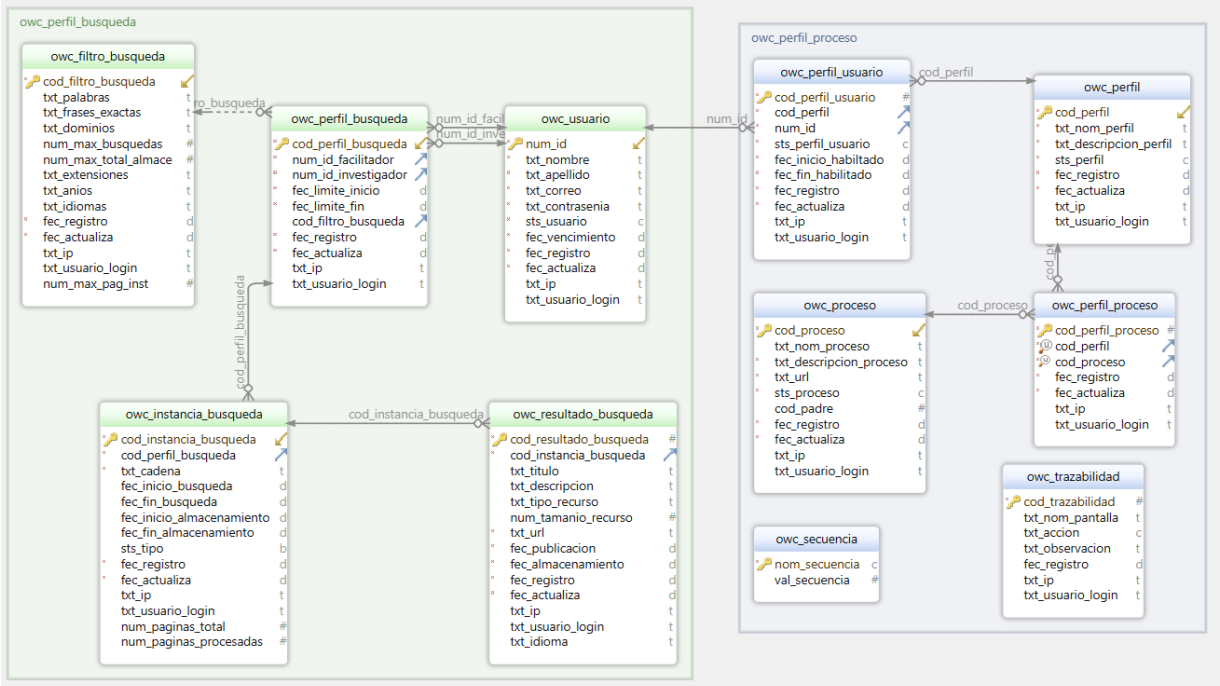
Debido al formato(.XLSX) y el tamaño del archivo se adjunta un minimizado. La versión completa siguiente enlace:

https://github.com/jhonnyjguevarac/OSINT_FORENSE_PRIVACIDAD/blob/main/Mapeo%20Forense%20OSINT%20Privacidad%202.0.rar

