

Grado en ingeniería informática

Grado 2021-2022

*Trabajo Fin de Grado*

# “Análisis forense de la aplicación Microsoft Your Phone”

---

Lucas González de Alba

Tutor/es

Pedro Peris López

Colmenarejo, 03/03/2022



Esta obra se encuentra sujeta a la licencia Creative Commons  
**Reconocimiento – No Comercial – Sin Obra Derivada**

## **DEDICATORIA**

Este trabajo ha sido posible gracias a mi tutor Pedro Peris López que me ayudó en la selección de temáticas y desarrollo de la investigación y Josué Tapia, con quien he podido discutir algunas de las metodologías de análisis que mejor se ajustaban al trabajo. Finalmente agradecer a la Universidad Carlos III y a la empresa Álvarez y Marsal el apoyo y los recursos de los que me han dotado.

# ÍNDICE DE CONTENIDOS

Dedicatoria	ii
Índice de contenidos	iii
Índice de figuras	vi
Índice de tablas	vii
Lista de abreviaturas	viii
Introducción	1
1.1 Estructura	1
1.2 Resumen	2
1.3 Motivación	3
1.4 Metodología de trabajo	4
1.5 Objetivos	6
Estado del arte	8
2.1 Definiciones	8
2.2 Recursos externos	9
2.3 Marco regulador	10
2.3.1 La legislación y normativa legal.	10
2.3.2 La figura del perito informático en los juzgados.	12
2.3.3 La cadena de custodia	13
2.4 Entorno socio-económico	15
Desarrollo del proyecto	17
3.1 Planificación	17
3.2 Presupuesto	17
3.3 Tecnologías empleadas	20
	iii

3.4 Análisis del problema	23
3.4.1 Aspectos generales de Your Phone	23
3.4.2 Análisis dinámico	25
3.4.3 Análisis estático	29
3.5 Diseño de la solución	35
3.5.1 Reconocimiento facial	38
3.6 Implementación de la solución	40
3.6.1 Contactos, teléfonos, llamadas, sms y mms	42
3.6.2 Tratamiento de imágenes	44
3.6.3 Salida del programa	45
3.7 Pruebas realizadas	47
3.8 Resultados obtenidos	52
Conclusiones y trabajos futuros	53
4.1 Objetivos cumplidos	53
4.2 Líneas futuras de trabajo	54
4.3 Conclusión	55
Bibliografía	57
Anexo I. English summary	61
Introduction	61
1.1 Abstract	61
1.2 Motivations	62
1.3 Work methodology	62
1.4 Goals	63
State of the art	63
2.1 External resources	63

2.2 Legislation	65
2.2.1 Legislation and legal regulations.	65
2.2.2 The figure of the computer expert in the courts.	65
2.2.3 Chain of custody	65
2.3 Socio-economic environment	66
Implementation	66
3.1 Planning	66
3.2 Budget	66
3.3 Your Phone forensic analysis	67
3.3.1 SQL tables	68
3.3.2 Image processing	69
3.3.3 Execution modes	69
3.4 Algorithm design	70
Conclusions	71
4.1 Accomplished objectives	71
4.2 Future lines of work	72
4.3 Conclusion	72
Anexo II. Legislación e informática forense	73

## ÍNDICE DE FIGURAS

Fig. 1 Diagrama de Gantt: Planificación del proyecto .....	17
Fig. 2 Vinculando PC y móvil .....	26
Fig. 3 HKU\sid\SOFTWARE\Microsoft\IdentityCRL\UserExtendedProperties\email .....	27
Fig. 4 Eliminar equipos vinculados. Izq Your Phone (Windows), dch Your Phone Companion (Android) .....	28
Fig. 5 Configuración de BBDD SQL .....	29
Fig. 6 Metadatos de la imagen almacenada en thumbnails .....	32
Fig. 7 Metadatos de la imagen almacenada en media .....	33
Fig. 8 Metadatos de la imagen original .....	33
Fig. 9 Tablas de settings.db y photos.db .....	33
Fig. 10 Tablas de deviceDataDB, sharedContentDB, notificationsDB y callingDB .....	34
Fig. 11 Tablas de Contacts.db .....	34
Fig. 12 Tablas de phone.db .....	35
Fig. 13 Diagrama de componentes de YourPhoneForensicAnalyzer .....	35
Fig. 14 Diagrama de clases .....	36
Fig. 15 Diagrama de flujo de información entre aplicación, programa y analizador ....	36
Fig. 16 Flujo de ejecución de YourPhoneForensicAnalyzer .....	42
Fig. 17 Llamadas y mensajes asociados a un contacto .....	43
Fig. 18 Llamadas y mensajes sin asociar .....	43
Fig. 19 Aplicaciones instaladas .....	43
Fig. 20 Rostros similares de un mismo sujeto .....	46
Fig. 21 Rostros femeninos de mujeres menores de 25 .....	46
Fig. 22 Rostros de raza negra .....	46
Fig. 23 Rostros sonrientes .....	46

## ÍNDICE DE TABLAS

Tabla 1 Matriz de costes de recursos humanos .....	18
Tabla 2 Matriz de costes para herramientas hardware .....	18
Tabla 3 Matriz de costes para herramientas software.....	19
Tabla 4 Presupuesto total del proyecto.....	20
Tabla 5 Características técnicas de los dispositivos empleados .....	20
Tabla 6 Permisos de Your Phone .....	24
Tabla 7 Permisos de Your Phone Companion.....	24
Tabla 8 Ficheros CSV para buscar en YourPhoneForensicAnalyzer.....	40
Tabla 9 Firma de archivo según la extensión .....	44
Tabla 10 Pruebas sobre modelos de reconocimiento facial.....	47
Tabla 11 Batería de pruebas para YourPhoneForensicAnalyzer.....	50
Tabla 12 Pruebas realizadas a los artefactos de Your Phone .....	51

## **LISTA DE ABREVIATURAS**

TFG: Trabajo final de grado

BBDD o db: Base de datos

DFIR: Digital Forensics and Incident Response

SMS: Short Message Service o Servicio de Mensajes Simples

MMS: Multimedia Messaging Service o Servicio de Mensajería Multimedia

IA: Inteligencia Artificial

UUID: Universally Unique IDentifier o Identificador único universal

TCP: Transport-Control-Protocol

ICP: Internet Control Protocol

TLS: Transport-Layer-Security

Mtcnn: Multi-task Cascaded Convolutional Networks



# INTRODUCCIÓN

El presente documento recoge el trabajo de fin de grado desarrollado por Lucas González de Alba, alumno de la Universidad Carlos III de Madrid. En él se presenta su proyecto de investigación sobre la estructura, comportamiento y artefactos de la aplicación Microsoft Your Phone.

## 1.1 Estructura

La organización del documento viene dada por la siguiente estructura.

- Introducción: En primer lugar, este apartado se compone de una breve sinopsis del proyecto, un resumen del cuerpo de trabajo donde se resaltan los aspectos clave del mismo. En segundo lugar, las motivaciones que originaron el mismo, así como la coyuntura y propósitos que justifican su desarrollo. En tercer lugar, se explicará la metodología de trabajo y por último los objetivos del estudio.
- Estado del arte: Esta sección se centra en el proceso de investigación. En ella se muestran los distintos estudios encontrados relacionados con la temática del proyecto, exponiendo sus resultados y limitaciones. Asimismo, se presenta el marco legal en el que se engloba. Dado que el campo de la informática forense está íntimamente ligado a la labor judicial, resulta muy relevante exponer el statu quo legal, así como indicar los distintos principios que se han seguido para garantizar la correcta protección, extracción y manipulación de la información sujeta a estudio ya que el criterio aquí recogido permite que la investigación conserve garantías legales.
- Desarrollo del proyecto: En este apartado se describe el cuerpo de trabajo, desde la planificación, preparación del presupuesto, aspectos técnicos de las tecnologías empleadas, su implementación y evaluación de resultados.
- Conclusiones y trabajos futuros: Esta sección está dedicada a la resolución de la investigación y el análisis de las soluciones propuestas. Finalmente, se plantean futuras vías de estudio para las cuales el trabajo desarrollado sirva de fundamento.

## 1.2 Resumen

Microsoft Your Phone es un servicio de Microsoft para sincronizar ordenador y móviles con el objetivo de facilitar al usuario el acceso a sus dispositivos integrando notificaciones, mensajes, fotos, llamadas y otras opciones del teléfono directamente en Windows. En las propias palabras de Microsoft:

“Obtén acceso instantáneo a todo lo que te gusta de tu teléfono, directamente desde tu PC. Vincula tu teléfono Android y tu PC para ver y responder a los mensajes de texto, hacer y recibir llamadas y mucho más, todo directamente en tu PC.” [1]

Dado que toda esta información puede resultar valiosa en el contexto de una investigación, el trabajo que aquí se presenta analizará la aplicación Your Phone desde la perspectiva forense en busca de sus artefactos y trazas digitales. Para ello se comenzó por indagar en el conocimiento ya existente sobre Microsoft Your Phone (artículos, publicaciones, blogs, foros...etc.) y al descubrir que el número de trabajos relacionados con el proyecto era escaso y atrasado se decidió profundizar sobre este.

El primer paso consistió en adquirir una visión general del funcionamiento de la app. Se recogieron varias muestras digitales mediante pruebas con distintos dispositivos que fueron variando su contenido e interacciones para observar el efecto que dichas evidencias producían en la muestra. Más adelante procedió a realizar un estudio formal del comportamiento de la aplicación utilizando técnicas de monitorización de procesos para encontrar aquellos artefactos del sistema relevantes. Sobre estos, se evaluó el formato y estructura y se estimó el valor forense de los datos que almacenaban. A continuación, se desarrolló un servicio que facilitase la extracción y búsqueda, un programa Python compuesto dedicado por un lado a parsear e indagar en el contenido textual-numérico del teléfono (registros de llamadas, conversaciones y aplicaciones instaladas), y por otro, al tratamiento de imágenes, en concreto el reconocimiento facial. Sobre este último se buscaron librerías que permitieran identificar múltiples caras dentro de una imagen, compararlas entre sí para buscar similitudes en base a un perfil descriptivo o una imagen de referencia. Se terminó escogiendo un híbrido entre OpenCV y DeepFace por sus resultados favorables y versatilidad. Finalmente se evaluó la calidad del software a través de una batería de pruebas donde se pudo comprobar su robustez y así como la configuración óptima, el modelo VGG-Face utilizando detección y alineamiento mtcnn.

### 1.3 Motivación

La investigación inicialmente surgió como un proyecto personal del autor dedicado a conocer más a fondo algunas de las técnicas de análisis utilizadas en ciberseguridad e informática forense. Rápidamente creció en extensión, complejidad y alcance por lo que se decidió a orientarlo en una dirección mayor, un trabajo final de grado. ¿Por qué informática forense? Porque desgraciadamente es una asignatura que no se llega a impartir en el grado de informática y su estudio puede resultar muy beneficioso para el currículo del alumno, ya que ofrece conocimientos transversales a muchas otras áreas de las tecnologías de la información (seguridad, sistemas y computación).

Otro aspecto motivador fue la colaboración interdisciplinaria entre la escuela politécnica y la empresa privada. Gracias a la labor del equipo de orientación y empleo de la UC3M el autor de la publicación pudo acceder al sector laboral a través de una beca en prácticas dentro del departamento de informáticos forenses de la empresa Álvarez & Marsal. Esta beca le permitió formarse en los conocimientos y metodologías específicas a la área forense, dotándolo así de herramientas y espacio de trabajo (copiadoras, discos, licencias software...) lo que en último término ha favorecido escoger la temática.

Respecto a la elección de la materia, se consideran muy positivos los beneficios que aportaría el trabajo de análisis de Microsoft Your Phone. Esta es una aplicación no demasiado estudiada, por lo que ahondar y expandir el conocimiento que se tiene de ella permitiría que futuros casos de estudio se valiesen del desarrollo realizado. Entre algunas de las ventajas se encuentra:

- Exponer el estudio formal de la estructura y artefactos determinando qué información se puede extraer de los artefactos de la aplicación. Con ello los analistas forenses podrían estimar qué información está a su alcance en las investigaciones.
- Vincular el entorno móvil (teléfonos, llamadas, chats y notificaciones) al de los computadores (usuario y programas) a través de la app. Conectar ambos permitiría obtener información del comportamiento de aquellos teléfonos vinculados con la aplicación, lo cual es de gran valor cuando se investiga una evidencia ya que exclusivamente a través de un PC no se tiene ninguna información de móviles. Cabe agregar que cada vez más son las personas que utilizan el móvil como

método no solo de comunicación, sino también para la gestión del entorno laboral, para el ocio y entretenimiento etc..., por lo que ampliar el alcance de una investigación a través de este medio puede suponer todo un punto de inflexión.

- Extracción automatizada de imágenes almacenadas por el sistema. Esto evitaría la tarea tediosa y repetitiva de buscar, seleccionar y guardar cada imagen de la app.
- Permitir análisis 'live' o 'en vivo' de evidencias. Ejecuciones en tiempo real sobre la evidencia agilizaría las investigaciones en las que no se dispone de evidencias para análisis pasivo. Gracias al software que específicamente se ha desarrollado se podrían procesar casos en vivo.

En resumen, estudiar desde la perspectiva forense la aplicación Microsoft Your Phone no solo aporta las ventajas previamente expuestas, sino que también resulta novedoso y ofrece una oportunidad al autor única para desarrollar un trabajo sustancial.

#### **1.4 Metodología de trabajo**

La metodología de trabajo se compone de dos grandes bloques, análisis y desarrollo.

Por un lado, el apartado dedicado al estudio del programa. Para adquirir la mayor cantidad de información posible del comportamiento del programa se establece un análisis incremental basado en las siguientes fases:

- Fase preliminar: se dispone el entorno de trabajo y las herramientas del análisis. Al instalar la aplicación tanto en móvil como en PC se recogen los permisos y propiedades de las instalaciones.
- Familiarización: esta fase determina una primera fuente de estudio y busca conocer intuitivamente qué requisitos y capacidades tiene la aplicación.
- Análisis dinámico: consiste en monitorizar en tiempo real la aplicación para conocer las interacciones de sus procesos y así poder recopilar aquellos artefactos del sistema que resulten relevantes.

- Análisis estático: consiste en estudiar las principales trazas y artefactos descubiertos previamente.

Por otro lado, el apartado centrado en el desarrollo de una solución software dedicada a facilitar el acceso a la información previamente analizada. De nuevo la metodología a seguir es progresiva e incremental, comenzando por:

- Identificación de requisitos: captación de las necesidades y restricciones del problema.
- Diseño: establecer cuál de todas las distintas arquitecturas y soluciones software es la que mejor se adapta al problema.
- Implementación: desarrollar el programa para la plataforma y arquitectura escogida.
- Pruebas: evaluar el correcto funcionamiento del programa y verificar el cumplimiento de los objetivos.

La principal desventaja de dividir el proyecto en dos grandes bloques sucesivos, análisis y desarrollo, es que el primero condiciona al segundo. Esto implica que se debe retrasar la implementación del programa hasta conocer exhaustivamente la forma en la que se almacena la información en la aplicación, y hacerlo no es demasiado recomendable ya que exige demorar el desarrollo. Para evitarlo, se ha optado por romper el problema en sub-problemas de forma que estos sí se puedan paralelizar. Para ello se realizará un estudio preliminar, no muy extenso, de Microsoft Your Phone para determinar cuáles son las trazas que ésta deja en el sistema, así como sus principales características. Una vez se tenga se analizará y desarrollará para cada una de esas características un módulo y así, de esta forma, se logrará adquirir la precisión del estudio a la par que los avances en la implementación.

Otra desventaja del esquema análisis-desarrollo es que exige “duplicar” las pruebas. En la primera fase de análisis es necesario interactuar con la aplicación para conocer su funcionamiento en los distintos casos de uso, y en la última fase de desarrollo, hay que

repetir las pruebas para comprobar que el código cumple con el comportamiento esperado. Este aspecto por desgracia resulta irreconciliable ya que no es recomendable construir un sistema basándose en una conjetura que debe probarse cierta en la última fase del proyecto. No obstante, el progreso escalonado presentado previamente permite construir software que rápidamente pase a ser funcional, acercando así la primera y última fase de test.

En resumen, la metodología escogida adopta tanto el análisis como el desarrollo de forma que ambos se complementen. Al combinarse con un avance incremental e iterativo se consigue una mayor cobertura del problema, además que permite agregar nuevas funcionalidades de forma modular. Este esquema se ajusta muy bien al programa que se busca desarrollar ya que se trata de construir un software multitarea (parseado, extracción y manipulación multimedia). Finalmente añadir que, aunque el trabajo de paralelización nunca es sencillo puesto que exige una extensa planificación y una laboriosa tarea de sincronización, se ha optado por organizar el proyecto siguiendo un diagrama de Gantt (Fig. 1 Diagrama de Gantt: Planificación del proyecto)

## 1.5 Objetivos

Este capítulo recoge la principal meta del trabajo y establece los distintos objetivos específicos que la componen.

Principales metas:

1. **Detallar los procesos que componen la aplicación Microsoft Your Phone y los artefactos que estos dejan en el sistema.** Recoger detalladamente las trazas digitales del programa e identificar qué información de valor se almacena en el mismo.
2. **Implementar una solución software que permita recoger, parsear y exportar la información que presente la aplicación Microsoft Your Phone.** A partir de los descubrimientos previos construir un sistema que facilite el acceso a la información almacenada y permita su extracción.

## Objetivos específicos:

1. **Extraer el contenido multimedia de la aplicación y aplicar sobre este detección y categorización de rostros.** Entornos de trabajo especializados para analistas menudo incluyen servicios de procesamiento de imágenes, entre los más comunes se encuentra la detección de rostros. Esto permite al analista ahorrar un ingente número de horas de revisión en busca de uno o varios individuos.
2. **Implementar un sistema de búsqueda por perfil facial.** Desarrollar una herramienta software para localizar personas dadas las principales características de su rostro.
3. **Evaluar el correcto funcionamiento de la solución software mediante un amplio espectro de evidencias y casos de prueba.** Comprobar si el programa funciona adecuadamente al enfrentarse a los distintos casos de uso de Microsoft Your Phone.
4. **Evaluar el contenido salvaguardado por la aplicación:** Siempre que sea posible, se valorará positivamente la capacidad de recuperar contenido almacenado automáticamente por el programa.
5. **Extraer contenido eliminado de la aplicación:** Siempre que sea posible, se valorará positivamente la capacidad de extraer contenido borrado. Para cumplirlo se utilizarán técnicas de carving sobre Unallocated Space

## ESTADO DEL ARTE

### 2.1 Definiciones

A continuación, se detallan los conceptos más importantes del proyecto con el objetivo de esclarecer ambigüedades y facilitar la comprensión y lectura del documento.

- Evidencia: En el contexto forense digital se entiende por evidencia un registro de información almacenada o transmitida a través de un sistema informático, que puede ser utilizado como prueba en procedimientos judiciales.
- Artefacto: “Todo aquello que puede obtener una evidencia, y cabe recalcar que son los diferentes ficheros, cadenas de registro, rutas de acceso y configuraciones que pueden determinar la actividad de un malware o de un usuario malicioso, así como las evidencias necesarias para una prueba.” [2]
- TLS: se trata de un protocolo de comunicaciones que utiliza criptografía asimétrica para proporcionar comunicaciones seguras y confidenciales a través de una red.
- File Carving: técnica utilizada para recuperar archivos eliminados. Consiste en localizar en un conjunto de bytes las cadenas de caracteres que identifican el comienzo y final de un fichero para posteriormente restaurarlo (copiar todo el contenido intermedio). Esta técnica se puede utilizar sobre cualquier cadena de bytes, es decir, se puede extraer contenido de disco, de la memoria RAM, de unallocated space o incluso dentro de un propio fichero.
- SMS y MMS: Un mensaje de texto de hasta 160 caracteres sin un archivo adjunto se conoce como SMS, mientras que un texto que incluye un archivo como una imagen, un vídeo, un emoji o un enlace a un sitio web- se convierte en un MMS.



## 2.2 Recursos externos

En el ámbito de la informática forense y ciberseguridad es común encontrar publicaciones relacionadas con el estudio y monitorización de programas de software libre, corporativo y malicioso. En dichos estudios se suele aplicar tanto análisis estático como dinámico y suelen involucrar hash análisis, carving, monitorización de red, reconocimiento de procesos y threads, rastreo del registro de Windows e investigación de artefactos. Comúnmente el trabajo de monitorización suele ser similar a todos los programas, y por tanto generalizable, pero cuando se trata de extraer de este, conocimiento e información propios de un programa, la labor se vuelve específica al caso. En otras palabras; puesto que cada aplicación es distinta, el análisis debe ser único para cada programa.

