UCASAL
UNIVERSIDAD CATÓLICA DE SALTA

z

Proyecto Final

"HERRAMIENTA FORENSE: RECONSTRUCCIÓN DE PAQUETES DE LA RED Y DETECCIÓN DE CONTENIDO INDEBIDO DE MENORES"

Alumno:

- **Soria, Luis Aaron Maximiliano**

Profesor:

- **Rivera Bernsdorff, Fernando Lucas**

Año: 2018

Facultad de Ingeniería

Ingeniería en Informática

"HERRAMIENTA FORENSE: RECONSTRUCCIÓN DE PAQUETES DE LA RED Y DETECCIÓN DE CONTENIDO INDEBIDO DE MENORES"

Firma profesor guía	Firma Alumno

Firma miembro Tribunal Evaluador	Firma miembro Tribunal Evaluador	Firma miembro Tribunal Evaluador

Fecha de Exposición de Trabajo _____

Agradecimientos

Mi profundo agradecimiento a mis padres que me han apoyado durante todo este trayecto, y que sin ellos nada de esto hubiera sido posible. Mis reconocimientos a la Universidad Católica de Salta y al Banco Macro por brindarme la oportunidad de pertenecer a esta casa de altos estudios. Mis gratitudes a todos mis docentes de la facultad de ingeniería, especialmente a mí director de tesis y amigo Fernando Rivera por haber impartido sus valiosos conocimientos, como así también a mi profesor y amigo Sergio Appendino por ayudarme en la elección de la temática de este trabajo.

Contenido

Introducción

Abstract

El presente trabajo deriva del proyecto de investigación "APLICACIÓN DE METODOLOGIAS, PROCESOS, Y TECNICAS FORENSES DIGITALES EN NUEVAS TECNOLOGIAS", realizado en la Universidad Católica de Salta. Actualmente en Argentina hay muy poca legislación aplicable a delitos informático, que dan lugar a ciertas áreas grises y vacíos legales ideales para delinquir, sumado a esto nos encontramos con el anonimato en las redes y el poco control de esta última. En este marco, el presente proyecto tiene por objeto desarrollar un software (con licencia libre) agilice la búsqueda de un tipo específico de evidencia digital y sea de fácil acceso en términos económicos.

Breve descripción del problema y su importancia

Anteriormente, en la legislación argentina se encontraba penada "*la producción, financiación, ofrecimiento, comercialización, publicación, facilitación, divulgación o distribución, por cualquier medio, representación un menor de dieciocho (18) años dedicado a actividades sexuales explicitas o toda representación de sus partes genitales con fines predominantemente sexuales, al igual que el que organizare espectáculos en vivo de representaciones sexuales explicitas en que participaren menores*"[1]. Actualmente, con la ley 27436 del 21 marzo de 2018 y publicada en el boletín oficial el 23 abril de 2018, la tenencia del contenido en cuestión se encuentra penada.

Ahora bien, el casi nulo control que existe en internet, el anonimato, usuarios inexpertos, entre otros factores, generan un lugar inseguro y por consiguiente, ideal para delinquir; por lo que el tráfico del contenido indebido de menores se realiza en distintos los medios digitales, *como son las redes peer-to-peer*[2].

Carácter del problema

La problemática abordada tiene una implicancia social, en efecto, el tráfico generado en la red es realizado por todo tipo de personas, teniendo en cuenta que se debe respetar la privacidad de estas. La existencia de cooperación internacional de los organismos de justica, como ser la comunicación de información, compartir recursos tanto hardware como software, servicio de profesionales de renombre, todos estos destinados a localización de estos ilícitos, nos indica que es una problemática global.

Como hemos mencionado anteriormente, ciertos factores en la red como: el anonimato, la inexperiencia de otros usuarios, y el desconocimiento general de los riesgos que hay en internet; generan un ambiente propicio para los ciberdelitos. Se puede ver en la actualidad, lo grave de la situación reflejada en los portales de noticias y medios de comunicación en la argentina y en el mundo:

[1]Artículo 128 del código penal procesal de la nación
[2]Amor M. 2016. Pornografía Infantil: Marco legal y herramientas peer to peer. Publicación on line. ISSN 2347-0372

Sitio:

https://www.infobae.com/2015/12/14/1776378-megaoperativo-contra-la-pornografia-infantil-mexico-chile-argentina-y-eeuu-60-detenidos/

Sitio

http://www.telam.com.ar/notas/201605/147655-detenidos-pedofilia-internet.html

Sitio:

https://www.eltribuno.com/jujuy/nota/2014-10-4-0-0-0-pedofilia-on-line-antesala-de-la-trata

Motivación

La mayoría de las alertas y anuncios, que comunican la detección de este tipo de delitos, son informados por entidades externas (Interpol – FBI – NcMEC, entre otros). En este sentido, es menester aumentar la agilizar en la detección de este tipo de delitos.

Existe software especializado en la detección de contenido sexual explícito de menores, pero en nuestro país es difícil el acceso a los mismos por el alto costo de las licencias.

Teniendo en cuenta estos puntos mencionados anteriormente, se persigue desarrollar un software que optimice los tiempos de búsqueda de este tipo de contenido y que sea de fácil acceso para los oficiales de la justicia.

Qué pasos se realizarán

El presente trabajo se encuentra organizado en 4 etapas, respetando lo establecido por el proceso unificado de desarrollo de software (PU), "*este es un marco de trabajo genérico que puede especializarse para una gran variedad de sistemas software, para diferentes áreas de aplicación, diferentes tipos de organizaciones, diferentes niveles de aptitud y diferentes tamaños de proyectos*"[3]; además se harán uso de las mejores prácticas que recomienda el PMBOK.

1. Primera Etapa
    a. Investigación
    b. Requerimientos
2. Segunda Etapa
    a. Análisis
    b. Diseño
3. Tercera Etapa
    a. Codificación
    b. Prueba
4. Cuarta Etapa
    a. Despliegue

En un ámbito mucho más específico, las fases de Análisis, Diseño, Codificación, Prueba e Implementación se apoyan en la metodología Orientada a Objetos, porque según el autor Ian Sommerville: "*los sistemas orientados a objetos son más fáciles de cambiar que aquellos sistemas desarrollados usando enfoques funcionales. Los objetos incluyen datos y operaciones para manipular dichos datos. En consecuencia, pueden entenderse y modificarse como entidades independientes. Cambiar la implementación de un objeto o agregar servicios no afectará a otros objetos del sistema.*"[4]

[3]Jacobson I, Booch G, Rumbaugh J. 2000. El proceso unificado de desarrollo de software. 464 paginas. Editorial Pearson Educación. ISBN 84-7829-036-2
[4]Sommerville I. 2011 Ingeniería de Software. 792 paginas. Editorial Pearson Educación. ISBN: 978-607-32-0603-7

Estado de la cuestión

Antecedentes relacionados con el problema

Disciplinas como la seguridad informática, auditorias de sistemas y pericias en el ámbito tecnológico cada vez se formalizan más, y las herramientas para llevar a cabo esta actividad son desarrolladas por entidades formales como ser:

- Kali Linux: Sistema operativo diseñado por Offensive Security Ltd, orientado a la seguridad informática y auditoria de sistemas, que cuenta con un amplio conjunto de herramientas para la realización de estos.
- Bugtraq-II Blackwidow: llevado a cabo por Bugtraq-Team, orientado al pentesting y la forensia, igual que kali Linux, es un sistema operativo que recopila un gran conjunto de herramientas para la realización de tales actividades.
- Griffeye: *es una plataforma de software versátil para investigaciones de medios digitales. Utilizada por las agencias de aplicación de la ley, defensa y seguridad nacional en todo el mundo, la plataforma ofrece a los profesionales de la aplicación de la ley y la inteligencia una ventaja sobre los volúmenes cada vez mayores de archivos de imágenes y video.*[5]
- EnCase Forensic Software: herramienta que brinda soporte en las investigaciones digitales. *Ayudar a recopilar y analizar la evidencia que necesita, de manera eficiente y precisa*.[6]
- Wireshark: *es el analizador de protocolo de red más importante y ampliamente utilizado en el mundo. Le permite ver lo que está sucediendo en su red a un nivel microscópico y es el estándar de facto en muchas empresas comerciales y sin fines de lucro, agencias gubernamentales e instituciones educativas. El desarrollo de Wireshark prospera gracias a las contribuciones voluntarias de expertos en redes en todo el mundo y es la continuación de un proyecto iniciado por Gerald Combs en 1998.*[7]
- Driffnet: Es un una herramienta diseñada para el sniffing de redes, enfocado en la obtención de imágenes jpg y gif

Se puede ver que hay muchas herramientas, tanto libre como pagas, con fines generales destinados a la forensia. Las aplicaciones mencionadas anteriormente solo realizan una de las siguientes tareas: extraer contenido de la red, reensamblarlo, analizarlo o presentar los resultados de dicho análisis, específicamente mostrar los resultados de detección de existencia de pornografía infantil en un conjunto de imágenes.

[5] griffeye.com/the-platform/
[6] guidancesoftware.com/encase-forensic?cmpid=nav_r
[7] wireshark.org

Fundamentos que cimientan la solución escogida

Aspectos Técnicos

El realizar un análisis del tráfico de la red permite la detección temprana de la distribución de contenido ilegal. Las aplicaciones examinadas anteriormente evalúan las imágenes siempre y cuando estas se encuentren en un medio de almacenamiento.

El tráfico de red se encuentra estructurado en hexadecimales, por lo que los programas convencionales de análisis de imágenes no pueden realizar su tarea debido al formato en el que se encuentra dicho tráfico de red. La solución propuesta es capaz de reconstruir imágenes a partir de los hexadecimales mencionados anteriormente.

En otras palabras, la solución propuesta, además de analizar imágenes en formatos conocidos, es capaz analizar la salida de un sniffer y determinar la existencia de pornografía infantil.

Aspectos Legales

Los siguientes artículos del código penal dan pauta de que en Argentina los delitos informáticos están considerados y tienen la misma importancia que cualquier otro tipo de delito. Por ejemplo:

El artículo 128 y posterior sustitución por la ley 27.436 en el año 2018 que penan la tenencia de pornografía infantil

"*ARTICULO 128.- Será reprimido con prisión de tres (3) a seis (6) años el que produjere, financiare, ofreciere, comerciare, publicare, facilitare, divulgare o distribuyere, por cualquier medio, toda representación de un menor de dieciocho (18) años dedicado a actividades sexuales explícitas o toda representación de sus partes genitales con fines predominantemente sexuales, al igual que el que organizare espectáculos en vivo de representaciones sexuales explícitas en que participaren dichos menores.*

Será reprimido con prisión de cuatro (4) meses a un (1) año el que a sabiendas tuviere en su poder representaciones de las descriptas en el párrafo anterior.

Será reprimido con prisión de seis (6) meses a dos (2) años el que tuviere en su poder representaciones de las descriptas en el primer párrafo con fines inequívocos de distribución o comercialización.

Será reprimido con prisión de un (1) mes a tres (3) años el que facilitare el acceso a espectáculos pornográficos o suministrare material pornográfico a menores de catorce (14) años.

Todas las escalas penales previstas en este artículo se elevarán en un tercio en su mínimo y en su máximo cuando la víctima fuere menor de trece (13) años."

El artículo 131 incorporado por la ley 26.904 en el año 2013, que penan las actividades conocidas como el grooming, es decir, que por cualquier medio electrónico/tecnológico se contacte a un menor de edad para cometes cualquier delito contra su integridad sexual

"*Artículo 131: Será penado con prisión de seis (6) meses a cuatro (4) años el que, por medio de comunicaciones electrónicas, telecomunicaciones o cualquier otra tecnología de transmisión de datos, contactare a una persona menor de edad, con el propósito de cometer cualquier delito contra la integridad sexual de la misma.*"

Los artículo 153 y 153 bis que sostienen violación de privacidad en todos su medios, incluidos los medios electrónicos/informáticos

"Artículo 153: Será reprimido con prisión de quince (15) días a seis (6) meses el que abriere o accediere indebidamente a una comunicación electrónica, una carta, un pliego cerrado, un despacho telegráfico, telefónico o de otra naturaleza, que no le esté dirigido; o se apoderare indebidamente de una comunicación electrónica, una carta, un pliego, un despacho u otro papel privado, aunque no esté cerrado; o indebidamente suprimiere o desviare de su destino una correspondencia o una comunicación electrónica que no le esté dirigida.

En la misma pena incurrirá el que indebidamente interceptare o captare comunicaciones electrónicas o telecomunicaciones provenientes de cualquier sistema de carácter privado o de acceso restringido.

La pena será de prisión de un (1) mes a un (1) año, si el autor además comunicare a otro o publicare el contenido de la carta, escrito, despacho o comunicación electrónica.

Si el hecho lo cometiere un funcionario público que abusare de sus funciones, sufrirá además, inhabilitación especial por el doble del tiempo de la condena."

"Artículo 153 bis: Será reprimido con prisión de quince (15) días a seis (6) meses, si no resultare un delito más severamente penado, el que a sabiendas accediere por cualquier medio, sin la debida autorización o excediendo la que posea, a un sistema o dato informático de acceso restringido.

La pena será de un (1) mes a un (1) año de prisión cuando el acceso fuese en perjuicio de un sistema o dato informático de un organismo público estatal o de un proveedor de servicios públicos o de servicios financieros."

Ley 26326 de habeas data, la misma en su artículo 2, en sus definiciones, contempla los datos informatizados

"ARTICULO 2° — (Definiciones).

A los fines de la presente ley se entiende por:

[...]Datos informatizados: Los datos personales sometidos al tratamiento o procesamiento electrónico o automatizado. [...]"

Teniendo en cuenta todo el desarrollo que se está teniendo en materia jurídica, es importante que se desarrolle una herramienta que se adecue a esta situación. Un software local propiciara un importante avance en este sentido.

Tratamiento de evidencia digital

Se debe partir por el concepto de evidencia digital, la siguiente definición establece:

"La evidencia digital es un tipo de evidencia física construida de campos magnéticos y pulsos electrónicos, que por sus características deben ser recolectadas y analizadas con herramientas y técnicas especiales."[8]

Es decir, se considera como evidencia digital a aquella que es almacenada y procesas por medios informáticos.

Las imágenes que se encuentran en computadoras, teléfonos inteligentes y otros dispositivos de almacenamiento comparten las características de evidencia digital, dichas características son las siguientes:

- Volátil: Puede perderse si no se recolecta en tiempo y forma.
- Duplicable: Pueden realizarse diversas copias sin poder reconocer el original.
- Alterable: Factible de su modificación y/o borrado y sin el registro de esas acciones.
- Anónima: En algunos casos no se pueden determinar el autor de las mismas.
- Posee datos adicionales no visibles por las herramientas tradicionales usadas por el usuario: Ultimo acceso, modelos de cámara usados en una fotografía, coordenadas de geo posicionamiento, autor, última impresión, secuencia de recorrido de un correo electrónico, etc.).[9]

Para proteger la cadena de custodia de la evidencia digital y poder determinar que no ha sido adulterada, un recurso muy utilizado es la función resumen o hash que consiste en: un algoritmo matemático que transforma cualquier bloque arbitrario de datos en una nueva serie de caracteres con una longitud fija. Independientemente de la longitud de los datos de entrada, el valor hash de salida tendrá siempre la misma longitud[10]. Por ejemplo:

Si se aplica la función hash a un archivo determinado se obtendrá una longitud de caracteres fija, luego si a ese mismo archivo se le introducen modificaciones y se guardan esos cambios, cuando se calcule nuevamente la función hash el resultado será distinto.

Desarrollo actual de las técnicas y herramientas requeridas en el tratamiento del problema.

Se sabe que existen múltiples técnicas de auditorías de redes y ethical hacking, que permiten la captura de contenido que se comparte en la web. Técnicas clásicas como Man in the Midle, phishing, ataques de botnet, configuración de firewalls, son utilizadas tanto por auditores, especialistas en seguridad y hackers de sombrero negro o blanco. Estos términos pueden ser consultados en el glosario de términos del anexo 2.

[8] Di Iorio, Ana Haydée et al. 2017. El rastro Digital del Delito. 554 paginas. Universidad FASTA Ediciones. ISBN 978-987-1312-81-8

[9] Guía de ciberdelitos. http://www.justiciasalta.gov.ar/images/uploads/ciberdelitos_web_v020518.pdf

[10] https://latam.kaspersky.com/blog/que-es-un-hash-y-como-funciona/2806/

La reconstrucción de contenido multimedia, específicamente las imágenes, es posible mediante la utilización de los compiladores de los lenguajes de programación como C# o JAVA. Estos lenguajes, mediante una cadena de caracteres hexadecimales, son capaces de retornar una imagen con las mismas características a la original.

Existen distintos métodos de analizar el contenido de una imagen. Por un lado, se cuenta con el reconocimiento de texturas de piel en una imagen, es una técnica rápida y eficiente para detectar nuestro objetivo en gran volumen de fotos.

Figura 1. Reconocimiento de texturas. Fuente de la imagen http://answers.opencv.org/upfiles/13506303131033303.png.

También existe la detección de objetos, este es mucho más popular y conocido, ya que se encuentra en el software de las cámaras de fotos de los smartphones, estamos hablando de la detección de rostros.

En este punto se debe aclarar la diferencia de los términos *detección* y *reconocimiento* per se. Por un lado el concepto *detección* se limita a buscar un objeto en cuestión en una foto. Como podemos ver en la siguiente imagen:

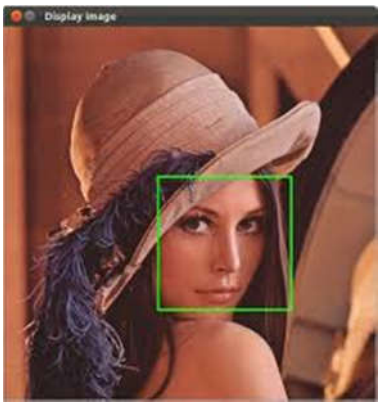


Figura 2. Detección facial. Fuente de la imagen https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcQYNNV8kaI_AeCTzO1yZFGLcHJlUU8eiQX78Fa-S39jtC0FyU56

La *detección* de objetos no se limita a caras, se pueden detectar ojos, cuerpos completos, como así el tren inferior o superior del mismo las personas.

Por otro lado el *reconocimiento*, consiste en aprender a reconocer un objeto es especifico y posteriormente buscar ese mismo objeto en otra imagen, por ejemplo Facebook, es uno de los que utiliza este tipo de técnica; ya que al subir una foto de varias personas, si estas son usuarios de la red social, puede reconocer a tales usuarios de la foto y etiquetarlos en la publicación.

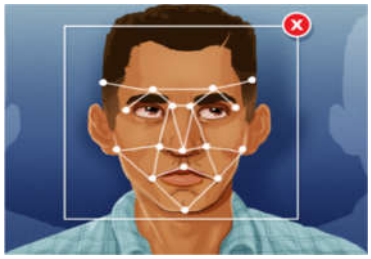


Figura 3. Reconocimiento facial. Fuente de la imagen https://i.amz.mshcdn.com/76SJtMLJ9op7PCOk8Nu7OeOas3Y=/950x534/filters:quality(90)/https%3A%2F%2Fblueprint-api-production.s3.amazonaws.com%2Fuploads%2Fcard%2Fimage%2F723865%2Fc58a78e7-a00b-41d5-a32a-eedb11098888.png

Definición del problema

Definir exactamente el problema

Nuestro país ha avanzado en materia de delitos informáticos como se ha expuesto, el problema es que en nuestra nación el desarrollo de software no está acompañando este avance en materia de pericias. Las herramientas de esta índole son realizadas por países como Estados Unidos e Israel, ambos potencia en esta materia. Es por ello que se proporcionará una herramienta de software libre y de fácil acceso que pueda ser utilizado por cualquier oficial de justicia para analizar el tráfico de la red y determinar la ilegalidad del contenido.

Objetivos

Objetivo General

Desarrollar una herramienta forense, que sea de fácil acceso en términos económicos, capaz de reconstruir de paquetes de la red detección de contenido indebido de menores

Objetivos Específicos

- Desarrollar un algoritmo capaz reconstruir de imágenes del tráfico TCP, HTML y no encriptados, fácilmente escalable a otros protocolos de red.
- Desarrollar un escanner de texturas, capaz de filtrar contenido inofensivo de contenido con desnudos.
- Desarrollar un escanner de objetos, capaz de separar imágenes de adultos y menores.
- Desarrollar un escanner capaz de reconocer un rostro específico en una colección fotos.

Restricciones o límites del trabajo

El trabajo se limitará a la reconstrucción de paquetes obtenidos a partir de la utilización de otra herramienta. Los paquetes son TCP, HTML y no encriptados.

No se incursionará en el desarrollo de nuevas técnicas o métodos de procesamiento de imagen, es decir que el objetivo de este proyecto no es proponer un nuevo método o procedimiento para analizar las mismas, si no que se hará una recopilación e implantación en software del conocimiento existente.

El contenido reconstruido solo se limitará a imágenes con extensión .jpeg

Se implementarán las últimas técnicas y se aprovecharán investigaciones previas, relacionadas con la problemática, realizadas por autores expertos en la materia/con dominio en la materia.

El tráfico es generado en una red LAN que intercambia datos e información con internet con el que se realizarán las pruebas, el mismo será realizado por el autor del presente proyecto, simulando la transmisión de imágenes con contenido normal, de dominio público, obtenido a partir de publicidades de playas y catálogos varios.

Alcance y procesos involucrados

El presente trabajo, iniciará haciendo una investigación de la situación actual respecto al desarrollo de software forense. Posteriormente, se buscarán proyectos similares y se formulará un modelo matemático acorde, que sustente el desarrollo del proyecto.

Luego, se hará uso de la técnica Man in the Middle y herramientas de sniffing, para la captura de paquetes TCP no encriptados, los mismos serán registrados en un archivo de texto plano para su posterior análisis; todo esto en un ambiente controlado y simulado para tal fin.

El análisis del tráfico resultante se realizará en frio, es decir, una vez que se considere que la información capturada es suficiente, se iniciará la reconstrucción de paquetes y el escaneo de contenido de las imágenes.

Para detectar las imágenes deseadas se hará uso de 3 tipos de escaners: de texturas, objetos y de rostro.

