



**TRABAJO DE GRADO**  
**Opción Seminario-Diplomado.**

**NullTrace: Sistema automatizado de inteligencia en fuentes abiertas  
para la consulta de datos públicos**

**Corporación Universitaria Remington**  
**Facultad de Ingenierías**  
**Ingeniería de Sistemas**

**Presentado por:**  
**Daniel Sehuanes**  
**Antony Reyes**

**Tutor: Luis Camargo Ortega**  
**Opción de Trabajo de grado Seminario.**

**2026.**

## **1 Dedicatoria**

Este trabajo de grado está dedicado a Dios, por la sabiduría y fortaleza brindada en cada etapa del proceso. A nuestras familias, cuyo apoyo incondicional fue el motor que nos impulsó a seguir adelante en los momentos más exigentes. A nuestros docentes, en especial a nuestro tutor Luis Camargo Ortega, por su guía, paciencia y compromiso con nuestra formación profesional. Y a la comunidad de desarrolladores de código abierto que, con su trabajo desinteresado, hace posible proyectos como el que aquí se presenta.

## **2 Agradecimientos**

Agradecemos a nuestro tutor Luis Camargo Ortega, guía del seminario de automatización con n8n, por su orientación técnica y disposición constante y a cada persona que aportó conocimiento, tiempo o palabras de aliento durante este proceso.

### 3 Tabla de contenido

1	Dedicatoria .....	2
2	Agradecimientos.....	3
4	Resumen .....	6
5	Palabras clave:.....	7
6	Introducción .....	8
7	Planteamiento del problema .....	10
8	Objetivo general .....	11
9	Objetivos específicos .....	11
10	Marco conceptual y contextual .....	12
10.1	OSINT (Open Source Intelligence).....	12
10.2	Datos públicos y fuentes abiertas.....	12
10.3	Web Scraping.....	12
10.4	APIs de datos públicos .....	13
10.5	Holehe .....	13
10.6	FastAPI.....	13
10.7	n8n.....	14
10.8	Telegram Bot API.....	14
10.9	Desarrollo e implementación del aprendizaje.....	15
11	Arquitectura general del sistema .....	15
12	Detección y clasificación del tipo de consulta .....	16
12.1	Correo electrónico.....	16

12.2	Número de teléfono colombiano.....	16
12.3	Cédula de ciudadanía.....	16
12.4	Placa vehicular colombiana.....	16
12.5	Nombre de empresa.....	16
12.6	Desconocido.....	16
13	Fuentes de datos integradas.....	17
13.1	Registros gubernamentales colombianos.....	17
13.2	APIs de validación internacional.....	17
13.3	Herramientas OSINT especializadas.....	18
14	Manejo de captchas.....	18
15	Resultados obtenidos y beneficios.....	19
15.1	Integración multifuente en una sola interfaz.....	19
15.2	Alcance internacional de la consulta telefónica.....	19
15.3	Análisis de presencia digital.....	19
15.4	Arquitectura modular y extensible.....	20
15.5	Aprendizaje en scraping avanzado y gestión de captchas.....	20
16	Figura del flujo de NullTrace en n8n y código del proyecto.....	21
17	Conclusión.....	22
18	Referencias bibliográficas.....	24

## 4 Resumen

El presente trabajo de grado describe el diseño, desarrollo e implementación de NullTrace, un sistema automatizado de inteligencia en fuentes abiertas (OSINT) accesible a través de un bot de Telegram. El proyecto surge de la necesidad de centralizar la consulta de datos públicos dispersos en múltiples fuentes digitales, combinando registros gubernamentales, APIs de validación internacional y herramientas de análisis de presencia digital en una sola interfaz conversacional.

NullTrace integra fuentes heterogéneas de información pública: registros mercantiles y de tránsito colombianos (RUES, SIMIT, RUNT), APIs de validación telefónica internacional (NumVerify), servicios de análisis de reputación de números (Telguarder) y la herramienta de código abierto Holehe para el análisis de presencia digital a partir de correos electrónicos. El sistema está construido sobre una arquitectura de microservicios con FastAPI como backend central y n8n como motor de orquestación de flujos.