CICLO OSINT	CICLO FORENSE	PRINCIPIOS ETICOS (PRECEPT)	PRINCIPIOS PRIVACIDAD (ISO_IEC_29100_2011) (PRECEPT)	PRINCIPIOS PARA UNA COMUNICACIÓN EFECTIVA DE CAPITAL INTELECTUAL (EFFAS CIC) (PRECEPT)	PRINCIPIOS ÉTICOS GENÉRICOS (RENAUD Y ZIMMERMANN) (PRECEPT)	LINEAMIENTOS PRESERVACIÓN DE PRIVACIDAD	LINEAMIENTOS TÉCNICOS DE PRESERVACIÓN DE PRIVACIDAD	REQUERIMIENTOS DE PRIVACIDAD	CONSIDERACIONES LEGALES	
1. PLANIFICACIÓN Y DIRECCIÓN El primer paso del ciclo OSINT consiste en planificar las prioridades y requisitos de la	**1.- IDENTIFICACIÓN** Identificar que se ha producido un incidente, por ejemplo: Analizar el informe de un delito, relación con otra investigación. Se debe:	**E1.- Delimitar mandato:** - Comience delineando cuidadosamente el ámbito de la investigación	**PP2.-** Legitimación y especificación de la finalidad **PP3.-** Limitación de recolección	N/A	- Respeto - Social - Responsabilidad	- Definir las fuentes de información de acuerdo al ámbito establecido para la investigación - Definir el volumen de	**- Implementar funcionalidades que permitan definir listas blancas de dominios y subdominios**	PR_01 Definir las fuentes y dominios de información de acuerdo al ámbito establecido para la investigación. Technical Guidelines	Capítulo I ÁMBITO DE APLICACIÓN INTEGRAL Art. 7.- Tratamiento legítimo de datos personas	DE LAS INFRACCIONES DEL RESPONSABLE DE PROTECCIÓN DE DATOS Art. 67.- Infracciones leves del Responsable de protección de datos Art. 68.- Infracciones graves del Responsable de protección de datos Art. 69.- Infracciones leves del Encargado de protec
2.- RECOLECCIÓN Una vez realizada la planificación adecuada, puede comenzar la recopilación de OSINT	**2.- ADQUISICION** Incautación física y almacenamiento de dispositivos y datos. Adquisición local:Datos almacenados localmente	**E2.- Respetar la privacidad del sujeto:** - Se debe proteger la privacidad del sujeto	**PP3.-** Limitación de recolección	**IC5.-** Compensación equilibrada entre divulgación y privacidad **IC7.-** Prevención del desbordamiento de información	- Justicia - Beneficio	- Verificar que la información que se almacenará provenga únicamente de fuentes del ámbito(dominios) establecido. - Verificar que la información	**- Implementar controles para filtrar y realizar perticiones únicamente a sitios**	PR_09 Verificar que la información que se almacenará provenga únicamente de fuentes del ámbito(dominios) establecido. Technical Guidelines	Capítulo IV CATEGORÍAS ESPECIALES DE DATOS Art. 25.- Categorías especiales de datos personales	
2.- RECOLECCIÓN Una vez realizada la planificación adecuada, puede	**3.- PRESERVACIÓN** Copia y verificación (suma de verificación) de medios, es decir, hacer	**E2.- Respetar la privacidad del sujeto** Se debe proteger la	**PP3.-** Limitación de recolección	**IC5.-** Compensación equilibrada entre divulgación y privacidad **IC7.-** Prevención del desbordamiento de información	- Justicia - Beneficio	- Establecer mecanismos para asegurar la seguridad de la información en cuanto a PRIVACIDAD, INTEGRIDAD Y	**- Implementar mecanismos para el aseguramiento de la seguridad de la**	PR_13 Registrar datos acerca del origen de la información. Technical Guidelines • 13.1 Implementar políticas y	Capítulo II PRINCIPIOS i) Conservación j) Seguridad de datos personales	
3. TRATAMIENTO Y EXPLOTACIÓN Una vez que haya adquirido sus datos, puede comenzar a	**4.- BUSQUEDA** Recuperar información de medios	**E2.- Respetar la privacidad del sujeto** Se debe proteger la	**PP3.-** Limitación de recolección	**IC5.-** Compensación equilibrada entre divulgación y privacidad **IC7.-** Prevención del desbordamiento de información	- Justicia - Beneficio	- Recuperar metadatos(procedencia de la información, fecha de	- Implementar funcionalidades que permitan recuperar metadatos de la información almacenada	PR_18 Previo a incluir a terceros en hipótesis considerar la evidencia encontrada. Technical Guidelines • 18.1 Definir criterios de inclusión		