El proceso se encuentra estructurado en cuatro etapas, cada una cuenta con dos fases que permiten el desarrollo de actividades en simultáneo. De esta manera resulta que tenemos:

1. Primera Etapa
    a. Análisis inicial de factibilidad: En la cual se realiza un diagnóstico de la situación actual y el estado del arte. Se identificarán a los distintos expertos en el tema. Se determinarán las métricas con las cuales se realizará el seguimiento del proyecto y se identificarán los interesados.
    b. Requerimientos: Se determinará los elementos necesarios para llevar a cabo el proyecto, tanto hardware, software, servicios de cloud y repositorios entre otros. Es aquí en donde se realiza el primer análisis formal de riesgos del proyecto, se inicia el trato con los interesados.
2. Segunda Etapa
    a. Análisis: Se realizan los primeros planos para la construcción del software, siguiendo lo establecido por proceso unificado y la metodología orientada a objetos, se obtienen nuevas métricas y estimaciones, además se realizará un seguimiento de los riesgos del proyecto, se mantiene el dialogo con los interesados.
    b. Diseño: Definición de interfaz gráfica, especificación de objetos encontrados y su interacciones.
3. Tercera Etapa
    a. Codificación: En esta se inicia la construcción formal de la aplicación; se realizan nuevas estimaciones y se consiguen nuevas métricas, se busca asegurar la calidad en el desarrollo de software y se mantienen relaciones con los interesados.
    b. Prueba: En esta fase se realizará la verificación de los estándares de calidad, le realización de distintas pruebas que garanticen el buen funcionamiento del software desarrollado
4. Cuarta Etapa

a. Despliegue: Esta es la fase final del proyecto, se compara laS estimaciones realizadas con las métricas obtenidas del proyecto, se realizarán las últimas interacciones con los interesados.

Alternativas tecnológicas de solución al problema

En una entrevista realizada al Ingeniero Sergio Appendino, Coordinador Gabinete Forense Digital del Ministerio Público de Salta, Cuerpo de Investigaciones Fiscales (CIF), este afirmó que de las herramientas utilizadas en su gabinete, la resulta similar a este proyecto es EnCase Forensic Software. Esta herramienta recupera información de dispositivos de almacenamientos (discos duros, pendrives), y este proyecto hará la recuperación de la información a partir del tráfico de red capturado mediante un sniffer.

La segunda diferencia entre EnCase y el presente proyecto es que, la primera solo hace un análisis general del contenido, la segunda se encuentra fuertemente enfocada a la detección de pedofilia y la búsqueda de una persona específica. La tercera diferencia es que la licencia de EnCase es propietaria y la licencia de este proyecto será GNU GPL versión 3.

En el siguiente cuadro comparativo se resumen las diferencias entre ambas herramientas

	EnCase Forensic Software	Proyecto propuesto en esta tesis
¿De dónde se recupera la información?	De dispositivos de almacenamiento	Trafico de la red capturado
Análisis de la información	Propósito general	Detección de pedofilia Búsqueda de una persona específica
Licencia	Propietario	GNU GPL versión 3

Solución propuesta

Justificación de la solución aportada.

Profesionales relacionados en el área de pericias y criminalística, ven necesario el uso de estas herramientas informáticas que les permita realizar un análisis más rápido y eficiente.

En un Encuentro realizado el día 7 de septiembre de 2018, en el marco de las "JORNADAS DE ACTUALIZACION DE CRIMINALISTICA E INVESTIGACION CRIMNAL" con los licenciados en criminalística Carlos Parraga y Marcelo Eber, se mostraron interesados en el proyecto, ya que este software en desarrollo les ayudará a analizar las imágenes de un ordenador de una manera mucho más ágil, en especial cuando la cantidad de fotos a analizar es abundante.

Tras tener una conversación realizada el 11 de septiembre de 2018 con la Ing. Maria Vignau, coordinadora de flisol resistencia, comunicó su apoyo para el desarrollo del proyecto. El aporte de una solución libre permitirá la disminución de los costos para los estudios criminalísticos que actúan de parte; como así también ayudará a ahorrar en recursos monetarios para las entidades estatales que decidan optar por ésta solución.

Desarrollo de software

Se deben tener en cuenta los siguientes aspectos:

I. El lenguaje a usar para la construcción del software es JAVA, bajo el entorno de desarrollo integrado NetBeans versión 8.2, ya que es menester que el software desarrollado posea una gran portabilidad.

II. Los datos con los que el software trabajará son no-estructurados, esto lleva a optar por un DBMS no-sql con datos organizados en forma de documentos; la orientación a documentos es más flexible a los cambios. Una estructura convencional de tablas, de los DBMS tradicionales no es óptima, ya que es estática e introducir cambios generará repercusión en el resto del software.

III. El algoritmo de búsqueda de paquetes se realiza en frio, es decir, que los paquetes obtenidos del sniffing son guardados en un archivo de texto plano. Luego se procede a buscar todos los paquetes que pertenezcan a una misma imagen.

Se ha optado por un acceso secuencial por sobre un acceso aleatorio a un archivo, esto es por:

1. Las herramientas de sniffeo de redes, disponen todos los paquetes que pertenecen a un archivo transmitido por la red uno a continuación de otros.
2. El acceso aleatorio, en archivos de gran tamaño, producen más operaciones de entrada-salida; lo que lleva a una perdida en la performance del algoritmo.

El sitio web qnap menciona las buenas prácticas para el rendimiento en almacenamiento, se puede trasladar el siguiente ejemplo a lo anteriormente planteado:

> "…*Para sistemas mecánicos basados en discos, donde la búsqueda en cada disco llevará alrededor de 10 ms. Escribir los datos de forma secuencial en ese mismo disco tarda alrededor de unos 30 ms por MB. Así que si usted escribe secuencialmente 100 MB de datos en un disco, tardará alrededor de 3 segundos. Pero si hace 100 escrituras aleatorias de 1 MB cada una, eso tardará un total de 4 segundos (3 segundos para la escritura real, y 10 ms * 100 = 1 segundo para toda las búsquedas)…*"[11]

IV. Así mismo, también se ha descartado la utilización de hilos y concurrencia, esto es porque: los paquetes de un solo archivo de imagen se encuentran uno a continuación de otro, generar un nuevo hilo para buscar un nuevo paquete, además de que la concurrencia solo se limita para el computo de un solo core, esto es redundante y poco eficiente.

V. Para la reconstrucción de una imagen a partir de sus paquetes, se utilizan las librerías graficas que son proporcionadas por el lenguaje, de esta manera el formato del archivo resulta poco relevante al momento de reconstruir una imagen. Sin embargo, es importante realizar una correcta transformación de sistema hexadecimal (contenido de un paquete capturado) a sistema decimal, para que las librerías anteriormente mencionadas puedan proceder de manera correcta.

VI. Posteriormente, implementando las librerías de OpenCV, el software realiza un meticuloso análisis de los elementos reconstruidos.

Desarrollo de los escáneres

Un escaner de texturas convencional trabaja a partir de un porcentaje definido por la siguiente expresión matemática.

$$\frac{cantidad\ de\ pixeles\ de\ color\ piel}{cantidad\ de\ pixeles\ que\ no\ son\ color\ piel}\ x\ 100$$

Este simple modelo, es implementado en aplicaciones como Nude de iOS, la API de ParallelDots Nudity Detection, NudityDetectioni2v entre otros. Este modelo también es mencionado por los siguientes autores.

- Marcial-Basilio, Jorge et al. 2011. Detección de pornografía en imágenes digitales. Revista Internacional De Computadoras Número 2, Volumen 5.
- Singh, Sanjay et al. 2003. Un algoritmo robusto de detección de rostro basado en el color de la piel. Revista Tamkang de Ciencia e Ingeniería. Número 4, Volumen. 6

El escáner de textura propuesto, funciona mediante un Coeficiente Mínimo de Desnudez (CMD) que se define de la siguiente manera

$$CMD = \frac{cantidad\ de\ pixeles\ de\ color\ piel}{cantidad\ de\ pixeles\ que\ no\ son\ color\ piel}\ (1)$$

[11] qnap.com/es-es/how-to/tutorial/article/las-mejores-pr%C3%A1cticas-para-el-rendimiento-de-almacenamiento-de-qnap/

Para simplificar el conteo de pixeles de la expresión (1), se toma la imagen en el espacio de colores RGB y se la binariza transformando los pixeles colores piel a pixeles blancos, y pixeles que no son colores piel se transformaron a pixeles de colores negros quedando de la siguiente manera:

$$CMD = \frac{cantidad\ de\ pixeles\ de\ color\ blanco}{cantidad\ de\ pixeles\ de\ color\ negro} \qquad (2)$$

Este último modelo propuesto, difiere del mencionado por otros autores ya que:

1. Marcial-Basilio trabaja con un conjunto de colores piel, y el modelo solo trabaja con dos colores (blanco y negro).
2. Marcial-Basilio propones utilizar un porcentaje, y el modelo propuesto no lo utiliza, ya que las operaciones de punto flotante afectan el desempeño de la aplicación en una computadora, es por ello que se reduce el uso de dichas operaciones de punto flotante.
3. Singh, Sanjay utiliza múltiples espacios de colores (RGB, YCbCr, HSI), mientras que este modelo solo trabaja con un solo espacio de colores (RGB). Se considera trabajar con un solo espacio de colores, ya que y los errores de clasificación pueden ser corregidos por el escáner de objetos a implementar.

A partir de 200 imágenes de desnudos, se ha determinado su coeficiente de desnudez, posteriormente se ha tomado la mediana, y es este valor el coeficiente mínimo de desnudez.

Ahora bien, para las imágenes que se quieran determinar que si son desnudos se realizan los siguientes pasos:

1. Calcular el coeficiente de desnudez de la imagen
2. Comparar con el coeficiente mínimo de desnudez
3. Si el coeficiente de la foto es mayor o igual al CMD entonces la imagen es un desnudo, caso contrario no es un desnudo.

Es una herramienta útil para un gran volumen de imágenes.

El escáner de objetos se hace uso de los resultados publicados de machine learning por OpenCV, los cuales son archivos de extensión xml. Este es capaz de detectar un cuerpo completo, el tren inferior o el tren superior de una persona en una foto. Este método corrige los errores del escáner de texturas, como por ejemplo la foto de un primer plano de una mano o de un rostro, pueden ser considerados como desnudos.

El reconocimiento facial se implementa mediante la utilización de técnicas de machine learning, se entrena a la aplicación para poder reconocer a una persona determinada, siendo un mínimo de dos fotos para que el mismo pueda reconocer, en un gran volumen de fotos, al individuo deseado; a medida que la cantidad de fotos aumenten para el entrenamiento, la precisión del programa será mayor.

El modelo orientado a documentos

Este modelo es uno no-SQL, es decir que:

- No se encuentra estructurado en tablas
- No son aplicables los conceptos de algebra relacional
- No contempla transacciones ACID (aunque algunos motores las incluyen)

Específicamente, el motor de base de datos elegido para llevar a cabo es MongoDB (como ya se ha nombrado anteriormente).

Un modelo orientado a documentos gestiona información semi-estructurada. La elección del mismo radica en esta cualidad del modelo, ya que el problema que se tiene se encuentra en el ámbito con datos desestructurados y/o semi-estructurados.

Antes de continuar, es importante aclarar los siguientes conceptos, que pueden llevar a la confusión:

- Colección: es un conjunto de documentos, su equivalente en el modelo relacional es una tabla común y corriente.
- Documento: consiste en un conjunto de datos, semejante a los registros de una tabla.

La ventaja de este tipo de tecnologías es que se pueden reestructurar las colecciones a medida que se lo desee, sin perder la estructuración anterior.

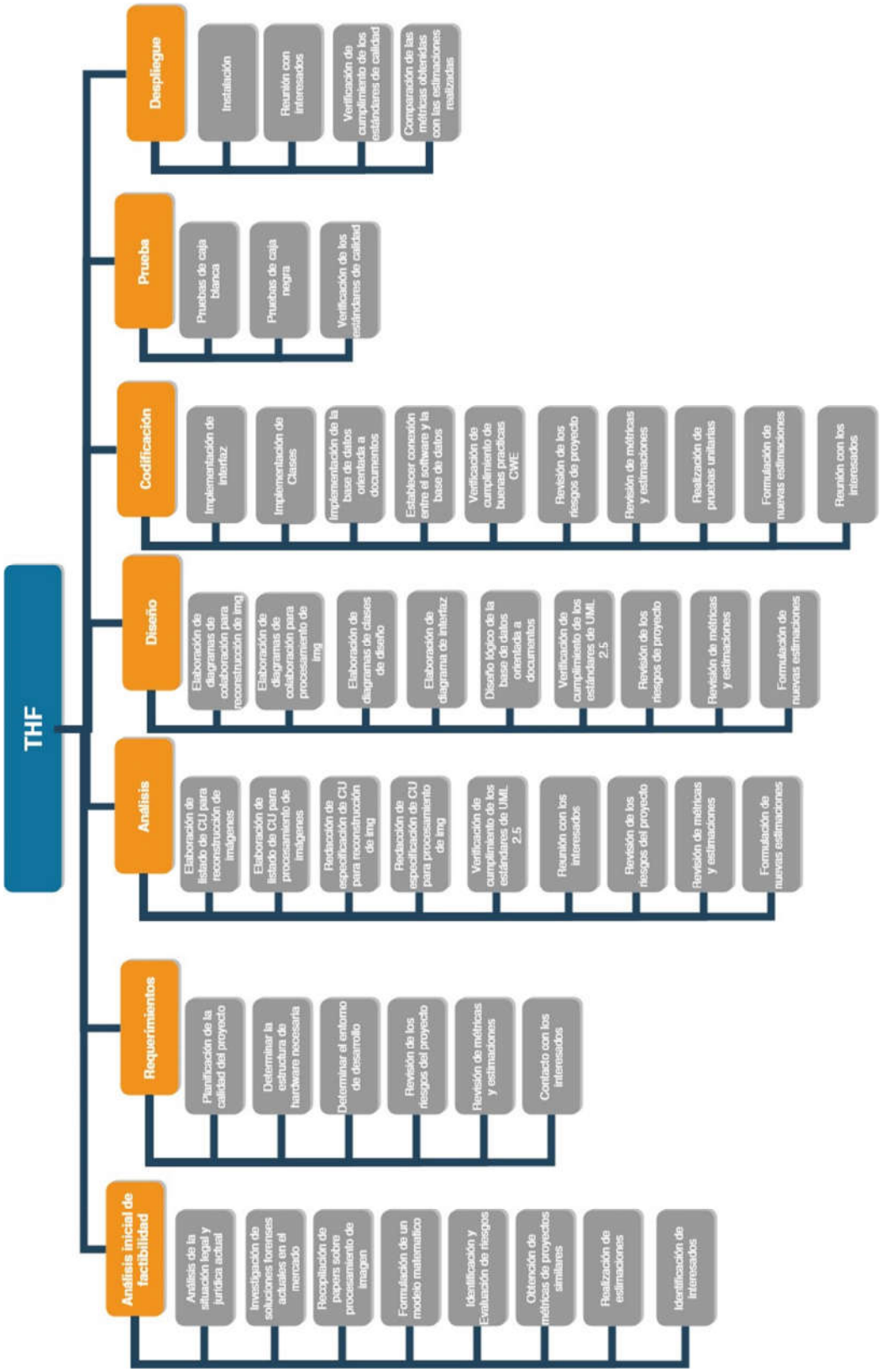
Por ejemplo: En un primer momento se crea una colección que almacena el nombre y la edad de la persona, con el transcurso del tiempo surge la necesidad de almacenar también el domicilio y el sexo de la persona, este cambio no causa estragos en los documentos anterior. Esta forma de estructuración es dinámica, es decir, se realiza en el momento de la carga de los datos como ser:

1ra carga: {nombre; apellido;}
2da carga: {nombre; apellido; domicilio; sexo}
3ra carga: {nombre; apellido; domicilio; sexo; teléfono}
4ta carga {DNI; apellido; teléfono}
Esta flexibilidad también permite modificar de manera más sencilla las aplicaciones, cuando se amplían los requerimientos o las necesidades del usuario final cambian.

El modelo orientado a documentos de la base de datos no-SQL se puede apreciar en el Anexo 3

Formulación del Proyecto

Siguiendo lo establecido por la metodología de proceso unificado, se puede ver claramente la planificación del proyecto software reflejado en el siguiente WBS. El mismo permite distinguir fácilmente los principales procesos asociados con los subprocesos necesarios para asegurar el cumplimiento del proyecto.


THF
Despliegue
Instalación
Reunión con interesados
Verificación de cumplimiento de los estándares de calidad
Comparación de las métricas obtenidas con las estimaciones realizadas
Prueba
Pruebas de caja blanca
Pruebas de caja negra
Verificación de los estándares de calidad
Codificación
Implementación de interfaz
Implementación de Clases
Implementación de la base de datos orientada a documentos
Establecer conexión entre el software y la base de datos
Verificación de cumplimiento de buenas practicas CWE
Revisión de los riesgos de proyecto
Revisión de métricas y estimaciones
Realización de pruebas unitarias
Formulación de nuevas estimaciones
Reunión con los interesados
Diseño
Elaboración de diagramas de colaboración para reconstrucción de img
Elaboración de diagramas de colaboración para procesamiento de img
Elaboración de diagramas de clases de diseño
Elaboración de diagrama de interfaz
Diseño lógico de la base de datos orientada a documentos
Verificación de cumplimiento de los estándares de UML 2.5
Revisión de los riesgos del proyecto
Revisión de métricas y estimaciones
Formulación de nuevas estimaciones
Analisis
Elaboración de listado de CU para reconstrucción de imágenes
Elaboración de listado de CU para procesamiento de imágenes
Redacción de especificación de CU para reconstrucción de img
Redacción de especificación de CU para procesamiento de img
Verificación de cumplimiento de los estándares de UML 2.5
Reunión con los interesados
Revisión de los riesgos del proyecto
Revisión de métricas y estimaciones
Formulación de nuevas estimaciones
Requerimientos
Planificación de la calidad del proyecto
Determinar la estructura de hardware necesaria
Determinar el entorno de desarrollo
Revisión de los riesgos del proyecto
Revisión de métricas y estimaciones
Contacto con los interesados
Análisis inicial de factibilidad
Análisis de la situación legal y jurídica actual
Investigación de soluciones forenses actuales en el mercado
Recopilación de papers sobre procesamiento de imagen
Formulación de un modelo matemático
Identificación y Evaluación de riesgos
Obtención de métricas de proyectos similares
Realización de estimaciones
Identificación de interesados

1. Análisis inicial de factibilidad

1.1. Análisis de la situación legal y jurídica actual: Se analiza que establece la legislación vigente, respecto a la problemática en cuestión. Se estudian los procedimientos actuales y las tecnologías involucradas que utilizan los profesionales en el área.

1.2. Investigación de soluciones forenses actuales en el mercado: En base a las tecnologías utilizadas en el mercado por los profesionales del área, se realizará una comparación entre estas últimas y la solución propuesta

1.3. Recopilación de papers sobre el procesamiento de imagen: Numerosos expertos han realizado investigaciones proponiendo nuevos algoritmos y más eficientes. En este aspecto, se compara lo propuesto con los más versados, definiendo que algoritmo es el que mejor se adapta para constatar la presencia de la desnudez, la detección de objetos y el reconocimiento facial

1.4. Formulación de un modelo matemático: Esto permite obtener una importante métrica acerca de la complejidad del software, como así también establece las bases con las cuales comparar el desempeño del software.

1.5. Identificación y Evaluación de riesgos: Aquí se evalúan todos aquellos factores que puedan inferir en el proyecto, es decir, se considera las oportunidades y amenazas que se presenten a lo largo del proyecto y como SE lidian con ellas.

1.6. Obtención de métricas de proyectos similares: Con el fin de evaluar el avance del proyecto, se identifican las variables más representativas que indiquen la situación actual del proyecto, como así también cuales son los valores deseables de los mismos, se realizarán intercambios con expertos para conseguir más conocimiento sobre qué factores se deben supervisar.

1.7. Realización de estimaciones: Se asigna un valor esperado para cada variable, con un rango aceptable de error, que permita comparar la situación actual del proyecto con la situación esperada. Para esto se realizan estimaciones tradicionales por puntos de función.

1.8. Identificación de interesados: Se reconocerá a aquellas personas y entidades que afectan y puedan verse interesados con el proyecto, y se puntualizará cuál es su preponderancia en el proyecto.

2. Requerimientos

2.1. Planificación de la calidad del proyecto: Se definen los planes que indican, qué requerimientos debe cumplir el producto software, y cuáles se serán los estándares de calidad con los que se debe verificar. En este proyecto se ha optado por la utilización de UML 2.5 como estándar para el de análisis y el diseño, las buenas prácticas de common weakness enumeration para la codificación (CWE) y para las pruebas se adopta el estándar ISO/IEC/IEEE 29119.

2.2. Determinar la estructura de hardware necesaria: Se determina el hardware necesario para poder desarrollar el proyecto. También se determina el proveedor más adecuado

2.3. Determinar el entorno de desarrollo: Se comparará el software base, lenguaje, plataforma de desarrollo y repositorio, necesario para concretar el proyecto, como así también la plataforma en la que se instalará el producto terminado.

2.4. Revisión de los riesgos del proyecto: Teniendo en cuenta la adquisición de los recursos tecnológicos necesarios, se revisar los riesgos que asociados a estos.

2.5. Revisión de métricas y estimaciones: Con el avance del proyecto es necesario controlar y comparar las estimaciones con las métricas obtenidas.

2.6. Contacto con los interesados: Se establecen las primeras comunicaciones con los interesados del proyecto, conferencias en la que participar y conocer cuáles serán sus expectativas y sus aportes.

3. Análisis

3.1. Elaboración de listado de CU para reconstrucción de imágenes: Se define qué debe hacer el software para reconstruir imágenes obtenidas a partir de los paquetes resultantes del proceso de sniffing.

3.2. Elaboración de listado de CU para procesamiento de imágenes: se define qué análisis realizará el software sobre las imágenes.

3.3. Redacción de especificación de CU para reconstrucción de imagen: A partir de los casos de uso, se determina con un alto nivel de abstracción la interacción del usuario con el sistema para reconstruir imágenes.