La interfaz principal es un bot de Telegram (@nulltrace\_osint\_bot) que permite al usuario realizar consultas introduciendo un dato de entrada — número de teléfono, correo electrónico, cédula, nombre de empresa o placa vehicular — y recibir información consolidada de múltiples fuentes en tiempo real. Los resultados demuestran la viabilidad de construir un sistema OSINT funcional y escalable usando herramientas de automatización modernas, con potencial de aplicación en verificación de identidad, investigación periodística y análisis de datos públicos.

## **5 Palabras clave:**

OSINT, inteligencia en fuentes abiertas, datos públicos, automatización, bot de Telegram, FastAPI, n8n, web scraping, registros públicos, análisis digital, Holehe, NumVerify.

## 6 Introducción

En el contexto actual de la sociedad de la información, una cantidad creciente de datos de interés público se encuentra disponible en fuentes digitales abiertas: portales gubernamentales, registros oficiales, APIs públicas, directorios de empresas, bases de datos de infracciones y plataformas digitales de todo tipo. Sin embargo, la consulta efectiva de esta información plantea un problema estructural: los datos están fragmentados en múltiples sistemas con interfaces distintas, formatos heterogéneos y barreras técnicas como captchas o estructuras de navegación complejas que dificultan su acceso ágil.

La disciplina OSINT (Open Source Intelligence) aborda precisamente esta problemática, proporcionando metodologías y herramientas para recopilar, procesar y analizar información proveniente de fuentes de acceso público con el propósito de obtener inteligencia accionable. NullTrace nace como una implementación práctica de estos principios: un sistema automatizado que integra múltiples fuentes de datos públicos — tanto nacionales como internacionales — en una interfaz conversacional accesible desde Telegram.

El sistema es capaz de procesar diferentes tipos de entrada — número de teléfono, correo electrónico, cédula de identidad, nombre de empresa o placa vehicular — y consultar automáticamente las fuentes más relevantes para cada tipo de dato, consolidando los resultados en una respuesta estructurada entregada al usuario en segundos. Esta capacidad de integración multifuente es el diferenciador central del proyecto frente a las soluciones existentes, que generalmente se enfocan en una sola fuente o tipo de consulta.

Desde el punto de vista técnico, el proyecto representa un ejercicio completo de ingeniería de software que abarca el diseño de APIs REST, arquitecturas de microservicios, técnicas avanzadas de web scraping, integración de herramientas OSINT especializadas, manejo

de captchas y orquestación de flujos con n8n. Este documento presenta el proceso completo de su desarrollo: desde la identificación del problema hasta los resultados obtenidos y las lecciones aprendidas.

## 7 Planteamiento del problema

El acceso a datos públicos digitales enfrenta una paradoja: la información existe y está disponible en principio, pero en la práctica su consulta requiere conocer múltiples plataformas, navegar interfaces poco intuitivas, resolver verificaciones de seguridad y consolidar manualmente la información obtenida de distintas fuentes. Esta fragmentación convierte una tarea que debería ser trivial — consultar quién es el propietario de un vehículo, verificar si un número de teléfono pertenece a una empresa o confirmar en qué plataformas está registrado un correo electrónico — en un proceso laborioso que puede tomar varios minutos o incluso no ser completable por usuarios no técnicos.

Esta problemática se presenta tanto en el ámbito de los datos gubernamentales colombianos — donde portales como el RUES, el SIMIT y el RUNT operan de forma completamente independiente — como en el ámbito de la información digital internacional, donde herramientas como la validación de números telefónicos o el análisis de presencia digital en plataformas requieren el uso de múltiples APIs y servicios distintos. La ausencia de un punto de acceso unificado para este tipo de consultas representa una ineficiencia significativa para usuarios que necesitan verificar información de forma frecuente.