En lo que respecta a Microsoft Your Phone solo se han encontrado tres publicaciones que realicen este análisis y son *Digital Forensics Tips & Tricks: «Your Phone» app Forensics* [3], *Digital forensic artifacts of the Your Phone application in Windows 10* [4] y su posterior revisión *Microsoft's Your Phone environment from a digital forensic perspective* [5]. El primero surgió como respuesta a la publicación del Insider Preview Build 18999 (20H1) de Windows 10 y se trata de un breve y superficial examen de los artefactos que la aplicación, recién introducida, producía. Por contraste, en el segundo los autores Patricio Domingues, Miguel Frade, Luis Miguel Andrade y Joao Victor Silva analizan las posteriores versiones 1.0.20453 y 3.4.4 de Your Phone para Windows 10 y la app para Android Your Phone Companion respectivamente. Además, su investigación propuso un script de Python diseñado para ejecutar en Autopsy. Por último, el tercero evalúa las actualizaciones 1.21011.127.0 (Windows) y 1.21021.81.0 (Android) y sigue la línea de desarrollo anterior, ampliando algunos aspectos que quedaron fuera del estudio previo y expandiendo las funcionalidades del programa propuesto. En comparación con el análisis que plantea Panov ambos trabajos son realmente reveladores ya que en gran medida presentan en mayor detalle cómo se organizaba la aplicación y como esta almacenaba los datos del usuario. En conjunto, los dos estudios hacen un análisis bastante completo y los autores consiguen cumplir con algunos de los objetivos que este trabajo persigue, pero no obstante dejan otros fuera. Algunos de los aspectos técnicos que la publicación no resuelve definitivamente son la monitorización de procesos, análisis del registro de Windows, descripción de la configuración de los artefactos. Otro problema adicional son las nuevas actualizaciones. La aplicación ha continuado renovándose,

incluyendo nuevas funcionalidades como la ampliación de su galería, mensajería instantánea, y compartición de pantalla. Esto ha hecho que aparezcan nuevos artefactos y que otros hayan ido variando su estructura interna, por lo que algunas partes de su investigación queden relegadas a versiones anteriores, y por tanto no se puedan utilizar como punto de partida. De la misma forma, el script de Python propuesto, aparte de resultar inservible para las nuevas versiones, está obsoleto ya que las nuevas versiones de Autopsy (el programa para el que se desarrolló) tampoco lo reconocen. Este fenómeno de obsolescencia ocurre constantemente en la informática, pero su efecto se siente especialmente en el campo forense.

En cuanto a fotos y videos se refiere son cada vez más comunes los servicios de procesamiento digital de imágenes embebidos en aplicaciones de carácter forense. Compañías como Belkasoft, Magnet y Cellbrite están apostando por programas con utilidades variadas, especialmente con servicios orientados al reconocimiento de imágenes. Las últimas versiones de productos consolidados como Belkasoft X, Axiom Cyber o Cellebrite Physical Analyzer incluyen reconocimiento óptico de caracteres (OCR), categorización de imágenes o auto detección de objetos y personas. En este sentido, se observa una tendencia a incorporar los últimos avances del machine learning e IA. Al hacerlo se expanden las capacidades del software y se facilita el trabajo de los analistas, pero ninguna de las anteriores publicaciones enfrenta este reto, lo cual abre la posibilidad a la innovación.

## **2.3 Marco regulador**

La informática forense es un sector ampliamente regulado puesto que en él se trata información sensible normalmente en un contexto judicial. Existen tres aspectos clave en lo que investigaciones forenses se refiere:

### **2.3.1 La legislación y normativa legal.**

El Convenio de la Ciberdelincuencia [6], elaborado en Budapest el 23 de noviembre de 2001 y ratificado por España en 2010 cataloga los ciberdelitos en cuatro ramas:

1. Delitos que atentan contra el derecho a la confidencialidad, integridad y la disponibilidad de sistemas informáticos (sea ataque, interceptación o interferencia)

2. Delitos de falsificación y fraude informático mediante introducción, alteración o destrucción de datos o sistemas informáticos
3. Delitos por tenencia, adquisición, producción o difusión de contenido pornográfico infantil.
4. Delitos contra la autoría y propiedad intelectual

Así en España, la ley prevé delitos contra la privacidad, el espionaje, robo, suplantación de la personalidad, fraudes, falsificaciones, malversación, manipulación de dispositivos, daños o alteraciones de programas de datos o archivos, etc., todo ello ejemplos en los que la informática forense interviene. Cabe mencionar dentro de este apartado las connotaciones éticas de la profesión, para la cual existen distintos códigos éticos o recomendaciones. Según la escuela internacional de informáticos forenses o la International Society of Forensics Computer Examiners [7] (ISFCE) algunos de los requisitos éticos necesarios para certificarse como profesional son:

- 1 Demostrar compromiso y diligencia en el desempeño de las funciones asignadas.
- 2 Demostrar integridad en la realización de tareas profesionales.
- 3 Mantener la máxima objetividad en todos los exámenes forenses y los hallazgos actuales con precisión.
- 4 Realizar exámenes basados en lo establecido, procedimientos validados.
- 5 El cumplimiento con los más altos estándares morales y éticos y cumplir con el Código de la ISFCE
- 6 Testificar en sinceridad en todos los asuntos ante cualquier junta, tribunal o procedimiento.
- 7 Evitar cualquier acción que pudiera presentar a sabiendas un conflicto de intereses.
- 8 Cumplir con todos los ordenamientos jurídicos de los tribunales

- 9 Examinar objetivamente y a fondo todas las pruebas dentro del alcance del trabajo.
- 10 Las personas certificadas son responsables de mantener la certificación en los más altos estándares éticos y demostrar integridad, imparcialidad, diligencia y profesionalidad.
- 11 No ser cómplice ni participar en conductas no éticas o ilegales

Otras instituciones como el instituto SANS, además de todas estas guías incluyen el respeto por la integridad y honestidad, la defensa de la propiedad intelectual, confidencialidad y los derechos y libertades individuales y en definitiva la profesionalidad y la salvaguarda de la verdad. Del mismo modo condena cualquier forma de corrupción (chantaje, soborno o comisión), actitud prevaricadora, atentado premeditado contra la privacidad o la discriminación por sexo, raza, religión, edad, etnia, política o cualquier otra condición.

En conclusión, la normativa y la ética profesional dentro de este campo buscan, ante todo, proteger y mantener la honradez y entereza, para que así sea posible esclarecer la verdad y legislar en base a ella.

### **2.3.2 La figura del perito informático en los juzgados.**

La ley define a esta persona como aquel profesional especializado en la informática y en las nuevas tecnologías cuya labor consiste en proveer asesoramiento técnico en procedimientos judiciales, así como contribuir a la mediación y resolución de conflictos. Puede ejercer varios roles, el de mediador y árbitro y el de auditor. Los primeros se toman cuando dos partes están en desacuerdo y el perito debe intervenir para resolver las diferencias. Más concretamente es mediador si dirige o interviene activamente en las negociaciones y árbitro si su papel es pasivo, objetivo e imparcial. El arbitraje se resuelve mediante el “laude arbitral”, el dictamen alcanzado tras peticiones, reivindicaciones y alegatos. Para poder ejercer y ser reconocido como perito, se debe disponer de titulación y pertenecer a un colegio de profesionales de Informática. De lo contrario si se ejerce sin titulación o sin estar colegiado se está cometiendo un delito de intrusismo profesional

(Art. 340 y Art 341 de la Ley 1/2000, de 7 de enero, de Enjuiciamiento Civil y Art 403 del Código Penal) [8]

El peritaje informático puede llevarse a cabo judicial o extrajudicialmente, siendo la motivación del análisis la principal diferencia. En uno, el procedimiento se centra en la obtención de pruebas para presentar ante el juez, mientras que, en el otro, las pruebas se recogen para esclarecer hechos o recabar mayor información. A menudo, si los abogados lo recomiendan, se puede presentar o ampliar una denuncia con el peritaje ratificado por el forense. Naturalmente también existe el contraperitaje, que consiste en rebatir el informe pericial de otra persona para poder impugnar mediante argumentos técnicos válidos. Esto último es importante ya que el forense debe testificar ante el juez que el informe es veraz y la evidencia del caso no ha sido alterada de ningún modo. Para asegurarse de que el procedimiento tiene garantías se sigue una extensa documentación y validación de cada interacción con la prueba (física o digital) y el cliente.

### **2.3.3 La cadena de custodia**

A la hora de presentar una prueba digital ante el juez existen una serie de requisitos previos que se deben cumplir para que ésta sea admitida. La normativa legal determina que para cualquier prueba recogida se debe preservar la evidencia original junto con su cadena de custodia. La cadena de custodia es un procedimiento de control que recoge el proceso de obtención, manipulación, transferencia, cesión y preservación de evidencias para asegurar de forma rigurosa que la prueba ha sido entregada y permanece inalterada (demostración mediante hash). Existen distintos tipos de cadenas de custodia según la evidencia (móvil, portátil, servidor, memoria etc...), pero todas ellas comparten los siguientes campos:

#### **1. Información general**

- ☐ n/a - Nombre del cliente
- ☐ n/a - Nombre del proyecto
- ☐ n/a - Nombre del custodio
- ☐ n/a - Número de la evidencia

## 2. Información del dispositivo original

☐ n/a - Nombre del fabricante

☐ n/a - Número de serie y modelo

☐ n/a - Tipo de dispositivo

Portátil, Sobremesa, Tablet, Móvil, Servidor, Otro

☐ n/a - Estado del dispositivo

Apagado, Encendido con sesión iniciada, Encendido sin sesión iniciada

☐ n/a - Tipo de información

Imagen forense, Correo electrónico

## 3. Información sobre el medio

☐ n/a - Nombre del fabricante

Seagate, Western Digital, Intel, Samsung, Toshiba, Hitachi, IBM, Otro

☐ n/a - Factor y forma

1.8", 2.5", 3.5", SATA, uSATA, USB

☐ n/a - Tipo de conexión

SATA, eSATA, SCSI, USB, IDE, ZIFF

☐ n/a - Tipo de almacenamiento

HDD, SSD, RAID, Memoria FLASH, Memoria RAM, Cinta, Floppy,  
CD/DVD, Nube/Web, FTP, Otro

☐ n/a - Número de serie y modelo

☐ n/a - Capacidad

☐ n/a - Encriptado

AES, RSA, Triple DES, Otro

## 4. Información sobre la adquisición

☐ n/a - Tipo

Imagen física, Imagen lógica, Copia lógica

- ☐ n/a - Versión
- ☐ n/a - Fecha y hora según dispositivo
- ☐ n/a - Tiempo de adquisición
- ☐ n/a - Localización
- ☐ n/a - Herramienta utilizada

Software (EnCase, Magnet Cyber, FTK Imager)

Hardware (Tableau, Cellebrite, Dossier, Disco de booteo Live...)

5. Información sobre la copia original y su salvaguardado

- ☐ n/a - Nombre de los fabricantes
- ☐ n/a - Números de serie y modelos

6. Cadena de custodia

- ☐ n/a - Fecha de recepción la evidencia
- ☐ n/a - Nombre y apellidos de emisor
- ☐ n/a - Fecha de devolución de la evidencia
- ☐ n/a - Nombre y apellidos de receptor
- ☐ n/a - Notas del proceso
- ☐ n/a - Imágenes tomadas

## 2.4 Entorno socio-económico

El impacto económico resulta difícil de estimar dado que en principio el producto se orienta a un nicho reducido, investigaciones forenses que involucren evidencias que contengan artefactos de Your Phone. No obstante, se podría esperar que la aplicación y el trabajo de investigación desarrollado traigan algunas mejoras en el desempeño de los analistas para los casos en los que se presenten dichas evidencias. Según vestigelt [9] los costes promedios de una investigación forense suelen rondar en media entre 5.000\$ a 15.000\$. No obstante, ninguna estimación es buena ya que depende en gran medida del tamaño del caso, su complejidad, estado de las evidencias, la premura con la que esta deba concluir y otro compendio de casuísticas como el tipo de actividad investigada, el

volumen de datos y restricciones según cada país. A pesar de ello según la agencia se puede considerar estándar un coste de 250\$ por hora de trabajo efectiva. En este sentido, el proyecto desarrollado podría tener un impacto positivo ya que lograría reducir el tiempo de análisis lo que podría repercutir en mayores beneficios y menos pérdida por sobrecostes debido a extensiones de plazos. Luego también sucede que las posibles aplicaciones del programa software desarrollado no se limitan exclusivamente al contexto de investigaciones forenses ya que el script podría servir como base para otros ejercicios. Por ejemplo, su funcionalidad de extracción de imágenes serviría para cualquier aplicación que busque interactuar (introducir o extraer) el contenido multimedia de Your Phone. Lo mismo ocurriría con el comparador y agrupador de rostros, que en caso de generalizarse y perfeccionarse podría comercializarse como módulo adicional para algunas de los programas forenses descritos previamente (Belkasoft, Magnet y Cellbrite).

En cuanto a las implicaciones sociales y éticas cabe mencionar que al tratarse de un programa destinado a descubrir y trabajar con datos personales (conversaciones, llamadas, imágenes) se debería mantener estricta confidencialidad. Esto es parte de la guía de buenas prácticas de todo investigador y esta herramienta no es más que otro medio eficaz de conocer y analizar la información. Los beneficios de su correcta utilización son directos, un análisis veraz y eficiente de las evidencias digitales de un caso.

En este aspecto también entra en juego la nueva normativa de la unión europea (UE), El Reglamento General de Protección de Datos o GDPR, que entró en vigor en 2018. La UE ha centrado su atención en la seguridad y privacidad de sus ciudadanos por lo que cualquier almacenamiento o utilización de información personal debe justificarse y consentirse por parte del usuario. Esto sumado a la sensibilidad del trabajo exigen rectitud moral ya que de lo contrario pueden suceder filtraciones, chantajes, prevaricación, daño reputacional u otras formas de extorsión y manipulación mediática.

En este sentido la labor forense no solo debe seguir un estricto cumplimiento normativo, sino que también debe salvaguardar el código ético de la profesión.



## DESARROLLO DEL PROYECTO

### 3.1 Planificación

Siguiendo la metodología escogida la planificación resultante es aproximadamente la siguiente; búsqueda de información, mes y medio, análisis y desarrollo, tres meses, documentación, mes y medio y evaluación y pruebas quince días. A

A partir de este esquema se deriva el siguiente diagrama de Gantt:

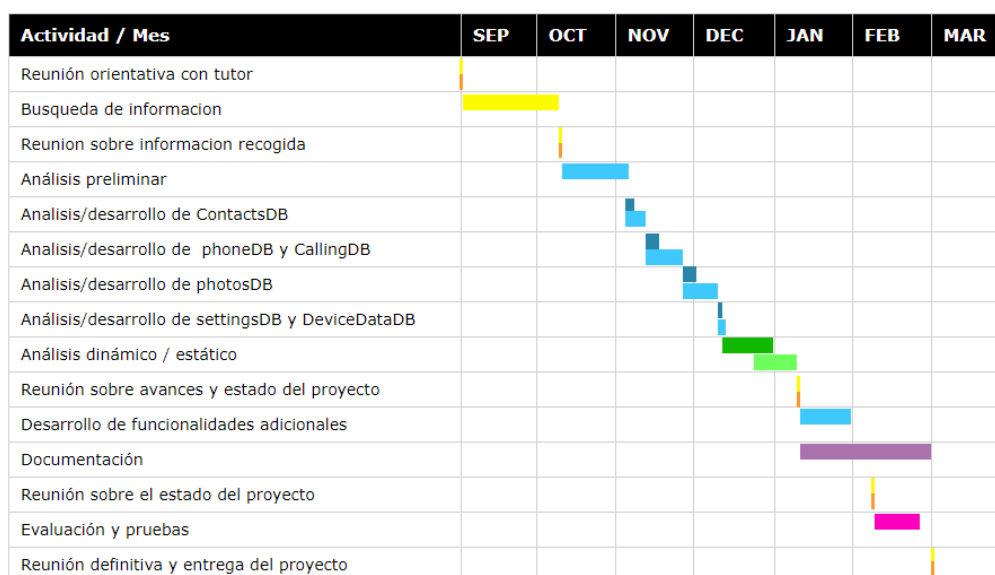


Fig. 1 Diagrama de Gantt: Planificación del proyecto

Puesto que el proyecto se compagina con la beca de estudios descrita previamente en el apartado de motivación, el número de horas asignadas al TFG se asemeja a una jornada parcial, por lo que el tiempo total estimado dedicado a la ejecución será de 290 horas.

### 3.2 Presupuesto

El presupuesto se divide en costes derivados del personal y herramientas de trabajo:

**Recursos humanos:** el número de horas incurridas en investigación y desarrollo. Dado que el Trabajo final de Grado se compone de tutor y alumno, para estimar el coste de su participación se han tomado como referencias los salarios medio de un programador

junior (novato) y un profesor universitario en España. Según glassdoor el promedio anual es de 19.745 [10] y 33.862 € [11] respectivamente, es decir aproximadamente 1.646 € y 2.822 € al mes, Partiendo de una jornada laboral completa (8 horas) a 21 días laborables por mes, el coste por hora es de alrededor de 10 y 18€. Si multiplicamos este valor por el número de horas de los participantes queda:

TABLA 3.2.1 - MATRIZ DE COSTES DE RECURSOS HUMANOS				
Puesto	Horas empleadas	Coste por hora	Subtotal	Total
Autor	310 h	10 €	3.100 €	3.244 €
Tutor	8 h	18 €	144 €	

*Tabla 1 Matriz de costes de recursos humanos*















**Herramientas de trabajo:** el gasto derivado de licencias y maquinaria.

- **Herramientas hardware:** el equipo y dispositivo sobre los que mayormente se ha trabajado o realizado pruebas han sido, un portátil DELL Latitude E7270 de 1.179,27 € y un móvil Samsung A20e, 145,20 €. Para calcular el coste total se ha calculado el porcentaje de uso (tiempo utilizado / tiempo de vida estimado) y se ha multiplicado por el precio de adquisición del recurso para así sacar su coste descontando su amortización. En el caso de teléfonos como el Samsung J3, A8 y HomTom HT20 Pro, o el portátil HP EliteBook 840 G60 no se ha imputado gasto ya que el uso ha sido ínfimo.

TABLA 3.2.2 - MATRIZ DE COSTES PARA HERRAMIENTAS HARDWARE					
Recurso	Precio	Horas de uso	Horas de vida útil	% uso	Subtotal
DELL Latitude E7270	1.179,27 €	310 h	~ 1.825 h	~16%	~190 €
Samsung A20e	145.2 €	125 h	~ 1.825 h	~7%	~10 €
HP EliteBook 840 G60	1.270 €	~ 0 h	~ 1.825 h	~ 0.0%	0 €
Samsung J3	124 €	~ 0 h	~ 1.825 h	~ 0.0%	0 €
Samsung A8	256 €	~ 0 h	~ 1.825 h	~ 0.0%	0 €
HomTom HT20 Pro	110 €	~ 0 h	~ 1.825 h	~ 0.0%	0 €
Coste total					200€

*Tabla 2 Matriz de costes para herramientas hardware*

- **Herramientas software:** la totalidad del trabajo se ha planteado desde la perspectiva del software de libre distribución por lo que no se incurre en gastos de licencias, no obstante, puesto que Your Phone es un programa desarrollado para Windows, sería necesario agregar el coste de uso del sistema operativo. Se comercializan cuatro versiones, una gratuita para estudiantes, por 145 euros Windows 10 Home, por 259 euros Windows 10 Pro, y a 439 euros Windows 10 Pro for Workstations, pero puesto que la Universidad Carlos III tiene acuerdo con Microsoft se utiliza una licencia educativa gratuita.