3.4. Redacción de especificación de CU para procesamiento de imagen: A partir de los casos de uso de procesamiento de imagen, se determina con un alto nivel de abstracción la interacción del usuario con el sistema para analizar imágenes.

3.5. Verificación de cumplimiento de los estándares de UML 2.5: Se evalúa que la documentación realizada cumpla con los estándares de calidad establecidos.

3.6. Reunión con los interesados: Se comunican los avances del proyecto, como así también se expresan nuevamente las expectativas en función de la información proporcionada.

3.7. Revisión de los riesgos del proyecto: Con el trabajo realizado hasta la fecha, se revisan los riesgos actuales, y si estos aún pueden afectar al proyecto. Se evalúan posibles nuevos riesgos.

3.8. Revisión de métricas y estimaciones: Con el trabajo realizado hasta la fecha, se compara las métricas obtenidas versus el avance del proyecto y se las compara con las estimaciones realizadas, para saber el estado de la situación actual.

3.9. Formulación de nuevas estimaciones: Este proceso iterativo de realizar nuevas estimaciones, permiten un mayor grado de precisión y confiabilidad al momento de realizar el seguimiento del proyecto.

4. Diseño

4.1. Elaboración de diagramas de colaboración para reconstrucción de imagen: Se define con mayor de detalle el proceso de reconstrucción de imagen. Se especifica cómo tiene actuar el programa para reconstruir imágenes y como deben interactuar los objetos para cumplir con esta tarea.

4.2. Elaboración de diagramas de colaboración para procesamiento de imagen: Se define con mayor detalle el procedimiento para el procesamiento de imagen. Se especifica cómo tiene actuar el programa para analizar imágenes y como deben interactuar los objetos para cumplir con estas tareas.

4.3. Elaboración de diagramas de clases de diseño: La documentación específica las relaciones existentes entre los objetos propuestos en la etapa de análisis.

4.4. Elaboración de diagrama de interfaz: Se determina cómo debe ser vista la herramienta para el usuario, la misma debe ser intuitiva y amigable.

4.5. Diseño lógico de la base de datos orientada a documentos: Se propone un esquema para la base de datos, la misma es no-SQL, por lo tanto, el diseño lógico será definido en pseudo código.

4.6. Verificación de cumplimiento de los estándares de UML 2.5: Se evalúa que la documentación realizada cumpla con los estándares de calidad establecidos.

4.7. Revisión de los riesgos del proyecto: Con el trabajo realizado hasta la fecha, se revisan los riesgos actuales, y si estos aún pueden afectar al proyecto. Se evalúan posibles nuevos riesgos.

4.8. Revisión de métricas y estimaciones: Con el trabajo realizado hasta la fecha, se compara las métricas obtenidas versus el avance del proyecto y se las compara con las estimaciones realizadas, para saber el estado de la situación actual.

4.9. Formulación de nuevas estimaciones: Este proceso iterativo de realizar nuevas estimaciones, permiten un mayor grado de precisión y confiabilidad al momento de realizar el seguimiento del proyecto.

5. Codificación

5.1. Implementación de interfaz: En base al diseño propuesto, se realiza la construcción de la interfaz.

5.2. Implementación de Clases: En función de los diagramas de comportamiento realizados anteriormente, se codificarán los mismos siguiendo las especificaciones de la etapa de diseño.

5.3. Implementación de la base de datos orientada a documentos: Con el esquema lógico de la base de datos definido, se implementa en un motor de base de datos orientado a documentos.

5.4. Establecer conexión entre el software y la base de datos: Se determinan e inician los servicios necesarios para que la aplicación pueda conectarse a la base de datos.

5.5. Verificación de cumplimiento de buenas prácticas CWE: Se controla que la aplicación realizada, en esta etapa, cumpla con las buenas prácticas establecidas.

5.6. Revisión de los riesgos del proyecto: Con el trabajo realizado hasta la fecha, se revisan los riesgos actuales, y si estos aún pueden afectar al proyecto. Se evalúan posibles nuevos riesgos.

5.7. Revisión de métricas y estimaciones: Con el trabajo realizado hasta la fecha, se compara las métricas obtenidas versus el avance del proyecto y se las compara con las estimaciones realizadas, para saber el estado de la situación actual.

5.8. Realización de pruebas unitarias: Primeras pruebas para controlar el correcto funcionamiento del producto software.

5.9. Formulación de nuevas estimaciones: Este proceso iterativo de realizar nuevas estimaciones, permiten un mayor grado de precisión y confiabilidad al momento de realizar el seguimiento del proyecto.

5.10. Reunión con los interesados: Se comunican los avances del proyecto, como así también se expresan nuevamente las expectativas en función de la información proporcionada.

6. Prueba

6.1. Pruebas de caja blanca: "*...diseño de casos de prueba que usa la estructura de control descrita como parte del diseño a nivel de componentes para derivar casos de prueba....* "[12]

6.2. Pruebas de caja negra: *"...se enfocan en los requerimientos funcionales del software; es decir, las técnicas de prueba de caja negra le permiten derivar conjuntos de condiciones de entrada que revisarán por completo todos los requerimientos funcionales para un programa…* "[13]

6.3. Verificación de los estándares de calidad: Se controla que las pruebas planificadas en esta etapa cumplan con el estándar ISO/IEC/IEEE 29119.

7. Despliegue

7.1. Instalación: Se implementan el producto software en las máquinas de los usuarios.

7.2. Reunión con interesados: Se comunican si las expectativas han sido cumplidas y el grado de satisfacción de los mismos

7.3. Verificación de cumplimiento de los estándares de calidad: Se realiza una revisión global del cumplimiento de todos los estándares establecidos a lo largo del proyecto.

7.4. Comparación de las métricas obtenidas con las estimaciones realizadas: Se cotejan las métricas obtenidas con las estimaciones realizadas, a fin de determinar la precisión de las estimaciones.

El progreso del proyecto se aprecia en el Anexo 3

Arquitectura de la propuesta tecnológica

Para llevar a cabo el proyecto es necesario contar son los recursos anteriormente definidos:

- Software base (sistema operativo, motor de base de datos)
- Lenguaje de programación
- Entorno de desarrollo integrado
- Repositorio y Sistema controlador de versiones
- Herramientas CASE
- Ordenadores

[12] [13] Pressman R. 2010. Ingeniería de software enfoque práctico. 736 páginas. Editorial McGraw-Hill ISBN: 978-607-15-0314-5

El producto software se deberá ejecutar en cualquier ordenador con cualquier sistema operativo, es decir, que el sistema desarrollado debe ser multiplataforma.

Gestión del Riesgo

Análisis de la situación actual del mercado

Para poder realizar esta tarea, utilizaremos el esquema propuesto por Michael Porter, el cual establece los siguientes factores que se deben tener en cuenta, Y las plantea como fuerzas para formulación de estrategias.

1. La competencia entre compañías.

2. La amenaza de nuevas compañías que entran al mercado.

3. La posibilidad de usar productos o servicios sustitutos.

4. El poder de negociación de los proveedores.

5. El poder de negociación de los compradores o clientes.[14]

En función a esto, a la fecha 20 de noviembre de 2018 se comprobó lo siguiente:

1. La competencia entre compañías:
a. Empresas extranjeras trabajan en el país mediante el sistema de partner (socios)
b. Empresas nacionales están logrando una aceptación en el mercado. Ej: Fundación Sadosky y asociados
2. La amenaza de nuevas compañías que entran al mercado.
a. En el país están surgiendo proyectos y emprendimientos de desarrollo forenses, ya que el costo de inversión inicial resulta bajo
b. El gobierno de la nación fomenta la investigación y desarrollo. Ej: Fonsoft
3. La posibilidad de usar productos o servicios sustitutos.
a. Existen herramientas de código abierto, si bien no cuentan con soporte oficial, tienen una comunidad que respalda el software.
4. El poder de negociación de los proveedores.
a. Existen gran variedad de proveedores que proporcionan entornos de desarrollo y librerías para la construcción de software forense.
5. El poder de negociación de los compradores o clientes.
a. Hay una gran cantidad de clientes dispuestos a comprar estos productos, ya que no solo se limita al mercado local, también es posible realizar negociaciones con clientes en el extranjero.

Como podemos ver; el mercado es emergente y próspero, sin embargo se debe tomar acciones rápidas para poder aprovechar la situación actual.

[14] Cannice M, Koontz H, Weihrich H. 2012. Administración Una perspectiva global y empresarial. 736 páginas. Editorial McGrawHill. ISBN 978-607-15-0759-4

Análisis de factores internos

Con el fin de conocer las fortalezas y debilidades involucradas que estarán presentes durante el desarrollo del proyecto, se ha desarrollado un relevamiento identificando a los mismos. A continuación, se presenta las fortalezas y debilidades descubiertas:

Fortalezas

- Factor de Innovación
- Recursos humanos indicados
- Conocimiento disponible

Debilidades

- Imposibilidad de formular un modelo matemático acorde.
- Subestimación del tamaño
- Rotación de personal

Análisis de factores externos

Amenazas

- Discontinuidad en el software base elegido.
- Discontinuidad en las librerías utilizadas.
- Documentación escasa sobre los algoritmos relacionados.
- Daños en los medios de almacenamiento donde se encuentra el código fuente.
- Daños en los medios de almacenamiento donde se encuentra la documentación realizada.

Oportunidades

- Presentación del software en foros de difusión informática.
- Profesionales interesados en colaborar en el proyecto.
- Profesionales interesados en invertir en el proyecto.
- Profesionales interesados en usar el producto software.

Proyección del riesgo

Se evalúa el riesgo, es decir las amenazas, oportunidades y debilidades que afecten la continuidad del proyecto; se realizará en función del impacto que puede llegar a afecar la continuidad del proyecto y probabilidad de que ocurra el resgo en cuestion, para ello se hace uso de la siguiente matriz sugerida en pmbok:

Probabilidad	Amenazas					Oportunidades				
0.9	0.05	0.09	0.18	0.36	0.72	0.72	0.36	0.18	0.09	0.05
0.7	0.04	0.07	0.14	0.28	0.56	0.56	0.28	0.14	0.07	0.04
0.5	0.03	0.05	0.1	0.2	0.4	0.4	0.2	0.1	0.05	0.03
0.3	0.02	0.03	0.06	0.12	0.24	0.24	0.12	0.06	0.03	0.02
0.1	0.01	0.01	0.02	0.04	0.08	0.08	0.04	0.02	0.01	0.01
Impacto	0.05	0.1	0.2	0.4	0.8	0.8	0.4	0.2	0.1	0.05
	muy bajo	Bajo	Moderado	Alto	muy alto	muy alto	Alto	moderado	bajo	muy bajo

En base a la matriz anterior, para cada amenaza le asignamos una probabilidad a partir de los valores que se encuentra en la columna de "Probabilidad" de la matriz. También asignamos un impacto a partir de la fila "Impacto". La calificación del riesgo es el resultado de seleccionar los valores para el impacto y la probabilidad. Dichos valores fueron elegidos en función de la experiencia.

Amenaza	Probabilidad	Impacto	Calificación
Discontinuidad en el software base optado.	0.1	0.4	0.04
Discontinuidad en las librerías utilizadas	0.1	0.4	0.04
Documentación escasa sobre los algoritmos relacionados	0.5	0.1	0.05
Imposibilidad de formular un modelo matemático acorde	0.5	0.2	0.1
Daños en los medios de almacenamiento donde se encuentra el código fuente	0.3	0.4	0.12
Daños en los medios de almacenamiento donde se encuentra la documentación realizada	0.3	0.4	0.12
Rotación de personal	0.1	0.8	0.08
Subestimación del tamaño	0.5	0.4	0.2

De igual modo para las Oportunidades se tiene:

Oportunidad	Probabilidad	Impacto	Calificación
Presentación del software en foros de difusión de informática.	0.7	0.1	0.07
Profesionales interesados en colaborar en el proyecto.	0.3	0.2	0.06
Profesionales interesados en invertir en el proyecto.	0.1	0.8	0.08
Profesionales interesados en usar el producto software	0.3	0.8	0.24

Planificación de la gestión del riesgo

Una vez que se ha logrado calificar correctamente los riesgos, se define una estrategia y un plan de acción para cada uno de ellos. Primero, se ordena de mayor a menor las amenazas en funcion de la datos obtenidos y se asigna una de las cuatro estrategias propuestas por PMBok

Amenaza	Calificación	Estrategia
Subestimación del tamaño	0.2	Evitar
Daños en los medios de almacenamiento donde se encuentra la documentación realizada	0.12	Transferir
Daños en los medios de almacenamiento donde se encuentra el código fuente	0.12	Transferir
Imposibilidad de formular un modelo matemático acorde	0.1	Evitar
Rotación de personal	0.08	Aceptar
Documentación escasa sobre los algoritmos relacionados	0.05	Evitar
Discontinuidad en el software base optado	0.04	Aceptar
Discontinuidad en las librerías utilizadas	0.04	Aceptar

En función a las estrategias pactadas, se define el plan de respuesta al riesgo. A continuación, se disponen los lineamientos de los mismos:

Plan para la subestimación del tamaño: Un continuo y correcto seguimiento de las métricas obtenidas y con la comparación con las estimaciones realizadas, se puede tomar a tiempo acciones necesarias, como incrementar las horas de trabajo retrasando otras tareas, ante demoras renegociar una nueva fecha de entrega pactada.

Plan para daños en los medios de almacenamiento donde se encuentra la documentación realizada: Con el uso de nuevas tecnologías como es el Cloud Storage, estas empresas se encargan del almacenamiento y protección de la información. Cada vez que se actualiza la documentación, deben subirse los cambios a la nube, esto es para mantener los respaldos actualizados.

Plan para daños en los medios de almacenamiento donde se encuentra el código fuente: Con las tecnologías como los sistemas controladores de versiones y repositorios web, se transfiere la responsabilidad a empresas especializadas. Por cada modificación realizada en el código, se deben subir los cambios en el repositorio.

Imposibilidad de formular un modelo matemático acorde: Si se prevé que ya se ha ocupado la mitad del tiempo asignado para formular un modelo matemático, y hay poco o nulo avance, será el momento de consultar las mejores prácticas y propuestas de otros autores.

Plan para la Rotación de personal: En función a la estrategia definida previamente, se revisará periódicamente la amenaza para asegurar que esta no varíe de manera significativa y se abordarán los riesgos cuando estos se materialicen. De igual manera, para evitar que con la salida de una persona el conocimiento se vaya, se organizaran reuniones en la que se comparta el progreso individual que haya conseguido el equipo.

Plan para documentación escasa sobre los algoritmos relacionados: Si se prevé que ya se ha ocupado la mitad del tiempo asignado para encontrar documentación y/o investigaciones por expertos, entonces las próximas acciones se orientarán a contactar A los expertos.

Discontinuidad en el software base optado: En función a la estrategia definida previamente, se revisará periódicamente la amenaza para asegurar que esta no varíe de manera significativa y se abordarán los riesgos cuando estos se materialicen.

Discontinuidad en las librerías utilizadas: En función a la estrategia definida previamente, se revisará periódicamente la amenaza para asegurar que esta no varíe de manera significativa y se abordarán los riesgos cuando estos se materialicen.

A continuación, ordenaremos de mayor a menor las oportunidades que se pueden presentar a lo largo del proyecto.

Oportunidad	Calificación	Estrategia
Profesionales interesados en usar el producto software	0.24	Explotar
Profesionales interesados en invertir en el proyecto	0.08	Aceptar
Presentación del software en foros de difusión informática	0.07	Mejorar
Profesionales interesados en colaborar en el proyecto	0.06	Aceptar

Plan para profesionales interesados en usar el producto software: Se asignaran todos los recursos disponibles y el mejor personal, destinados a conseguir una versión "lite" del producto software, completamente funcional para mantener las relaciones con aquellos potenciales usuarios, hasta que la versión full esté disponible.

Plan para profesionales interesados en invertir en el proyecto: Una vez que la oportunidad se presente, se iniciará la gestión de la misma para que los interesados puedan invertir en el proyecto

Plan para presentación del software en foros de difusión informática: Se tendrá preparada la documentación lista para presentar, junto con artículos académicos para presentarlo en tales foros.

Plan para profesionales interesados en colaborar en el proyecto: Una vez que la oportunidad se presente, se iniciará la gestión de la misma para que los interesados puedan realizar sus aportes para la consecución de los objetivos del proyecto.

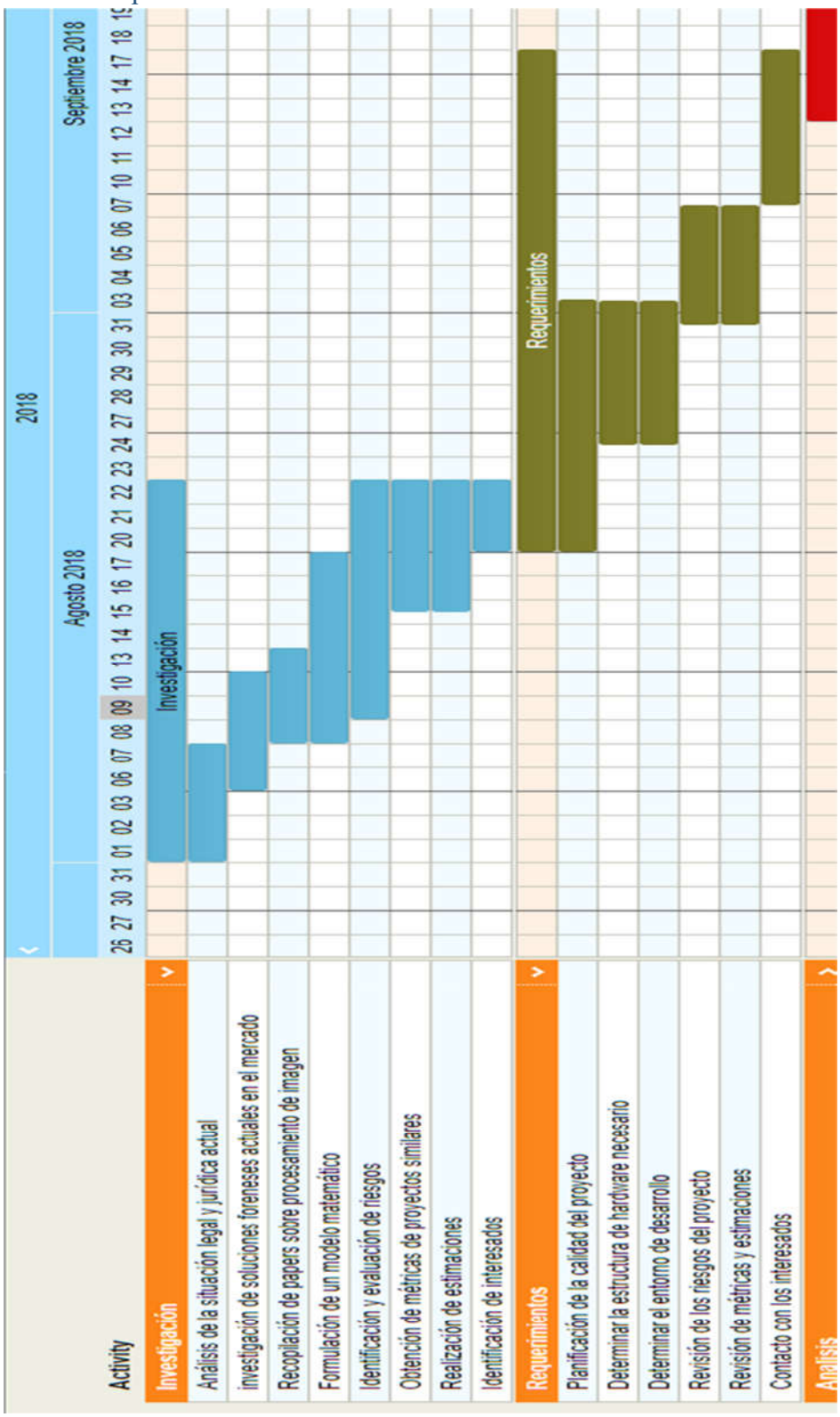
Monitorización del riesgo

Para monitorear los riesgos anteriormente nombrados, se revisará las tendencias en la ejecución del proyecto, teniendo como referencia la información sobre el desempeño, comparando con los resultados planificados. De igual manera otros indicadores, son el tiempo disponible para la realización de la tarea.