Frente a esta realidad surge la pregunta de investigación que orienta el presente proyecto: ¿Es posible construir un sistema automatizado, accesible desde una plataforma de mensajería masiva, que centralice la consulta de datos públicos provenientes de fuentes heterogéneas — gubernamentales, APIs internacionales y herramientas OSINT especializadas — de manera eficiente, confiable y usable por cualquier persona?

## 8 Objetivo general

Diseñar e implementar un sistema OSINT automatizado, accesible mediante un bot de Telegram, que permita la consulta centralizada de datos públicos provenientes de fuentes heterogéneas a partir de diferentes tipos de datos de entrada.

## 9 Objetivos específicos

Integrar fuentes de datos públicos gubernamentales colombianos (RUES, SIMIT, RUNT) mediante técnicas de web scraping con Playwright y BeautifulSoup.

Incorporar APIs de validación y análisis internacional (NumVerify, Telguarder) para la consulta de información telefónica global.

Integrar la herramienta de código abierto Holehe para el análisis de presencia digital a partir de correos electrónicos.

Implementar una arquitectura de microservicios con FastAPI como backend central del sistema.

Orquestar los flujos de consulta, enrutamiento y respuesta mediante la plataforma de automatización n8n.

Desarrollar una interfaz conversacional funcional a través del bot de Telegram @nulltrace\_osint\_bot.

## **10 Marco conceptual y contextual**

Para la comprensión del presente proyecto es necesario definir y contextualizar los conceptos sobre los cuales se sustenta su diseño e implementación.

### **10.1 OSINT (Open Source Intelligence)**

OSINT hace referencia a la práctica de recopilar, procesar y analizar información proveniente de fuentes de acceso público con el propósito de obtener inteligencia accionable. Las fuentes OSINT incluyen registros gubernamentales, bases de datos públicas, redes sociales, sitios web, publicaciones académicas, APIs abiertas y cualquier dato disponible sin necesidad de autorización especial (Bazzell, 2021). En el contexto de NullTrace, el OSINT se aplica tanto sobre fuentes colombianas de dominio público como sobre servicios internacionales de acceso abierto o de bajo costo.

### **10.2 Datos públicos y fuentes abiertas**

Los datos públicos son aquellos producidos, recopilados o custodiados por entidades gubernamentales o privadas que los ponen a disposición de cualquier persona sin restricción de acceso. En el contexto digital, estas fuentes incluyen desde portales de registros mercantiles hasta APIs de validación telefónica internacional y herramientas de análisis de presencia en plataformas digitales. La integración de fuentes heterogéneas de este tipo es el reto técnico central que NullTrace aborda.

### **10.3 Web Scraping**

El web scraping es la técnica mediante la cual se extrae información de sitios web de forma automatizada. NullTrace emplea dos estrategias complementarias según el tipo de sitio: para portales con HTML estático se utilizan las librerías requests y BeautifulSoup; para sitios que dependen de JavaScript o implementan captchas dinámicos se utiliza Playwright, un

framework de automatización de navegadores que simula la interacción de un usuario real con la página.

#### **10.4 APIs de datos públicos**

Una API (Application Programming Interface) es un mecanismo que permite a sistemas externos consultar datos de forma programática. NullTrace consume múltiples APIs públicas o de acceso abierto: NumVerify para la validación y geolocalización de números telefónicos a nivel internacional, y Telguarder para el análisis de reputación y reportes de spam asociados a números de teléfono. Estas APIs permiten obtener información estructurada sin necesidad de scraping, simplificando la integración.

#### **10.5 Holehe**

Holehe es una herramienta de código abierto diseñada para determinar en qué plataformas y servicios digitales está registrado un correo electrónico dado, sin necesidad de acceder a la cuenta. Funciona consultando el flujo de recuperación de contraseña de más de 120 servicios y analizando las respuestas para inferir si el correo está registrado. En NullTrace, Holehe se integra como un módulo del backend FastAPI que recibe el correo como parámetro y retorna la lista de plataformas donde fue detectado.