4. ANÁLISIS Y PRODUCCIÓN Después del procesamiento inicial de los datos recopilados,	**5.- ANÁLISIS** Juicios de relevancia, organización de hechos de bajo nivel en evidencia. Considerar cuatro preocupaciones semi-ortogonales: temporal, espacial,	**E2.- Respetar la privacidad del sujeto** Se debe proteger la privacidad del sujeto	**PP3.-** Limitación de recolección	**IC5.-** Compensación equilibrada entre divulgación y privacidad **IC7.-** Prevención del desbordamiento de información	- Justicia - Beneficio	- Establecer mecanismos para el aseguramiento de la seguridad de la información.	**- Implementar mecanismos para el aseguramiento de la seguridad de la información.**	-PR_14 Establecer mecanismos para el aseguramiento de la seguridad de la información. Technical Guidelines • 14.1 Definir políticas de control de	Capítulo II PRINCIPIOS g) Confidencialidad j) Seguridad de datos personales	
5.- DIFUSIÓN E INTEGRACIÓN El paso final en el ciclo OSINT implica entregar la inteligencia recopilada y analizada a las	**6.- RECONSTRUCCIÓN** Inducción y comprobación de hipótesis. **A.** ¿Qué crees que pasó? **B.** ¿Cómo puede sustentarse en los hechos?	**E2.- Respetar la privacidad del sujeto** Se debe proteger la privacidad del sujeto	**PP3.-** Limitación de recolección	**IC5.-** Compensación equilibrada entre divulgación y privacidad **IC7.-** Prevención del desbordamiento de información	- Justicia - Beneficio	- Establecer mecanismos para el aseguramiento de la seguridad de la información.	**- Implementar mecanismos para el aseguramiento de la seguridad de la información.**	PR_20 Definir tipos de informes a presentar en función de: Tipo de información que contiene (Nivel de sensibilidad), Perfil del lector, Relevancia en cuanto a la	TRANSFERENCIA O COMUNICACIÓN Y ACCESO A DATOS PERSONALES POR TERCEROS Art. 33.- Transferencia o comunicación de datos personales.	
5. DIFUSIÓN E INTEGRACIÓN El paso final en el ciclo OSINT implica entregar la inteligencia	**7.- INFORMES** Informe judicial, comparecencia ante el tribunal.	**E2.- Respetar la privacidad del sujeto** Se debe proteger la	**PP3.-** Limitación de recolección	**IC5.-** Compensación equilibrada entre divulgación y privacidad **IC7.-** Prevención del desbordamiento de información	- Justicia - Beneficio	- Definir tipos de informes a presentar en función de: Tipo de información que contiene(Nivel de sensibilidad)	- Implementar funcionalidades que permitan generar reportes en base al nivel de acceso	PR_20 Definir tipos de informes a presentar en función de: Tipo de información que contiene (Nivel de sensibilidad), Perfil del lector,	Capítulo VI SEGURIDAD DE DATOS PERSONALES Ar. 37.- Seguridad de datos personales.	
5. DIFUSIÓN E INTEGRACIÓN El paso final en el ciclo OSINT implica entregar la inteligencia recopilada y analizada a las	**8.- REFLEXION Y REVISIÓN :** Consideración del desempeño y lecciones a aprender. A. Genere un documento que detalle qué investigadores forenses estuvieron	**E10.- Se debe mantener la integridad y confidencialidad de la información**	**P11.-** Cumplimiento de la privacidad	**IC10.-** Colocación y oportunidad efectiva de la divulgación	- Integridad	- Analizar la revocación de privilegios a ciertos perfiles en cuanto a acceso de información. - Analizar la conservación la	- Implementar funcionalidades que permitan revocación de privilegios de acceso a información.	PR_22 Analizar la revocación de privilegios a ciertos perfiles en cuanto a acceso de información. Technical Guidelines • 22.1 Definir políticas para	Capítulo II PRINCIPIOS g) Confidencialidad j) Seguridad de datos personales	