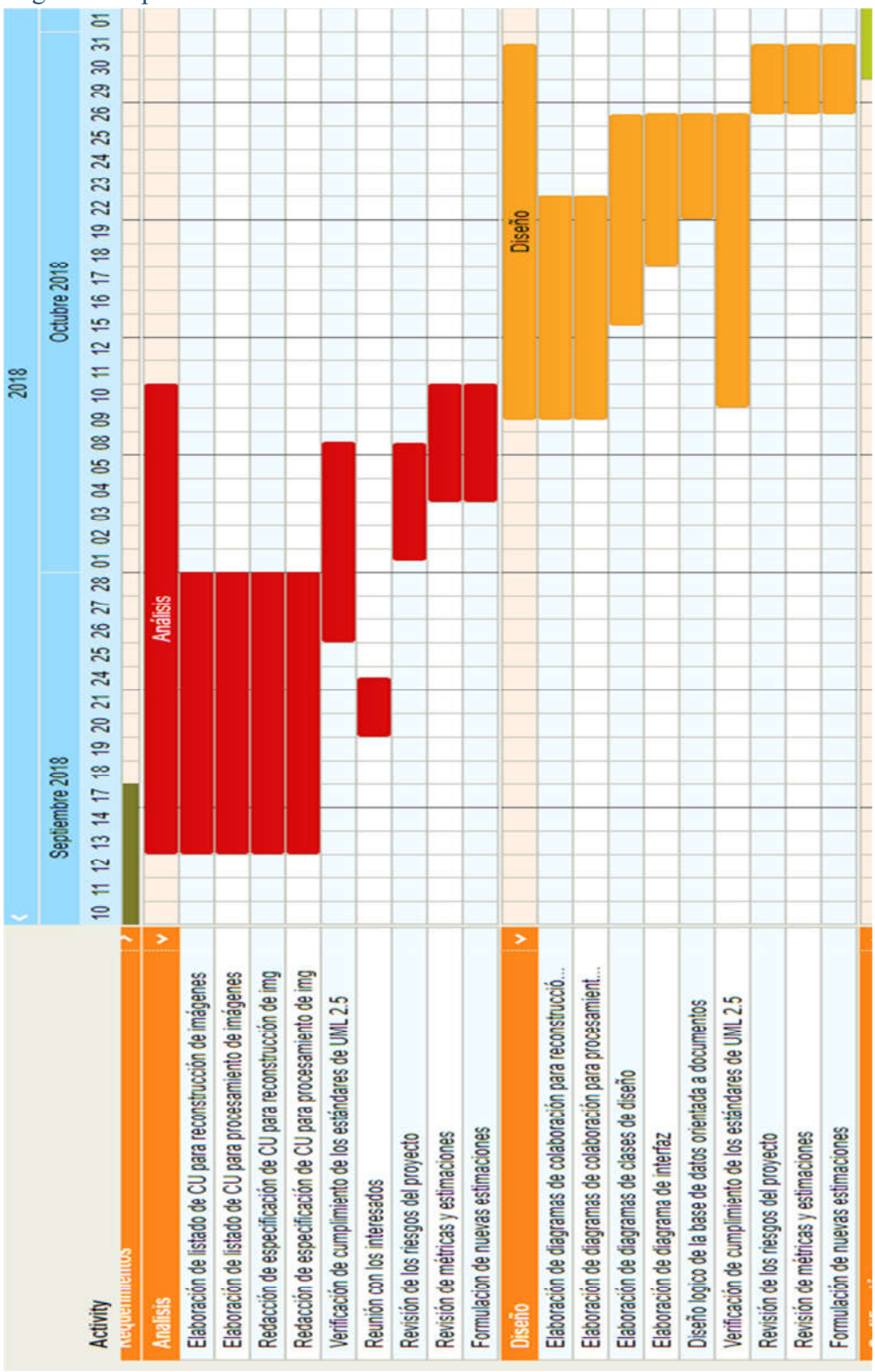
Asignación de tiempos

A partir de los procesos y subprocesos definidos en el WBS, se puede confeccionar el siguiente diagrama de Gantt, el mismo especifica de manera clara los tiempos asignados para cada componente del WBS, a saber:

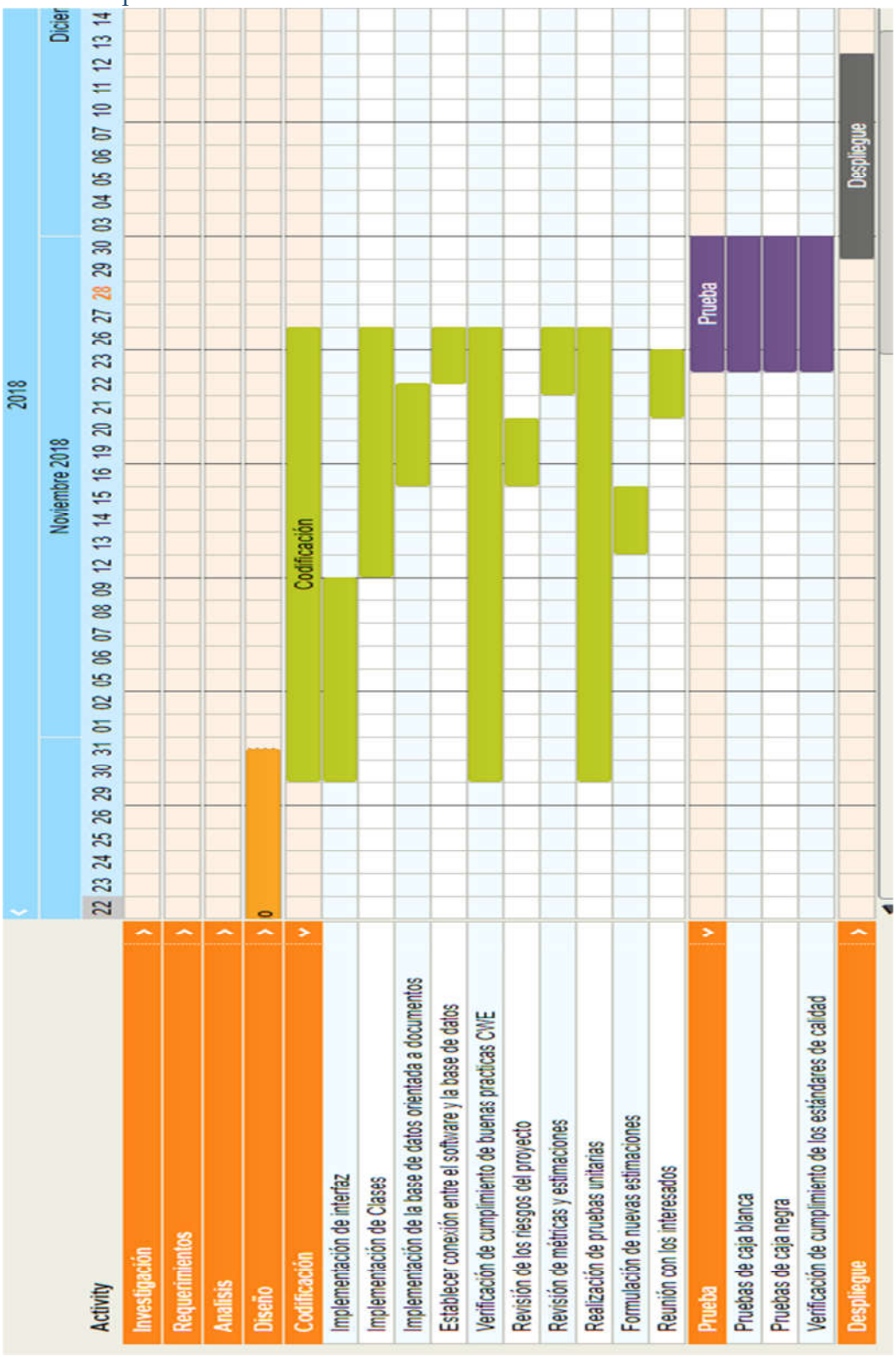
Primera etapa



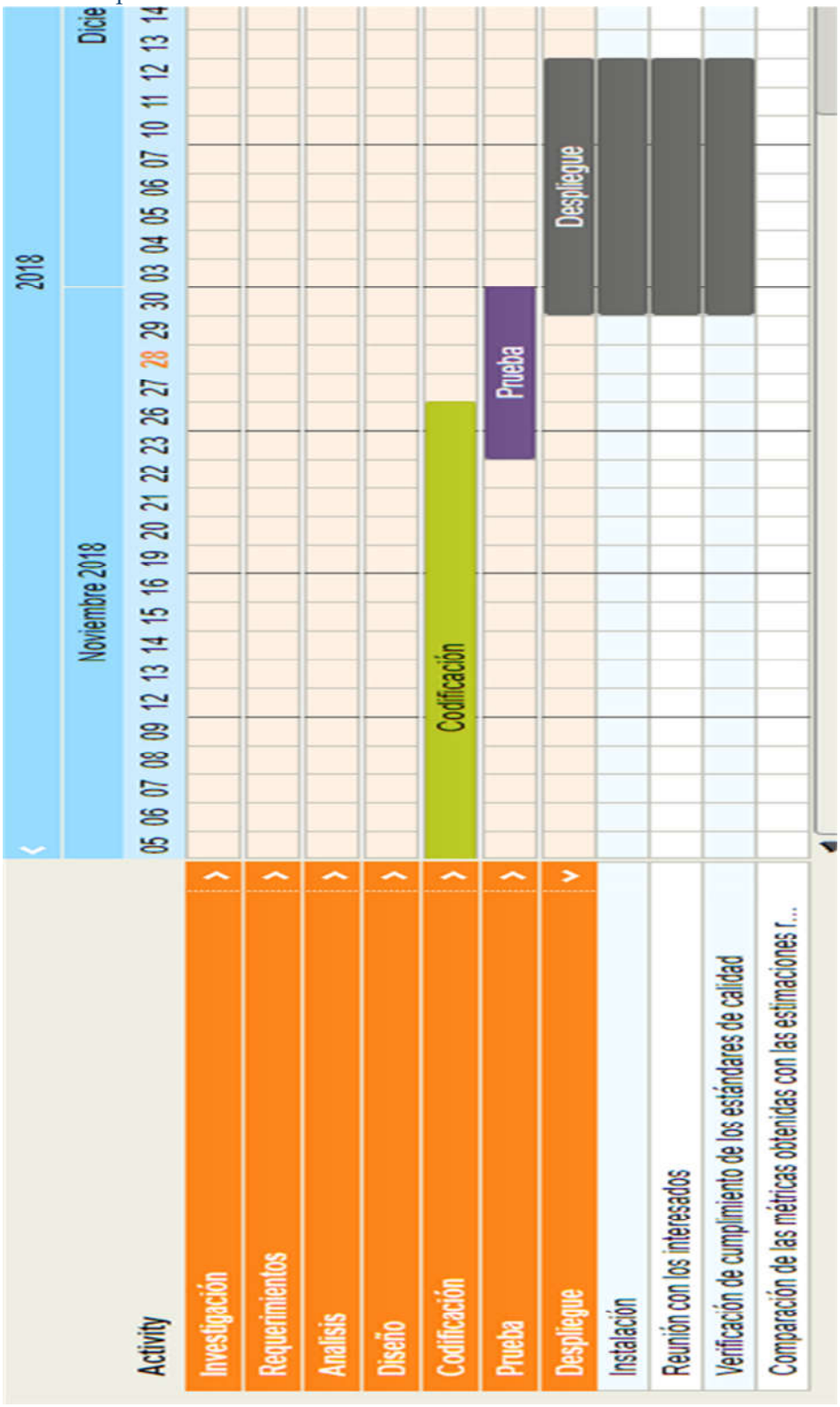
Segunda Etapa



Tercera etapa

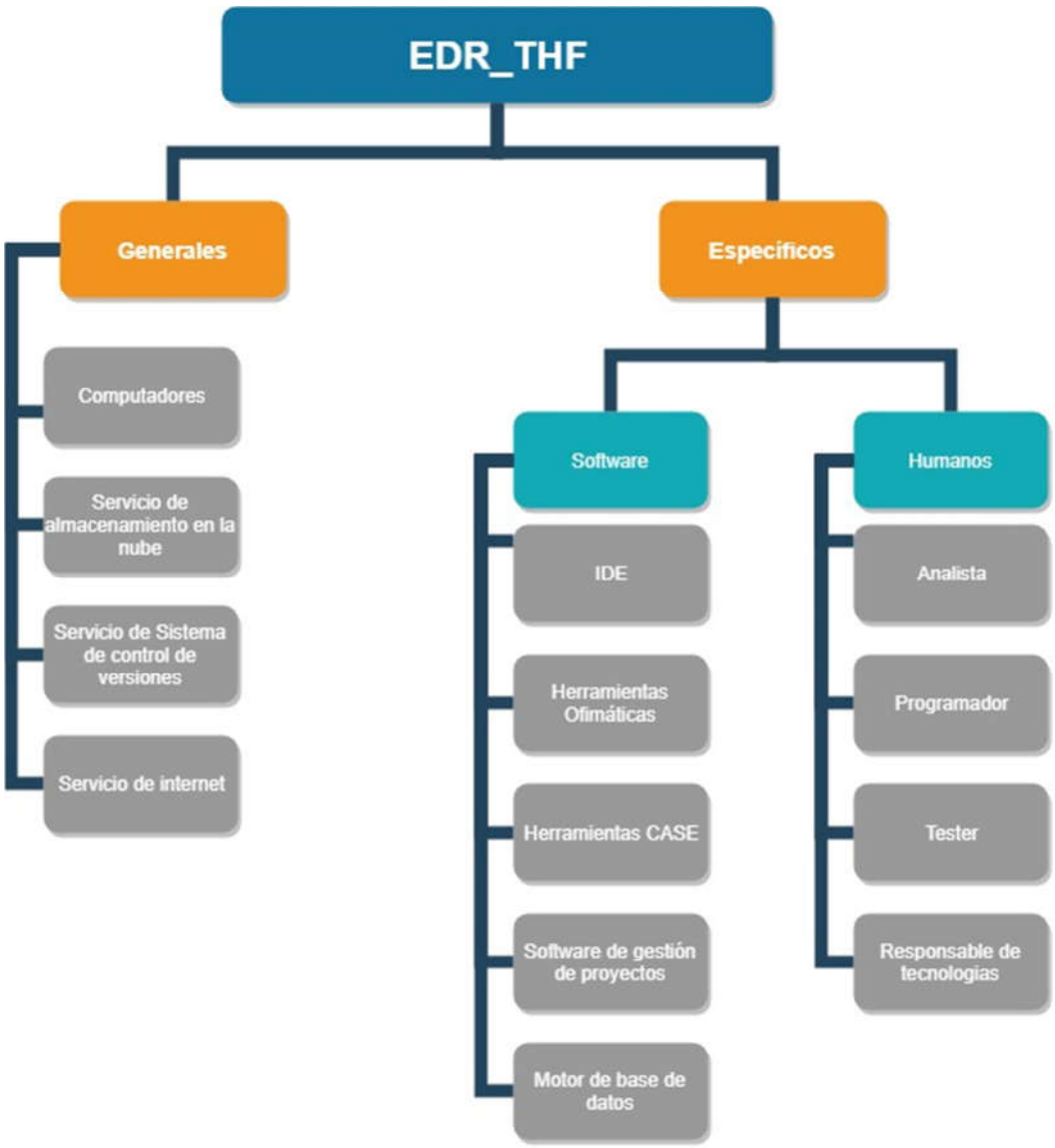


Cuarta etapa

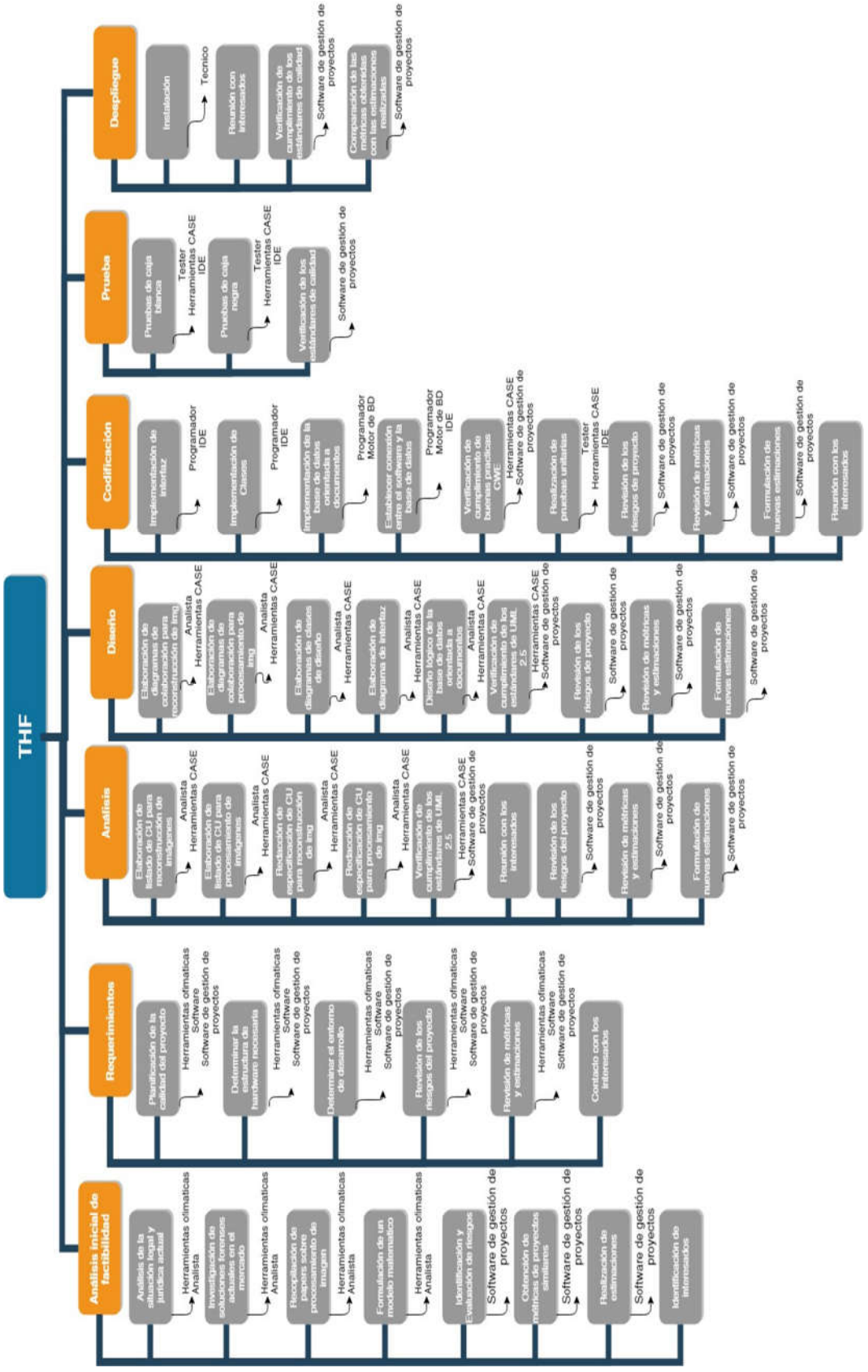


Asignación de Recursos

En el siguiente grafico se pueden ver la estructura de desglose de recursos que se necesitaran a lo largo del proyecto.



Y en el grafico a continuación; se pueden ver los recursos específicos asignados para cada una de las tareas a desarrollar en el proyecto.


THF
Análisis inicial de factibilidad
Análisis de la situación legal y jurídica actual
Herramientas ofimáticas
Analista
Investigación de soluciones forenses actuales en el mercado
Herramientas ofimáticas
Analista
Recopilación de papers sobre procesamiento de imagen
Herramientas ofimáticas
Analista
Formulación de un modelo matemático
Herramientas ofimáticas
Analista
Identificación y Evaluación de riesgos
Software de gestión de proyectos
Obtención de métricas de proyectos similares
Software de gestión de proyectos
Realización de estimaciones
Software de gestión de proyectos
Identificación de interesados
Requerimientos
Planificación de la calidad del proyecto
Herramientas ofimáticas
Software
Software de gestión de proyectos
Determinar la estructura de hardware necesaria
Herramientas ofimáticas
Software
Software de gestión de proyectos
Determinar el entorno de desarrollo
Herramientas ofimáticas
Software
Software de gestión de proyectos
Revisión de los riesgos del proyecto
Herramientas ofimáticas
Software
Software de gestión de proyectos
Revisión de métricas y estimaciones
Herramientas ofimáticas
Software
Software de gestión de proyectos
Contacto con los interesados
Análisis
Elaboración de listado de CU para reconstrucción de imágenes
Analista
Herramientas CASE
Elaboración de listado de CU para procesamiento de imágenes
Analista
Herramientas CASE
Redacción de especificación de CU para reconstrucción de img
Analista
Herramientas CASE
Redacción de especificación de CU para procesamiento de img
Analista
Herramientas CASE
Verificación de cumplimiento de los estándares de UML 2.5
Herramientas CASE
Software de gestión de proyectos
Reunión con los interesados
Software de gestión de proyectos
Revisión de los riesgos del proyecto
Software de gestión de proyectos
Revisión de métricas y estimaciones
Software de gestión de proyectos
Formulación de nuevas estimaciones
Software de gestión de proyectos
Diseño
Elaboración de diagramas de colaboración para reconstrucción de img
Analista
Herramientas CASE
Elaboración de diagramas de colaboración para procesamiento de img
Analista
Herramientas CASE
Elaboración de diagramas de clases de diseño
Analista
Herramientas CASE
Elaboración de diagrama de interfaz
Analista
Herramientas CASE
Diseño lógico de la base de datos orientada a documentos
Analista
Herramientas CASE
Verificación de cumplimiento de los estándares de UML 2.5
Herramientas CASE
Software de gestión de proyectos
Revisión de los riesgos del proyecto
Software de gestión de proyectos
Revisión de métricas y estimaciones
Software de gestión de proyectos
Formulación de nuevas estimaciones
Software de gestión de proyectos
Codificación
Implementación de interfaz
Programador IDE
Implementación de Clases
Programador IDE
Implementación de la base de datos orientada a documentos
Programador Motor de BD
Establecer conexión entre el software y la base de datos
Programador Motor de BD IDE
Verificación de cumplimiento de buenas practicas CWE
Herramientas CASE
Software de gestión de proyectos
Realización de pruebas unitarias
Tester
Herramientas CASE IDE
Revisión de los riesgos del proyecto
Software de gestión de proyectos
Revisión de métricas y estimaciones
Software de gestión de proyectos
Formulación de nuevas estimaciones
Software de gestión de proyectos
Reunión con los interesados
Prueba
Pruebas de caja blanca
Tester
Herramientas CASE IDE
Pruebas de caja negra
Tester
Herramientas CASE IDE
Verificación de los estándares de calidad
Software de gestión de proyectos
Despliegue
Instalación
Técnico
Reunión con interesados
Verificación de cumplimiento de los estándares de calidad
Software de gestión de proyectos
Comparación de las métricas obtenidas con las estimaciones realizadas
Software de gestión de proyectos

Matriz de responsabilidades

A continuación; se presenta la matriz de responsabilidades RACI. En función del Gantt anterior, las tareas son organizadas según ese criterio

R: Responsable

A: Aprueba

C: Consultado

I: Informado

Tarea	Líder de proyecto	Analista	Programador	Tester	Responsable de tecnologías	Especialista Jurídico
Análisis de la situación legal y jurídica actual	A					R
Investigación de soluciones forenses actuales en el mercado	A	R				
Recopilación de papers sobre el procesamiento de imagen		R				
Formulación de un modelo matemático		R	C			
Identificación y Evaluación de riesgos	R					
Obtención de métricas de proyectos similares	R					
Realización de estimaciones	R					
Identificación de interesados	R					
Planificación de la calidad del proyecto	A	C		R		
Determinar la estructura de hardware necesaria	A				R	
Determinar el entorno de desarrollo	A		R		C	
Revisión de los riesgos del proyecto	R					
Revisión de métricas y estimaciones	R					
Contacto con los interesados	R					
Elaboración de listado de CU para reconstrucción de imágenes		R	C			
Elaboración de listado de CU para procesamiento de imágenes		R				
Redacción de especificación de CU para reconstrucción de imagen		R				
Redacción de especificación de CU para procesamiento de imagen		R	C			
Verificación de cumplimiento de los estándares de UML 2.5	A	I		R		
Reunión con los interesados	R					

Revisión de los riesgos del proyecto	R					
Revisión de métricas y estimaciones	R					
Formulación de nuevas estimaciones	R					
Elaboración de diagramas de colaboración para reconstrucción de imagen		R	C			
Elaboración de diagramas de colaboración para procesamiento de imagen		R				
Elaboración de diagramas de clases de diseño		R	C			
Elaboración de diagrama de interfaz		R				
Diseño lógico de la base de datos orientada a documentos		R				
Verificación de cumplimiento de los estándares de UML 2.5	A	I		R		
Revisión de los riesgos de proyecto	R					
Revisión de métricas y estimaciones	R					
Formulación de nuevas estimaciones	R					
Implementación de interfaz		C	R			
Implementación de Clases			R			
Implementación de la base de datos orientada a documentos			R			
Establecer conexión entre el software y la base de datos			R			
Verificación de cumplimiento de buenas prácticas CWE	A	C		R		
Revisión de los riesgos de proyecto	R					
Revisión de métricas y estimaciones	R					
Realización de pruebas unitarias		C		R		
Formulación de nuevas estimaciones	R				R	
Reunión con los interesados	R					
Pruebas de caja blanca				R		
Pruebas de caja negra				R		
Verificación de los estándares de calidad	A	R		I		
Instalación						
Reunión con interesados	R					
Verificación de cumplimiento de los estándares de calidad	A			R		

Comparación de las métricas obtenidas con las estimaciones realizadas	R					

Análisis Económico-Financiero

En la siguiente Tabla, se presenta los costos iniciales al momento de la inversión, se ha calculado su monto en dólares, cuando el valor del mismo estaba en $ 38.07 respecto de la moneda nacional, cifra proporcionada por el banco central de la república Argentina, en el día 02 de octubre de 2018.