#### **10.6 FastAPI**

FastAPI es un framework moderno de Python para la construcción de APIs web de alto rendimiento, basado en los estándares OpenAPI y JSON Schema. Su arquitectura asíncrona lo hace adecuado para actuar como capa middleware en sistemas que consultan múltiples fuentes externas de forma simultánea. En NullTrace, FastAPI actúa como el backend central que expone los endpoints de consulta al motor de orquestación n8n.

## 10.7 n8n

n8n es una plataforma de automatización de flujos de trabajo de código abierto que permite conectar aplicaciones, APIs y servicios mediante una interfaz visual de nodos. En NullTrace, n8n recibe los mensajes del bot de Telegram, determina el tipo de consulta mediante expresiones regulares en un nodo de código, enruta la solicitud al endpoint correspondiente del backend FastAPI y formatea la respuesta para el usuario. Esta arquitectura permite modificar o extender los flujos sin alterar el código del backend.

## 10.8 Telegram Bot API

Telegram ofrece una API pública para la creación de bots automatizados que pueden recibir y enviar mensajes de forma programática. NullTrace utiliza Telegram como interfaz principal de usuario, permitiendo que cualquier persona con acceso al bot `@nulltrace_osint_bot` realice consultas desde su dispositivo sin necesidad de instalar ninguna aplicación adicional.

## 10.9 Desarrollo e implementación del aprendizaje

El desarrollo de NullTrace se estructuró en torno a tres capas principales que trabajan de forma coordinada: el backend de consultas (FastAPI), el motor de orquestación (n8n) y la interfaz de usuario (Telegram). A continuación, se describe el proceso de implementación de cada componente.

## 11 Arquitectura general del sistema

NullTrace adopta una arquitectura de microservicios desacoplada. El backend FastAPI se ejecuta en un contenedor Docker en el puerto 5000 y expone endpoints REST para cada tipo de consulta soportada. El motor n8n, también dockerizado, opera en el puerto 5679 y se comunica con el backend utilizando la IP del gateway del bridge de Docker para superar las limitaciones de red entre contenedores. El bot de Telegram actúa como punto de entrada: los mensajes del usuario llegan vía webhook a n8n, que los procesa y devuelve la respuesta formateada al chat.

Esta separación de responsabilidades garantiza que cada capa pueda modificarse, reemplazarse o escalarse de forma independiente. Agregar una nueva fuente de datos implica únicamente implementar un nuevo endpoint en FastAPI y crear el nodo correspondiente en n8n, sin afectar el resto del sistema.

## **12 Detección y clasificación del tipo de consulta**

El flujo principal de n8n incluye un nodo de código denominado Expresiones regulares, que analiza el texto recibido desde Telegram e identifica el tipo de dato ingresado por el usuario.

Las reglas de detección son las siguientes:

### **12.1 Correo electrónico.**

Patrón estándar de email, convertido a minúsculas para normalización.

### **12.2 Número de teléfono colombiano.**

Cadena de 10 dígitos iniciando en 3, o número con prefijo +57.

### **12.3 Cédula de ciudadanía.**

Cadena numérica de 6 a 10 dígitos que no coincide con el patrón telefónico.

### **12.4 Placa vehicular colombiana.**

Patrón de 3 letras seguidas de 3 números (automóviles) o 3 letras, 2 números y 1 letra (motocicletas).

### **12.5 Nombre de empresa.**

Activado mediante el comando /negocio seguido del término de búsqueda.

### **12.6 Desconocido.**

Cualquier entrada que no coincida con los patrones anteriores genera una respuesta orientativa al usuario.

Una vez identificado el tipo, el flujo enruta la solicitud a la rama correspondiente del Switch, donde se construye la llamada al endpoint FastAPI adecuado.

## 13 Fuentes de datos integradas

NullTrace integra fuentes de naturaleza muy distinta, agrupadas en tres categorías.