Anexo III – Diagrama físico de la base de datos y privilegios de usuario

Modelo físico de base de datos



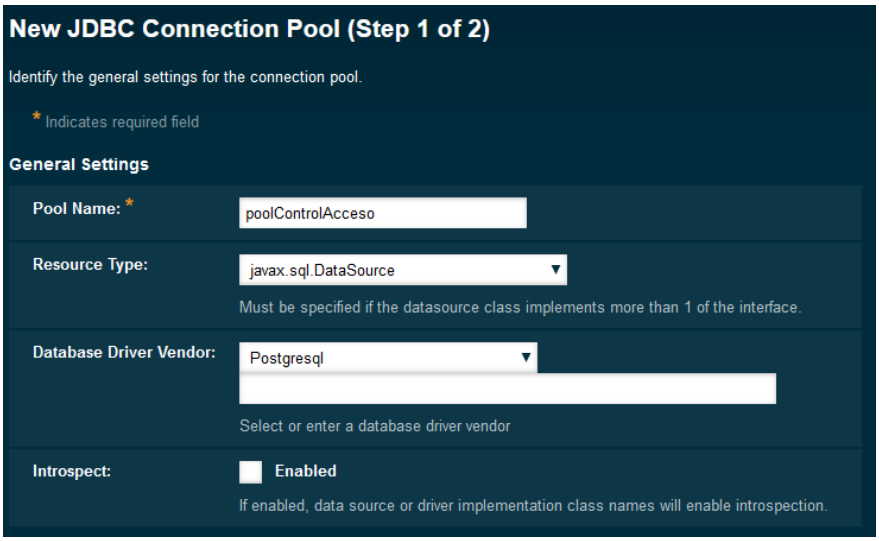
Creación y configuración de permisos de usuarios

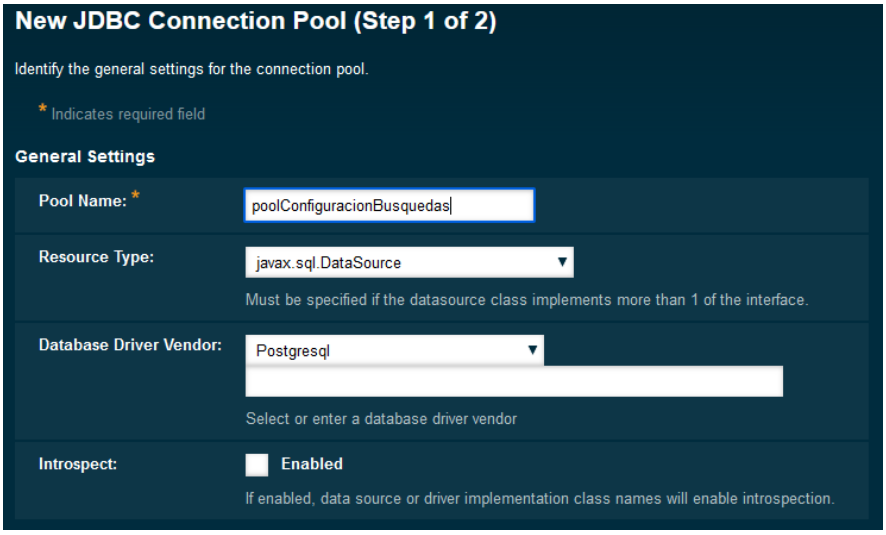
```
CREATE ROLE administrador_user LOGIN PASSWORD 'xxxxx';
GRANT CONNECT ON DATABASE owc TO administrador_user;
GRANT USAGE ON SCHEMA owc TO administrador_user;

CREATE ROLE investigador_user LOGIN PASSWORD 'xxxxx';
GRANT CONNECT ON DATABASE owc TO investigador_user;
GRANT USAGE ON SCHEMA owc TO investigador_user;

CREATE ROLE facilitador_user LOGIN PASSWORD 'xxxxx';
GRANT CONNECT ON DATABASE owc TO facilitador_user;
GRANT USAGE ON SCHEMA owc TO facilitador_user;


GRANT SELECT,INSERT,UPDATE ON owc_perfil              administrador_user;
```