Recursos Materiales						
	Monto	Descripción	Cantidad	unidad		precio unitario
Computadoras	735.49	INTEL CORE I5 3,1GHZ 4 NUCLEOS, DISCO SATA 320GB 4GB MEMORIA RAM DDR3, Monitor Lg Flatron 19 W1943se, teclado y mouse noganet	2	equipos		367.74363
Servicio de almacenamiento en la nube	12.5	12.50 dólares precio anual	1	año		12.501182
Servicio de Sistema de control de versiones	45.01	9 dólares convertido a pesos, precio por mes	5	meses		9.00131337
Servicio de internet	94.04	servicio de arnet por mes	5	meses		18.8074599
Software						
IDE	0	Licencia GPLv				
Herramientas ofimáticas	0	Licencia GPLv				
Herramientas CASE	0	Licencia GPLv				
Software de gestión de proyectos	0	Licencia GPLv				

Recursos Humanos	
Rol	Remuneración
Analista	1607.57
Programador	925.93
Tester	761.75
Responsable de tecnologías	525.35
Líder de proyecto	1930.65
Especialista jurídico	47.28

En la tabla 1 del anexo 1 se encuentra el cálculo del monto de los recursos humanos.

Desembolso Mensual

	mes 1	mes 2	mes 3	mes 4	mes 5	Totales
Computadoras	735.49					735.49
Servicio de almacenamiento en la nube	12.5					12.5
Servicio de Sistema de control de versiones	9.002	9.002	9.002	9.002	9.002	45.01
Servicio de internet	18.808	18.808	18.808	18.808	18.808	94.04
Analista	535.856667	535.856667	535.856667			1607.57
Programador				925.93		925.93
Tester			380.875	380.875		761.75
Responsable de tecnologías					525.35	525.35
Líder de proyecto	386.13	386.13	386.13	386.13	386.13	1930.65
Especialista Jurídico	47.28					
Totales	1745,06667	949,796667	1330,67167	1720,745	939,29	6685,57

Tras investigaciones y contacto con los proveedores, un software forense de propósitos generales como Encase Forensic tiene el valor de aproximadamente 2300 dólares; otra empresa desarrolladora como Gryffeye, que realiza análisis de imágenes (software con funcionas más similares a este proyecto), no está trabajando en Argentina, y tampoco tiene interés en ingresar al mercado argentino por el 2018-2019.

Para nuestro análisis usaremos como referencia el precio del software Encase. Para una organización cualquiera que desee usar una herramienta forense, implica que todos los años debe invertir aproximadamente 2300 dólares anuales por la licencia. El uso del producto final del proyecto, lleva al ahorro de esta suma de dinero.

Desembolso Anual

Inflación anual = 0% (los montos trabajados se encuentran en dólares)

TNA = 6%

Periodo	**año 0**	**año 1**	**año 2**	**año 3**	**año 4**	**Año 5**
Ahorro	2300	2300	2300	2300	2300	
Costo	6638.29	0	0	0	0	
Total	-4338,29	**2300**	**2300**	**2300**	**2300**	**0**
Saldo	-4338,29	-2038,29	261,71	2561,71	4861,71	4861,71
Total (ajustado)	-4338,29	2169,81	2046,99	1931,12	1821,81	0
Saldo (Ajustado)	-4338,29	-2168,47	-121,487	1809,63	3631,45	3631,45

El valor actual neto es:

$$VAN = 3631{,}45$$

Para este proyecto calculando el valor de la TIR (usando Excel) tenemos como resultado

$$TIR = 39\%$$

Resultados o verificación experimental

Experimentos y pruebas realizadas

Con el primer prototipo funcional y listo para su uso, se evaluará:

1. Reconstruir imágenes
2. Clasificar imágenes:
    a. Por objeto
    b. Por textura
    c. Por capacidad de reconocer un rostro

Reconstrucción de imágenes

Para proceder con las pruebas se realizó la captura de paquetes, los mismos fueron generados en una red LAN privada, navegando por el sitio web 4chan.org; se eligió el mismo porque no cuenta con certificados de navegación segura. La herramienta utilizada para el sniffing de paquetes es tcpdump, haciendo uso del siguiente comando

```
Tcpdump -A -vv -XX -i wlan0 -l > datos.txt
```

Las características de la computadora utilizada para realizar las pruebas son las siguientes:

- Sistema operativo debían 9 64 bits
- Procesador Intel Core i3 de cuarta generación
- 3GB de memoria RAM
- 80 GB para el sistema de archivos

Se realizaron cuarto capturas, de las cuales las primeras dos se le destino toda la capacidad de procesamiento del ordenador a capturar paquetes, asignándole prioridad mediante el comando `nice`. Las dos capturas restantes, se realizaron con la prioridad asignada por defecto por el sistema operativo.

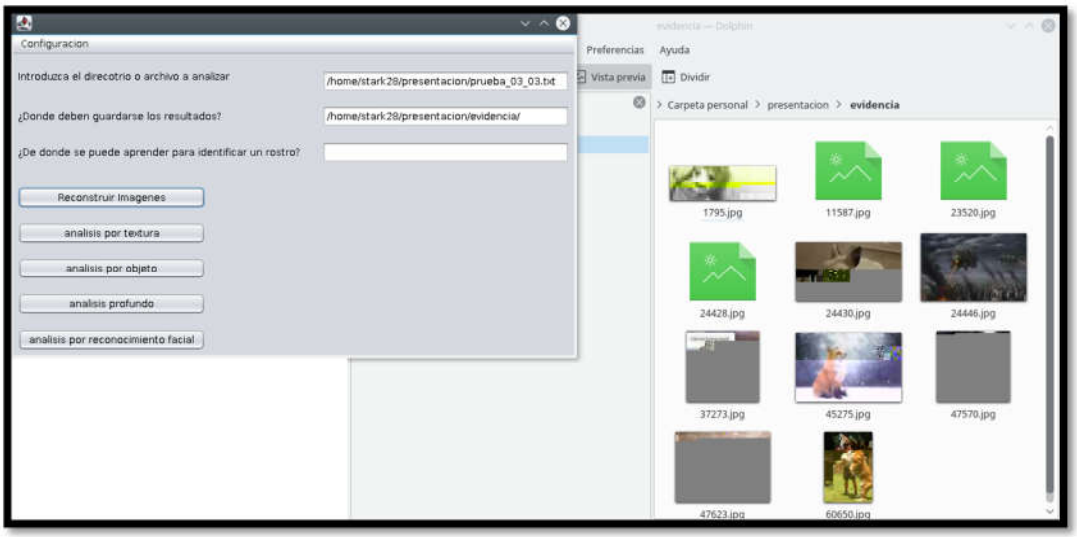


Figura 4 Primera captura con todos los recursos asignados al sniffing

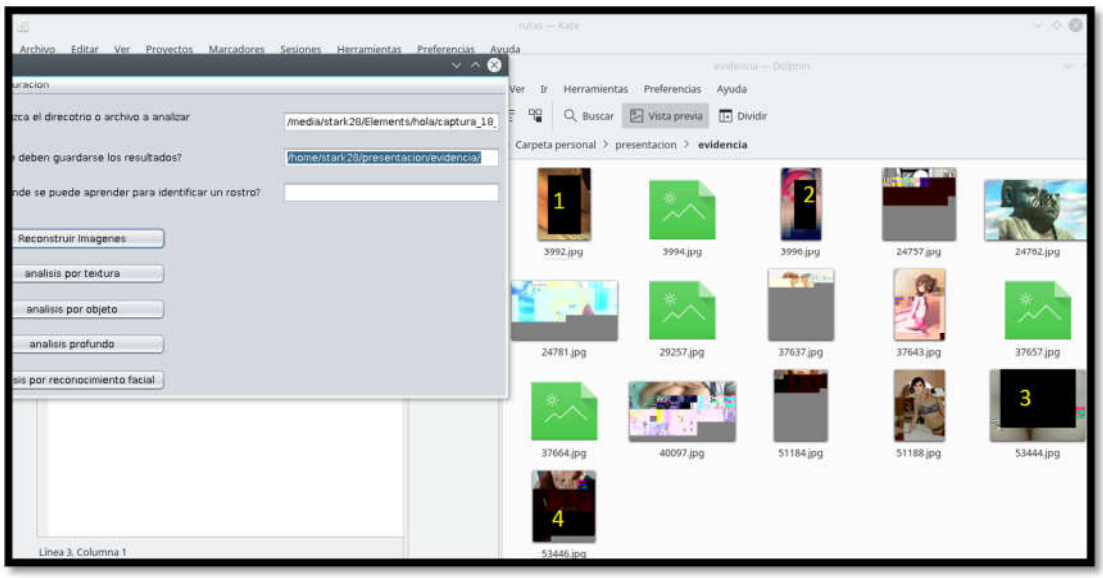


Figura 5 Segunda captura con todos los recursos asignados al sniffing Imágenes 1,2,3,y 4 censuradas

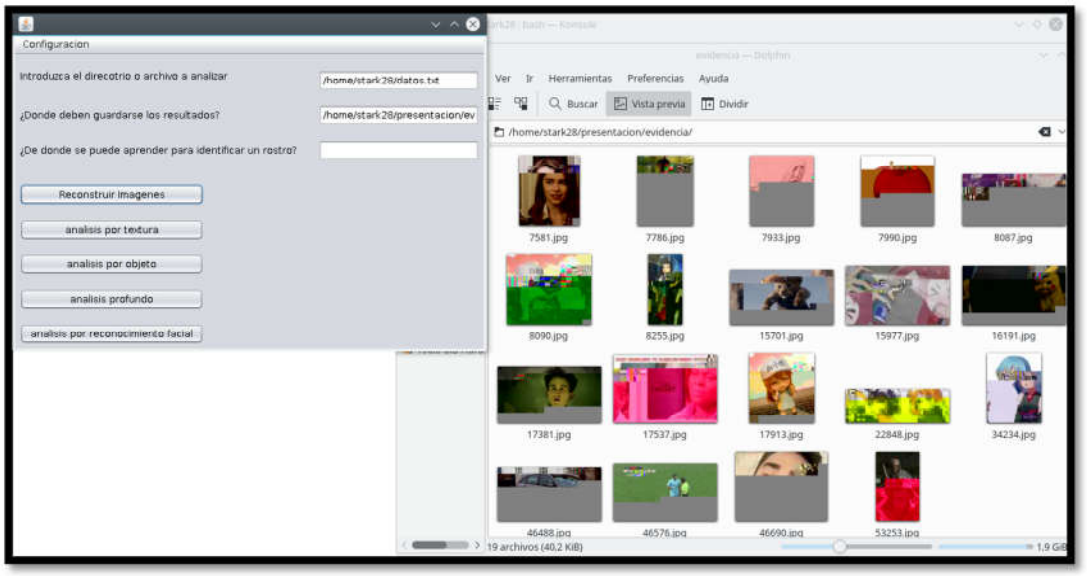


Figura 6 Primera captura con prioridad estándar asignada por el Sistema Operativo

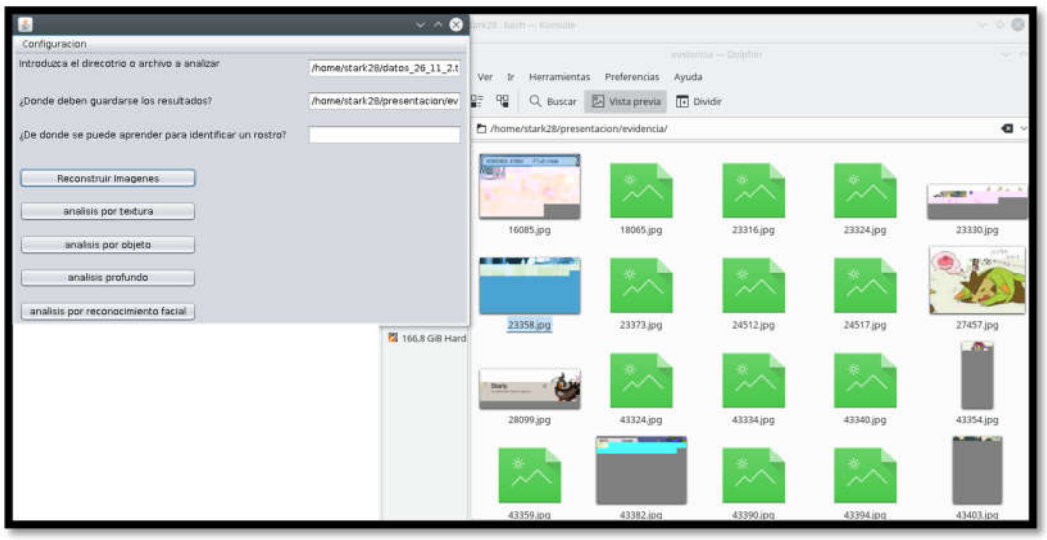


Figura 7 Segunda captura con prioridad estándar asignada por el Sistema Operativo

Capacidad de Clasificación

Con imágenes obtenidas de distintas fuentes públicas como ser: 4chan.org, google imágenes y fotos propias, se realizaron las pruebas obteniendo los siguientes resultados:

Clasificación por textura

Se recuerda que la clasificación por textura determina el porcentaje de piel en una foto. Y si ese porcentaje obtenido es mayor al previamente definido se considera a la imagen como sospechosa. Cabe destacar que para estas pruebas, se ha optado por imágenes de personas en trajes de baño, para evitar la exposición a contenido explícito y/o herir susceptibilidades.

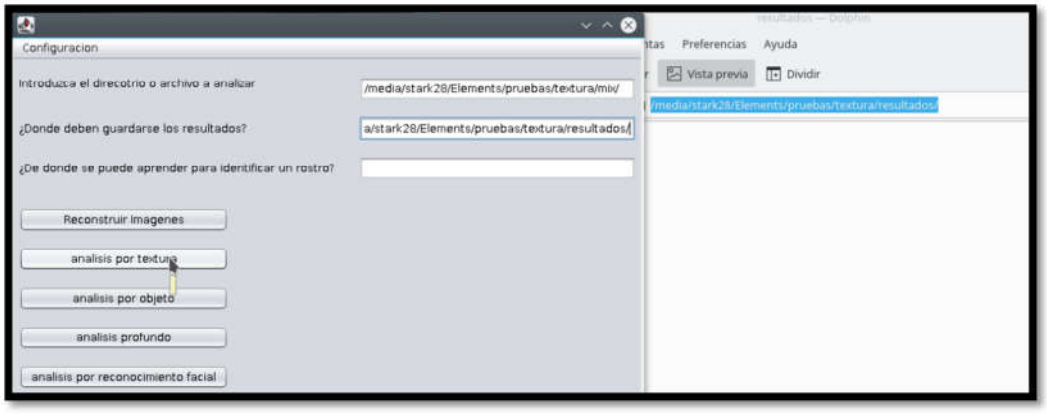


Figura 8 carpeta de destino antes de realizar el análisis

Figura 9. Conjunto de fotos de prueba para análisis de texturas

Figura 10. Resultados obtenidos

Precisión=42/47= 89%

	True no	True si
Pred. No	27	2
Pred. Si	3	15

Clasificación por objeto

Se recuerda que el análisis por objeto está centrado en la detección de cuerpos completos, esto se debe a que el analizador anterior puede confundirse con un primer plano de una parte del cuerpo por ejemplo una mano, una pierna, etc. Así mismo, el analizador anterior también puede confundir ciertos tonos de arena con la piel humana, es aquí que este segundo clasificador soluciona esos problemas.

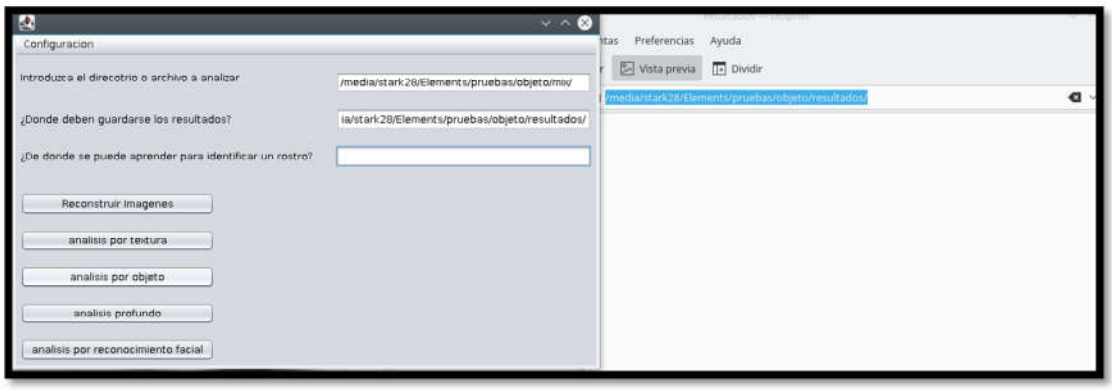


Figura 11. Carpeta de destino antes de realizar el análisis

Figura 9. Conjunto de fotos de prueba para análisis de objetos

Figura 12. Resultados obtenidos del análisis de objetos

Precisión=29/34= 85%

	True no	True si
Pred. No	14	1
Pred. Si	4	15

Clasificación por reconocimiento facial

Existen situaciones en las que se desea buscar una persona específica en un conjunto voluminoso de fotos, es por esta razón; que existe esta funcionalidad en la aplicación. A continuación, se presentan los resultados obtenidos.

Figura 13. Carpeta de destino antes de realizar el análisis

Figura 14. Conjunto de fotos para que la aplicación reconozca a la persona deseada

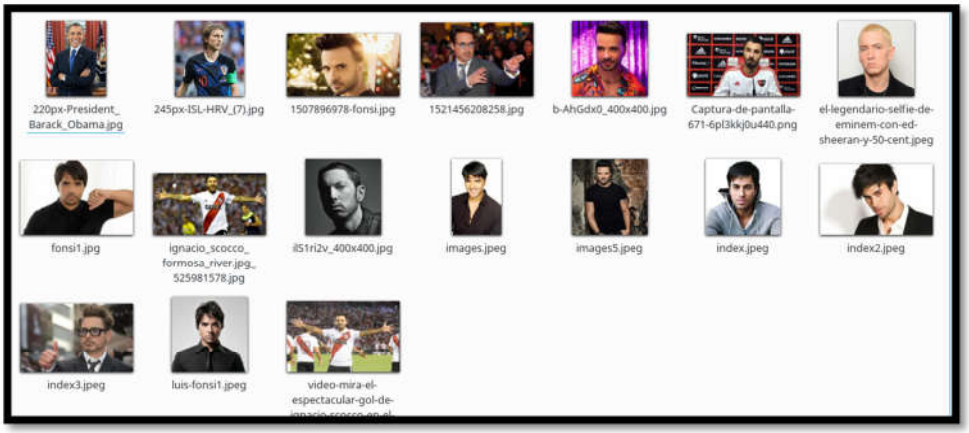


Figura 15. Conjunto de fotos en donde se busca a la persona

Figura 16. Resultados esperados

Figura 17. Resultados obtenidos

Precisión=12/17= 71%

	True no	True si
Pred. No	9	3
Pred. Si	2	3

Conclusiones acerca de los resultados obtenidos

Respecto a los resultados obtenidos en la reconstrucción de imágenes, como primera observación se puede establecer que:

1. Para un buen proceso de sniffing, es necesario contar con buenos recursos de hardware

Un equipo, ya sea un router, firewall, o cualquier computadora cuyo fin sea administra una red, debe ser capaz de realizar sus funciones sin comprometer el desempeño de esta última. Tareas comunes para estos equipos como el *nateo,* enrutamiento y conmutacion requieren de mucha capacidad de procesamiento y memoria disponible, el añadir una actividad adicional como el sniffing ralentiza el funcionamiento del equipo.

Un router de servicios integrados Cisco de la serie 1000, ideal para una empresa pequeña, cuenta con hasta 8 núcleos y 8 GB de memoria RAM[15], solo para poder realizar tareas propias de un router. Por lo que si se desea hacer sniffing y no afectar el desempeño de la red, los requisitos mínimos para un ordenador deben ser los mencionados anteriormente.

2. La calidad de la reconstrucción está en función del proceso de sniffing anterior

El sniffing es un proceso que es controlado por el sistema operativo, y este último puede asignarles menos recursos de procesamiento con el fin garantizar el correcto funcionamiento del ordenador. Por lo tanto si se resta prioridad al proceso, este capturara menos paquetes y por consiguiente menos imágenes podrán ser reconstruidas.