### 13.1 Registros gubernamentales colombianos

RUES (Registro Único Empresarial y Social): permite consultar información de personas naturales con actividad comercial y empresas registradas en cámara de comercio. Se accede mediante scraping directo por URL parametrizada.

SIMIT (Sistema Integral de Información sobre Multas y Sanciones por Infracciones de Tránsito): retorna información de multas, comparendos e infracciones de tránsito asociadas a una cédula o placa. Utiliza un captcha personalizado (qxcaptcha) que se resuelve mediante Playwright interceptando los tokens dinámicos.

RUNT (Registro Único Nacional de Tránsito): provee datos del propietario de un vehículo a partir de la placa. El captcha de este portal es una imagen base64 que se resuelve mediante OCR con pytesseract.

### 13.2 APIs de validación internacional

NumVerify: API internacional que valida números telefónicos de cualquier país del mundo, retornando información sobre el operador, tipo de línea, país de origen y formato estándar del número.

Telguarder: servicio de análisis de reputación telefónica que indica si un número ha sido reportado como spam, scam o llamadas no deseadas, junto con el conteo de denuncias y la fecha del último reporte.

### 13.3 Herramientas OSINT especializadas

Holehe: herramienta de código abierto integrada directamente en el backend FastAPI. Analiza más de 120 plataformas y servicios digitales para determinar en cuáles está registrado un correo electrónico, sin necesidad de acceder a las cuentas.

## 14 Manejo de captchas

Uno de los principales retos técnicos del proyecto fue superar los mecanismos de protección anti-bot de los portales gubernamentales. NullTrace implementa estrategias diferenciadas según el tipo de captcha

SIMIT (qxcaptcha): Playwright intercepta la respuesta de la API del captcha que retorna los tokens dinámicos question, time y nonce, los cuales se incluyen en la solicitud de consulta.

RUNT (imagen base64): la imagen del captcha se extrae del atributo src del elemento img como cadena base64, se decodifica y se procesa con pytesseract y Pillow para obtener el texto por OCR.

Policía Nacional (reCAPTCHA Enterprise): este portal usa reCAPTCHA Enterprise de Google. La estrategia prevista es la integración con el servicio 2captcha, que permite resolver este tipo de captchas mediante una API externa.

## **15 Resultados obtenidos y beneficios**

Al cierre del presente trabajo, NullTrace cuenta con cinco módulos de consulta operativos y completamente funcionales desde el bot de Telegram: análisis de correo electrónico (Holehe), consulta de número telefónico (NumVerify + Telguarder), consulta de cédula (RUES + SIMIT), búsqueda de empresa por nombre (RUES) y consulta de placa vehicular (SIMIT). Estos módulos están integrados con el flujo de n8n y disponibles para consulta en tiempo real.

### **15.1 Integración multifuente en una sola interfaz**

El resultado más significativo del proyecto es la capacidad de consolidar en una sola interacción de Telegram información proveniente de fuentes completamente distintas: desde registros gubernamentales colombianos hasta APIs de validación internacional y herramientas OSINT de código abierto. El usuario no necesita conocer ninguna de estas fuentes ni acceder a ellas directamente.

### **15.2 Alcance internacional de la consulta telefónica**

La integración con NumVerify permite consultar números telefónicos de cualquier país del mundo, no solo Colombia. El sistema identifica automáticamente el prefijo internacional, el operador, el tipo de línea y el país de origen, lo que amplía significativamente el alcance del sistema más allá del contexto local.

### **15.3 Análisis de presencia digital**

La integración de Holehe añade una dimensión de análisis que va más allá de los registros gubernamentales: a partir de un correo electrónico, el sistema determina en qué plataformas digitales, redes sociales, servicios de streaming, herramientas de productividad, plataformas de freelance, entre otras, está registrado ese correo. Esta capacidad es especialmente valiosa para casos de investigación de identidad digital.