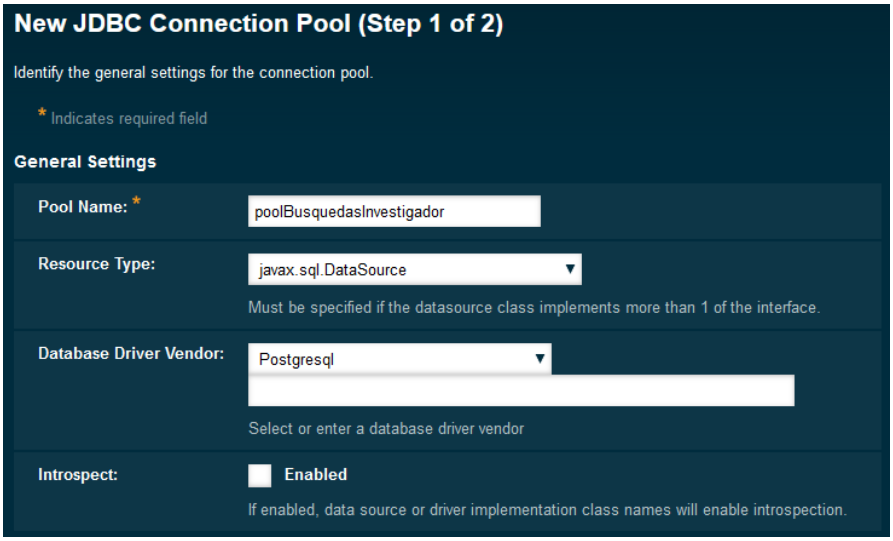
```
GRANT SELECT,INSERT,UPDATE ON owc_usuario             administrador_user;
GRANT SELECT,INSERT,UPDATE ON owc_perfil_proceso      administrador_user;
GRANT SELECT,INSERT,UPDATE ON owc_perfil_usuario      administrador_user;
GRANT SELECT,INSERT,UPDATE ON owc_proceso             administrador_user;
GRANT INSERT,INSERT        ON owc_trazabilidad        administrador_user;


GRANT INSERT,UPDATE        ON owc_filtro_busqueda     facilitador_user;
GRANT SELECT,INSERT,UPDATE ON owc_perfil_busqueda     facilitador_user;
GRANT SELECT,INSERT,UPDATE ON owc_secuencia           facilitador_user;
GRANT INSERT               ON owc_trazabilidad        facilitador_user;

GRANT SELECT               ON owc_filtro_busqueda     investigador_user;
GRANT SELECT,INSERT        ON owc_instancia_busqueda  investigador_user;
GRANT SELECT,UPDATE        ON owc_perfil_busqueda     investigador_user;
GRANT SELECT,INSERT        ON owc_resultado_busqueda  investigador_user;
GRANT SELECT,INSERT,UPDATE ON owc_secuencia           investigador_user;
GRANT INSERT               ON owc_trazabilidad        investigador_user;
```

Anexo IV – Configuración de pools de conexión

Creación de pools de conexión







Configuración de pool hacia usuarios de base de datos

Select	Name	Value
☐	password	p@ssw0rd
☐	databaseName	owc
☐	driverClass	org.postgresql.Driver
☐	serverName	localhost
☐	user	administrador_user
☐	portNumber	5432

Select	Name	Value
☐	password	p@ssw0rd
☐	databaseName	owc
☐	driverClass	org.postgresql.Driver
☐	serverName	localhost
☐	user	investigador_user
☐	portNumber	5432

Select	Name	Value
☐	password	p@ssw0rd
☐	databaseName	owc
☐	driverClass	org.postgresql.Driver
☐	serverName	localhost
☐	user	facilitador_user
☐	portNumber	5432

Anexo V –Tabla de búsquedas

#	Cadena búsqueda	Dominio	Filtros
S01	Julio Jaramillo Ecuador	https://ec.ebay.com/	Formato: html, png, jpg ; Idioma: español;Vol.Max:0.5 GB; Max.Pag:400; Años: 2022,2023
S02	Julio Jaramillo	https://ec.ebay.com/	Formato: html, png, jpg ; Idioma: español ;Vol.Max:0.5 GB;Max.Pag:400
S03	Julio Jaramillo colección	https://ec.ebay.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB;Max.Instancias: 15; Max.Pag:400
S04	Julio Jaramillo colección 1980	https://ec.ebay.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB;Max.Pag:400
S05	Julio Jaramillo colección LP	https://ec.ebay.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB;Max.Pag:400
S06	Julio Jaramillo colección LP	https://ec.ebay.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB;Max.Pag:400
S07	Julio Jaramillo Ecuador	https://www.amazon.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.2 GB;Max.Pag:250
S08	Julio Jaramillo	https://www.amazon.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.2 GB;Max.Pag:250
S09	Julio Jaramillo colección	https://www.amazon.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.2 GB;Max.Pag:250
S10	Julio Jaramillo colección 1980	https://www.amazon.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.2 GB;Max.Pag:250