En cuanto a los resultados de los analizadores, se consideran a los mismos que son satisfactorios y contundentes para un primer prototipo funcional y listo para ser usado, tal y como lo indica la precisión que alcanzada en las pruebas anteriores.

Dificultades no previstas encontradas y sus soluciones.

A lo largo del proceso desarrollo de software, se presentaron distintos problemas que dificultaron la construcción de la herramienta, entre ellas resolución de condiciones lógicas y utilización de recursos, que obligaron a la migración del entorno .NET al entorno java. Cabe destacar que, en un principio, no se había considerado que la posibilidad de que la herramienta sea multiplataforma, la capacidad de ejecutarse en distintos entornos es una ventaja a considerar con respecto a las otras aplicaciones. La migración a otra plataforma resulto sencilla, ya que el código con el que se disponía en ese momento no hacía uso de funciones, procedimientos y librerías propias del lenguaje.

Otro factor que llevo a la migración, fue la disponibilidad de librerías y documentación para el procesamiento de imagen.

Evolución de la herramienta

En un principio, con el objetivo de comprobar si el primer algoritmo planteado funcionaba, se planteó una interfaz rudimentaria como se puede ver en la figura 16.

[15] https://www.cisco.com/c/en/us/products/routers/1000-series-integrated-services-routers-isr/index.html#~stickynav=1

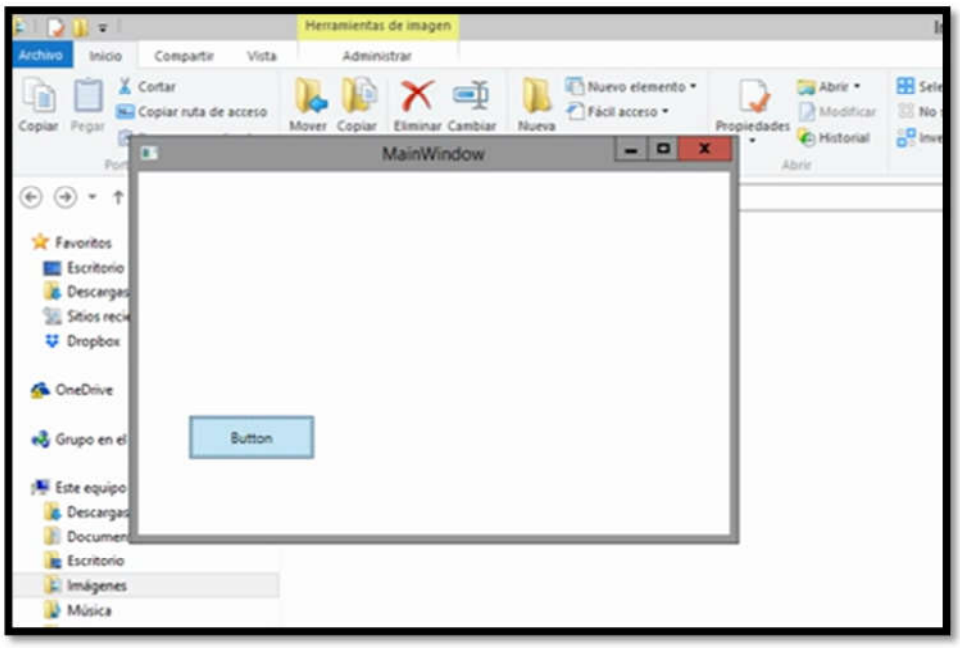


Figura 18. Primera interfaz

En la figura 17. se pueden apreciar los resultados obtenidos.

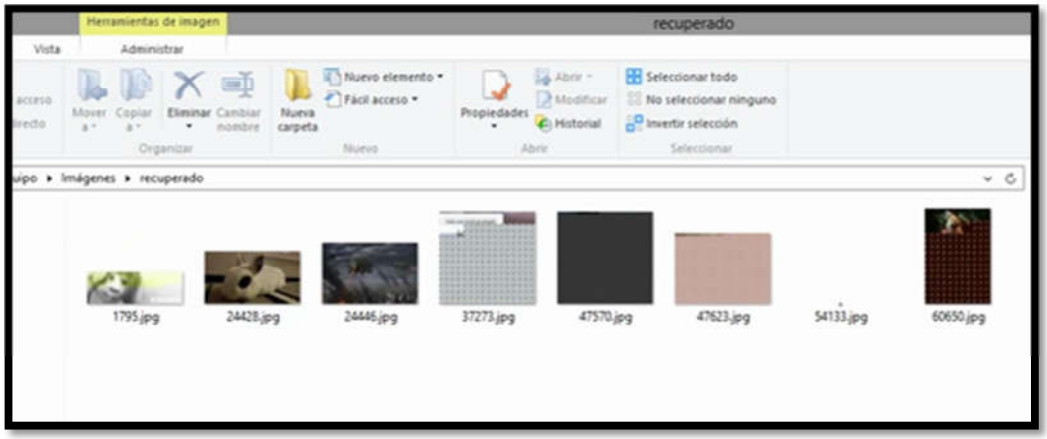


Figura 19. Resultados obtenidos de en .NET

Tanto la figura 1 como la figura 17, son resultados del mismo archivo. Las diferencias en los resultados se deben a la capacidad del compilador para reconstruir una imagen, como así también, la segunda diferencia radica que a lo largo del desarrollo algoritmo ha sido modificado para buscar mucho más rápido imágenes completas.

Una vez que, se consiguieron resultados deseados como los de la figura 1, la interfaz del aplicativo lucía de la siguiente manera:

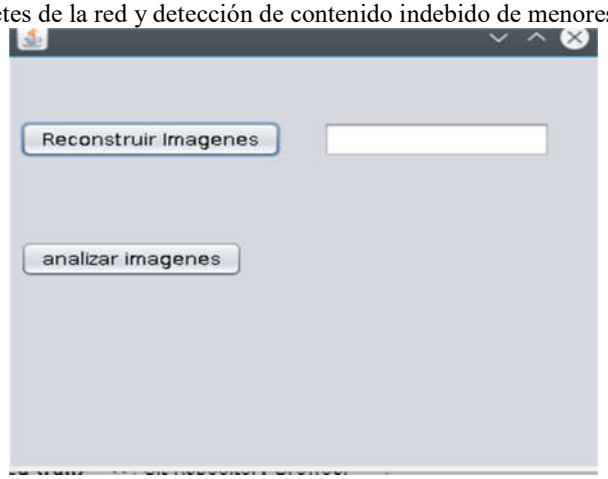


Figura 20. Interfaz migrada a JAVA y algoritmo mejorado

Se procedió a añadir más funcionalidades al software, los distintos analizadores se desarrollaron uno a continuación de otro, es decir, una vez que un clasificador/analizador lograba resultados deseados, se continuaba con el siguiente, así hasta lograr la apariencia que posee ahora.

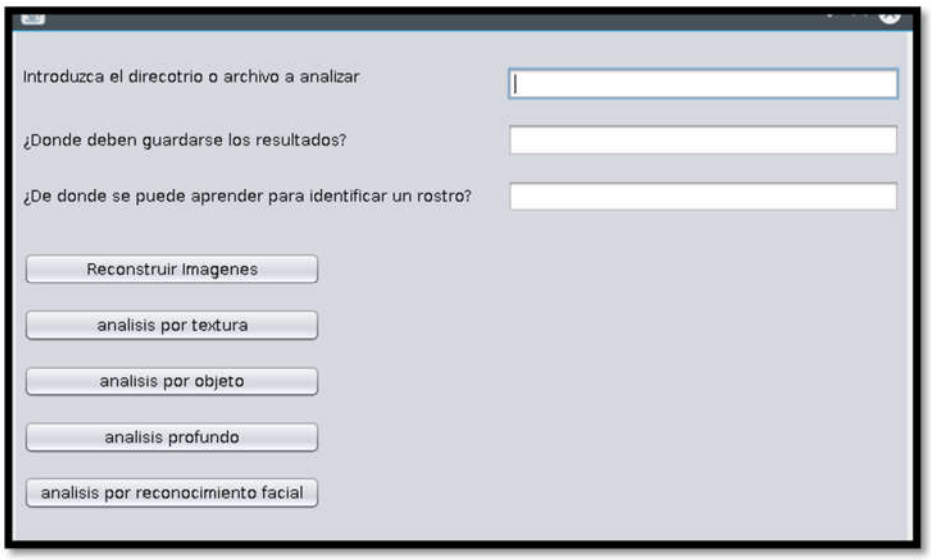


Figura 21. Interfaz migrada a JAVA y algoritmo mejorado.

Problemas con haar_casacade y pruebas con Python y JAVA

La librería de openCV tiene pre-instalados clasificadores del tipo Haar Cascade: es un método efectivo de detección de objetos propuesto por Paul Viola y Michael Jones en su artículo, "Detección rápida de objetos usando una cascada aumentada de características simples" en 2001. Es un enfoque basado en el machine learning donde el algoritmo es

entrenado a partir de muchas imágenes positivas y negativas. Luego se utiliza para detectar objetos en otras imágenes.[16]

Los clasificadores son ampliamente utilizados en detección de rostros, sin embargo, cuando se trata de cuerpos completos la situación cambia. Para proceso de reconocimiento de un cuerpo humano, se debe tener en cuenta la posición en la que este se encuentra (no debemos limitarnos a si este se encuentra horizontal o vertical), también debemos considerar la distancia a la que se encuentra la persona fotografiada, estos factores influenciarán en la detección de un cuerpo.

Muchos desarrolladores proponen en sus blogs; cambiar el espacio de colores de RGB a escala de grises, y en base a esta última realizar el proceso de detección. Sin embargo, esto último no proporcionaba resultados confiables. En primera instancia, se había considerado que esto podría ser a raíz de una mala instalación de las librerías, y se decidió hacer un script en Python para comprobar esto.

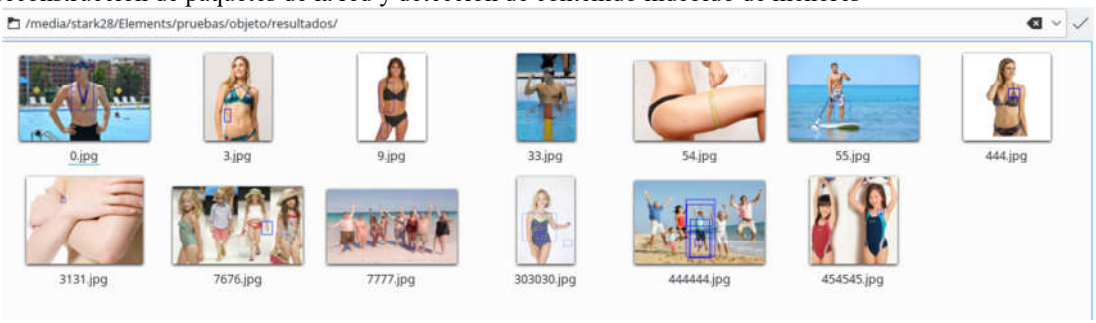
```
import os

face_cascade = cv.CascadeClassifier('haarcascade_fullbody.xml')
lower_cascade = cv.CascadeClassifier('haarcascade_lowerbody.xml')
upper_cascade = cv.CascadeClassifier('haarcascade_upperbody.xml')
carpeta='/media/stark28/Elements/pruebas/objeto/mix/'
i=0

for archivo in os.listdir(carpeta):
        if archivo!=".directory":
                name=os.path.join(carpeta,archivo)
                print(name)
                img = cv.imread(os.path.join(carpeta,archivo))
                gray = cv.cvtColor(img, cv.COLOR_BGR2GRAY)
                faces = face_cascade.detectMultiScale(gray, 1.1, 3)
                for (x,y,w,h) in faces:
                        cv.rectangle(img,(x,y),(x+w,y+h),(255,0,0),2)
                        roi_gray = gray[y:y+h, x:x+w]
                        roi_color = img[y:y+h, x:x+w]
                        print(os.path.join(carpeta,archivo))
                        cv.imwrite("/media/stark28/Elements/pruebas/objeto/resu
                        i=i+1
                lower = lower_cascade.detectMultiScale(gray, 1.1, 3)
                for (x,y,w,h) in lower:
                        cv.rectangle(img,(x,y),(x+w,y+h),(255,0,0),2)
                        roi_gray = gray[y:y+h, x:x+w]
                        roi_color = img[y:y+h, x:x+w]
                        print(os.path.join(carpeta,archivo))
                        cv.imwrite("/media/stark28/Elements/pruebas/objeto/resu
                        i=i+1
                upper = upper_cascade.detectMultiScale(gray, 1.1, 3)
```

```
stark28@JARVIS:~/python_projects/body_detection$ python detection.py
```

[16] docs.opencv.org/3.3.0/d7/d8b/tutorial_py_face_detection.html



Los resultados fueron contundentes, las salidas se debían a la variedad de poses y distancias a las que se encontraban las personas.

Se decidió trabajar con las siluetas que producían el binarizar las imágenes. El proceso de binarizado es el que se utiliza en el análisis de texturas, el mismo consiste en transformar pixeles blancos aquellos pixeles que tengan un color de piel y el resto de colores se los transforma en pixeles de color negros. De esta forma se consiguió mejores resultados.

Implementación de la solución propuesta

Base de Datos

El motor de base de datos optado para almacenar imágenes obtenidas de la reconstrucción de paquetes es MongoDB, este es un motor de base de datos orientado a documentos. Para realizar una correcta implementación, debemos garantizar que el servicio de base de datos disponga de al menos de 2 cores de CPU[17]. En lo que respecta a la utilización de memoria, esta puede ser configurable según el hardware que se disponga, sin embargo se considera que por cada 100 mil documentos en la base de datos se necesita 1 GB de memoria.

MongoDB puede ser utilizado en sistemas operativos Windows, MacOS y sistemas operativos con núcleo Linux. Se recomienda una arquitectura de 64 bits para que la implementación del sistema sea fácilmente escalable.

Hardware necesario

Se necesitan de dos equipos:

Una computradora con:

- Procesador Intel Core i3 de séptima generación
- 8GB de memoria RAM
- 1 TB para el sistema de archivos

Una computadora con:

- Procesador Intel Core i7 de séptima generación
- 8GB de memoria RAM
- 1 TB para el sistema de archivos

[17] https://docs.mongodb.com/manual/administration/production-notes/#hardware-considerations

El primer equipo cuenta con las características ampliadas de la computadora con la que se hicieron las pruebas. El ordenador de prueba que se utilizó durante el análisis de imágenes funciono de manera aceptable, por lo que si se escala de manera vertical el rendimiento será mejor.

El segundo equipo es para hacer uso del sniffer, estas características son necesarias para no perjudicar la performance del tráfico en la red.

Validez judicial y formalización del proceso de reconstrucción y análisis

En el convenio de Budapest, el cual la república Argentina se ha adherido, se define:

- Delitos contra la confidencialidad, la integridad y la disponibilidad de los datos y los sistemas informáticos
- Delitos relacionados a la falsificación y el fraude
- Delitos relacionados con el contenido
- Delitos relacionados con infracciones de la propiedad intelectual y derechos afines

Los delitos relacionados con el contenido, específicamente, la tenencia de toda representación de un menor de dieciocho años dedicado a actividades sexuales se encuentra tipificada en el artículo 128 del código penal.

Con respecto al sniffing de paquetes:

Para que el trafico capturado y posteriormente analizado se válido como evidencia digitial, la legislación del país establece que el responsable de la red debe notificar, mediante un acuerdo de confidencialidad, que los usuarios que están siendo monitoreados. Esta es la única manera de que los resultados de sniffing sean considerados como evidencia.

Para garantizar que el archivo resultante de la captura de paquetes no ha sido modificado, el mismo debe ser protegido contra escritura, y una vez finalizada la captura se debe calcular el hash del archivo resultante.

Con el archivo protegido contra escritura, se realizará una copia forense y sobre dicha copia se procederá a realizar la reconstrucción de imágenes a partir de los paquetes capturados.

Finalizada la reconstrucción, se calculará el hash del archivo para corroborar que el archivo no ha sido modificado.

Las imágenes resultantes serán destinadas a un directorio específico del sistema de archivos, se calculará un hash de la respectiva carpeta.

Se realizará una copia forense de las imágenes, y sobre esta copia se realizarán los respectivos análisis de textura, objeto y reconocimiento facial.

Al finalizar los respectivos análisis, se calculará el hash para determinar si se ha preservado la cadena de custodia.

Para la protección de la información sensible, se encriptará usando LUKS (Linux Unified Key Setup) y cryptsetup.

Con respecto a las imágenes obtenidas de medios de almacenamientos:

Las imágenes que se hayan conseguido de medios de almacenamiento, siendo estos últimos obtenidos según lo establecido el procedimiento pericial correspondiente, con el fin de proteger la cadena de custodia se realizará una copia forense y sobre ella se harán los análisis de textura, objeto y reconocimiento facial. Al finalizar los respectivos análisis, se calculará el hash para determinar si se ha preservado la cadena de custodia. Para la protección de la información sensible, se encriptará usando LUKS

Política de seguridad

Objetivo

Definir las medidas de seguridad informática para el uso adecuado de los recursos tecnológicos, proteger la integridad y confidencialidad la información sensible, garantizando la disponibilidad apropiada de los recursos del sistema.

Alcance

Las políticas se orientan a proteger y resguardar la información que es almacenada, procesada y transmitida; esto también aplica para todos los equipos y recursos tecnológicos. Las medidas deben ser conocidas y acatadas por todos los miembros de la organización, como así también los terceros que interactúan de manera habitual u ocasional, que accedan a información y/o a los recursos informáticos.

Sobre la información

Toda la información con la que trabajen la aplicación debe encontrarse debidamente encriptado.

La información siempre debe estar protegida contra escritura.

Se debe calcular un hash antes y después de trabajar con la información, con el fin de garantizar que esta no ha sido adulterada.

Se conservará la información sensible según el tiempo establecido por la ley, una vez finalizado el mismo se procederá a eliminar todo registro de la misma.

Sobre los usuarios

Cada usuario debe contar con credenciales que identifiquen al usuario

Las credenciales de identidad de usuario son intransferibles.

Las credenciales deben ser renovadas periódicamente según corresponda.

El usuario es responsable directo por las alteraciones causadas en la información.

Sobre los equipos

El uso de uso de dispositivos de almacenamiento extraíble, debe ser previamente autorizado.

Los equipos deben encontrarse debidamente identificados y nomeclados.

Los dispositivos de almacenamiento de respaldo de información deben encontrarse debidamente resguardados.

Los equipos de procesamiento de información deben contar con un dispositivo de protección eléctrica.

Los medios de almacenamiento que deban ser reutilizados deben ser debidamente wipeados, es decir, se debe eliminar correctamente todo el contenido para su posterior uso.

Sobre la aplicación

El correcto uso de la aplicación es responsabilidad del usuario

El usuario debe respetar la licencia de la aplicación.

Conclusiones y futuras líneas de investigación

Las funcionalidades de la aplicación son comunes en redes sociales y dispositivos inteligentes realizan reconocimiento facial para identificar a sus usuarios, o reconocimiento de objeto para usarla el control parental.

Aprovechando lo investigado hasta el momento por otros especialistas, se lo ha trasladado a otro ámbito y se ha construido un software con un objetivo forense, con funcionalidades que agiliza el proceso pericial, reduciendo la cantidad de herramientas que se pueden utilizar para este proceso.

El presente trabajo, se limita al monitoreo y análisis de paquetes TCP no encriptados, por lo que se propone como línea futura de investigación, el análisis del trafico encriptado como así también los paquetes UDP, siendo estos comúnmente utilizados en herramientas P2P como Ares.

Las cabeceras de los paquetes TCP conforman un gran volumen de información, el cual permitiría la aplicación de técnicas de minería de datos, y así predecir desde que IP se consume y/o distribuye el material y detenerlo a tiempo.

Otro ámbito en el que puede continuarse la investigación, es en la generación de nuevos resultados de machine learning que ayuden a mejorar los actuales publicados por OpenCV.

Mediante la utilización de Cloud Computing, se puede implementar un servicio en la nube que se encargue de realizar el análisis de imágenes. De esta forma, ya no sería necesario contar con hardware para correr el software en máquinas locales.

Cabe destacar que, con la llegada de la programación paralela, se puede aumentar la performance del software, y por consiguiente, evaluar imágenes de una manera mucho más rápida.