#### **15.4 Arquitectura modular y extensible**

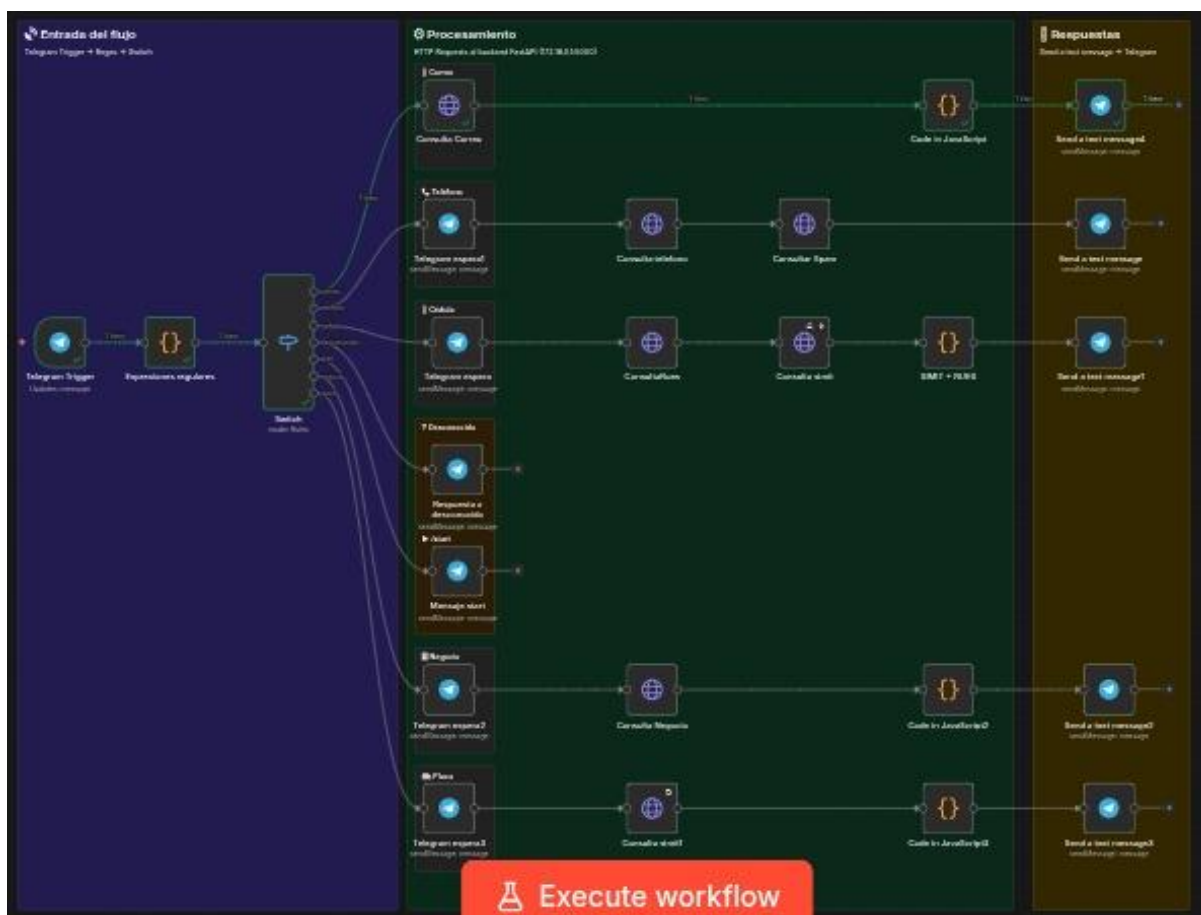
El diseño basado en microservicios con FastAPI y n8n demostró ser altamente escalable durante el desarrollo. La incorporación de cada nueva fuente de datos siguió el mismo patrón: implementar un endpoint en FastAPI, crear la rama correspondiente en el Switch de n8n y conectar el nodo de respuesta. Esta modularidad facilita la incorporación futura de nuevas fuentes sin afectar las existentes.