TABLA 3.2.3 - MATRIZ DE COSTES DE HERRAMIENTAS SOFTWARE			
	Software	Versión	Precio
	Windows 10 Pro	21H1 / 19043.1415 (Licencia estudiante)	Gratuito
	Microsoft Word 2016	MSO 16.0.4498.1000 (Licencia estudiante)	Gratuito
	Visual Studio	1.63.2	Gratuito
	Your Phone	1.21113.36.0	Gratuito
	Your phone companion	1.21113.85.0	Gratuito
	Android	5.1.1 / 6 / 11	Gratuito
	Sysinternals	-	Gratuito
	procDot	1.22	Gratuito
	WireShark	3.6.2	Gratuito
	Python	3.8.2	Gratuito
	Autopsy	4.19.2	Gratuito
	FTK Imager	4.5.03	Gratuito
	DB Browser	3.12.2	Gratuito
	DBBeaver Lite	21.3.0.202112052011	Gratuito
Coste total			0€

*Tabla 3 Matriz de costes para herramientas software*

El presupuesto total del proyecto es:

TABLA 3.2.4 PRESUPUESTO TOTAL DEL PROYECTO	
Recursos humanos	3.244 €
Herramientas hardware	200 €
Herramientas software	0 €
Total	3500 €

Tabla 4 Presupuesto total del proyecto

### 3.3 Tecnologías empleadas

El proyecto se sirvió de los siguientes dispositivos:







TABLA 3.3.1 - CARACTERÍSTICAS TÉCNICAS DE LOS DISPOSITIVOS EMPLEADOS							
Tipo	Modelo	Red	Pantalla	CPU	Memoria	Discos	Soportado
 PC	DELL Latitude E7270	Ethernet, Bluetooth Wi-Fi	12.5 pulgadas 1366x768 píxeles	i5-6300 2-Cores 2.4 GHz	DDR4 SDRAM 16GB 2133 MHz	SATA SSD SKhynix SC311 256GB	Your Phone
 PC	HP EliteBook 840 G60	Ethernet, Bluetooth Wi-Fi	14 pulgadas 1920x1080 píxeles	i7-8565U 4-Cores 1.8 GHz	DDR4 SDRAM 16GB 2400 MHz	SSD NVMe KBG30ZMV512G KIO 512GB	Your Phone
 Móvil	Samsung A20e SM-A202	Bluetooth, Wi-Fi,GSM 4G/3G/2G, NFC	5.8 pulgadas 720x1560 píxeles	Exynos 7884 2-Cores 1.6 GHz	LPDDR4 3GB 2133 MHz	eMMC 5.1 32GB MicroSD 1TB	Your Phone Companion
 Móvil	Samsung A8 SM-A530FZKDPHN	Bluetooth, Wi-Fi,GSM 4G/3G/2G, NFC	5.6 pulgadas 1080x2220 píxeles	Exynos 7885 7-Cores 2.2 GHz	LPDDR4 4GB 2133 MHz	eMMC 5.1 256 GB MicroSD 32GB	Your Phone Companion
 Móvil	Samsung J3 SM-J320FN	Bluetooth, Wi-Fi, GSM 4G/3G/2G,	5 pulgadas 720x1280 píxeles	ARM CortexA7 4-Cores 1.5 GHz	LPDDR3 1.5GB 800 MHz	eMMC 4.5 8GB MicroSD 4GB	No
 Móvil	HomTom HT20 Pro	Bluetooth, Wi-Fi,GSM 4G/3G/2G	4.7 pulgadas 1280x720 píxeles	ARM CortexA5 8-Core 1.3 GHz	LPDDR3 3GB 800 MHz	32GB microSD -	No

Tabla 5 Características técnicas de los dispositivos empleados

Microsoft lista los dispositivos soportados en la página oficial de Your Phone [12]

Para el software desarrollado se utilizó como lenguaje Python, versión 3.8.2. A continuación se recogen las librerías empleadas de acuerdo al siguiente formato:

Tipo de librería

- *Librería:*

Justificación

- Sistema

- *Io* [13]: consiste en una librería para entradas y salidas de texto, binario, y contenido sin formato. Necesaria para manipular bytes de las imágenes y texto (ej. `io.BytesIO`).
- *time* [14]: librería diseñada para manipular formatos, fechas, horas, minutos y segundos. Necesaria para recoger el tiempo de ejecución del programa
- *datetime* [15]: librería diseñada para manipular fechas, Necesaria para parsear los timestamps en formato LDAP a dd, mm, yyyy – h, m, s.
- *os* [16]: Librería diseñada para interactuar con el sistema operativo y facilitar la portabilidad de programas entre plataformas. Necesaria para manipular ficheros y directorios (e.j `open`, `close`, `makedirs`, `getcwd` etc...).
- *sys* [17]: consiste en una librería por defecto de Python y sirve para ofrecer acceso a algunas variables y funciones del intérprete, por ejemplo, la funcionalidad de escape del programa (`sys.exit()`).
- *uuid* [18]: consiste en una librería para generar UUID (identificadores universalmente únicos) según la norma RFC 4122, lo cual permite nombrar aquellos ficheros cuyo nombre es desconocido de forma que no exista colisión con el resto.

- Imágenes

- *PIL* [19]: Python Imaging Library, también abreviada como Pillow, es una biblioteca diseñada en colaboración por Alex Clark orientada a abrir, manipular y guardar imágenes en distintos formatos. Necesaria para operar con las imágenes.
- *Deepface* [20]: es un framework ligero de reconocimiento facial y análisis de atributos faciales diseñado por Serengil, Sefik Ilkin and Ozpinar, Alper. Necesario para detectar y comparar rostros.
- *Cv2* [21]: consiste en una librería de OpenCV de Bradski, para Python. Permite manipular imágenes y transformar sus formatos.
- Consola
  - *argparse* [22]: librería diseñada por Steven J. Bethard, steven.bethard@gmail.com, para recoger y validar los argumentos de un programa. Facilita la recogida de argumentos opcionales y posicionales además de generar automáticamente el mensaje de ayuda (-h, --help).
  - *Termcolor* [23]: librería diseñada por Konstantin Lepa, konstantin.lepa@gmail.com, para imprimir en la terminal código ANSI coloreado. Facilita la comprensión de la salida por consola del programa (SMS enviados y recibidos verde, errores rojo, llamadas morado, etc...).
  - *Halo* [24]: librería diseñada por Manraj Singh, manrajsinghgrover@gmail.com, para generar símbolos de carga (espirales rotativas) en la terminal. Favorece la comprensión de la salida por pantalla al indicar al usuario que la ejecución continua e imprimiendo su resultado al concluir.
- Ficheros
  - *csv* [25]: consiste en una librería para leer, escribir o manipular ficheros csv. Permite exportar el contenido en formato separado por coma
  - *Sqllite3* [26]: consiste en una interfaz DB-API 2.0 diseñada por Gerhard Häring, gh@ghaering.de, para facilitar las interacciones con bases de datos

SQLite (versiones superiores a la 3.7.15). Necesaria para extraer los datos de las bases de datos.

- Matemática
  - *Numpy* [27]: librería diseñada para cálculo numérico y análisis de datos. Necesario para el cálculo matricial en el que se basa el procesamiento de imágenes.
- Expresiones regulares
  - *RE* [28]: librería de Secret Labs AB para manipular expresiones regulares basadas en strings de 8 bits o UNICODE. Necesaria para evitar discrepancias al enlazar los teléfonos de la agenda con los teléfonos de las llamadas, sms... (Ejemplo +34 123 45 67 89, +34123456789, 123456789...etc.).

### 3.4 Análisis del problema

Este apartado se compone de tres secciones, una dedicada a las características generales, otra al análisis dinámico y otra al estático.

#### 3.4.1 Aspectos generales de Your Phone

El servicio de Microsoft Your Phone se compone de dos partes, el programa para Windows Your Phone y la aplicación para Android Your Phone Companion. Desde la versión 1809.7 de Windows 10, Your Phone pasó a ser un programa por defecto, por lo que viene preinstalado en *C:\ProgramFiles\WindowsApps*, donde se encuentran una serie de directorios que identifican las aplicaciones estándar de Windows. Estas se catalogan según versión, arquitectura y proveedor, por lo que en este caso se tiene *Microsoft.YourPhone\_1.21121.256.0\_x64\_\_8wekyb3d8bbwe* donde *8wekyb3d8bbwe* es el PublisherId de Microsoft, que identifica todos los nombres de paquetes de sus aplicaciones. El programa se instala independientemente en cada usuario y guarda sus datos en *%Appdata%* concretamente en local, ya que está asociada a un único ordenador.

Los requisitos de instalación de la aplicación móvil y su versión para computador son:






TABLA 3.4.1 - PERMISOS DE YOUR PHONE			
	Tiene acceso a todas las líneas telefónicas del dispositivo		Se comunica con dispositivos Bluetooth ya emparejados
	Usa todos los recursos del sistema Administra otras aplicaciones directamente		Acceso a la conexión de Internet y a la red doméstica o de trabajo
	Detectar e iniciar aplicaciones en otros dispositivos en los que se ha iniciado sesión. Se cierra y cierra sus ventanas y retrasa el cierre		
	* Requiere como mínimo un sistema operativo Windows 10 versión 18362.0 o superior Xbox y una arquitectura ARM, ARM 64, x64 o x86		

Tabla 6 Permisos de Your Phone










TABLA 3.4.2 - PERMISOS DE YOUR PHONE COMPANION			
	Cámara: Realizar fotos y videos		Consultar agenda de contactos
	Leer SMS o MMS y enviar SMS		Consultar registro de llamadas, la identidad y estado del teléfono
	Ubicación: Acceso solo en primer plano		Acceder a ajustes Bluetooth y vincular dispositivos.
	Leer o modificar contenido de la tarjeta SD y almacenamiento compartido		Ver acceso a las conexiones de red y recibir datos de internet
	Ejecución en segundo plano, al inicio, aparecer sobre otras apps y recuperar aplicaciones en ejecución. Denegar optimización de batería y modo suspensión e inhabilitar bloqueo de pantalla. Permite establecer alarmas, leer notificaciones Api install Referrer de Play		
	* Las llamadas requieren conexión Bluetooth y un PC con Windows 10 y una versión posterior a mayo de 2019 ** La funcionalidad de arrastrar y suelta, pantalla de teléfono y aplicaciones requieren un dispositivo compatible		

Tabla 7 Permisos de Your Phone Companion

\* La autoría de todos los iconos utilizados en las tablas 6 y 7 pertenece al repositorio de libre distribución FlatIcon [33].

Según se ha podido observar, Your Phone requiere tanto permisos de conexión wifi/datos móviles como bluetooth. La conexión a internet se utiliza para sincronizar el contenido del teléfono con el ordenador (transferir imágenes, logs, notificaciones...). Por defecto la app Your Phone Companion desactiva el envío de datos ya que puede suponer gastos elevados al usuario. Esto se puede editar en los ajustes de la aplicación. Según Microsoft, pc y móvil han de estar conectados a la misma red, aunque posteriores pruebas demostraron que no es necesario. No obstante, para poder gestionar llamadas salientes y entrantes desde el ordenador se necesita vincular los dispositivos mediante una conexión bluetooth.



### 3.4.2 Análisis dinámico

Para estudiar el comportamiento del programa se han utilizado los servicios del paquete de aplicaciones sysinternals, concretamente uno llamado Procmon. Procmon es un programa de seguimiento de procesos muy utilizado en ciberseguridad que sirve para capturar lecturas y escrituras a ficheros o registros, eventos de Windows, creación y finalización de threads etc.... todo ello filtrando según el tipo de operación o valor y ajustando el periodo de monitorización. Paralelamente se ejecuta WireShark para atender las entradas y salidas de red. Posteriormente se utilizó Procdot, un servicio de diagramado para ilustrar gráficamente y así facilitar el estudio de las operaciones capturadas. Con ello se estudiaron los principales casos de transferencia de datos:

#### **Emparejar ordenador y móvil**

En un primer momento el contenido se encuentra exclusivamente en el teléfono, y para extraerlo es necesario iniciar sesión con una cuenta Outlook o MS tanto en móvil como en PC. A continuación, para emparejar los dispositivos el usuario debe vincularlos, o bien a través de un código enviado a su cuenta de correo, o mediante un código QR que se autogenera periódicamente. En el primero el usuario deberá introducir sus datos de inicio de sesión en el teléfono y validarse mediante el código de siete dígitos enviado a su dirección de correo electrónico. A continuación, como método de segunda verificación aparece en la pantalla del PC donde se inició sesión un código de nueve dígitos que el usuario deberá insertar en el teléfono. En el segundo, el PC crea un QR y Your Phone Companion lo escanea. Ya sea a través de uno u otros se transfieren por la red los datos del teléfono, y luego en el PC se crea una nueva carpeta llamada *Indexed* en el directorio `C:\Users\username\AppData\Local\Packages\Microsoft.YourPhone_8wekyb3d8bbwe` para almacenar los datos del dispositivo. Cada dispositivo tiene una carpeta con un uuid propio, con lo que se evitan colisiones de datos. A continuación, se muestra un ejemplo de los pasos descritos previamente, así como la estructura de carpetas una vez se ha vinculado uno o más dispositivos, por ejemplo, el Samsung A20e género el identificador `0857d319-bd80-495f-a54a-34c472f351aa` y Samsung A8 el `c2a273c1-0af6-487e-b4c3-6963db1d2caf`.

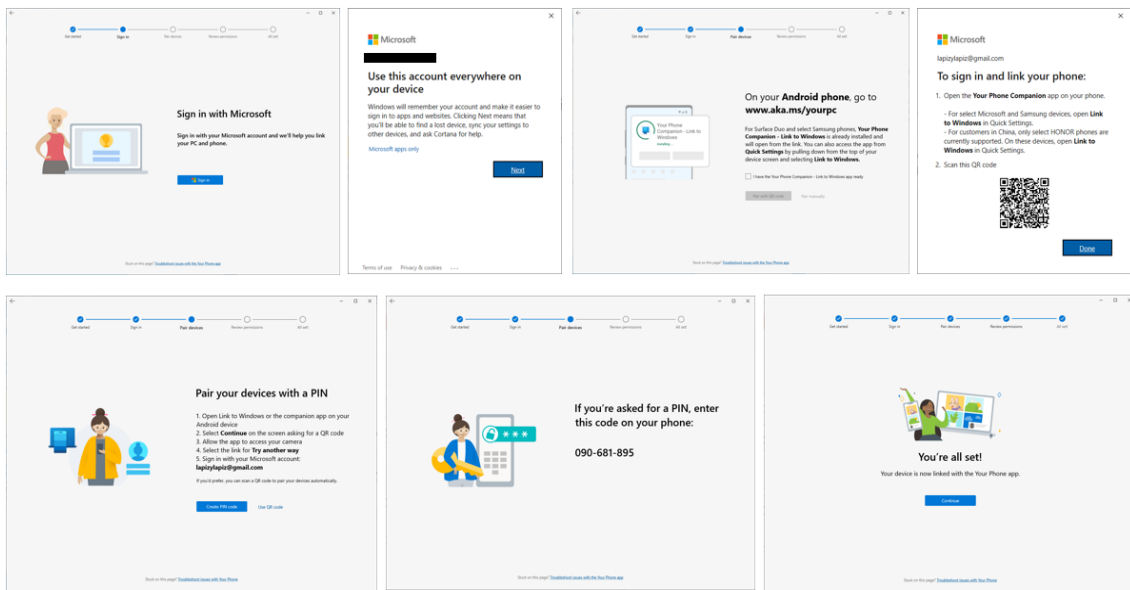


Fig. 2 Vinculando PC y móvil

El siguiente árbol ha sido generado con `tree /a /f` y ejecutado sobre el directorio `C:\Users\username\AppData\Local\Packages\Microsoft.YourPhone_8wekyb3d8bbwe` para ilustrar cómo queda la estructura de carpetas una vez se agregan los dos dispositivos Samsung citados en el ejemplo anterior. Se han agrupado los ficheros `*.db *.db-wal y *.db-shm` para facilitar su lectura.

```
+---AC
|   +---Temp
|   \---TokenBroker
|       \---Cache
+---AppData
+---LocalCache
|   |   YppCryptoTrustRelationships
|   |
|   +---Indexed
|   |   +---0857d319-bd80-495f-a54a-34c472f351aa
|   |   |   \---System
|   |   |       \---Database
|   |   |           *.db
|   |   |           *.db-shm
|   |   |           *.db-wal
|   |   |
|   |   |---c2a273c1-0af6-487e-b4c3-6963db1d2caf
|   |   |   \---System
|   |   |       \---Database
|   |   |           *.db
|   |   |           *.db-shm
|   |   |           *.db-wal
```

```

|      +---Local
|      |      \---Microsoft
|      |      \---Roaming
|      |          \---Microsoft
|      |              \---Windows
|      |                  \---Start Menu
|      |                      \---Programs
+---LocalState
+---RoamingState
+---Settings
|      roaming.lock
|      settings.dat
|
+---SystemAppData
|      \---Helium
|              User.dat
|              UserClasses.dat
|
\---TempState
      \---amplicons

```

Mediante ProcDot observamos como YourPhone.exe, comienza a leer y escribir en algunos registros y, al seleccionar el inicio de sesión, se lanza svchost.exe. Este inicia un enlace TCP (icp 4) dirigido a login.live.com, dirección 20.190.160.7 y más adelante TLSv1.2 inicia el intercambio de claves criptográficas para establecer una comunicación segura con el servidor. Esta valida usuario y contraseña e intercambia certificados con el PC. Algunos de los datos del correo electrónico de Microsoft con el que se está iniciando sesión en Windows se van almacenando en el registro de Windows (Figura 3).



Fig. 3 HKU\sid\SOFTWARE\Microsoft\IdentityCRL\UserExtendedProperties\email

Más adelante se genera en AC\TokenBroker\Cache\ un fichero con extensión .tbres (en este caso se trató de f22b36b69223efe531c87b229e75ddc69c1f0246.tbres.) Este tipo de extensión .tbres es un fichero de caché utilizado para administrar permisos de programas descargados desde Microsoft Store. Microsoft Windows TokenBroker se encarga de controlar que todas las aplicaciones cumplan con sus permisos y no realizan ningún tipo de acciones maliciosas. Según avanza la ejecución Your Phone guarda en el registro

\REGISTRY\A\{5472abfe-5b7e-89ff-d098-98132f6eed9d} \ LocalState\ Devices\ 0857d319-bd80-495f-a54a-34c472f351aa\ algunos de los datos del teléfono como DevicePlatform, FriendlyName, Capabilities, Locale, CountryIso, PhotosDeviceermission, ContentDevicePermission, ContactsPcPermission, MessagesPcPermission, NotificationsPcPermission, MessagesDevicePermission, MessageMMSSendPermission...etc. Luego se generan los principales artefactos del programa; ocho bases de datos sobre las que se escribe de manera recursiva. Para cada uno de ellos se crean también dos ficheros con extensión .db-wal y .db-shm. En el análisis se observa como los ficheros generados son inspeccionados por MsMpEng.exe, el principal proceso de la aplicación antimalware Windows Defender. Entre sus funciones está generar y almacenar en *C:\ProgramData\Microsoft\Windows Defender\Scans\mpenginedb.db-wal* hashes de la información del disco para así monitorizar actividad maliciosa.

## Desemparejar ordenador y móvil

También es posible desvincular los dispositivos vinculados en Your Phone, para ello simplemente hay que entrar en Your Phone, 'Mis dispositivos' y seleccionar la opción 'Eliminar' sobre el teléfono en cuestión. Al hacerlo la aplicación borra del listado el dispositivo, pero no elimina sus artefactos. Es por esto que incluso tras haber quitado un teléfono es posible encontrar su contenido en el PC, lo cual supone una oportunidad para que los analistas obtengan conocimiento valioso. El programa también advierte que se debe eliminar manualmente el enlace al PC del móvil (Figura 4) No se han estudiado artefactos en Android por lo que no se puede estimar si se genera o perdura algún tipo de rastro en el móvil.

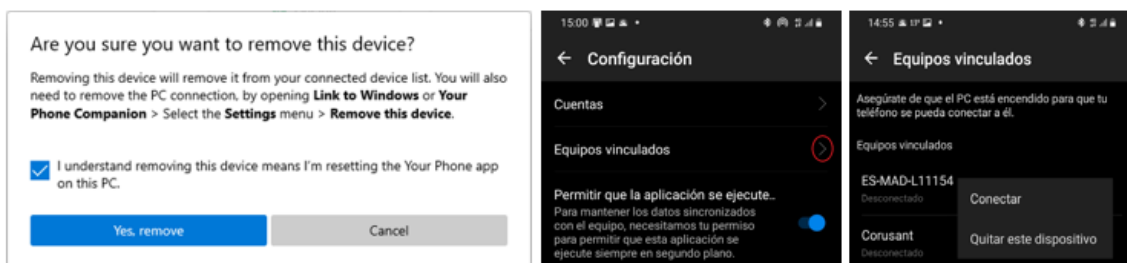


Fig. 4 Eliminar equipos vinculados. Izq Your Phone (Windows), dch Your Phone Companion (Android)

### 3.4.3 Análisis estático

Una vez se identificaron las trazas de Your Phone en el sistema, se procedió a estudiarlas individualmente. Para conocer la estructura interna de las bases de datos, sus tablas y registros almacenados, se ha utilizado DB Browser y DBeaver Lite. Estas herramientas permiten operar y visualizar cualquier formato, y particularmente SQLite, que es el de Your Phone.