Bibliografía citada

1 Artículo 128 del código penal procesal de la nación

2 Amor M. 2016. Pornografía Infantil: Marco legal y herramientas peer to peer. Publicación on line. ISSN 2347-0372

3 Jacobson I, Booch G, Rumbaugh J. 2000. El proceso unificado de desarrollo de software. 464 paginas. Editorial Pearson Educación. ISBN 84-7829-036-2

4 Sommerville I. 2011 Ingeniería de Software. 792 paginas. Editorial Pearson Educación. ISBN: 978-607-32-0603-7

5 griffeye.com/the-platform/

6 guidancesoftware.com/encase-forensic?cmpid=nav_r

7 wireshark.org

8 Di Iorio, Ana Haydée et al. 2017. El rastro Digital del Delito. 554 paginas. Universidad FASTA Ediciones. ISBN 978-987-1312-81-8

9 Guía de ciberdelitos. http://www.justiciasalta.gov.ar/images/uploads/ciberdelitos_web_v020518.pdf

10 https://latam.kaspersky.com/blog/que-es-un-hash-y-como-funciona/2806/

11 qnap.com/es-es/how-to/tutorial/article/las-mejores-pr%C3%A1cticas-para-el-rendimiento-de-almacenamiento-de-qnap/

12 Pressman R. 2010. Ingeniería de software enfoque práctico. 736 páginas. Editorial McGraw-Hill ISBN: 978-607-15-0314-5

13 Pressman R. 2010. Ingeniería de software enfoque práctico. 736 páginas. Editorial McGraw-Hill ISBN: 978-607-15-0314-5

14 Cannice M, Koontz H, Weihrich H. 2012. Administración Una perspectiva global y empresarial. 736 páginas. Editorial McGrawHill. ISBN 978-607-15-0759-4

15 https://www.cisco.com/c/en/us/products/routers/1000-series-integrated-services-routers-isr/index.html#~stickynav=1

16 docs.opencv.org/3.3.0/d7/d8b/tutorial_py_face_detection.html

17 https://docs.mongodb.com/manual/administration/production-notes/#hardware-considerations

18 https://www.kaspersky.es/blog/que-es-un-botnet/755/

19 Tanenbaum A, Wetherall D. 2012 Redes de Computadora. 816 paginas. Editorial Pearson Educación. ISBN: 978-607-32-0817-8

20 Tanenbaum A, Wetherall D. 2012 Redes de Computadora. 816 paginas. Editorial Pearson Educación. ISBN: 978-607-32-0817-8

21 https://www.icann.org/news/blog/que-es-un-ataque-de-intermediarios

22 Tanenbaum A, Wetherall D. 2012 Redes de Computadora. 816 paginas. Editorial Pearson Educación. ISBN: 978-607-32-0817-8

23 https://www.avast.com/es-es/c-phishing

24 RFC 5694 https://tools.ietf.org/rfc/rfc5694.txt

Anexo 1

Tabla 1

Tarea	Líder de proyecto	Analista	Programador	Tester	Responsable de tecnologías	Especialista Juridico
Análisis de la situación legal y jurídica actual	5					2
Investigación de soluciones forenses actuales en el mercado		5				
Recopilación de papers sobre el procesamiento de imagen		4				
Formulación de un modelo matemático		8	1			
Identificación y Evaluación de riesgos	5					
Obtención de métricas de proyectos similares	3					
Realización de estimaciones	3					
Identificación de interesados	1					
Planificación de la calidad del proyecto				2		
Determinar la estructura de hardware necesaria					6	
Determinar el entorno de desarrollo			1		5	
Revisión de los riesgos del proyecto	3					
Revisión de métricas y estimaciones	3					
Contacto con los interesados	3					

Elaboración de listado de CU para reconstrucción de imágenes		6				
Elaboración de listado de CU para procesamiento de imágenes		6				
Redacción de especificación de CU para reconstrucción de imagen		6	1			
Redacción de especificación de CU para procesamiento de imagen		6				
Verificación de cumplimiento de los estándares de UML 2.5				2		
Reunión con los interesados	1					
Revisión de los riesgos del proyecto	3					
Revisión de métricas y estimaciones	2					
Formulación de nuevas estimaciones	2					
Elaboración de diagramas de colaboración para reconstrucción de imagen		5	1			
Elaboración de diagramas de colaboración para procesamiento de imagen		5				
Elaboración de diagramas de clases de diseño		5	1			
Elaboración de diagrama de interfaz		5				

Diseño lógico de la base de datos orientada a documentos		3				
Verificación de cumplimiento de los estándares de UML 2.5				2		
Revisión de los riesgos de proyecto	1					
Revisión de métricas y estimaciones	1					
Formulación de nuevas estimaciones	1					
Implementación de interfaz		2	5			
Implementación de Clases			15			
Implementación de la base de datos orientada a documentos			3			
Establecer conexión entre el software y la base de datos			2			
Verificación de cumplimiento de buenas prácticas CWE				7		
Revisión de los riesgos de proyecto	1					
Revisión de métricas y estimaciones	1					
Realización de pruebas unitarias				3		
Formulación de nuevas estimaciones	1					
Reunión con los interesados	1					
Pruebas de caja blanca				6		
Pruebas de caja negra				6		
Verificación de los estándares de		6				

calidad						
Instalación					9	
Reunión con interesados	4					
Verificación de cumplimiento de los estándares de calidad				1		
Comparación de las métricas obtenidas con las estimaciones realizadas	4					
Total Dias	49	72	30	29	20	2
Total Meses	2.45	3.6	1.5	1.45	1	0.1
Honorarios por mes	788.0220646	446.545837	617.2839506	525.348043	525.3480431	472,813239
Total Honorarios	1930.654058	1607.56501	925.9259259	761.754662	525.3480431	47,2813239

*Cabe destacar que se ha pactado que el trabajo es por objetivos, sin embargo; para cada actividad se le ha asignado un tiempo estimado de realización.

Anexo 2

Glosario de términos

ACID:es un acrónimo en ingles que que proviene de las palabras Atomicity (Atomicidad), Consistency (Consistencia), Isolation (Aislamiento) y Durability (Durabilidad). Dichas características son necesarias para una sistema un motor de base de datos que realiza transacciones.

Ataques de botnet: *Botnet es el nombre genérico que denomina a cualquier grupo de PC infectados y controlados por un atacante de forma remota.*[18]

Binarizacion: Recurso de procesamiento de imagen que consiste en modificar una imagen a blanco y negro. Se determinan los conjuntos de colores que se transformaran a negro, de igual modo se procede con el blanco. De esta forma se consigue separa los objetos que posteriormente se desean analizar.

Espacio de Colores: es una atracción matemática que permite representar los colores en función de vectores n dimensional.

Firewalls: *actúa como un filtro de paquetes. Inspecciona todos y cada uno de los paquetes entrantes y salientes. Los paquetes que cumplen cierto criterio descrito en reglas formuladas por el administrador de la red se reenvían en forma normal. Los que fallan la prueba simplemente se descartan.*[19]

Hackers de sombrero negro o blanco: *la prensa popular llama "hackers" a las personas que irrumpen en las computadoras*[20]. El concepto de sombrero jerga popular, hace referencia a la conducta moral del hacker, identificando con el sombrero color negro a aquellos ciberdelicuentes con grandes conocimientos. Por el contrario, aquellos identificados con el sombrero blanco popularmente son asociados a los especialistas en seguridad y hacking ético.

HSV: es un espacio de color que se encuentra en función del Hue (matiz), Saturation (saturacion) y Value (valor).

HTML: HyperText Markup Language por sus siglas en ingles (lenguaje de marcas de hipertexto en español), su función consiste en definir la estructura de una pagina web.

LAN: Local Area Network por sus siglas en inglés (red de área local en español). Consiste en una red de computadoras de abarca áreas pequeñas.

[18] https://www.kaspersky.es/blog/que-es-un-botnet/755/
[19] Tanenbaum A, Wetherall D. 2012 Redes de Computadora. 816 paginas. Editorial Pearson Educación. ISBN: 978-607-32-0817-8
[20] Tanenbaum A, Wetherall D. 2012 Redes de Computadora. 816 paginas. Editorial Pearson Educación. ISBN: 978-607-32-0817-8

Man in the Midle: *involucra a un atacante (o un dispositivo) que puede interceptar o alterar las comunicaciones entre dos partes que normalmente no son conscientes de que el atacante está presente en sus comunicaciones*[21]

Modelo OSI: *Desarrollada por la Organización Internacional de Normas (iso) como el primer paso hacia la estandarización internacional de los protocolos utilizados en las diversas capas. Se le llama Modelo de referencia OSI (Interconexión de Sistemas Abiertos, del inglés Open Systems Interconnection) de la iso puesto que se ocupa de la conexión de sistemas abiertos; esto es, sistemas que están abiertos a la comunicación con otros sistemas.*[22]

Modelo TCP/IP: Es un conjunto de protocolos para la comunicación den redes desarrollado por Vinton Cerf y Robert E. Kahn.

Nateo: derivado de NAT, por sus siglas en ingles Network Address Translation, que español significa traducción de direcciones de red. Es un mecanismo que usan los routers para intercambiar paquetes entre dos redes distintas.

Paquetes: Unidad de transporte de información en una red de computadoras.

Phishing: *es un método que los ciberdelincuentes utilizan para engañar y conseguir que se revele información personal, como contraseñas o datos de tarjetas de crédito y de la seguridad social y números de cuentas bancarias*[23]

Peer-to-Peer: Se considera que un sistema es Peer-to-Peer si los elementos que forman el sistema comparten sus recursos para proporcionar el servicio que el sistema ha sido diseñado para proporcionar. Los elementos del sistema proporcionan servicios a otros elementos y solicitan servicios de otros elementos.[24]

RGB: es un espacio de color que se encuentra en función del Red (rojo), Green (verde) y Blue (azul).

Sniffers: Es una herramienta informática que captura paquetes los paquetes de una red, la misma permite analizar el tráfico en esta última. También es una herramienta ampliamente utilizada en la administración de sistemas, para conocer los protocolos que se están usando en la red.

YCbCr: Es un espacio de color que se encuentra en función de Y que representa la luminosidad de la imagen, Cb que consisten en la crominancia en azul y Cr la crominancia en rojo

[21] https://www.icann.org/news/blog/que-es-un-ataque-de-intermediarios
[22] Tanenbaum A, Wetherall D. 2012 Redes de Computadora. 816 paginas. Editorial Pearson Educación. ISBN: 978-607-32-0817-8
[23] https://www.avast.com/es-es/c-phishing
[24] RFC 5694 https://tools.ietf.org/rfc/rfc5694.txt

Anexo 3

Documentación de análisis

Listado de casos de uso

1- Reconstruir imágenes a partir de paquetes
2- Determinar la presencia de desnudos en una imagen
3- Determinar la presencia de cuerpos completos en una imagen
4- Encontrar a una persona en una foto

Especificaciones de casos de uso

Caso de uso: Reconstruir imágenes a partir de paquetes

- Actor: Perito
- Pre-condiciones:
  - La reconstrucción se realiza en frio
  - Los paquetes ya están almacenados
- Post-Condiciones:
  - Se reconstruyen las imágenes

Escenario principal de éxito

Actor	Sistema
1. Iniciar proceso de reconstrucción de imágenes	
	2. Solicitar paquetes para reconstruir
3. Ingresar paquetes	
	4. Buscar paquetes que pertenezcan a una imagen
	5. Extraer metadatos del paquete
	6. Reconstruir imagen
	7. Almacenar imagen
	8. Informar final del proceso
	9. Fin del caso de uso

Caso de uso: Determinar la presencia de desnudos en una imagen

- Actor: Perito
- Pre-condiciones:
  - La imagen es un archivo valido
- Post-Condiciones:
  - Determinar si es un desnudo

Escenario principal de éxito

Actor	Sistema
1. Iniciar proceso de detección de desnudo	
	2. Solicitar la imagen a analizar
3. Ingresar imagen	
	4. Calcular el porcentaje de piel
	5. Comparar con el porcentaje promedio
	6. Informar que la imagen está por encima del porcentaje promedio con alto contenido de piel
	7. Informar final del proceso
	8. Fin del caso de uso

Escenario alternativo:

5. Comparar con el porcentaje promedio (el porcentaje está por debajo del promedio)

Actor	Sistema
	5.1 Comparar con el porcentaje promedio
	5.2 Informar que la imagen está por debajo del porcentaje promedio con alto contenido de piel
	5.3 Informar final del proceso
	5.4 Fin del caso de uso

Caso de uso: Determinar la presencia de cuerpos completos en una imagen

- Actor: Perito
- Pre-condiciones:
  - La imagen es un archivo valido
- Post-Condiciones:
  - Determinar si hay un cuerpo completo

Escenario principal de éxito

Actor	Sistema
1. Iniciar proceso de búsqueda de cuerpos	
	2. Solicitar la imagen a analizar
3. Ingresar imagen	
	4. Buscar cuerpo en la imagen
	5. Informar que se ha detectado un cuerpo completo
	6. Informar final del proceso
	7. Fin del caso de uso

Escenario alternativo:

4. Buscar cuerpo en la imagen (no se ha encontrado un cuerpo)

	4.1 Buscar cuerpo en la imagen
	4.2 Informar que se ha detectado un cuerpo completo
	4.3 Informar final del proceso
	4.4 Fin del caso de uso

Caso de uso: Encontrar a una persona en una foto

- Actor: Perito
- Pre-condiciones:
  - La imagen es un archivo valido
  - Proporcionar una buena imagen con la persona que se desea buscar
- Post-Condiciones:
  - Determinar si hay un cuerpo completo

Escenario principal de éxito

Actor	Sistema
1. Iniciar proceso de búsqueda de una persona en una foto	
	2. Solicitar imagen de persona a buscar
3. Ingresar imagen	
	4. Solicitar imagen en donde buscar a la persona
5. Ingresar imagen	
	6. Buscar persona
	7. Informar que se ha encontrado a la persona
	8. Informar final del proceso
	9. Fin del caso de uso

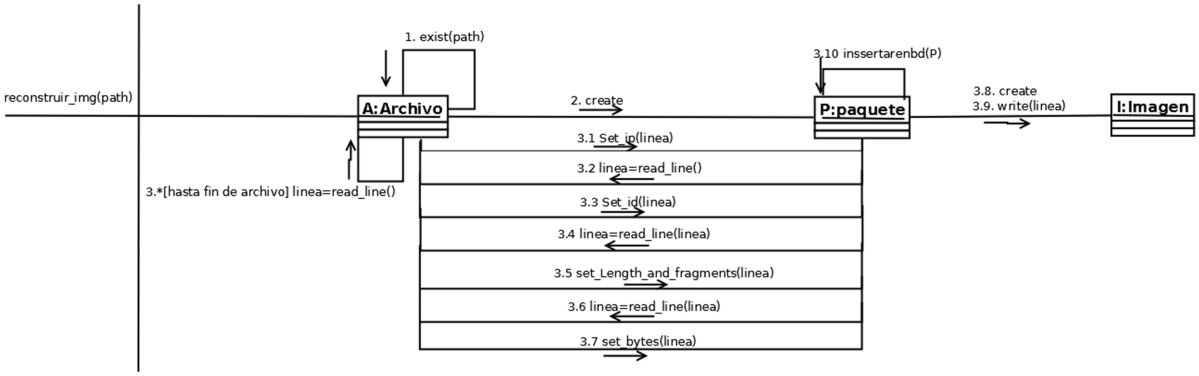
Escenario alternativo:

6. Buscar persona (no se ha encontrado un cuerpo)

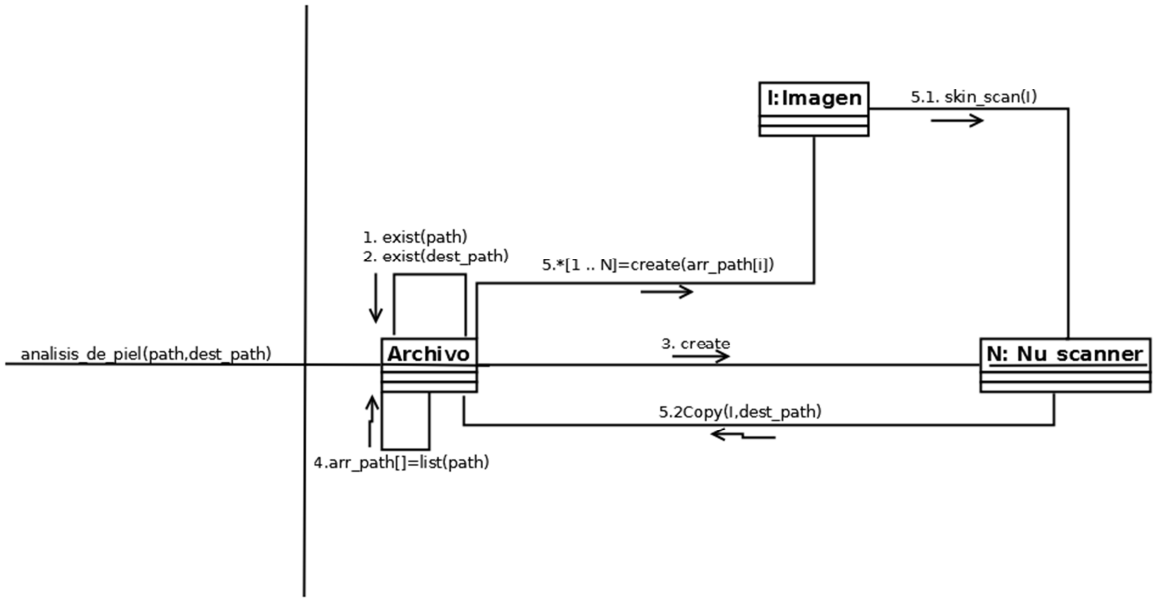
	6.1 Buscar persona
	6.2 Informar que se ha encontrado a la persona
	6.3 Informar final del proceso
	6.4 Fin del caso de uso

Documentación de diseño

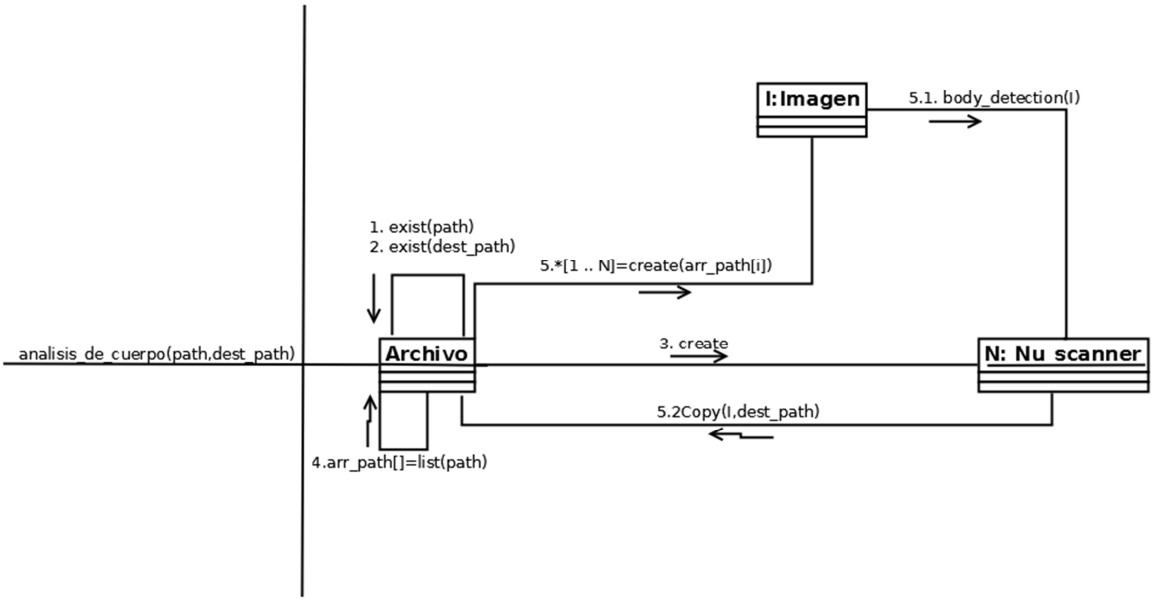
Reconstruir imágenes a partir de paquetes



Determinar la presencia de desnudos en una imagen



Determinar la presencia de cuerpos completos en una imagen



Encontrar a una persona en una foto

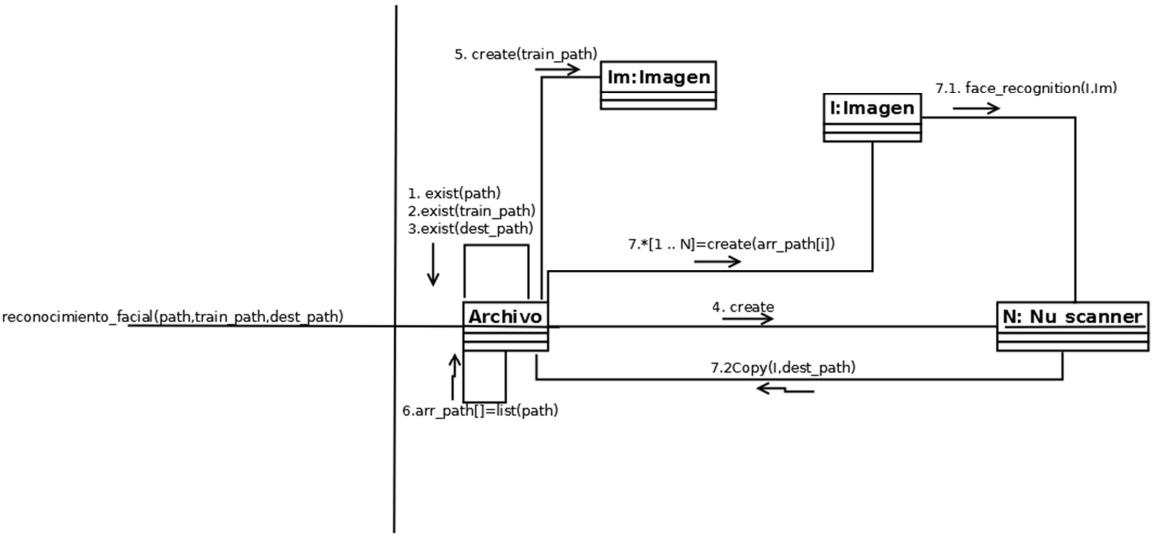
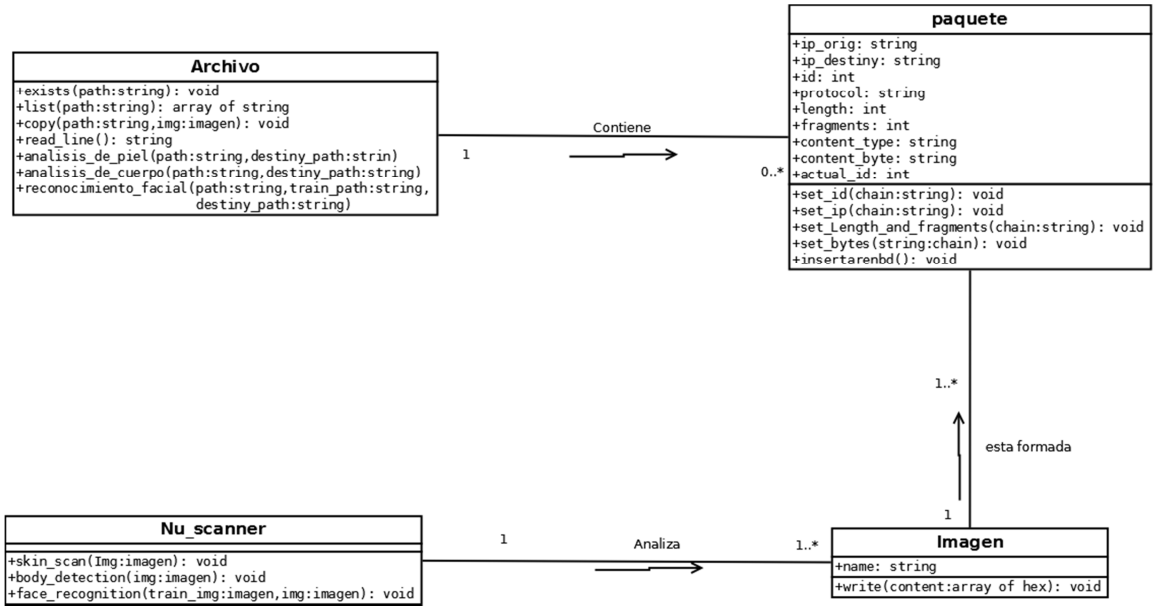
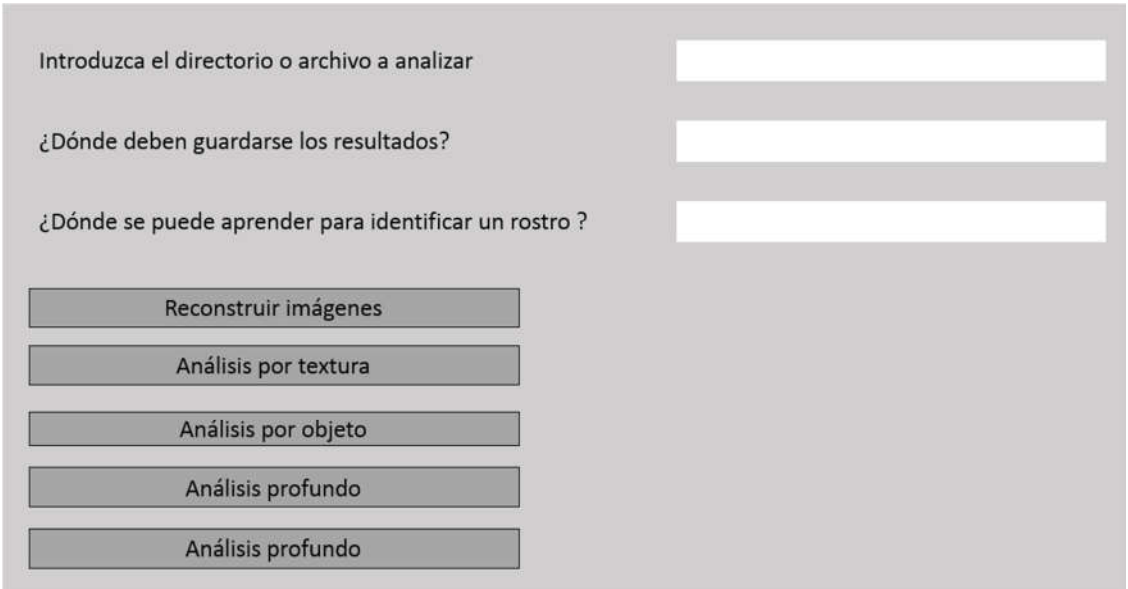