S11	Julio Jaramillo colección LP	https://www.amazon.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.2 GB;Max.Pag:250
S12	Julio Jaramillo colección LP	https://www.amazon.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.2 GB;Max.Pag:250
S13	Julio Jaramillo Ecuador	https://www.mercadolibre.com.ec/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.8 GB;Max.Pag:500
S14	Julio Jaramillo	https://www.mercadolibre.com.ec/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.8 GB Max.Pag:500
S15	Julio Jaramillo colección	https://www.mercadolibre.com.ec/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.8 GB Max.Pag:500
S16	Julio Jaramillo colección 1980	https://www.mercadolibre.com.ec/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.8 GB Max.Pag:500
S17	Julio Jaramillo colección LP	https://www.mercadolibre.com.ec/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.8 GB Max.Pag:500
S18	Julio Jaramillo colección LP	https://www.mercadolibre.com.ec/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.8 GB Max.Pag:500
S19	Julio Jaramillo Ecuador	https://www.mercadolibre.com.ec/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.8 GB Max.Pag:500
S20	Julio Jaramillo Ecuador	https://www.discogs.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.6 GB Max.Pag:350
S21	Julio Jaramillo	https://www.discogs.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.6 GB Max.Pag:350
S22	Julio Jaramillo colección	https://www.discogs.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.8 GB

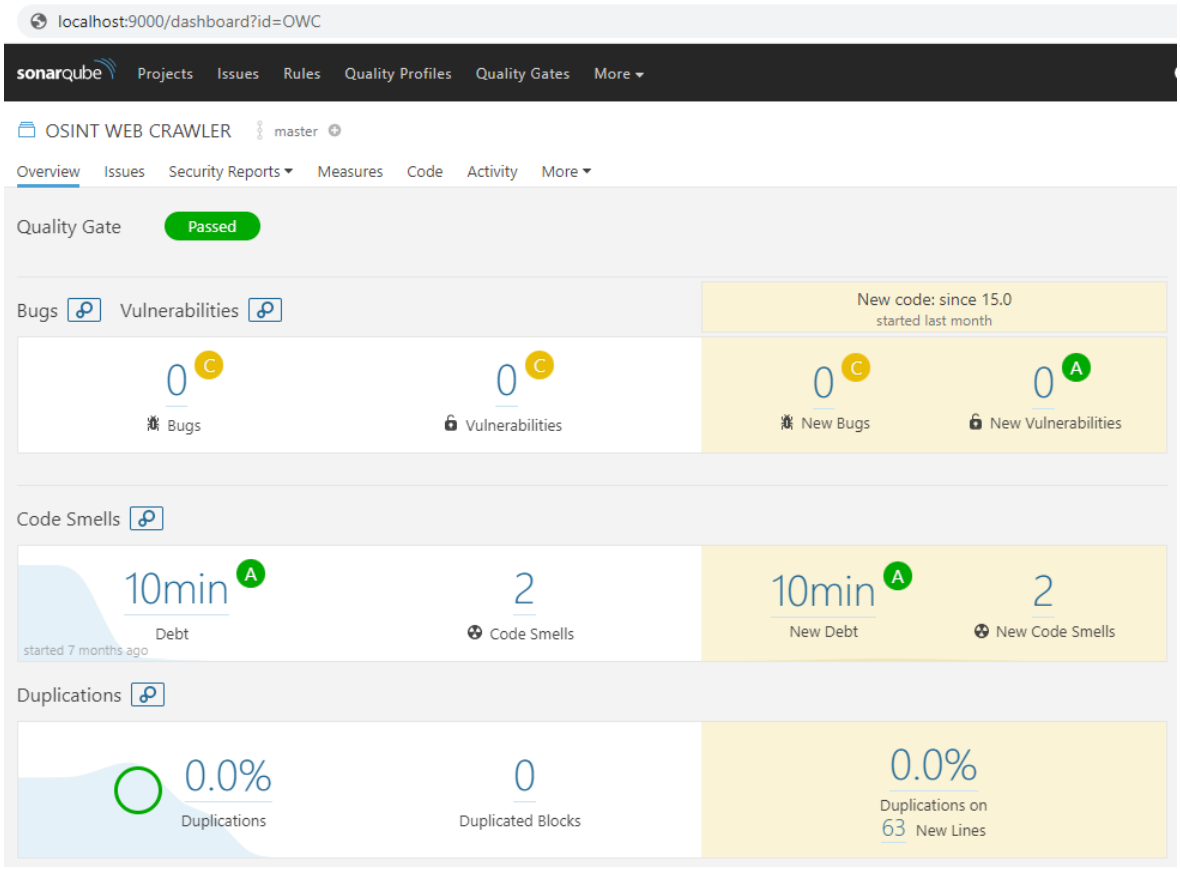
			Max.Pag:350
S23	Julio Jaramillo colección 1980	https://www.discogs.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.8 GB Max.Pag:350
S24	Julio Jaramillo colección LP	https://www.discogs.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.8 GB Max.Pag:350
S25	Julio Jaramillo colección LP	https://www.discogs.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.8 GB Max.Pag:350
S26	Collector 1975 Cassette Bob Dylan, Blood on The Tracks	https://www.mercadolibre.com.ec/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB Max.Pag:350
S27	Collector 1975 Cassette Bob Dylan, Blood on The Tracks	https://www.mercadolibre.com.ec/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB Max.Pag:350
S28	Collector 1975 Cassette Bob Dylan, Blood on The Tracks	https://www.mercadolibre.com.ec/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB Max.Pag:350
S29	Collector 1975 Cassette Bob Dylan, Blood on The Tracks	https://www.mercadolibre.com.ec/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB Max.Pag:350
S30	Collector 1975 Cassette Bob Dylan, Blood on The Tracks	https://www.mercadolibre.com.ec/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB Max.Pag:350
S31	Collector 1975 Cassette Bob Dylan, Blood on The Tracks	https://www.discogs.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB Max.Pag:350
S32	Collector 1975	https://www.discogs.co	Formato: html, png, jpg ; Idioma:

	Cassette Bob Dylan, Blood on The Tracks	m/	español; Vol.Max:0.5 GB Max.Pag:350
S33	Collector 1975 Cassette Bob Dylan, Blood on The Tracks	https://www.discogs.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB Max.Pag:350
S34	Collector 1975 Cassette Bob Dylan, Blood on The Tracks	https://www.discogs.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB Max.Pag:350
S35	Collector 1975 Cassette Bob Dylan, Blood on The Tracks	https://www.discogs.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB Max.Pag:350
S36	LP KISS – TheBest of Vinyl 7, Vintage Soul	https://www.mercadolibre.com.ec/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB Max.Pag:350
S37	LP KISS – TheBest of Vinyl 7, Vintage Soul	https://www.mercadolibre.com.ec/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB Max.Pag:350
S38	LP KISS – TheBest of Vinyl 7, Vintage Soul	https://www.mercadolibre.com.ec/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB Max.Pag:350
S39	LP KISS – TheBest of Vinyl 7, Vintage Soul	https://www.discogs.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB Max.Pag:350
S40	LP KISS – TheBest of Vinyl 7, Vintage Soul	https://www.discogs.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB Max.Pag:350

S41	LP KISS – TheBest of Vinyl 7, Vintage Soul	https://www.discogs.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB Max.Pag:350
S42	LP KISS – TheBest of Vinyl 7, Vintage Soul	https://www.discogs.com/	Formato: html, png, jpg ; Idioma: español; Vol.Max:0.5 GB Max.Pag:350

Anexo VI – Análisis de vulnerabilidades de código estático

Reporte análisis código (SonarQube)




```
sonar.projectKey=OWC
sonar.projectName=Osint Web Crawler
sonar.projectVersion=1.0
sonar.modules=Web
sonar.sources=src
sonar.java.binaries=target/classes
sonar.binaries=target/classes
sonar.sourceEncoding=UTF-8
modWeb.sonar.projectName=owc-web
modEjb.sonar.projectBaseDir=owc-crawler
```