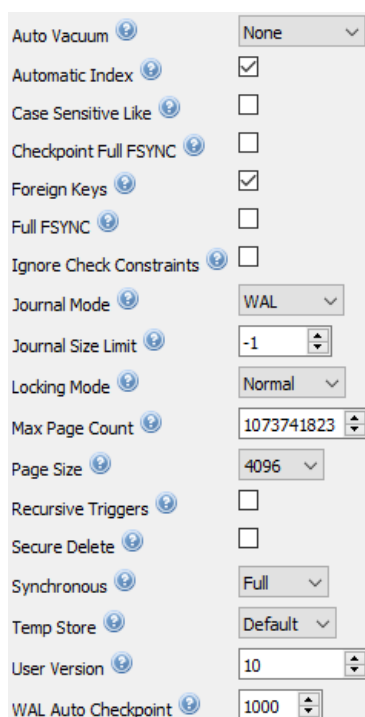


Fig. 5 Configuración de BBDD SQL

Su estudio reveló que las BBDD están dispuestas siguiendo la configuración de sentencias Pragma que muestra la figura 5 de la izquierda. Pragma es una extensión SQL específica para el formato SQLite que permite para alterar el funcionamiento de la librería, así como consultar sus metadatos. En este caso resulta relevante que la opción “Auto Vacuum” o auto-vaciado se encuentre inhabilitada dado que implica que la base de datos no reduce su tamaño tras aplicar operaciones de borrado, sino que las páginas del archivo db no utilizadas se añaden a una "lista libre" y se reutilizan para las siguientes inserciones. Asimismo, que “Secure Delete” esté también desactivada implica que las operaciones de borrado no sobrescriban el contenido con ceros. Finalmente está “Journal Mode” que toma la configuración Wal de entre Delete, Truncate Persist

Memory y Off. La sentencia pragma especifica que el tratamiento de conflictos en la BBDD ha de ser Write-Ahead Log, lo que explica por qué se generan dos ficheros .db-wal y .db-shm por cada archivo db.

Estos ficheros, son archivos temporales que se generan o actualizan cada vez que se abre una conexión y sirven para dar respaldo a las operaciones internas en caso de que se produzcan fallos. Concretamente los Write-Ahead Logs son registros/diarios pensados para conflictos de escritura en commits y rollbacks, mientras que los archivos de memoria compartida, shared memory file o shm se utilizan cuando dos o más conexiones comparten el mismo archivo db y deben actualizar la misma ubicación de memoria. Esta última sólo está presente cuando SQLite se ejecuta en modo WAL, como es el caso. El hecho es que al tener las BBDD configuradas en este modo se puede intentar recuperar

algunos registros mediante carving. UnDark es una herramienta que examina base de datos SQLite y vuelca en csv las filas de datos que encuentra intactas (tanto las actuales como las eliminadas) sin diferenciar entre los datos actuales y los borrados. Utilizando UnDark se pudo comprobar cómo es posible recuperar filas borradas, lo cual es muy positivo para la tarea de forense digital. No obstante, siendo pragmáticos, hay que reconocer que el alcance de estas metodologías es reducido ya que a menudo solo aparece información parcial o corrupta. De la misma forma tampoco se descarta que exista algún registro en unallocated space.

## Tablas SQL

Respecto a las tablas y su contenido se comprobó que:

En primer lugar, la agenda se almacena en *contacts.db*, donde reside la tabla principal *contact* que se relaciona a través de un identificador único *contact\_id* con otras tablas *phonenummer*, *postaladress*, *emailaddresss*, *contactDate*, *contactUrl*. La experimentación llevada a cabo indica que mayoritariamente estas se encuentran vacías, pero podría deberse a las condiciones de prueba (modelo de teléfono y versión de Android). De lo que si se dispone es de teléfonos, localizados en *phonenummer*. Además, se ha observado que para cada tabla expuesta anteriormente existe una virtual con prefijo 'fts' así como que la base de datos cuenta con tres disparadores (trigger AFTER {DELETE, UPDATE, INSERT} ON) por cada una de estas tablas. En un primer momento se pensó que su función era la de salvaguardar y actualizar el contenido a medida que la aplicación agrega o elimina contactos, pero resultó ser una hipótesis incorrecta. Se trata de Full-Text-Search, un método de búsqueda optimizado por el cual a través de tablas virtuales se puede examinar eficazmente una o más instancias de un término de búsqueda en la BDDD.

En segundo lugar, el registro de llamadas, que se encuentra en *call\_history* en *calling.db*. De aquí se averiguó que los únicos valores que ofrecen información útil son el teléfono establece la llamada, si esta fue entrante o saliente, si fue aceptada o declinada, su duración y fecha. Your Phone registra números enteros en estos campos por lo que la información requiere análisis y parseo. Para ello se realizaron pruebas con las distintas casuísticas, lo que permitió descifrar el significado de cada valor en los campos (Página

42). Así mismo, se pudo comprobar que las observaciones de las publicaciones con respecto a anteriores versiones permanecían al día en la actual. El resto de campos o bien se repiten por defecto o son directamente nulos.

En tercer lugar, tenemos *phone.db* donde se almacena la información del dispositivo. Por un lado, en *subscription*, se tiene el contrato de telecomunicaciones (teleoperadora, país, sim, características de la tarifa, condiciones de roaming, mensajes multimedia, rcs y limitaciones en el envío). Sobre la actividad relativa al registro de sms y mms todos ellos se asocian a un chat (*conversation*) mediante un identificador denominado *Thread\_id* donde los primeros se guardan en *messages* mientras que los segundos en *mms*. Al igual que sucedía con las llamadas, los campos relativos a mensajes (tiempo del envío, tipo, status) contienen valores aparentemente cifrados, que requirieron pruebas y análisis para parseo.

En cuarto lugar, *settings.db*, que contiene la información de las aplicaciones, instaladas y recientes, del dispositivo, así como *phone\_request*. Por desgracia las evidencias recogidas no muestran ninguna traza de apps recientes o *phone\_requests* por lo que no se puede determinar si en algún caso traerán alguna información relevante. Sin embargo, sí se puede de aplicaciones donde lo relevante es nombre y versión. El resto de campos no ofrece ningún valor adicional ya que conociendo la app siempre se puede obtener su id o el de su paquete, por ejemplo 2926454279753648442 com.whatsapp o 2534083964653685902 com.google.

En quinto lugar, *notifications.db*, que almacena los avisos que las aplicaciones envían al usuario del teléfono. En un primer momento se creyó que se trataría de una fuente muy valiosa de datos, pero pronto se desestimó ya que el análisis mostró que tan solo se almacena la cola de notificaciones del teléfono. Your Phone actualiza *notifications.db* y su pantalla cada vez que una aplicación envía un aviso, pero si el usuario las elimina o bien desde el teléfono o bien desde el PC esto se ve reflejado en la base de datos. Por ello tan solo queda registrado el último estado de la cola.

En sexto puesto, *photos.db* y *deviceData.db*, dos bases de datos que contienen la galería de imágenes (*media*, *photos*) y el fondo de pantalla del teléfono móvil (*wallpaper*). Complementando al estudio de P. Domingues, L. M. Andrade y M. Frade [5] se ha confirmado que la tabla *photos*, ahora obsoleta, era parte del funcionamiento de versiones

anteriores donde se almacenaban las últimas 25 imágenes del dispositivo. Ahora *media* almacena en sus registros hasta 2000 imágenes del teléfono. en forma de blob (cadena de bytes). Photos.db también agrega información útil como el timestamp de la última vez que se actualizó o abrió, su extensión, tamaño, altura y anchura, y su uri. Al estudiar Your Phone se observó que la galería del teléfono no se previsualiza completa, sino que el programa carga en sus registros las imágenes a medida que el usuario se desplaza por la galería. Por tanto, a la hora de estudiar una evidencia las imágenes almacenadas dependen del uso que haya hecho el usuario de Your Phone. Respecto a estas imágenes de la BBDD, se ha examinado la persistencia de metadatos (la configuración de la cámara utilizada al tomar las fotografías, la información sobre la fecha, la ubicación, y las miniaturas) mediante un visor en línea denominado Jeffrey Friedl's Image Metadata Viewer [29]. Este ha permitido comprobar que Your Phone no almacena la imagen original en sus registros, sino que guarda, por un lado, una previsualización (la miniatura o thumbnail del original) sin metadatos EXIF, y una copia de menor tamaño con una sección reducida de estos. En caso de que el usuario seleccione y realice una acción de guardado de una imagen de la galería se genera una copia idéntica a la original exceptuando timestamps.

**Metadatos:**

A continuación, se muestran los metadatos almacenados por la base de datos de Your Phone (thumbnail y media) en comparación con el fichero original.

**Thumbnail**

File:	384 × 512 JPEG \$3,156 bytes (81 kilobytes)
Color Encoding:	Embedded color profile: "sRGB"

**JFIF**

JFIF Version	1.01
Resolution	1 pixels/None

**File** — basic information derived from the file.

File Type	JPEG
File Type Extension	.jpg
MIME Type	image/jpeg
Encoding Process	Baseline DCT, Huffman coding
Bits Per Sample	8
Color Components	3
File Size	81 kB
Image Size	384 × 512
YCbCr Sub Sampling	YCbCr4:2:0 (2 2)

**Composite**  
This block of data is computed based upon other items. Some of it may be wildly incorrect, especially if the image has been resized.

Megapixels 0.197

**ICC\_Profile** — this block of data describes the color space used to encode pixel colors.  
[\[ click to show profile data \]](#)

Fig. 6 Metadatos de la imagen almacenada en thumbnails



## Fichero original

Camera:	samsung SM-A202F
Lens:	1.1 mm (Max aperture f2.2) (shot wide open)
Exposure:	Auto exposure, Program AE, 1 sec, f/2.2, ISO 320
Flash:	none
Date:	<b>July 19, 2021 1:12:33PM</b> (timezone not specified) (4 months, 27 days, 11 hours, 31 minutes, 46 seconds ago, assuming image timezone of US Pacific)
File:	<b>1,932 × 2,576 JPEG (5.0 megapixels)</b> 1,401,658 bytes (1.3 megabytes)
Color Encoding:	<b>WARNING:</b> Color space tagged as sRGB, without an embedded color profile. Windows and Mac browsers and apps treat the colors randomly. Images for the web are most widely viewable when in the sRGB color space and with an embedded color profile. See my <a href="#">Introduction to Digital Image Color Spaces</a> for more information.

EXIF		MakerNotes	
Make	samsung	Samsung Trailer (Initial) Name	Image_UTC_Data
Camera Model Name	SM-A202F	Time Stamp	2021:07:19 13:12:33.00 4 months, 27 days, 11 hours, 31 minutes, 46 seconds ago
Orientation	Rotate 90 CW	Samsung Trailer (Initial) Name	MCC_Data
Software	A202FXXU1CUE9	Samsung Trailer (Initial)	(3 bytes binary data)
Modify Date	2021:07:19 13:12:33 4 months, 27 days, 11 hours, 31 minutes, 46 seconds ago	Samsung Trailer (Initial) Name	LiveVideo_PhotoDate_Data
Y Cb Cr Positioning	Centered	Samsung Trailer (Initial)	(13 bytes binary data)
Exposure Time	1/24	File — basic information derived from the file.	
F Number	2.20	File Type	JPEG
Exposure Program	Program AE	MIME Type	image/jpeg
ISO	320	Exif Byte Order	Little-endian (Intel, II)
Exif Version	0220	Encoding Process	Baseline DCT, Huffman coding
Date/Time Original	2021:07:19 13:12:33 4 months, 27 days, 11 hours, 31 minutes, 46 seconds ago	Bits Per Sample	8
Create Date	2021:07:19 13:12:33 4 months, 27 days, 11 hours, 31 minutes, 46 seconds ago	Color Components	3
Offset Time	+02:00	File Size	1389 kB
Other Time Original	+02:00	File Type Extension	jpg
Shutter Speed Value	1	Image Size	2,576 × 1,932
Aperture Value	2.20	Y Cb Cr Sub Sampling	YCbCr4:2:0 (2,2)
Exposure Compensation	0	Composite	
Max Aperture Value	2.2	This block of data is computed based upon other items. Some of it may be wildly incorrect.	
Metering Mode	Center-weighted average	Aperture	2.20
Flash	No Flash	Megapixels	5.0
Focal Length	1.1 mm	Scale Factor To 35 mm Equivalent	11.3
Color Space	sRGB	Shutter Speed	1/24
Exposure Mode	Auto	Date/Time Original	2021:07:19 13:12:33.00 4 months, 27 days, 11 hours, 31 minutes, 46 seconds ago
Digital Zoom Ratio	1	Modify Date	2021:07:19 13:12:33.00 4 months, 27 days, 11 hours, 31 minutes, 46 seconds ago
Focal Length In 35mm Format	11 mm	Circle Of Confusion	0.003 mm
Image Size	2,576 × 1,932	Field Of View	193.3 deg
White Balance	Auto	Focal Length	1.1 mm (35 mm equivalent: 13.0 mm)
Scene Capture Type	Standard	Hyperfocal Distance	0.22 m
Image Unique ID	2018F3A60N04	Light Value	5.2
Image Width	192	ExifTool	
Image Height	144	Warning: [samsung] Unknown APP1 segment	
Compression	JPEG (old-style)		
Resolution	72 pixels/inch		
Thumbnail Length	43,180		
Thumbnail Image	(43,180 bytes binary data)		

Fig. 8 Metadatos de la imagen original

## Media

Camera:	samsung SM-A202F
Date:	<b>July 19, 2021 1:12:33PM</b> (timezone not specified) (4 months, 27 days, 11 hours, 24 minutes, 1 second ago, assuming image timezone of US Pacific)
File:	<b>1,382 × 1,842 JPEG (2.5 megapixels)</b> 879,200 bytes (0.84 megabytes)
Color Encoding:	Embedded color profile: "sRGB"

EXIF	
Image Size	1,842 × 1,382
Camera Model Name	SM-A202F
Make	samsung
Date/Time Original	2021:07:19 13:12:33 4 months, 27 days, 11 hours, 24 minutes, 1 second ago
Light Source	Unknown
Orientation	Rotate 90 CW
Modify Date	2021:07:19 13:12:33 4 months, 27 days, 11 hours, 24 minutes, 1 second ago

JFIF	
JFIF Version	1.01
Resolution	1 pixels/None

File — basic information derived from the file.

File Type	JPEG
MIME Type	image/jpeg
Exif Byte Order	Big-endian (Motorola, MM)
Encoding Process	Baseline DCT, Huffman coding
Bits Per Sample	8
Color Components	3
File Size	859 kB
File Type Extension	jpg
Image Size	1,842 × 1,382
Y Cb Cr Sub Sampling	YCbCr4:2:0 (2,2)

Composite  
This block of data is computed based upon other items. Some of it may be wildly incorrect, especially if the image has been resized.

Megapixels	2.5
------------	-----

ICC\_Profile — this block of data describes the color space used to encode pixel colors.

[[click to show profile data](#)]

Fig. 7 Metadatos de la imagen almacenada en media

## Bases de datos de Your Phone

### SettingsDB

phone_apps	recent_apps
123 app_id	123 app_id
ABC app_name	ABC package_name
ABC package_name	ABC app_name
ABC version	123 checksum
123 etag	123 recent_app_rank
123 favorite_rank	ABC intent_action
blob	123 task_id
123 last_updated_time	blob
123 checksum	ABC web_uri
	ABC desktop_uri
settings	content_sequence
ABC setting_group_id	123 content_sequence_id
ABC setting_key	ABC content_type
123 setting_type	123 sequence
ABC setting_value	

### PhotosDB

media	content_sequence
123 id	123 content_sequence_id
ABC name	ABC content_type
123 last_updated_time	123 sequence
123 taken_time	
ABC orientation	
123 last_seen_time	
ABC mime_type	
123 height	
123 width	
123 size	
ABC uri	
thumbnail	
media	
123 checksum	

Fig. 9 Tablas de settings.db y photos.db

## deviceDataDB

content_sequence	wallpaper
123 content_sequence_id	123 wallpaper_type
ABC content_type	123 wallpaper_id
123 sequence	blob

## sharedContentDB

sharedcontent	shareduri
123 data_id	123 uri_id
123 group_id	ABC original_uri
123 item_id	123 shared_time
ABC mime_type	123 type
123 size	123 is_deleted
123 timestamp	ABC uri
data	ABC title
alt_data	ABC description
ABC name	image
123 custom_type	icon

## notificationsDB

notifications
123 id
ABC notification_id
ABC package_name
ABC json
123 post_time
123 state
ABC anonymous_id

## callingDB

content_sequence	call_history
123 content_sequence_id	123 call_id
ABC content_type	ABC phone_number
123 sequence	123 duration
	123 call_type
	123 start_time
	123 is_read
	ABC phone_account_id
	123 last_updated_time

content_sequence
123 content_sequence_id
ABC content_type
123 sequence

Fig. 10 Tablas de deviceDataDB, sharedContentDB, notificationsDB y callingDB

## ContactsDB

fts_contact_config	fts_contact_data	fts_contact_docsize	fts_contactdate	fts_contactdate_config	fts_contactdate_data	fts_contact
123 k	123 id	123 id	123 contact_id	123 k	123 id	123 contact_id
123 v	blob	sz	123 display_date	123 v	blob	123 display_name
fts_emailaddress_docsize	fts_phonenumber	fts_phonenumber_data	fts_phonenumber_docsize	fts_postaladdress_data		123 nickname
123 id	123 contact_id	123 id	123 id	123 id		123 company
sz	123 display_phone_number	blob	sz	blob		123 job_title
fts_postaladdress_config	fts_phonenumber_config	fts_contacturl_docsize	fts_contactdate_docsize	fts_postaladdress_docsize	fts_phonenumber_idx	123 notes
123 k	123 k	123 id	123 id	123 id	123 segid	
123 v	123 v	sz	sz	sz	123 term	
fts_contacturl_config	fts_contacturl_data	fts_emailaddress	fts_emailaddress_config	fts_contacturl	fts_emailaddress_data	123 pgno
123 k	123 id	123 contact_id	123 k	123 contact_id	123 id	
123 v	blob	123 address	123 v	123 url_address	blob	fts_contacturl_idx
content_sequence	fts_contact_idx	fts_contactdate_idx	fts_emailaddress_idx	contact	postaladdress	123 segid
123 content_sequence_id	123 segid	123 segid	123 segid	123 contact_id	123 postal_address_ic	123 term
ABC content_type	123 term	123 term	123 term	ABC display_name	123 contact_id	123 pgno
123 sequence	123 pgno	123 pgno	123 pgno	ABC nickname	123 type	
				ABC last_updated_time	ABC label	
				ABC thumbnail	ABC street	
				ABC company	ABC city	
				ABC job_title	ABC region	
				ABC notes	ABC postal_code	
				ABC name_prefix	ABC country_code	
				ABC name_suffix	ABC sub_administrative_area	
				ABC given_name	ABC display_address	
				ABC middle_name	123 checksum	
				ABC family_name		
				123 checksum		
fts_postaladdress	contactdate	contacturl	emailaddress			phonenumber
123 contact_id	123 contact_date_id	123 contact_url_id	123 email_address_id			123 phone_number_id
123 street	123 contact_id	123 contact_id	123 contact_id			123 contact_id
123 city	123 date_type	123 type	123 type			ABC phone_number
123 region	ABC label	ABC label	ABC label			ABC display_phone_number
123 postal_code	ABC display_date	ABC url_address	ABC address			123 phone_number_type
123 country_code	123 checksum	123 checksum	123 checksum			ABC label
123 sub_administrative_area						123 checksum
123 sub_locality						

Fig. 11 Tablas de Contacts.db

## PhoneDB

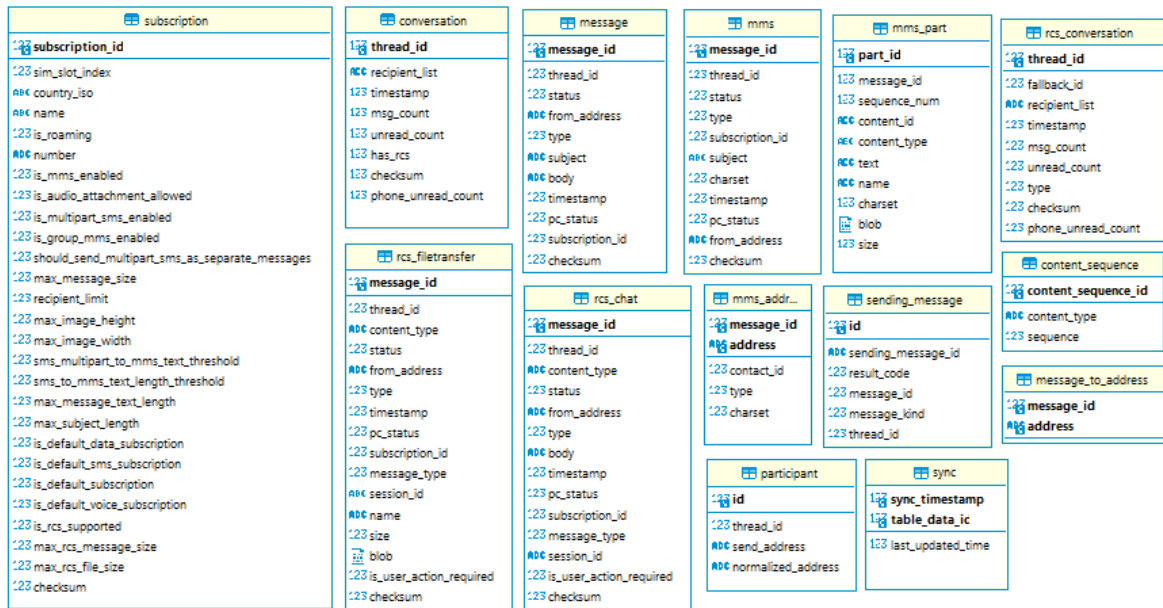


Fig. 12 Tablas de phone.db

### 3.5 Diseño de la solución

Una vez conocida la forma en que el programa almacena la información se procedió a desarrollar un script que permitiese extraerla y estudiarla. Para ello, primero se organizó la funcionalidad y luego la estructura del programa. La figura 13 ilustra los componentes que lo conforman mientras que la figura 14 revela cómo se organizan las clases. Por último, la figura 15 muestra cómo fluye la información desde el origen (aplicación de Android), al ordenador (programa de Windows), hasta el script de análisis forense. Es decir, el diagrama de componente y clases exponen la arquitectura software y sus dependencias y el diagrama de flujo indica cómo la información del usuario viaja entre la app y el programa de Your Phone, además de cómo está se recoge en el script para analizar su contenido.