Diagrama de clases de diseño



Diseño de interfaz



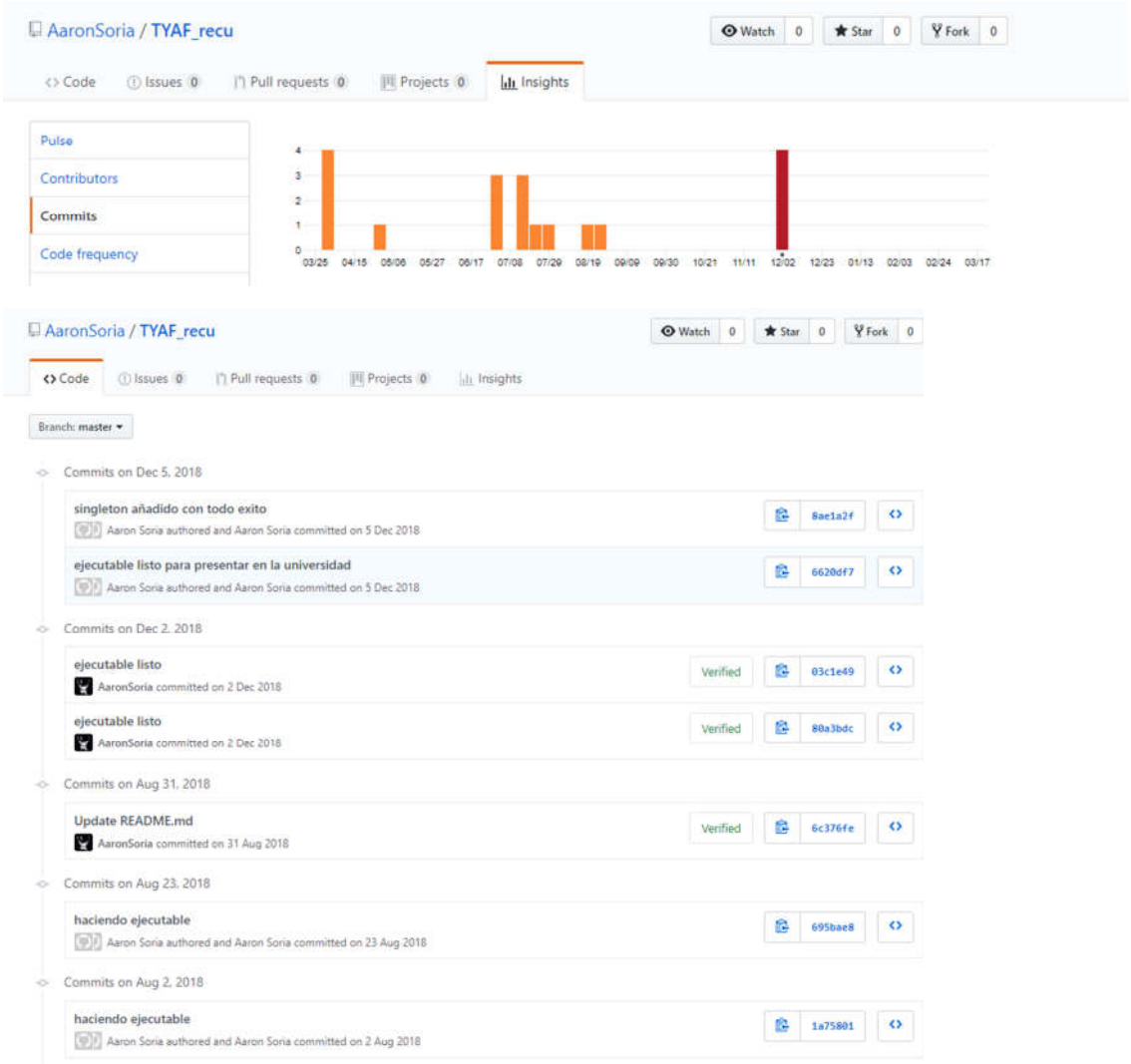
Diseño del esquema lógico de la base de datos

La base de datos solo contiene una colección y el esquema lógico es el siguiente

```
{
“id”
“ip_origin”
“ip_destiny”
“protocol”
“lenght”
“fragments”
“content_type”
“content_byte”
}
```

Código Fuente

Mediante el uso de GitHub, se puede apreciar el progreso del desarrollo de software.



Para consultar el historial completo de commits se puede acceder a los siguientes enlaces:

https://github.com/AaronSoria/TYAF_recu/graphs/commit-activity

https://github.com/AaronSoria/TYAF_recu/commits/master

Si desea consultar el código fuente de la aplicación, lo puede hacer dirigiéndose al siguiente enlace

https://github.com/AaronSoria/TYAF_recu

Anexo 4

Instalación y uso

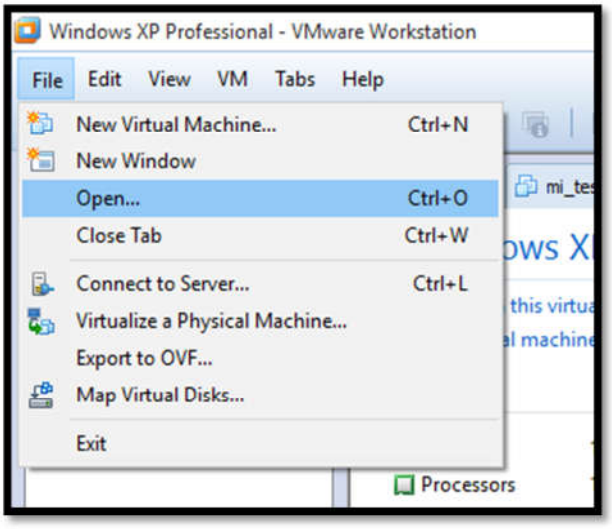
En el DVD con el que se presenta este trabajo, se encuentra un archivo de extensión .ovf, este es un archivo de máquina virtual. El mismo puede ser usado instalando:

VMware worksation

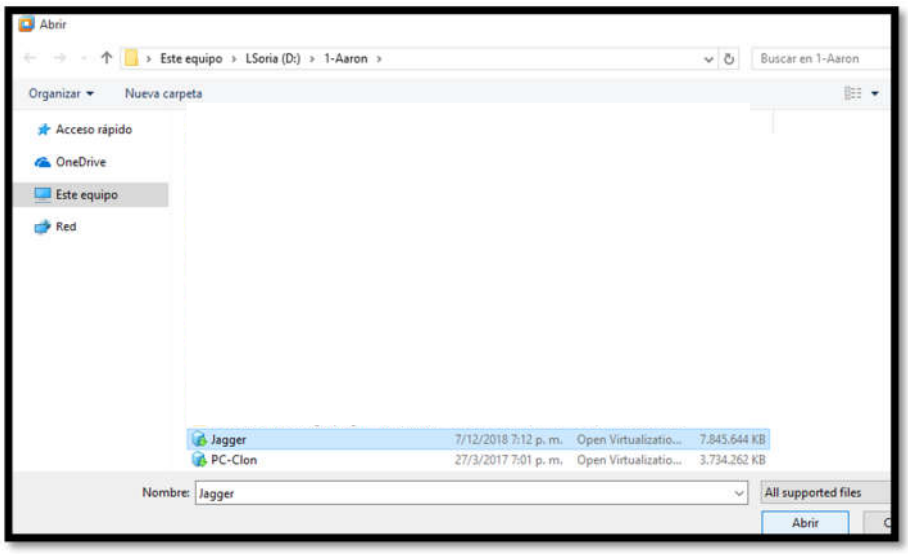
https://www.vmware.com/products/workstation-player/workstation-player-evaluation.html

El proceso para importar una máquina virtual es el siguiente:

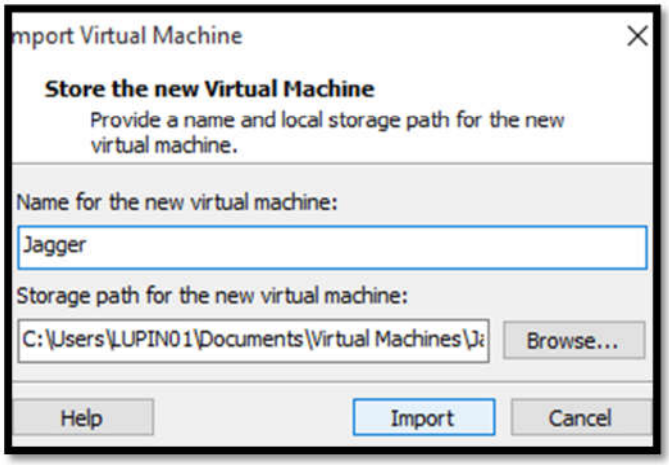
1. Luego de instalar VMware worksation, se ejecuta y se hace clic en File > Open…



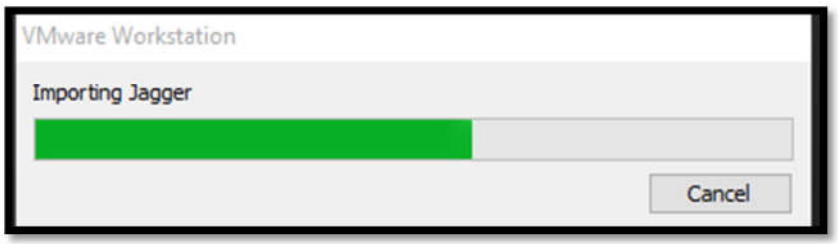
2. Se abrirá una nueva ventana, se debe buscar el archivo con extensión .ovf



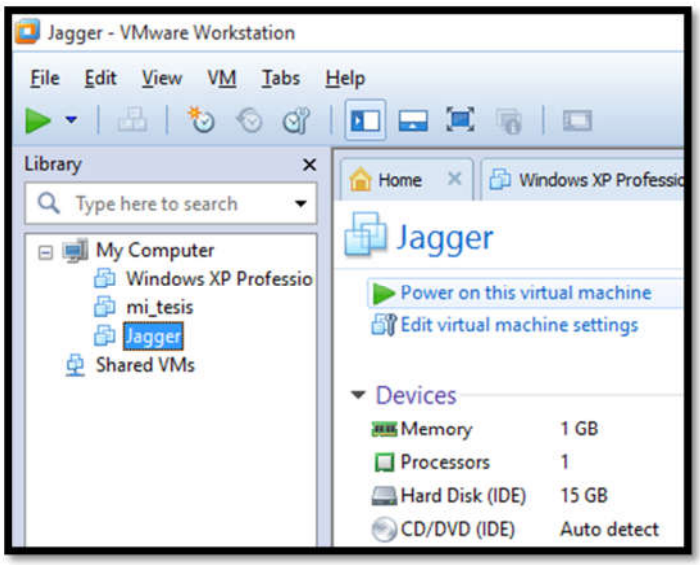
3. Una vez que se ha ubicado el archivo, hacer click en abrir. Se desplegará una nueva ventana que indica el nombre que tendrá la nueva máquina virtual, como así también la ruta en donde se alojará dicha máquina virtual. Dejar los valores por defecto y hacer click en Import



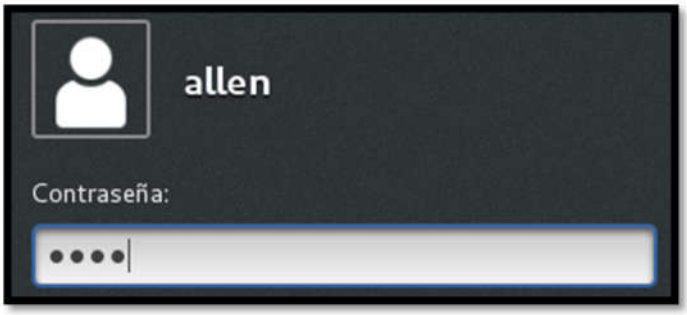
4. Esperar hasta que finalice el proceso de importación



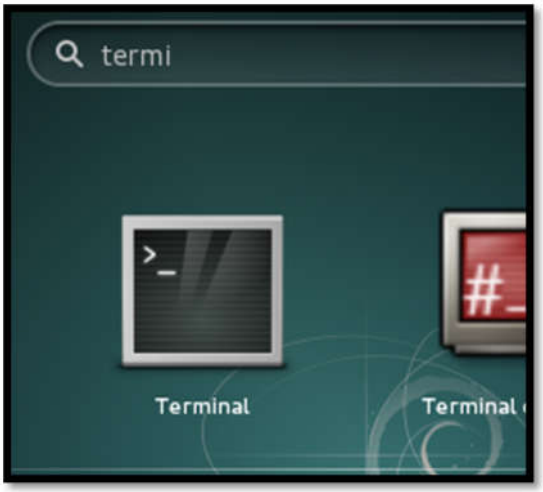
5. Seleccionar la máquina virtual y hacer click en: Power on this virtual machine



6. Iniciar sesión con el usuario allen, pass: 1234



7. Presionar la tecla inicio del teclado, y teclear “terminal”, luego hacer click en el icono de Terminal.



8. Se desplegará una consola. Se debe iniciar el servicio de la base de datos, ya que este solo se activa de forma manual. Teclear `su`, luego presionar enter, la contraseña que se solicita es: asdf. Después se debe iniciar el servicio de la base de

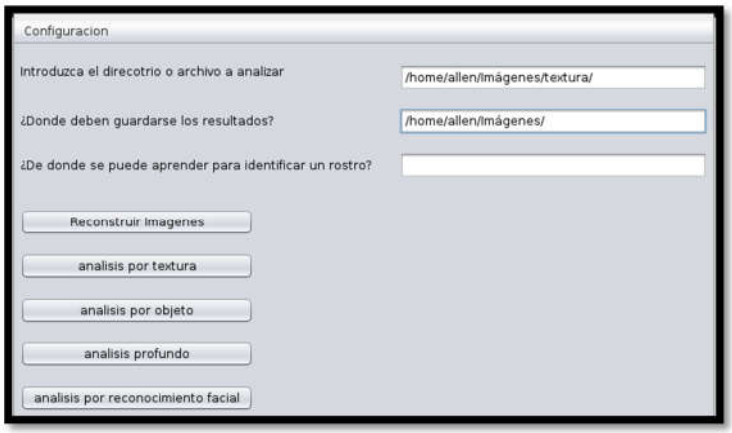
datos de la siguiente forma: `service mongod start.` Por último, teclear exit y dar enter.

9. Para iniciar la aplicación, solo se debe teclear `./tyaf_recu`

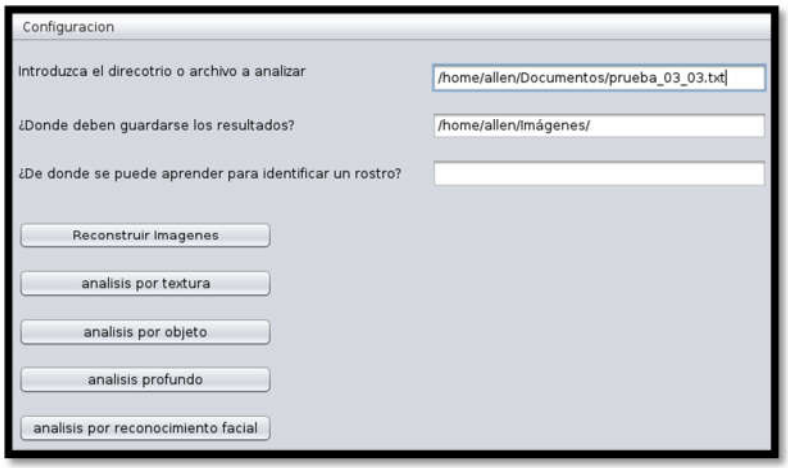
10. Ya se cuenta con la aplicación iniciada.

Si se desea analizar una imagen por textura, por objeto o un análisis profundo; solo se debe copiar la ruta completa del directorio que se desea analizar en el campo "introduzca el directorio o archivo a analizar", y luego se le debe definir en el campo

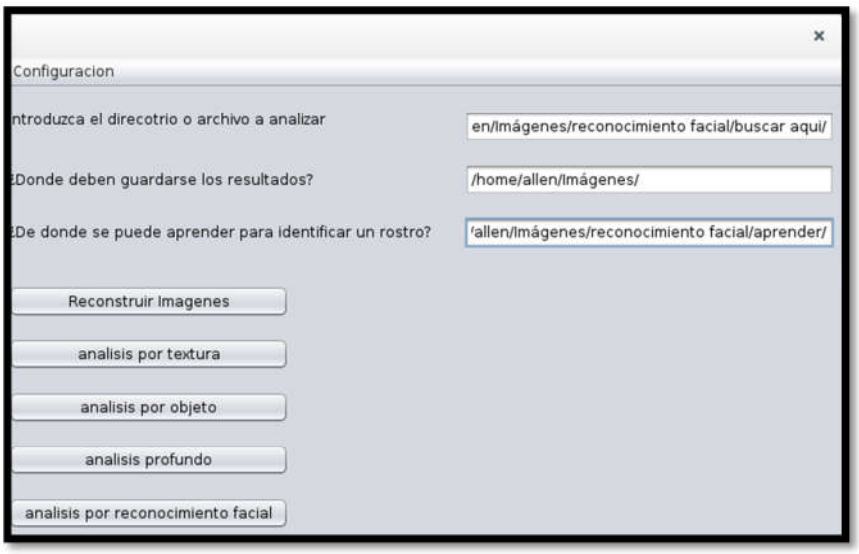
"¿Dónde deben guardarse los resultados?" la ruta completa del directorio en donde se desea guardar los resultados. Tal y como indica la imagen. Nótese que las ruta finalizan con el carácter /



Para reconstruir el tráfico de la red se procede de manera similar. Solo se debe copiar la ruta completa del archivo del que se desea analizar en el campo "introduzca el directorio o archivo a analizar", y luego se le debe definir en el campo "¿Dónde deben guardarse los resultados?" la ruta completa del directorio en donde se desea guardar los resultados. Tal y como indica la imagen. Notese que solo la ruta del segundo campo finaliza con el carácter /



Para un análisis de reconocimiento facial, lo que se debe hacer es lo siguiente. Se debe copiar la ruta completa del directorio que se desea analizar en el campo "introduzca el directorio o archivo a analizar", y luego se le debe definir en el campo "¿Dónde deben guardarse los resultados?" la ruta completa del directorio en donde se desea guardar los resultados. En el campo "¿De dónde se puede aprender para identificar un rostro?", se debe suministrar la ruta completa de un directorio que contenga únicamente fotos de la persona que debemos buscar. Tal y como se ve en la imagen. Nótese que las ruta finalizan con el carácter /



Los archivos para reconstrucción de imágenes se encuentran en /home/allen/Documentos/, se recomienda discreción al momento de reconstruir las imágenes.

Para realizar análisis de imágenes, se ha proporcionado un conjunto de fotos que se encuentran en la ruta: /home/allen/Imágenes/

La carpeta textura y la carpeta objeto tienen la siguiente estructura

./aquí -> contiene todas las imágenes que si se deben reconocer

./aquí no -> contiene todas las imágenes que no se deben reconocer

Estas dos carpetas permiten identificar fácilmente cuales son los resultados deseados y no deseados

./mix contiene las imágenes de las dos carpetas anteriores. Es sobre esta carpeta que se realiza el análisis de imágenes

./resultados es la carpeta en donde se guardan los resultados.

./resultados esperados -> contiene todas las imágenes que si se deben reconocer

./no esperado -> contiene todas las imágenes que no se deben reconocer

Estas dos carpetas permiten identificar fácilmente cuales son los resultados deseados y no deseados

./buscar aqui contiene las imágenes de las dos carpetas anteriores. Es sobre esta carpeta que se realiza el análisis de imágenes

./aprender contiene las fotos de una persona en particular