#### **15.5 Aprendizaje en scraping avanzado y gestión de captchas**

El proyecto implicó el análisis detallado del comportamiento de múltiples portales con tecnologías y protecciones distintas, desarrollando competencias en la selección de la estrategia de scraping adecuada según el tipo de sitio, la interceptación de peticiones de red con Playwright, el procesamiento de imágenes con OCR y la resolución automatizada de captchas de imagen.

## 16 Figura del flujo de NullTrace en n8n y código del proyecto

A continuación, se presenta el diagrama del flujo principal de NullTrace implementado en n8n. El flujo está organizado horizontalmente por secciones: la entrada (Telegrama Trigger → Expresiones regulares → Switch), el procesamiento por rama (HTTP Requests al backend FastAPI y nodos de código para formateo), y la respuesta al usuario (Send a text message → Telegram).



Flujo principal de NullTrace en n8n. Elaboración propia.

**Código fuente de n8n (repositorio de github)**

[NullTraceBot](#)

## 17 Conclusión

El desarrollo de NullTrace permitió demostrar que es posible construir un sistema OSINT funcional, modular y accesible combinando herramientas de código abierto, APIs de datos públicos y plataformas de automatización modernas. La integración exitosa de fuentes, desde registros gubernamentales colombianos hasta APIs de validación telefónica internacional y herramientas de análisis de presencia digital, en una única interfaz conversacional de Telegram constituye el aporte técnico central del proyecto.

Desde el punto de vista del aprendizaje, el proyecto representó un recorrido completo por múltiples áreas de la ingeniería de software: diseño de APIs REST, arquitecturas de microservicios con Docker, técnicas de web scraping estático y dinámico, manejo de captchas mediante Playwright y OCR, consumo de APIs internacionales, integración de herramientas OSINT especializadas y orquestación de flujos con n8n. Cada uno de estos componentes aportó competencias transferibles a contextos profesionales reales.

El enfoque en datos públicos, en lugar de limitarse a un solo país o tipo de registro, resultó ser la decisión arquitectónica más importante del proyecto. Esta orientación permite que NullTrace sea aplicable en distintos contextos: verificación de identidad, debida diligencia empresarial, análisis de seguridad, investigación periodística o cualquier escenario donde se requiera consultar información de fuentes abiertas de forma ágil y centralizada.

Como trabajo futuro, se identifican las siguientes líneas de desarrollo: integración del módulo de antecedentes de Policía Nacional mediante 2captcha, incorporación del RUNT con resolución automática de captcha por OCR, expansión a nuevas fuentes de datos internacionales,

implementación de caché para optimizar tiempos de respuesta en consultas frecuentes y generación de reportes exportables en PDF desde el propio bot de Telegram.

## 18 Referencias bibliográficas

Bazzell, M. (2021). Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information (8th ed.). IntelTechniques.

Federación Colombiana de Municipios. (2024). Sistema Integral de Información sobre Multas y Sanciones por Infracciones de Tránsito — SIMIT. <https://www.fcm.org.co/simit/>

Holehe. (2024). Holehe — OSINT tool to check accounts by email. GitHub. <https://github.com/megadose/holehe>

Microsoft. (2024). Playwright documentation. <https://playwright.dev>

Ministerio de Transporte de Colombia. (2024). Registro Único Nacional de Tránsito — RUNT. <https://www.runt.com.co>

n8n GmbH. (2024). n8n workflow automation documentation. <https://docs.n8n.io>

NumVerify. (2024). NumVerify — Phone number validation API. <https://numverify.com>

Ramírez, S. (2024). FastAPI documentation. Tiangolo. <https://fastapi.tiangolo.com>

Richardson, L. (2024). BeautifulSoup documentation. <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

Superintendencia de Industria y Comercio. (2024). Registro Único Empresarial y Social — RUES. <https://www.rues.org.co>

Telegram. (2024). Telegram Bot API documentation. <https://core.telegram.org/bots/api>

Telguarder. (2024). Telguarder — Phone reputation and spam detection. <https://www.telguarder.com>