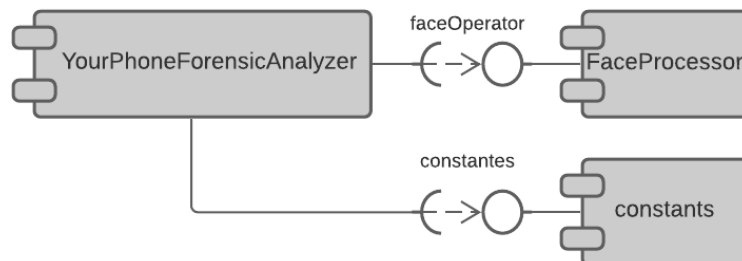


Fig. 13 Diagrama de componentes de YourPhoneForensicAnalyzer

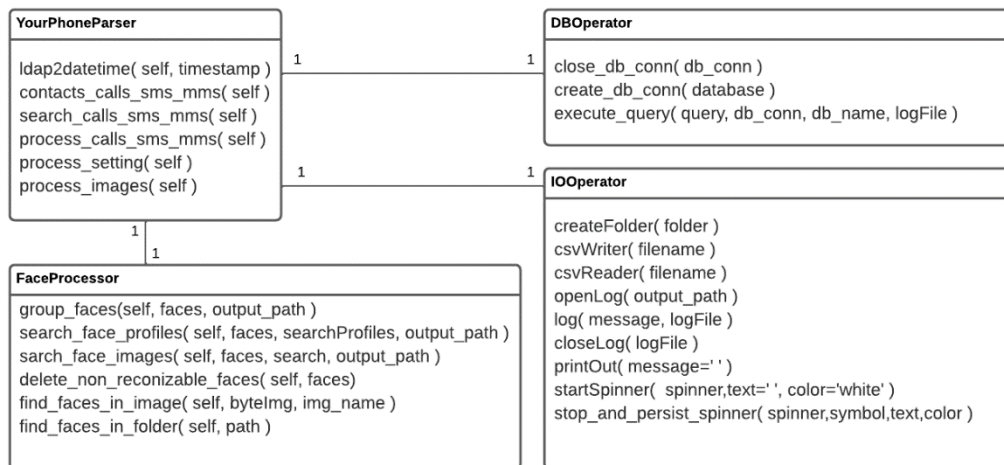


Fig. 14 Diagrama de clases

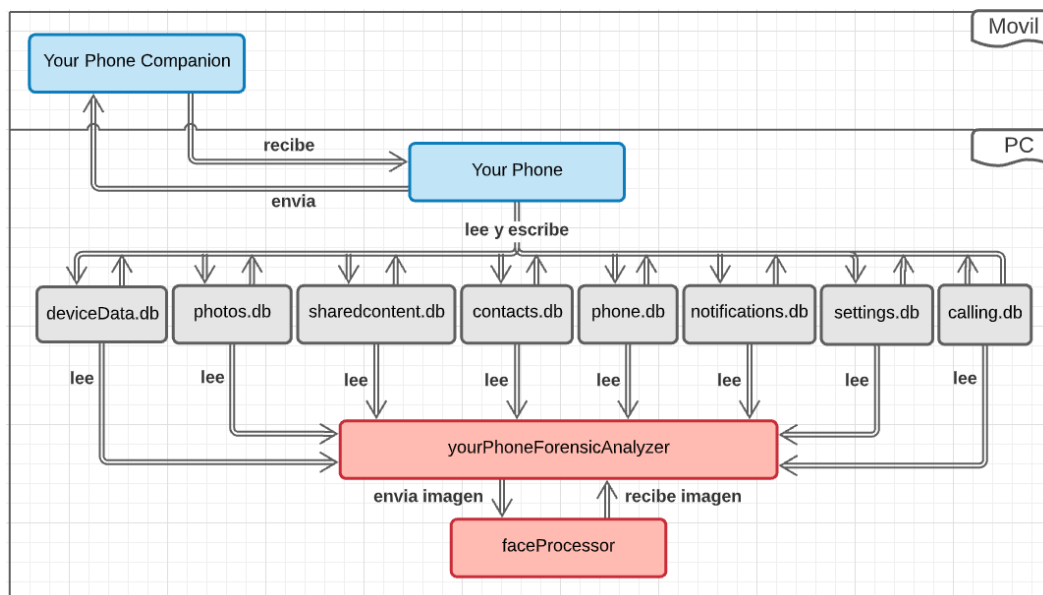


Fig. 15 Diagrama de flujo de información entre aplicación, programa y analizador

Una vez quedó comprendido el flujo de datos y se organizó la estructura de clases y componentes el siguiente paso consistió en desarrollar el código del programa. Esta etapa se dividió en dos fases; por un lado, la extracción y parseo de datos de índole numérica o textual y por otro el de obtención y reconocimiento facial en imágenes.

Para transformar los campos se estudió el formato y la casuística con la que se presentaban llamadas, mensajes y notificaciones.

#### Llamadas de teléfono

- Teléfono: Casa 1, Móvil 2, Trabajo 3, trabajo 4, Principal 5, Otros 6 y Escuela 7
- Tipo: Entrante 1, Saliente 2, Perdida 3, Desconocido 4, Declinada 5 y Bloqueada 6
- Estado: Leído 1, No leído 2

#### Sms – mms

- Estado: Leído 1, No leído 2
- Estado móvil: Recibido 1, No recibido 2

#### Timestamps

- Windows NT Time format: Utiliza el formato de tiempo Win32 filetime también llamado systemtime. Son marcas de tiempo de 18 dígitos que corresponden al número de intervalos de 100 nanosegundos desde el 1 de enero de 1601 UTC.

#### Notificaciones (JSON format)

- {"id": number,
- "key": string,
- "groupKey": string,
- "tag": string,
- "packageName": string,
- "appName": string,
- "isClearable": true/false,
- "isGroup": true/false,
- "isOngoing": true/false,
- "featureFlags": number,
- "platform": number,
- "version": number,
- "category": string
- "tickerText": string,

- "flags": number,
- "eventCount": number,
- "priority": number,
- "postTime": Windows NT time format,
- "timestamp": Windows NT time format,
- "notificationClass": number,
- "notificationActions": [
  - {
    - "actionName": String,
    - "isActionInlineReply": true/false,
    - "actionIndex": number},
    - "template": string,
    - "text": string,
    - "title" string:
    - "showWhen": true/false,
    - "messages": string,
    - "senderNames" string:
    - "importance" string:

### 3.5.1 Reconocimiento facial

El tratamiento de las imágenes se basa en dos modelos, OpenCV [30] y DeepFace [20]. El primero se utiliza para detectar rostros humanos, aislarlos y recortar un rectángulo alrededor suyo. El segundo se utiliza para aplicar reconocimiento facial sobre los rostros obtenidos. ¿Por qué se combinan ambos? Si bien es cierto que ambas librerías proporcionan por separado los medios para lograr las tareas de clasificación y comparación, existían ciertas limitaciones a la hora de implementarlas. Para OpenCV se observó que el modelo importado resultaba menos preciso a la par que menos versátil que DeepFace. Esto provocaba que, a la hora de extraer las características identificativas de los rostros (encodings de 128 bits) y comparar mediante distancia euclídea, se obtuviese una mayor métrica del error. OpenCV utiliza el algoritmo haar-cascade que no se basa en

técnicas de aprendizaje profundo por lo que es más rápido, pero su rendimiento es relativamente bajo. En contraposición DeepFace permite corregir este error ajustando el modelo (VGG-Face, Facenet, Facenet512, OpenFace, DeepFace, DeepID, ArcFace, Dli) y la distancia a usar (coseno, euclídea, euclidean\_l2) por lo que resultaba más beneficioso utilizarlo como comparador. No obstante, a diferencia de DeepFace, OpenCV es capaz de aislar múltiples rostros dentro de una imagen, lo que permite separar las caras. Combinando ambas alternativas se logra identificar, aislar, recortar y evaluar los rostros de las imágenes presentes en Your Phone. Al tratarse de dos modelos trabajando conjuntamente puede ocurrir que primero, OpenCV, detecte y recorte un rostro, y después DeepFace no reconozca ninguna cara. En estos casos se descarta el recorte ya que, una imagen sin rostro reconocible es una imagen sobre la que no se pueden realizar comparaciones.

Sobre el reconocimiento facial se tomó la decisión de ofrecer tres servicios al analista:

- Agrupador de rostros.

Consiste en buscar y juntar rostros similares a partir del set de imágenes extraídas, con el objetivo de facilitar el estudio de los individuos identificados. Con ello se busca agrupar los rostros según individuo, es decir, juntar los rostros más similares. Para hacerlo se compara cada rostro con las caras detectadas, si supera el umbral de similitud se escoge, exporta y elimina del listado. El algoritmo se repite hasta haber vaciado la lista de caras a agrupar.

- Comparador de rostros.

Consiste en encontrar a partir de una cara rostros similares dentro del set de imágenes extraídas. La diferencia con el agrupador de rostros es que la tarea no trata de juntar individuos parecidos entre sí a base de excluir rostros ya evaluados, sino que busca aquellos individuos cuya similitud sea superior a un determinado umbral sin realizar descartes.

- Buscador de perfiles faciales.

Consiste en buscar y extraer aquellos rostros que encajen con una determinada descripción basada en edad, género, gesto y raza de un sujeto. Permite concatenar características para afinar las búsquedas. El programa localizará todos los rostros

reconocibles y obtendrá un perfil aproximado para cada uno. La búsqueda devuelve todos los rostros coincidentes con los perfiles generados para cada perfil proporcionado.

La siguiente tabla recoge las variables que componen los CSVs de las búsquedas:

TABLA 3.5.1 - FICHEROS CSV		
Búsqueda	Campo	Valor
Perfiles faciales	Teléfonos	Número
		Formato E.164 de la UIT [31]
	Signo	Comparador <, <=, ==, !=, >=, o null (Null) para ignorar
	Edad	Entero o null (Null) para ignorar
	Género	Hombre (Man), mujer (Woman) o null (Null) para ignorar
	Gesto	Enfado (angry), miedo (fear), neutral (neutral), tristeza (sad), asco (disgust), felicidad (happy) y sorpresa (surprise) o null (Null) para ignorar
	Raza	Asiático (asian), Blanco (white), árabe (middle eastern), indio (indian), latino (latino), negro (black) o null (Null) para ignorar

*Tabla 8 Ficheros CSV para buscar en YourPhoneForensicAnalyzer*

### 3.6 Implementación de la solución

El primer paso hacia la implementación de YourPhoneForensicAnalyzer consistió en desarrollar el control de ejecución. Como mínimo el script exige facilitar la ruta al directorio que contiene las bases de datos (–i o –input entrada). De no hacerlo o en caso de suministrar una dirección incorrecta se aborta la ejecución. Si se desea también se puede añadir una ruta para dirigir su salida al directorio que se desee (–o o –output salida). Si estos son correctos el programa continúa evaluando el resto de modos de ejecución proporcionados por el analista. Estos son:

- Modos sin parámetros de entrada.
  - Modo ayuda: -h o –help
  - Modo verbose: -v o –verbose
  - Modo exportar imágenes: -e o –export
  - Modo agrupar caras similares: -gfi o –groupFaceImages



- Modos con parámetros de entrada.
  - Modo buscar mediante caras: `-sfi --searchFaceImages rutaImágenes`
    - *rutaImágenes*: directorio con la o las imágenes a buscar (cada imagen se considera como una búsqueda). Si el directorio no existe se notifica y aborta la ejecución.
  - Modo buscar mediante perfiles faciales: `-sfp --searchFaceProfiles rutaCSV`
    - *rutaCSV*: fichero csv con el o los perfiles a buscar. Cada fila del csv se considera como una búsqueda. Si se proporciona una ruta incorrecta o un fichero con un formato incorrecto se finaliza la ejecución lanzando un mensaje de error.
  - Modo buscar mediante teléfonos: `-spn --searchPhoneNumbers rutaCSV`
    - *rutaCSV*: ficheros csv con el o los teléfonos a buscar. Cada fila del csv se considera como una búsqueda. Si se proporciona una ruta incorrecta o un fichero con formato incorrecto se finaliza la ejecución lanzando un mensaje de error.

Tal y como se muestra en la figura 16, una vez se validan los parámetros de ejecución se procede a explorar las bases de datos en busca de la información del usuario. Lo primero que hace el algoritmo es preparar la base de queries o consultas sql que obtienen los registros de las bases de datos. Todas las consultas a una misma tabla comparten un cuerpo común y constante (`SELECT atributos FROM tabla`), por lo que esta parte de la consulta se guarda en la clase constants mientras que la parte variable (`WHERE condición`) se va modificando a lo largo del programa. Si se ha establecido uno o varios modos de búsqueda, el sondeo se limitará a los criterios de búsqueda proporcionados, es decir, no se realizará un rastreo en amplitud, sino que se impondrá el criterio de búsqueda sobre las instancias de la base de datos que cumplan con los parámetros proporcionados.

El siguiente esquema ilustra las etapas que sigue el algoritmo del programa:

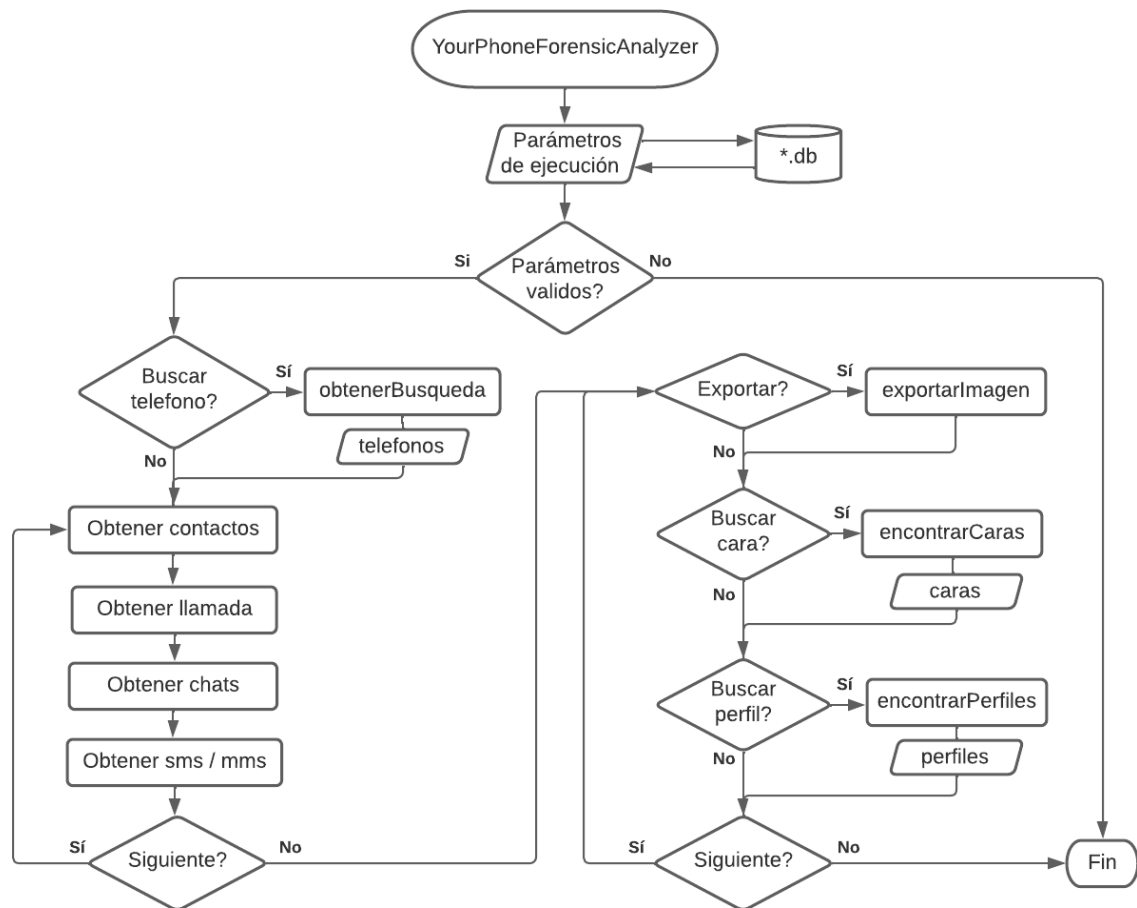


Fig. 16 Flujo de ejecución de YourPhoneForensicAnalyzer

### 3.6.1 Contactos, teléfonos, llamadas, sms y mms

El algoritmo comienza por la agenda de contactos, de donde extrae el identificador de contacto. Para cada una de las entradas de la BBDD se buscan los teléfonos asociados al identificador y con estos se explora el registro de llamadas y a su vez se rastrean las conversaciones. Cuando un teléfono aparece en el registro de llamadas se muestra por pantalla sus características, duración, tipo, leído o no leído, fecha de comienzo y actualización e id del teléfono. Similarmente, si en un chat interviene uno de los teléfonos se muestra el número total de mensajes que colecciona, cuantos no han sido leídos desde el PC y cuantos no han sido leídos desde el móvil, si rcs está soportado, y el timestamp del último mensaje. Para facilitar la visualización de la salida sobre las llamadas se utiliza el color morado y sobre el chat se implementa un código de colores para cada mensaje, siendo verdes salientes y azules entrantes. Cada mensaje contiene quien lo envía, cuando

lo envía y de nuevo cuantos no han sido leídos desde el PC y cuantos no han sido leídos desde el móvil. Cabe agregar que no todos los mensajes de la BBDD tienen relación con los teléfonos agregados a la agenda, algunos pueden ser conversaciones con teléfonos desconocidos, mensajes automáticos, spam, etc... Todos ellos se agrupan en una salida denominada “Unnasociated phones, calls, sms and mms” es decir, teléfonos, llamadas, sms y mms sin asociar. Por último, se muestran las aplicaciones instaladas en el dispositivo, nombre, versión, orden de favoritos, y última fecha de actualización. Al finalizar se muestra un reporte de todas las operaciones. Estos son algunos ejemplos:

```

? Search: +34[REDACTED]
['Joseluis Gonzalez Prieto', '', '', '', '', '', '', '', '2022-02-13 11:07:01']
['+34[REDACTED]', '+34[REDACTED]', 'Mobile phone number']
Call -> ['676065588', 310, 'Incoming', 'Taken', '2022-02-08 20:05:33', '2022-02-13 11:38:16', '8934222007210679220F']
Call -> ['676065588', 0, 'Declined', 'Taken', '2022-02-12 17:39:52', '2022-02-13 11:38:16', '8934222007210679220F']
Call -> ['+34676065588', 120, 'Outgoing', 'Taken', '2022-02-12 17:42:21', '2022-02-13 11:38:16', '8934222007210679220F']
Call -> ['+34676065588', 27, 'Outgoing', 'Taken', '2022-02-12 19:54:59', '2022-02-13 11:38:16', '8934222007210679220F']
Call -> ['676065588', 0, 'Missed', 'Taken', '2022-02-13 13:52:09', '2022-02-13 13:53:11', '8934222007210679220F']
Conv -> ['+34[REDACTED]', 5, 0, 0, 0, '2022-02-13 13:55:46']
+ ['2022-02-13 13:53:20', 'Read', 'Read', 'Hola']
+ ['2022-02-13 13:53:52', 'Read', 'Read', 'Que tal?']
+ ['2022-02-13 13:54:47', 'Read', 'Read', 'Bien. Trabajando en el TFG?']
+ ['2022-02-13 13:55:38', 'Read', 'Read', 'Si. Analisis forense de la aplicación Microsoft Your Phone']
+ ['2022-02-13 13:55:46', 'Read', 'Read', 'Animo!']
Conv -> ['+34[REDACTED]', '+34[REDACTED]', '+34[REDACTED]', 6, 0, 0, 0, '2022-02-13 13:57:55']
+ ['2022-02-13 13:57:31', 'Read', 'Read', 'Este es un mensaje grupal']
+ ['2022-02-13 13:57:33', 'Read', 'Read', 'Este es un mensaje grupal']
+ ['2022-02-13 13:57:34', 'Read', 'Read', 'Este es un mensaje grupal']
+ ['2022-02-13 13:57:51', 'Read', 'Read', 'Se envia mediante una lista de distribución']
+ ['2022-02-13 13:57:53', 'Read', 'Read', 'Se envia mediante una lista de distribución']
+ ['2022-02-13 13:57:55', 'Read', 'Read', 'Se envia mediante una lista de distribución']

```

Fig. 17 Llamadas y mensajes asociados a un contacto

```

Conv -> ['CORPO', 0, 0, 0, 0, '1601-01-01 00:00:00']
Conv -> ['DEGIRO', 0, 0, 0, 0, '1601-01-01 00:00:00']
Conv -> ['Finizem', 1, 0, 0, 1, '2022-02-10 23:16:41']
+ ['2022-02-10 23:16:41', 'Unread', 'Unread', 'Tu código de verificación para Finizem es 2532']
Conv -> ['Signaturit', 0, 0, 0, 0, '1601-01-01 00:00:00']
Conv -> ['SegSocial', 2, 0, 0, 0, '2022-01-16 21:28:07']
+ ['2022-01-16 21:21:18', 'Read', 'Read', '720992 es tu código de seguridad para acceder a los servicios electrónicos de la Seguridad Social']
+ ['2022-01-16 21:28:07', 'Read', 'Read', '757573 es tu código de seguridad para acceder a los servicios electrónicos de la Seguridad Social']
Conv -> ['SEUR', 0, 0, 0, 0, '1601-01-01 00:00:00']
Conv -> ['SEUR', 0, 0, 0, 0, '1601-01-01 00:00:00']
Conv -> ['TheFork', 0, 0, 0, 0, '1601-01-01 00:00:00']
Conv -> ['SPRINGFIELD', 0, 0, 0, 0, '1601-01-01 00:00:00']
Conv -> ['SEUR', 0, 0, 0, 0, '1601-01-01 00:00:00']
Call -> ['918561309', 332, 'Outgoing', 'Taken', '2022-02-10 14:20:33', '2022-02-11 22:26:27', '8934222007210679220F']
Call -> ['916240606', 4, 'Outgoing', 'Taken', '2022-02-11 12:34:22', '2022-02-11 12:53:21', '8934222007210679220F']
Call -> ['916240606', 3, 'Outgoing', 'Taken', '2022-02-11 12:51:38', '2022-02-11 12:53:21', '8934222007210679220F']
Call -> ['918561309', 0, 'Outgoing', 'Taken', '2022-02-11 13:02:43', '2022-02-11 22:26:27', '8934222007210679220F']
Call -> ['918561351', 16, 'Outgoing', 'Taken', '2022-02-11 15:38:08', '2022-02-11 22:25:43', '8934222007210679220F']

```

Fig. 18 Llamadas y mensajes sin asociar

```

Parsing settings:
['YouTube', '17.04.35', -1, '1601-01-02 21:40:39']
['Radio', '12.0.00.46', -1, '1601-01-02 21:31:27']
['Madrid MBC', '6.9.0.gab62', -1, '1601-01-02 21:39:06']
['Kit herramientas SIM', '11', -1, '1601-01-02 21:30:37']
['Contactos', '12.7.05.26', -1, '1601-01-02 21:31:27']
['Internet', '16.0.6.23', -1, '1601-01-02 21:39:00']
['Play Store', '29.1.10-21 [0] [PR] 425080933', -1, '1601-01-02 21:39:57']
['Calendario', '2022.02.0-420616974-release', -1, '1601-01-02 21:39:00']
['Gmail', '2022.01.09.423451891-release', -1, '1601-01-02 21:39:06']
['LinkedIn', '4.1.664', -1, '1601-01-02 21:39:57']
['OneDrive', '6.47.1', -1, '1601-01-02 21:39:00']
['Compañero de Tu Teléfono', '1.21121.354.0', -1, '1601-01-02 21:39:06']
['Samsung Pay', '4.1.88', -1, '1601-01-02 21:39:00']
['Drive', '2.22.017.4.90', -1, '1601-01-02 21:40:39']
['Escáner QR', '2.7.5-L', -1, '1601-01-02 21:30:37']
['Google', '13.2.19.23.arm64', -1, '1601-01-02 21:39:00']

```

Fig. 19 Aplicaciones instaladas

Cabe mencionar que los timestamps de las aplicaciones instaladas de las evidencias evaluadas presentan horas, minutos y segundos propias, pero traen el mes y día por defecto de Windows NT Timestamps. (Fig. 19)

### 3.6.2 Tratamiento de imágenes

Una vez concluido el parseo se obtiene la galería de imágenes de photos.db. Para ello, el algoritmo recorre todas las filas de la tabla *Media* y evalúa si los campos thumbnail y media se encuentran vacíos. El primero almacena una miniatura y el segundo una versión reducida del original, por lo que siempre que exista se preferirá la extracción del segundo ya que mantiene una resolución superior. Asimismo, se extrae el fondo de pantalla que guarda deviceData.db. Aquí cabe mencionar que la tabla de la que procede, wallpaper, no guarda ninguna información de la imagen original, por lo que el script le genera y adjudica un uuid como nombre y analiza la firma de archivo (file signature) de la cadena de bytes para adivinar su extensión.

TABLA 3.6.1 – EXTENSIONES DE IMÁGENES	
Extensión	Firma de archivo
.png	0x89, 0x50, 0x4E, 0x47, 0x0D, 0x0A, 0x1A, 0x0A
.jpg	0xFF, 0xD8, 0xFF, 0xE0
.gif	0x47, 0x49, 0x46, 0x38

*Tabla 9 Firma de archivo según la extensión*

A cada imagen disponible se le aplica detección de rostros con OpenCV. Este utiliza el modelo VGG-Face y si la predicción es mayor de 0.5 el módulo recorta un rectángulo alrededor de los píxeles analizados. Dado que una imagen puede contener varias caras se le asocia internamente el nombre de la imagen al prefijo face\_N, siendo ‘N’ un número entero que se va incrementando conforme se detectan caras en la imagen. A continuación, se estudia el recorte mediante DeepFace. Los modelos que utiliza DeepFace son redes neuronales convolucionales con entradas de tamaño estándar, por lo que la librería las normaliza antes de enviarlas al modelo. Después de la detección y la alineación añade píxeles negros de relleno para evitar deformar la entrada. Luego se propagan por el modelo y, o bien la librería produce una excepción al no detectar ningún rostro, o devuelve un diccionario de Python con las distintas categorías que identifican la cara. Finalmente, nombre, recorte del rostro y perfil son almacenados en memoria para posteriormente aplicar clustering y búsquedas. El clustering agrupa rostros del listado con un índice de similitud mayor o igual a 0.6, y va descartando en función de esto las

instancias de la lista seleccionadas. Las búsquedas siguen el mismo razonamiento que el apartado anterior; se elabora un sondeo para cada criterio proporcionado, recogiendo los perfiles o las fotos y realizando el mismo proceso de análisis. Como resultado se obtiene otro listado de rostros reconocidos sobre el que comparar similitudes. Aquellos que superen el umbral de similitud se exportan como coincidencias de la búsqueda.

### 3.6.3 Salida del programa

En función del modo de ejecución el programa produce una salida distinta. Si se indica exportar (-e) las imágenes de las bases de datos de Your Phone se envían tres subcarpetas dentro de exported según sea el origen (media, thumbnail, wallpaper). Si también se indicaron criterios de búsqueda, ya fueran por imagen o perfil, se enviarán a las respectivas carpetas search\_face\_images y search\_face\_profiles. Por último, si se añadió la tarea de agrupar (-gfi) se crearán dentro de group\_face\_images tantos subdirectorios con el prefijo suspect como individuos se categoricen.

```
+---exported
|   +---media
|   +---thumbnails
|   \---wallpaper
|
+---search_face_images
|   +---search_img_0
|   ...
|   \---search_img_n
+---search_face_profiles
|   +---search_profile_0
|   ...
|   \---search_profile_n
+---group_face_images
|   +---suspect0
|   ...
|   \---suspectn
```

La librería DeepFace ofrece ocho motores de reconocimiento facial, VGG-Face (98.78%), Facenet (99.20%), Facenet512 (99.65%), OpenFace (93.80%), DeepFace (97.25%), DeepID (99.15%), ArcFace (99.41%), Dlib (99.38 %). Todas las funciones de deepface aceptan un argumento opcional, el modelo de detección y alineación de rostros, OpenCV, SSD, Dlib, MTCNN, RetinaFace y MediaPipe. Más adelante en la sección 3.7 Pruebas realizadas y 3.8 Resultados obtenidos se determina cual de entre los modelos de reconocimiento y detección resulta óptimo atendiendo al rendimiento y tiempo de



ejecución. A continuación, se muestran ejemplos de la salida que generaría el programa al realizar algunas búsquedas, por ejemplo, la Figura 20 muestra todos los rostros identificados de un mismo sujeto. La Figura 21 recoge aquellos rostros que cumplen con el perfil facial de mujer menor de 25 años. Por último, la imagen inferior derecha, Fig. 23, recoge algunos de los rostros localizados al buscar caras con gestos de sonrisa mientras que la imagen a su izquierda, Fig. 22, recoge rostros de hombres de raza negra.

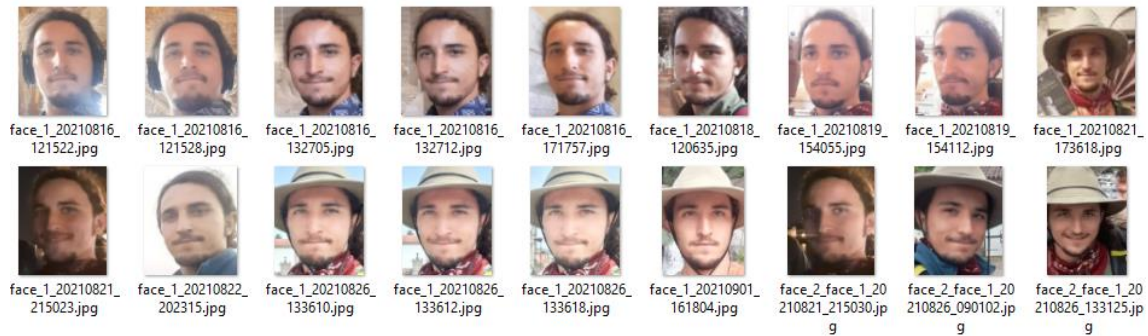


Fig. 20 Rostros similares de un mismo sujeto



Fig. 21 Rostros femeninos de mujeres menores de 25

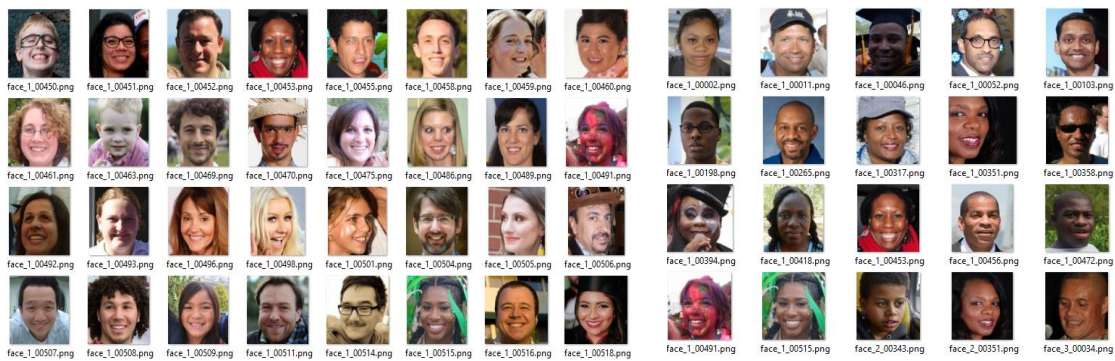


Fig. 22 Rostros de raza negra

Fig. 23 Rostros sonrientes

### 3.7 Pruebas realizadas

A continuación, las pruebas realizadas. En primer lugar la Tabla 10 Tabla 10 Pruebas sobre modelos de reconocimiento facial ilustra el número de rostros reconocibles de un total de 1562 en la BBDD. Las imágenes han sido tomadas del repositorio de libre acceso y distribución (Flickr, 2022). La Tabla 11, agrupa los casos de uso con los que se ha testado el script desarrollado y por último la Tabla 12 expone las evidencias sobre las que se ha basado el análisis forense.

TABLA 3.7.1 – MODELOS DE RECONOCIMIENTO FACIAL						
Arquitectura de reconocimiento	Medida de distancia	Modelo de detección				
		opencv	ssd	dlib	mtcnn	RetinaFace
VGG-Face	euclidean_l2	448	256	743	908	511
	euclidean	448	256	743	908	511
	cosine	448	256	744	908	511
Facenet	euclidean_l2	448	256	743	908	511
	euclidean	448	256	743	908	511
	cosine	448	256	745	908	511
Facenet512	euclidean_l2	448	256	744	908	511
	euclidean	448	256	743	908	511
	cosine	448	256	743	908	511
OpenFace	euclidean_l2	448	256	744	908	511
	euclidean	448	256	743	908	511
	cosine	448	256	743	908	511
DeepFace	euclidean_l2	448	256	745	908	511
	euclidean	448	256	743	908	511
	cosine	448	256	744	908	511
DeepId	euclidean_l2	448	256	745	908	511
	euclidean	448	256	743	908	511
	cosine	448	254	744	908	511
ArcFace	euclidean_l2	448	257	743	908	511
	euclidean	448	256	744	908	511
	cosine	448	256	743	908	511
Dlib	euclidean_l2	448	256	744	908	511
	euclidean	448	256	743	908	511
	cosine	448	256	743	908	511

Tabla 10 Pruebas sobre modelos de reconocimiento facial

**TABLA 3.7.2 – BATERÍA DE PRUEBAS PARA EL SCRIPT**

	Descripción	Resultado esperado	Salida	Test
Formato de los parámetros	Parámetros inexistentes	Aborta ejecución y produce un mensaje de error	Unrecognized arguments	✓
	Sin parámetros	Aborta ejecución y produce un mensaje de error	Invalid input path.	✓
	Parámetros repetidos	Completa el algoritmo con el último duplicado	-	✓
Input	Parámetros válidos: -input	Completa el algoritmo procesando el número esperado de Contactos, llamadas, sms e imágenes	-	✓
	Parámetros válidos: -input BBDD en uso por otro programa	Completa el algoritmo procesando el número esperado de Contactos, llamadas, sms e imágenes	-	✓
	Parámetros válidos: -input BBDD con valores NULL	Lanza una excepción y continúa con las BBDD existentes	✗Contacts, calls, sms and mms: unsupported operand type(s) for /: 'NoneType' and 'int'	✓
	Parámetros válidos: -input Directorio vacío o falta alguna BBDD	Lanza una excepción si no existe la BBDD que esperaba e intenta continuar el análisis con las siguientes.	✗Contacts, calls, sms and mms: 'NoneType' object is not iterable ✗Installed apps: 'NoneType' object is not iterable	✓
Verbose	Parámetros válidos : -v	Muestra cada operación por pantalla	-	✓
Output	Parámetros válidos: -output Directorio ya existente	Envía la salida al directorio ya existente	-	✓
	Parámetros válidos: -output Directorio inexistente	Envía la salida al directorio recién creado	-	✓
Export	Parámetros válidos: -export Directorio inexistente	Genera una carpeta denominada exported en el directorio de ejecución	-	✓
	Parámetros válidos: -output -export Directorio output inexistente o existente	Genera una carpeta denominada exported en el directorio indicado por output	-	✓
	Parámetros válidos: -output -export Directorio 'exported' existe dentro de output	Envía la salida a la carpeta 'exported' en el directorio indicado por output	-	✓



Search phone numbers	Parámetros válidos: -searchPhoneNumbers	Se muestran exclusivamente las búsquedas	-	✓
	Parámetros inválidos: csv vacío -searchPhoneNumbers	Al no poder realizar la búsqueda no se muestra ningún resultado por pantalla	-	✓
	Parámetros inválidos: csv con header y sin body -searchPhoneNumbers	Al no poder realizar la búsqueda no se muestra ningún resultado por pantalla	-	✓
	Parámetros inválidos: csv incorrecto -searchPhoneNumbers	Al no poder realizar la búsqueda no se muestra ningún resultado por pantalla	✗Search phones	✓
	Parámetros inválidos: csv inexistente -searchPhoneNumbers	Aborta ejecución y produce un mensaje de error	Phone search csv provided doesn't exist.	✓
Imágenes	Parámetros válidos: directorio con imágenes válidas -searchFaceImages	Se muestra el número de imágenes y de rostros reconocibles en la BBDD.  Se muestra el número de imágenes, de rostros reconocibles y de coincidencias con el set de imágenes de búsqueda.  Se genera un directorio por cada criterio sobre el que se exportan los rostros coincidentes	<b>✓ Processing images:</b>   Found faces ..   Recognizable faces .. <b>✓ Search face images</b>   Found faces from search set: ..   Recognizable faces from search set: ..   Matched face images ..	✓
	Parámetros válidos: directorio sin imágenes -searchFaceImages -searchFaceProfiles -groupFaceImages	Se muestra el número de imágenes y de rostros reconocibles en la BBDD.  Se muestra el número de imágenes, de rostros reconocibles y de coincidencias con el set de imágenes de búsqueda.  Se genera un directorio por cada criterio sobre el que se exportan los rostros coincidentes  Se muestra el número de individuos agrupados.	<b>✓ Processing images:</b>   Found faces ..   Recognizable faces .. <b>✓ Search face images</b>   Found faces from search set: ..   Recognizable faces from search set: ..   Matched face images .. <b>✓ Search face profiles</b>   Matched face profiles .. <b>✓ Group faces</b>   Grouped faces ..	✓
	Parámetros válidos: -groupFaceImages	Se muestra el número de imágenes y el número de rostros reconocibles en la BBDD.  Se muestra el número de individuos agrupados.	<b>✓ Processing images:</b>   Found faces ..   Recognizable faces .. <b>✓ Group faces</b>   Grouped faces ..	✓
	Parámetros válidos: DDBB sin imágenes	Se muestra el número de imágenes y el número de	<b>✓ Processing images:</b>   Found faces 0	✓

-searchFaceImages -searchFaceProfiles -groupFaceImages	rostros reconocibles en la BBDD.  Se muestra el número de imágenes, número de rostros reconocibles y número de coincidencias con el set de imágenes de búsqueda.  Se muestra el número de individuos agrupados.	Recognizable faces 0 <b>✓ Search face images</b>   Found faces from search set: ..   Recognizable faces from search set: ..   Matched face images 0 <b>✓ Search face profiles</b>   Matched face profiles 0	
Parámetros válidos: DDBB o directorio con rostros irreconocibles -searchFaceImages -searchFaceProfiles -groupFaceImages	Se muestra el número de imágenes y de rostros reconocibles en la BBDD.  Se muestra el número de imágenes, de rostros reconocibles y de coincidencias con el set de imágenes de búsqueda.  Se genera un directorio por cada criterio sobre el que se exportan los rostros coincidentes  Se muestra el número de individuos agrupados.	<b>✓ Processing images:</b>   Found faces ..   Recognizable faces .. <b>✓ Search face images</b>   Found faces from search set: ..   Recognizable faces from search set: ..   Matched face images .. <b>✓ Search face profiles</b>   Matched face profiles .. <b>✓ Group faces</b>   Grouped faces ..	✓
Parámetros válidos: -searchFaceProfiles	Se muestran exclusivamente las búsquedas	-	✓
Parámetros inválidos: csv vacío -searchFaceProfiles	Al no poder realizar la búsqueda no se muestra ningún resultado por pantalla	-	✓
Parámetros inválidos: csv con header y sin body -searchFaceProfiles	Al no poder realizar la búsqueda no se muestra ningún resultado por pantalla	-	✓
Parámetros inválidos: csv incorrecto -searchFaceProfiles	Al no poder realizar la búsqueda no se muestra ningún resultado por pantalla	✗Search phones	✓
Parámetros válidos: directorio inexistente -searchFaceImages	Aborta ejecución y produce un mensaje de error	Face search folder doesn't exist.	✓
Parámetros inválidos: csv inexistente -searchFaceProfiles	Aborta ejecución y produce un mensaje de error	Face search folder doesn't exist.	✓

Tabla 11 Batería de pruebas para YourPhoneForensicAnalyzer

**TABLA 3.7.2 – BATERÍA DE PRUEBAS REALIZADAS A YOUR PHONE**

Prueba	Condiciones	Artefactos	Salida	Test
Vincular dispositivo	A través de la misma red	calling.db, contacts.db, deviceData.db, notifications.db, phone.db, photos.db, settings.db, sharedcontent.db	Se generan los artefactos en una nueva estructura de ficheros correspondiente al dispositivo.	✓
	A través de una red distinta			
Desvincular dispositivo	A través de la misma red	calling.db, contacts.db, deviceData.db, notifications.db, phone.db, photos.db, settings.db, sharedcontent.db	Los artefactos permanecen en la estructura de ficheros correspondiente al dispositivo.	✓
	A través de una red distinta			
Abrir galería de imágenes	A través de la misma red	photos.db	Se agrega a photos.db las nuevas filas de imágenes previusualizadas (imagen miniatura o <i>thumbnail</i> ).	✓
	A través de una red distinta			
Abrir imagen de la galería	A través de la misma red	photos.db	Se agrega en el campo <i>media</i> , de la instancia dentro de photos.db, la misma imagen con mayor resolución.	✓
	A través de una red distinta			
Guardar imagen de la galería	A través de la misma red	photos.db	Se transfiere el fichero original seleccionado al directorio indicado por el usuario	✓
	A través de una red distinta			
Eliminar notificaciones desde PC y móvil	A través de la misma red	notifications.db	Se borran las filas de la base de datos notifications.db	✓
	A través de una red distinta			
Interactuar con notificaciones	A través de la misma red	notifications.db	Se actualiza el contenido del campo JSON de la notificación seleccionada en la base de datos notifications.db	✓
	A través de una red distinta			
SMS desde PC y móvil	A través de la misma red	phone.db	Se agregan nuevas filas a la base de datos phone.db	✓
	A través de una red distinta			
Forzar actualización	A través de la misma red	calling.db, contacts.db, deviceData.db, notifications.db, phone.db, photos.db, settings.db, sharedcontent.db	Se agregan nuevas filas si han entrado mensajes / llamadas / notificaciones / imágenes y se actualizan los timestamps de las instancias ya existentes de la BBDD	✓
	A través de una red distinta			
Llamadas desde PC y móvil	Conexión bluetooth	calling.db	Se agregan nuevas filas si han entrado mensajes/llamadas/notificaciones/ imágenes y se actualizan los timestamps de las instancias ya existentes de la BBDD	✓
Intentar manipular un dispositivo sin conexión	Teléfono desconectado de la red o directamente con Your Phone Companion desinstalado	-	Se muestra un mensaje de error: “No podemos conectarnos a su dispositivo Android”	✓
Desinstalar Your Phone Companion	-	-	Los artefactos permanecen en la estructura de ficheros correspondiente al dispositivo.	✓

*Tabla 12 Pruebas realizadas a los artefactos de Your Phone*

### 3.8 Resultados obtenidos

La selección del modelo para reconocimiento facial se ha basado en los resultados de las pruebas expuestas en la Tabla 10, de donde se ha podido comprobar que se cumplen las estimaciones de DeepFace en lo que a consistencia se refiere (el número de caras permanece casi constante a través de los modelos). Respecto al modelo seleccionado se ha tomado como criterio maximizar los rostros detectados así como optimizar el tiempo que lleva hacerlo. En este sentido la arquitectura VGG-Face utilizando distancia euclídea el modelo mtcnn obtiene las mejores métricas de reconocimiento y tiempo, por lo que ha sido la opción seleccionada.

Las pruebas realizadas sobre YourPhoneForensicAnalyzer (véase Tabla 11) muestran una alta robustez en el código ya que ninguno de los casos testados ha provocado un comportamiento inesperado, indicativo de que se recogen correctamente las excepciones que producen parámetros inválidos o datos incorrectos. Cabría añadir que en caso de producirse se generan mensajes de error específicos al fallo y se registra la información previamente procesada en un log (ej. las imágenes extraídas). Asimismo, se ha verificado que la totalidad de filas de las bases de datos se parsean, es decir, no quedan mensajes, llamadas o contactos sin procesar.

Sobre las pruebas realizadas a Your Phone, Tabla 12, cabe mencionar que el estudio del comportamiento de la aplicación se ha centrado en rango de acciones aplicado a los artefactos encontrados y no sobre el registro. Según se ha comprobado durante el análisis dinámico, en el registro se almacenan constantemente métricas de la aplicación, pero no información del usuario, lo que reduce el interés forense de su extracción. Respecto a la capacidad de obtener información del usuario el estudio demuestra que no solo es factible, sino que la información almacenada resulta valiosa por su contenido y timestamps. No obstante, se ha verificado que tal y como vaticinaban estudios anteriores los registros SQL se ven limitados a un rango temporal de un mes. Esto reduce las posibilidades de encontrar datos fuera de este rango, aunque gracias a las pruebas de recuperación de datos borrados sabemos que existe la posibilidad de hacerlo.

## CONCLUSIONES Y TRABAJOS FUTUROS

### 4.1 Objetivos cumplidos

Este trabajo final de grado ha cumplido con las metas propuestas en el apartado 1.5 Objetivos Concretamente los objetivos principales, “Detallar los procesos que componen a la aplicación Microsoft Your Phone y los artefactos que estos dejan en el sistema” e “Implementar una solución software que permita recoger, parsear y exportar la información que presente la aplicación Microsoft Your Phone”, han sido llevados a cabo satisfactoriamente. Concretamente del conjunto de retos específicos se consigue:

- **Extraer el contenido multimedia de la aplicación y aplicar sobre este detección y categorización de rostros.** La manipulación de imágenes cumple con los objetivos propuestos y produce resultados claramente favorables, aunque sufre ciertas limitaciones como:
  - a. La resolución de las imágenes presentes en Your Phone. Ninguna imagen en las BBDD supera los 1.5 MiB, lo cual no facilita la detección puesto que a menor resolución menor información para analizar.
  - b. El número de rostros reconocibles sea menor que el de rostros detectados (0.58%). Idealmente se busca que todos los rostros detectables fuesen manipulables, es decir, que de ellos se pudiesen extraer características identificativas para comparar. No obstante, esto no siempre es posible ya que son tareas de complejidad distinta. La primera consiste en categorizar la imagen como rostro humano, la segunda consiste en analizar y extraer sus características.
  - c. La funcionalidad de agrupamiento de rostros produce resultados ambiguos. El algoritmo propuesto para esta funcionalidad flaquea ya que dependiendo de las imágenes no siempre agrupa todas las fotos de una persona como un único sujeto. La tarea de clusterizar caras resulta ser mucho más compleja de lo que en un principio se esperaba, por lo que el algoritmo de selección y descarte no

basta. Este problema hoy por hoy continúa siendo investigando por lo que podrían considerar soluciones específicamente diseñadas para esta tarea como pueden ser algoritmos de k-Medias o Spectral.

- **Implementar un sistema de búsqueda por perfil facial.** Los casos de uso demuestran que la funcionalidad cumple con los estándares establecidos y permiten encontrar una persona dadas su descripción facial. Así mismo existen algunos retos, como reconocer con alta precisión la edad del individuo, especialmente en niños.
- **Evaluar el correcto funcionamiento de la solución software mediante un amplio espectro de evidencias y casos de prueba.** Se ha generado y testado mediante el set de pruebas (Tabla 11 Pruebas realizadas a los artefactos de Your Phone) variadas que recogen la casuística de situaciones (válidas e inválidas) a las que el Script deberá enfrentarse al ejecutar.
- **Evaluar el contenido salvaguardado por la aplicación.** Según se ha comprobado, la aplicación produce artefactos con información valiosa desde la perspectiva forense. Esto ha permitido extraer datos adicionales como el estado de las conversaciones (llamadas y sms) o fechas de lectura y actualización
- **Extraer contenido eliminado de la aplicación:** A pesar de haberse demostrado como una alternativa limitada se ha logrado recuperar mediante carving con UnDark algunas de filas eliminadas por la base de datos.

## 4.2 Líneas futuras de trabajo

Las líneas de desarrollo futuro tendrán dos partes. Por un lado, el trabajo de optimización en la implementación y por otro la extensión del estudio forense.

Sobre el primero, se deberá perfeccionar el análisis facial para expandir su precisión en el reconocimiento de rostros. Concretamente deberá encontrar una alternativa fiable para el algoritmo de clustering el cual permita descartar falsos positivos y evitar que un individuo sea categorizado múltiples veces. Asimismo, estudiar también otros modelos con mayor precisión de comparación de rostros dada la resolución de la que se dispone.

Sobre el segundo, se deberá ampliar el estudio de las funcionalidades que no se han evaluado en este trabajo dado el limitado acceso a recursos y el condicionante del tiempo. Sobre la actual versión se podría estudiar los artefactos del servicio de compartir pantalla con el PC, solamente soportado para los últimos modelos móviles. También y debido a la extensión del trabajo, ha quedado fuera el análisis del dispositivo Android. Resultaría muy interesante aplicar la misma metodología sobre el dispositivo. Asimismo, al igual que ha hecho esta investigación, se deberán estudiar las futuras actualizaciones de Your Phone ya que si Microsoft continúa en la línea anunciada es muy probable que profundicen en la sincronización de móviles y PCs, lo cual podría dar a acceso a nueva información relevante. Si es así sería conveniente agregar nuevos módulos al programa desarrollado para que la extracción de estos nuevos datos no se vuelva manual.

Adicionalmente, cabe agregar que existen posibles mejoras a realizar sobre el trabajo presentado. Por ejemplo, sería positivo introducir sistemas de censura de material sensible dado que, en investigaciones de índole criminalista se tratan con frecuencia imágenes sobre pornografía infantil, homicidios o violencia doméstica. Soluciones como Axiom Cyber de Magnet o X platform de Belkasoft aplican filtros de difuminado sobre este tipo de contenido, por lo que sería interesante agregar medios similares para censurar total o parcialmente, y así evitar la sobreexposición del analista. Otro aspecto que podría ampliarse es el tratamiento de imágenes. El reconocimiento de imágenes no tiene por qué limitarse exclusivamente a la detección y clasificación de rostros, también puede ampliarse a detección de objetos.

### **4.3 Conclusión**

Tal y como se puede comprobar, se han abarcado temáticas dentro de la ciberseguridad, el diseño de sistemas y la inteligencia artificial, lo que hace del proyecto uno creativo y multidisciplinar. La finalidad, en cualquier caso, no ha sido la de centrarse únicamente en una de los ámbitos expuestos, sino que desde un primer momento se ha buscado ofrecer un estudio en amplitud de las características definitorias de Your Phone, así como un servicio distinguido por sus múltiples funcionalidades

Como conclusión señalar que el campo de la informática forense es un ámbito extenso, interdisciplinar y siempre en constante avance donde cada día surgen nuevas tecnologías

y las que ya existen, cambian a lo largo del tiempo. Los retos a los que se enfrentan los analistas también cambian por lo que en este sector es de gran importancia mantener la discusión e investigación abierta, ya que retribuye en una comunidad actualizada y formada. Siguiendo estas líneas, la última meta de este trabajo es expandir el conocimiento y las herramientas de las que dispone la comunidad. Para ello se este trabajo así como el software desarrollado ha sido publicado libremente en un repositorio público en GitHub <https://github.com/groongra/Microsoft-Your-Phone-parser>.



## BIBLIOGRAFÍA

- [1] Microsoft Corporation, "Microsoft," [Online]. Available: <https://www.microsoft.com/en-us/p/your-phone/9nmpj99vjbwv#activetab=pivot:overviewtab>. [Accessed 09 02 2020].
- [2] S. Cordero, *Análisis forense informático*, Barcelona, 2015.
- [3] N. Panov, "Digital Forensics Tips&Tricks," 2019. [Online]. Available: <https://habr.com/en/post/470952/>. [Accessed 01 12 2020].
- [4] L. M. A. y. M. F. y. J. V. S. Patricio Domingues, "Digital forensic artifacts of the Your Phone application in Windows 10," *elsevier*, vol. 30, no. 1742-2876, pp. 32-42, 2019.
- [5] L. M. A. y. M. F. Patricio Domingues, "Microsoft's Your Phone environment from a digital forensic perspective," *elsevier*, vol. 38, no. 2666-2817, 2021.
- [6] R. L. Rivera, "peritoit," [Online]. Available: <https://peritoit.com/delitos-informaticos/>. [Accessed 01 01 2022].
- [7] International Society of Forensics Computer Examiners, "isfc," [Online]. Available: <https://www.isfce.com/ethics2.htm>. [Accessed 05 01 2022].
- [8] E. Peris, "informaticajudicial," [Online]. Available: <https://informaticajudicial.es/normativa/>. [Accessed 09 02 22].
- [9] B. Betsy Mikalacki, "How much does digital forensic services cost," Vestige, [Online]. Available: <https://www.vestigeltd.com/thought-leadership/digital-forensic-services-cost-guide-vestige-digital-investigations/>. [Accessed 2022 02 09].
- [10] glassdoor, "glassdoor," [Online]. Available: <https://www.glassdoor.es/Sueldos/madrid-programador-junior-sueldo->

SRCH\_IL.0,6\_IM1030\_KO7,25.htm?clickSource=searchBtn/. . [Accessed 01 01 2022].

- [11] glassdoor, "glassdoor," [Online]. Available:  
[https://www.glassdoor.es/Sueldos/madrid-profesor-universitario-sueldo-SRCH\\_IL.0,6\\_IM1030\\_KO7,29.htm?clickSource=searchBtn](https://www.glassdoor.es/Sueldos/madrid-profesor-universitario-sueldo-SRCH_IL.0,6_IM1030_KO7,29.htm?clickSource=searchBtn). [Accessed 01 01 2022].
- [12] Microsoft, "Supported devices for Your Phone app experiences," [Online]. Available: <https://support.microsoft.com/en-us/topic/supported-devices-for-your-phone-app-experiences-cb044172-87aa-9e41-d446-c4ac83ce8807>. [Accessed 09 02 2022].
- [13] Python, "GitHub," [Online]. Available:  
<https://github.com/python/cpython/blob/main/Lib/io.py>. [Accessed 09 02 2022].
- [14] Python, "Python Documentation," [Online]. Available:  
<https://docs.python.org/3/library/time.html>. [Accessed 09 02 2022].
- [15] Python, "GitHub," [Online]. Available:  
<https://github.com/python/cpython/blob/main/Lib/datetime.py>. [Accessed 09 02 2022].
- [16] Python, "GitHub," [Online]. Available:  
<https://github.com/python/cpython/blob/main/Lib/os.py>. [Accessed 09 02 2022].
- [17] Python, "GitHub," [Online]. Available:  
<https://github.com/python/cpython/blob/main/Python/sysmodule.c>. [Accessed 09 02 2022].
- [18] Python, "GitHub," [Online]. Available:  
<https://github.com/python/cpython/blob/main/Lib/uuid.py>. [Accessed 09 02 2022].
- [19] A. C. a. Contributors., "Python Imaging Library," [Online]. Available:  
<https://github.com/python-pillow/Pillow/>. [Accessed 09 02 2022].

- [20] S. I. y. O. A. Serengil, "GitHub," [Online]. Available: <https://github.com/serengil/deepface>. [Accessed 09 02 2022].
- [21] G. Bradski, "GitHub," [Online]. Available: <https://github.com/opencv/opencv/wiki/CiteOpenCV>. [Accessed 09 02 2022].
- [22] S. J. Bethard, "GitHub," [Online]. Available: <https://github.com/python/cpython/blob/3.10/Lib/argparse.py>. [Accessed 09 02 2022].
- [23] K. Lepa, "GitHub," [Online]. Available: <https://github.com/hfeeki/termcolor>. [Accessed 09 02 2022].
- [24] M. Singh, "GitHub," [Online]. Available: <https://github.com/manrajgrover/halo>. [Accessed 09 02 2022].
- [25] Python, "GitHub," [Online]. Available: <https://github.com/python/cpython/blob/main/Lib/csv.py>. [Accessed 09 02 2022].
- [26] G. Häring, "GitHub," [Online]. Available: <https://github.com/python/cpython/blob/main/Doc/library/sqlite3.rst>. [Accessed 09 02 2022].
- [27] Numpy, "GitHub," [Online]. Available: <https://github.com/numpy/numpy>. [Accessed 09 02 2022].
- [28] Secret Labs, "GitHub," [Online]. Available: <https://github.com/python/cpython/blob/main/Lib/re.py>. [Accessed 09 02 2022].
- [29] J. Friedl, "Jeffrey's Image Metadata Viewer," [Online]. Available: <http://exif.regex.info/exif.cgi>. [Accessed 21 12 2022].
- [30] K. Bhanot, "GitHub," [Online]. Available: <https://github.com/kb22/Create-Face-Data-from-Images>. [Accessed 11 02 2022].

- [31] Union Internacional de telecomunicaciones, "itu," [Online]. Available: [itu.int/rec/T-REC-E.164/es](https://www.itu.int/rec/T-REC-E.164/es). [Accessed 04 01 2022].
- [32] Viso, "viso.ai," [Online]. Available: <https://viso.ai/computer-vision/deepface/>. [Accessed 20 02 2022].
- [33] Flaticon, "flaticon.es/," [Online]. Available: <https://www.flaticon.es/>. [Accessed 20 02 2022].

## **ANEXO I. ENGLISH SUMMARY**

### **INTRODUCTION**

This document contains the final year project developed by Lucas González de Alba, a student from Universidad Carlos III de Madrid. This research project analyzes from a digital forensics standpoint, the structure, behavior, and artifacts created by Microsoft's Your Phone application.

#### **1.1 Abstract**

Microsoft's Your Phone is a service that facilitates user access to phone devices by integrating its notifications, messages, photos, and calls directly into Windows. All this data can have huge value on forensic investigation, so the work we present analyzes Your Phone app on the look for digital artifacts.

The project began with large research of existing knowledge; that is articles, publications, blogs, forums...etc. After a brief overview of the app's behavior, several digital samples were collected. With this in mind, a formal study of the app's behavior was conducted using process monitoring. Then both format and structure from any relevant system artifact were analyzed. With the obtained knowledge, a Python script to facilitate extraction and search analysis was developed. This program parses call logs, sms and mms messages, settings, and a list of installed applications and supports facial recognition. The identity verification module uses OpenCV and DeepFace and its objective is to detect one or multiple faces within an image, allow for attributes comparison, and support the search for similarities based on a descriptive profile. A hybrid between OpenCV and DeepFace was chosen for its favorable results and versatility. Finally, the quality of the developed software was evaluated using a large test set.

## **1.2 Motivations**

This project was first conceived as the author's personal research on analysis techniques used in cybersecurity and computer forensics. However, since it quickly grew in length and complexity, a larger goal was set. Why computer forensics? Because, unfortunately, it is a subject that is not taught in the computer science degree, but offers transversal knowledge to many areas such as cybersecurity, software engineering, and data science.

The analysis of Microsoft's Your Phone application can offer very positive benefits to the forensic community. A few are:

- Exposing the formal study of the structure and artifacts by determining what information can be extracted from the artifacts of the application.
- Creating a linked mobile and computer environment through Your Phone.
- Automating extraction of the images stored by the system. Avoid the tedious and repetitive task of searching, selecting, and saving each image in the application.
- Enabling "live" evidence analysis

In summary, studying Microsoft's Your Phone application from a forensic perspective not only provides the advantages outlined above, but it also offers a unique opportunity to develop innovative research with substantial work.

## **1.3 Work methodology**

The first stage, analysis, was dedicated to the study of the program behavior. To acquire as much information as possible about the program behavior, an incremental examination based on the following phases was established:

- Preliminary phase: the working environment and the analysis tools are set up.
- Functionalities exploration: getting acquainted with the app's capabilities.
- Dynamic analysis: real-time monitoring to identify and collect system artifacts.
- Static analysis: study the main digital traces and artifacts previously discovered.

The second stage, development, focused on the implementation of a software solution dedicated to facilitating access to the previously analyzed information.

- Requirements identification: capturing the needs and constraints of the problem.

- Design: study and establish the best possible software architecture
- Implementation: develop the program for the selected architecture
- Testing: evaluate results and verify compliance with the requirements.

## **1.4 Goals**

This chapter sets out the main goal of the work and establishes the various specific objectives that comprise it.

Primary goals:

1. Analyze and report Your Phone's artifacts identifying any forensically valuable information they might contain.
2. Implement a software solution to collect, parse and export the information presented by the Microsoft Your Phone application

Secondary goals

1. Extract the multimedia content of the application and apply face detection and categorization
2. Implement a facial profile search system. Develop a software tool to locate people given the main characteristics of their faces.
3. Evaluate the correctness and accuracy of the software solution that was developed
4. Evaluate if there exists content safeguarded by the application
5. Extract deleted content from the application

# **STATE OF THE ART**

## **2.1 External resources**

In the field of computer forensics and cybersecurity, it is common to find publications related to the study and monitoring of free, corporate, and malicious software programs. Such studies usually apply both static and dynamic analysis and usually involve hash analysis, carving, network monitoring, process and thread recognition, Windows registry

tracing, and artifact examination. As far as Microsoft Your Phone is concerned, only three publications have been found. These are Digital Forensics Tips & Tricks: "Your Phone" app Forensics [3], Digital forensic artifacts of the Your Phone application in Windows 10 [4] and its subsequent review of Microsoft's Your Phone environment from a digital forensic perspective [5]. The first came just after Windows 10's Insider Preview Build 18999 (20H1) was released and it briefly reviewed the functionalities of the newly introduced application. By contrast, in the second article the authors Patricio Domingues, Miguel Frade, Luis Miguel Andrade and Joao Victor Silva analyze versions 1.0.20453 and 3.4.4.4 of Windows 10's Your Phone and Android's app Your Phone Companion respectively. In addition, they proposed a Python script designed to run on Autopsy. Finally, the third article examines updates 1.21011.127.0 (Windows) and 1.21021.81.0 (android) and follows the previous line of development, extending some aspects that were left out of the previous study and expanding the functionalities of the proposed program. Compared to the first analysis proposed by Panov the last two publications really expand the knowledge and explain with greater detail how the application was organized and how it stored the user's data. Overall, these three studies manage to meet some of the objectives of this paper, but nevertheless, leave others out.

Some of the technical issues they do not solve are process monitoring, Windows registry analysis, and description of artifact configurations and that is without adding the challenge of new updates. The application has continued to renew itself, including new features such as the expansion of its gallery, instant messaging, and screen sharing which renders some of the previous work obsolete.

As far as photos and videos are concerned, digital image-processing services embedded in forensic applications are becoming the norm. Companies such as Belkasoft, Magnet, and Cellebrite with their consolidated products such as Platform X, Axiom Cyber, or Physical Analyzer are betting on programs with a variety of utilities, especially image recognition. This trend acknowledges the power of IA by incorporating it into the forensic toolkit to facilitate the work of analysts. Interestingly, none of the previous publications face this challenge, which opens the possibility for innovation.



## **2.2 Legislation**

### **2.2.1 Legislation and legal regulations.**

In 2010 Spain ratified The Cybercrime Convention [A.8], drawn up in Budapest on 23 November 2001. Along with this convention and the Spanish code of law, crimes against privacy, espionage, theft, impersonation, fraud, forgery, embezzlement, manipulation of devices, damage, or alteration of data programs or files are contemplated. It is worth mentioning that the forensic profession relies heavily on ethical integrity. There are various codes of ethics and recommendations such as the ISCF code of conduct from the International School of Computer Forensics or the SANS institute work-ethic guidelines. These documents signify the importance of honesty, defense of intellectual property, confidentiality, and individual rights and freedoms. In summary professionalism and commitment to safeguarding the truth. Likewise, it condemns any form of corruption (blackmail, bribery), prevaricating attitude, a premeditated attack against privacy, or discrimination based on sex, race, religion, age, ethnicity, politics, or any other condition.

### **2.2.2 The figure of the computer expert in the courts.**

The law defines this role as a professional specialized in computer science whose work consists of providing technical advice in judicial proceedings, as well as contributing to mediation and conflict resolution. He/she can exercise several roles, mediator, arbitrator, and auditor, all of which are overseen by judges.

### **2.2.3 Chain of custody**

When presenting a piece of digital evidence to the judge there are several prerequisites that must be met for admission. For any evidence collected, the original evidence must be preserved together with its chain of custody. The chain of custody is a control procedure that covers the process of obtaining, handling, transferring, assigning, and preserving evidence to rigorously ensure that the evidence has been delivered and remains unaltered (as demonstrated by hash compliance).

## 2.3 Socio-economic environment

The economic impact is difficult to estimate given that in principle the product is oriented to a narrow niche, forensic investigations involving evidence containing Your Phone artifacts. According to vestige ltd [9], the average costs of a forensic investigation are usually around \$5,000 to \$15,000 on average. According to the agency, a cost of \$250 per hour of effective work can be considered standard. In this sense, the developed project could have a positive impact since it would reduce the analysis time. The software solution developed could also be exported to different contexts such as image extraction functionality. Any application that seeks to interact (input or extract) multimedia content from Your Phone. The same would be true for the face comparator and face grouper, which if generalized and refined could be marketed as an add-on module for some of the forensic programs described above. As for the social and ethical implications, it is worth mentioning that since this is a program designed to discover and work with personal data (conversations, calls, images), strict confidentiality should be maintained

# IMPLEMENTATION

## 3.1 Planning

Following the chosen methodology, the resulting planning is as follows: information gathering, one and a half months, analysis and development, three months, documentation, one and a half months, and testing, fifteen days. The total estimated time dedicated to the execution of the project is 290 hours.

## 3.2 Budget

**Human resources:** Since the final year project depends on both tutor and student, the average salaries of a junior programmer (novice) and a university professor in Spain have to be taken into account. According to glassdoor, the annual average salary is 19.745 € [10] and 33.862 € [11] respectively, i.e. approximately 1.646 € and 2.822 € per month. Working full time (8 hours) 21 days per month is equivalent to 10 and 18 € per hour. If

we multiply this value by the number of hours of the participants, the result is 3.100 € and 144 €.

**Working tools:** the cost of licenses and hardware.

- **Hardware tools:** to estimate the total hardware cost, the percentage of usage (time used / estimated lifetime) has been calculated and later multiplied by the acquisition price. Tests were carried out on a DELL Latitude E7270 laptop worth 1,179.27 € and a Samsung A20e cell phone 145.20 € with a 16%, 7% of lifetime usage respectively. Thus, the total cost is approximately 200€
- **Software tools:** since the project believes on open-source software, only non-licensed products were used, which mean 0€ on expenses

The project's budget is  $3.244 \text{ €} + 200\text{€} = 3500 \text{ €}$

### 3.3 Your Phone forensic analysis

DB Browser and DBeaver were used to explore the artifacts previously found. The study revealed that the database files were based on a specific Pragma configuration. Pragma is a SQL extension specific to the SQLite format that allows customization of databases configuration. To begin with, there was the “Auto Vacuum” setting disabled, which indicates that the database does not reduce its size after delete operations. Instead, unused pages of the database file are added to a “free list” and reused for the following insertions. Likewise, the “Secure Delete” was also disabled so delete operations do not overwrite the content with zeros. Finally, there is “Journal Mode” which takes the value Write-Ahead. This explains why for each .db file two additional files get generated .db-wal and .db-shm files. These files are temporary files meant to support internal operations in case of failures, specifically, Write-Ahead Logs are logs/diaries intended for write conflicts on commits and rollbacks, while shared memory files are used when two or more connections share the same .db file and must update the same memory. With the given configuration configured it is possible to recover some records by carving tools such as UnDark. However, these methodologies usually yield limited results since often only partial or corrupted information can be found. In the same way, it is not excluded that some records exist in unallocated space.

### 3.3.1 SQL tables

Regarding the tables and their contents, it was found that:

The address book is stored in `contacts.db`, where the main table `contact` resides, which is related through a unique identifier `contact_id` with other tables such as `phonenumber`, `postaladdress`, `emailaddresss`, `contactDate`, `contactURL`. Tests carried out indicates that these are mostly empty, but this could be due to device conditions

The call log is located in `call_history` in `calling.db`. This table provides useful information on phone call number, duration, type (incoming or outgoing), accepted or declined, and date.

Then there is `phone.db` where the device information is stored. On one hand, in subscription, we have the telemarketer settings, but on the other hand, there is the activity related to conversations sms, and mms messages.

There is also `settings.db`, which contains information about the device's applications, installed and recent, as well as `phone_request`.

`Notifications.db` stores the applications messages sent to the user. At first, it was conceived as a valuable source of data, but the idea was soon dismissed as the analysis showed that only the phone's notification queue was stored. Your Phone updates the notifications screen every time an application sends a warning. If the user deletes these (from either the phone or PC) the database records are emptied, therefore, one can only access the last state of the notifications queue.

Lastly, `photos.db` and `deviceData.db`, two databases containing the image gallery (media, photos) and the device's wallpaper. As the study by P. Domingues, L. M. Andrade and M. Frade [5] indicated the *photos* table is still obsolete. Instead, the application relies on the table `media`, which can store up to 2000 image records in the form of blob (byte string). Testing showed that Your Phone does not load the phone's entire gallery at once; instead, the program offers images previews as the user scrolls throughout the view. Therefore, when studying a piece of evidence what images are stored will depend on the user's usage. Regarding image metadata, persistence was examined using an online viewer called Jeffrey Friedl's Image Metadata Viewer [29]. Your Phone does not store the original image in its records but loads a thumbnail as a preview. This low-resolution

copy dismisses the original EXIF metadata. In case the user selects a larger resolution is transferred and when a save operation is performed, an identical copy (except timestamps) of the original is transferred.

### 3.3.2 Image processing

Image processing is based on two models, OpenCV [30] and DeepFace [20]. The first one is used to detect human faces, while the second one is used to apply face recognition on the obtained faces. The second one is used to apply facial recognition on the obtained faces. OpenCV uses the haar-cascade algorithm, which is not based on deep learning techniques and is, therefore, faster, but its performance is relatively low. In contrast, DeepFace allows to correct this error by adjusting the model (VGG-Face, Facenet, Facenet512, OpenFace, DeepFace, DeepID, ArcFace, Dli) and comparing distance (cosine, euclidean, euclidean\_l2) so it was more beneficial to use it as a comparator. However, unlike DeepFace, OpenCV can detect multiple faces within an image. By combining both alternatives, it is possible to identify, isolate, crop, and evaluate each face present in Your Phone. Since the two models work together, it may happen that OpenCV first detects and cuts out a face, and then DeepFace does not recognize any face. In these cases, cropping is discarded since an image without a recognizable face is an image on which no comparisons can be made.

Regarding face recognition, the decision was made to offer three services to the analyst:

- **Suspect clustering:** group similar faces from the set of extracted images
- **Face comparer search engine.** Given a suspect's face image, find similar faces within the set of extracted images.
- **Facial profile search engine.** Find faces that fit a certain description based on age, gender, gesture, and race of a given subject.

### 3.3.3 Execution modes

The first step towards the implementation of YourPhoneForensicAnalyzer was to develop the execution control. At a minimum, the script requires the path to the directory containing the databases (--i or --input). Optionally the user can indicate an output folder

(-o or --output). The program continues to evaluate the execution modes provided by the analyst:

- Modes without input parameters.
  - Help mode: -h or -help
  - Verbose mode: -v or -verbose
  - Image export mode: -e or -export
  - Group similar faces mode: -gfi or -groupFaceImages
- Modes with input parameters.
  - Search by faces mode: -sfi -searchFaceImages pathToImages
  - Search mode using face profiles: -sfp -searchFaceProfiles pathToCSV
  - Phone search mode: -spn -searchPhoneNumbers pathToCSV

Once the execution parameters have been validated, the database files are scanned in search of the user's information. If one or more search modes have been set, the search set will be limited to the provided criteria. No amplitude crawl will be performed, as the search criteria will be imposed on the database instances that meet the provided parameters.

### 3.4 Algorithm design

The algorithm retrieves a contact identifier from each contact present in the address book. If the selected identifier is associated with a phone number every call or conversation related to it gets retrieved. Nonetheless, not every message or phone call is related to the address book. These are processed and output later as "Unassociated phones, calls, sms, and mms".

Once the parsing is finished, the algorithm tackles images present in photos.db and deviceData.db by evaluating if the database rows contain any data stored in thumbnail and media fields. Then face detection is performed using OpenCV's VGG-Face model. If the prediction is greater than 0.5 the module crops a rectangle around the analyzed pixels. Next DeepFace analyzes the cropped image using convolutional neural networks. These models use standard size inputs, so the library normalizes and aligns them before sending them to the model. Then they are propagated through the model, and either the

library produces an exception when no face is detected, or returns a Python dictionary with the different categories identifying the face. These results are used for face clustering and image searching.

Figure 1 shows an example of database parsing (Contact, phone call and SMS chat) while Figure 2 displays a face profile search (Black faces profiles)

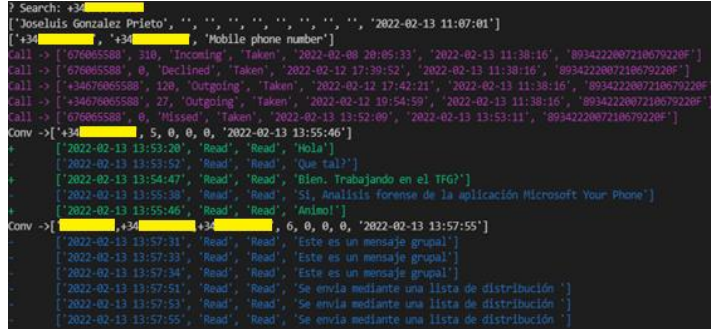


Fig. 1 Associated phones, calls, sms, and mms".

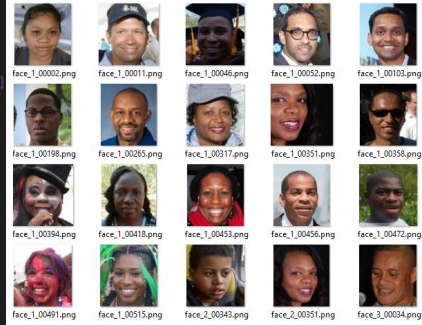


Fig. 2 Black faces profiles

## CONCLUSIONS

### 4.1 Accomplished objectives

This project has fulfilled all of the goals proposed in section. Some of the challenges have been:

1. The image's resolution in Your Phone. No image stored in the databases exceeds 1.5 MiB, which does not facilitate detection since the lower the resolution, the lesser the information available.
2. The number of recognizable faces is lower than the number of detected faces. Ideally, it should be possible to extract identifying characteristics of every detected face; however, since these tasks vary in complexity it is not always achieved.
3. The face clustering functionality produces ambiguous results. The proposed algorithm for this functionality is weak, as it does not produce reliable results. The task of clustering faces turns out to be much more complex than initially expected, so the selection and discard algorithm is not sufficient.

## 4.2 Future lines of work

Due to the limited access to resources and the time constraints, there are some aspects of the project left uncovered. Future development will likely have to work on face detection optimization as well as expand the forensic study.

On one hand, the face recognition module should improve the current accuracy in face recognition by extending the experimentation to all models with all distance measures supported by DeepFace. Moreover, a reliable alternative to the clustering algorithm must be found. On the other hand, the study of Your Phone can grow even larger with the inclusion of the screen sharing service, only supported for the latest mobile models, and Android device artifacts. Similarly, just like this research has done, future updates of Your Phone will require similar analysis techniques since it is very likely that Microsoft will expand the synchronization of cell phones and PCs. Lastly, it would be neat to add new parsing modules to the current program so that extraction continues to be easy and accessible.

To conclude, one last improvement proposal. Given that criminal investigations often deal with images of child pornography, homicides, or domestic violence it would be helpful to introduce censorship systems for sensitive material to prevent overexposure of the analyst. Another aspect of the script that leaves room for enhancement is the image-processing module, which could introduce object detection techniques.

## 4.3 Conclusion

The project covers multiple topics ranging from cybersecurity, software design, to artificial intelligence. The purpose was not to focus exclusively on one of the exposed areas but to offer a multidisciplinary comprehensive study of the defining characteristics of Your Phone. In addition, software solutions with multiple functionalities are presented as well as tested. In conclusion, it should be noted that the field of computer forensics is a vast, interdisciplinary, and constantly advancing field where new technologies emerge every day and those that already exist change over time. Hence, the challenges faced by analysts also change, so it is crucial to keep the discussion and research open by the community. As a result, this work will be available for free as an open-source repository at <https://github.com/groongra/Microsoft-Your-Phone-parser>.



## **ANEXO II. LEGISLACIÓN E INFORMÁTICA FORENSE**

[A.1] Ley 25/2007, de 18 de octubre, de conservación de datos relativos a las comunicaciones electrónicas y a las redes públicas de comunicaciones. [www.boe.es/aeboe/consultas/bases\\_datos/doc.php?id=BOE-A-2007-18243](http://www.boe.es/aeboe/consultas/bases_datos/doc.php?id=BOE-A-2007-18243)

[A.2] Ley 1/2000, de 7 de enero, de Enjuiciamiento Civil (LEC). [www.boe.es/aeboe/consultas/bases\\_datos/doc.php?id=BOE-A-2000-323](http://www.boe.es/aeboe/consultas/bases_datos/doc.php?id=BOE-A-2000-323)

[A.3] Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal (LOPD). [www.boe.es/aeboe/consultas/bases\\_datos/doc.php?id=BOE-A-1999-23750](http://www.boe.es/aeboe/consultas/bases_datos/doc.php?id=BOE-A-1999-23750)

[A.4] Ley Orgánica 19/1994, de 23 de diciembre, de protección a testigos y peritos en causas criminales. [www.boe.es/aeboe/consultas/bases\\_datos/doc.php?id=BOE-A-1994-28510](http://www.boe.es/aeboe/consultas/bases_datos/doc.php?id=BOE-A-1994-28510)

[A.5] Real Decreto 704/2011, de 20 de mayo, por el que se aprueba el Reglamento de Protección de las Infraestructuras Críticas (PIC). [www.boe.es/diario\\_boe/txt.php?id=BOE-A-2011-8849](http://www.boe.es/diario_boe/txt.php?id=BOE-A-2011-8849)

[A.6] Orden PRE/2740/2007, de 19 de Septiembre, por la que se regula el Reglamento de Evaluación y Certificación de Seguridad de las Tecnologías de la Información. [www.boe.es/boe/dias/2007/09/25/pdfs/A38781-38805.pdf](http://www.boe.es/boe/dias/2007/09/25/pdfs/A38781-38805.pdf)

[A.7] Protocolo Adicional, 28 Enero de 2003, al Convenio sobre la Ciberdelincuencia relativo a la penalización de actos de índole racista y xenófoba cometidos por medio de sistemas informáticos.

[www.gdt.guardiacivil.es/webgdt/media/Legislacion/Protocolo\\_adicional\\_convencion\\_ciberdelincrimen.pdf](http://www.gdt.guardiacivil.es/webgdt/media/Legislacion/Protocolo_adicional_convencion_ciberdelincrimen.pdf)

[A.8] BOE 14221/ 2010, de 17 Septiembre, Instrumento de Ratificación del Convenio de la Ciberdelincuencia, redactado en Budapest el 23 de noviembre de 2001.  
[www.boe.es/boe/dias/2010/09/17/pdfs/BOE-A-2010-14221.pdf](http://www.boe.es/boe/dias/2010/09/17/pdfs/BOE-A-2010-14221.pdf)

[A.9] Real Decreto Legislativo 1/1996, de 2 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia.  
[www.boe.es/aeboe/consultas/bases\\_datos/doc.php?id=BOE-A-1996-8930](http://www.boe.es/aeboe/consultas/bases_datos/doc.php?id=BOE-A-1996-8930)

[A.10] Ley 11/2011, de 20 de mayo, de reforma de la Ley 60/2003, de 23 de diciembre, de Arbitraje y de regulación del arbitraje institucional en la Administración General del Estado. [www.boe.es/aeboe/consultas/bases\\_datos/doc.php?id=BOE-A-2011-8847](http://www.boe.es/aeboe/consultas/bases_datos/doc.php?id=BOE-A-2011-8847)

[A.11] Ley Orgánica 5/2011, de 20 de mayo, complementaria a la Ley 11/2011, de 20 de mayo, de reforma de la Ley 60/2003, de 23 de diciembre, de Arbitraje y de regulación del arbitraje institucional en la Administración General del Estado para la modificación de la Ley Orgánica 6/1985, de 1 de julio, del Poder Judicial.  
[www.boe.es/aeboe/consultas/bases\\_datos/doc.php?id=BOE-A-2011-8846](http://www.boe.es/aeboe/consultas/bases_datos/doc.php?id=BOE-A-2011-8846)