Aplicación de web scraping a la red social facebook, para las investigaciones realizadas en la FGN

Wolfgang Mauricio Muñoz

Trabajo de grado para optar al título profesional:

Maestría en Ciberseguridad y Ciberdefensa

Escuela Superior de Guerra "General Rafael Reyes Prieto"
Bogotá D.C., Colombia

2019

MINISTERIO DE DEFENSA NACIONAL
COMANDO GENERAL DE LAS FUERZAS MILITARES
ESCUELA SUPERIOR DE GUERRA "GENERAL RAFAEL REYES PRIETO"

APLICACIÓN DE WEB SCRAPING A LA RED SOCIAL FACEBOOK, PARA LAS INVESTIGACIONES REALIZADAS EN LA FGN.

WOLFGANG MAURICIO MUÑOZ PROFESIONAL INVESTIGADOR I FISCALÍA GENERAL DE LA NACIÓN

DIRECTOR: FABIÁN VALERO DUQUE

MAESTRÍA EN CIBERSEGURIDAD Y CIBERDEFENSA
TRABAJO DE GRADO PARA OPTAR TÍTULO DE MAGISTER
BOGOTÁ - COLOMBIA
2019

Resumen

Todo documento, en cualquier tipo de formato, que contengan información que permita segregar las características económicas, sociales, culturales, geográficas, psicológicas, etc, se constituye en objetivo de gran interés para la búsqueda de información de carácter investigativo. Los expertos en la búsqueda y análisis de información están de acuerdo en que las fuentes abiertas (OSINT), gestionadas correcta y adecuadamente, son un medio valioso y fundamental para la consecución de información pertinente en los casos que incluyen medios tecnológicos entre los procesos investigativos. Durante las etapas del desarrollo de este documento se logra la implementación de una herramienta tecnológica con un costo mucho más pequeño que el costo de las herramientas licenciadas y con mayor seguridad y conocimiento adquirido que las herramientas libres. Se logra más eficiencia y efectividad en la búsqueda y procesamiento de la información obtenida en la red social Facebook, mediante la técnica de WEB SRAPING, y agilizar de este modo los procedimientos implicados en la investigación penal.

Palabras Claves: OSINT, Fuentes Abiertas, Redes Sociales, Web Scraping.

Abstract

Any document, in any type of format, that allows Identifying economic, social, cultural, geographic, or psychological factors, is very useful in character investigations. Experts in the search and analysis of information agree that open sources, properly managed, are valuable and essential for obtaining relevant information in cases where they also include technological means complementing the investigative processes. During the stages of development, this document achieves the implementation of a technological tool with a

much smaller cost than the cost of licensed tools and with greater security and knowledge acquired than free tools. It achieves more efficiency and effectiveness in the search of and processing of the information obtained in the open sources (OSINT), especially for the social network Facebook, and thus streamlining the procedures involved in a criminal investigation.

Keywords: OSINT, Open Sources, Social Networks.

TABLA DE CONTENIDO

1. Capítulo I

En el capítulo inicial de este documento se desarrollan las etapas introductorias del proyecto, se identifica la problemática y la importancia de utilizar las redes sociales como fuente de información para las labores investigativas de la FGN. Se determinan los límites de la propuesta a desarrollar y se estudia el estado del arte y las ofertas comerciales que dan solución en parte o en gran medida a la problemática identificada. Se realiza un comparativo de las capacidades y características de las ofertas comerciales contra una propuesta de desarrollo In-House.

1.1. Introducción

En el presente documento se estudia la importancia de la información de las fuentes abiertas para las investigaciones de la FGN, se evalúa el estado del arte, las capacidades técnicas y la percepción que posee los investigadores de la Fiscalía en herramientas que apoyan las investigaciones, utilizando las fuentes abiertas como insumo. Se muestra como un desarrollo In-house, de una plataforma que extraiga la información pública de la red social Facebook, es más favorable en términos de recursos y conocimientos que las herramientas pagas y las propuestas Open Source del mercado.

Se realiza una valoración de los riesgos técnicos, tecnológicos y regulatorios de un desarrollo in-house. Y se muestra los procesos de diseño, desarrollo y prueba de CRAWLERFB, plataforma para extraer la información pública de un usuario o un grupo de usuarios de la red social Facebook, aportando información conducente a los investigadores de la Fiscalía, mejorando así sus capacidades de análisis y estudio de casos con vínculos en la red social.

1.2. Planteamiento del problema

Durante las etapas investigativas en la FGN y más específicamente en la recolección de elementos materiales probatorios y en el entendimiento general de los fenómenos y modus operandi, existen actividades de carácter estratégico desarrolladas por los analistas, que exigen el tratamiento de información, y cuyos análisis, en aras de una justicia oportuna, son requeridos rápidamente y en corto tiempo. Las implementaciones que se están realizando y los proyectos actuales planteados por las direcciones de Planeación y Políticas Públicas y las diferentes policías judiciales, indican un conjunto de posibilidades y escenarios en donde la información será cada vez más abundante, y, por lo tanto, su tratamiento será mucho más complejo. Por otra parte, La FGN, como el ente investigativo del Estado y enmarcado en los derechos constitucionales de todos los ciudadanos, debe acceder a la mayor cantidad de información posible con la intención de construir hipótesis fuertes, fundamentadas en las circunstancias de tiempo, modo y lugar. Hoy la Internet es un medio muy expedito para la consecución de este tipo de información, especialmente a través de las redes sociales, en donde se exponen públicamente las actividades, comportamientos, lugares y personas relacionadas con algún tipo de investigación. Aunque la gran mayoría de las personas que trabajan en la FGN como investigadores, analistas y fiscales son conscientes de esta situación, aún no se ha desarrollado algún tipo de estrategia para obtener provecho de la información contenida y expuesta en el universo de las redes sociales. Estas actividades requieren de tiempo y paciencia, por su carácter manual, porque requieren de algún tipo de conocimiento técnico específico, o porque sus resultados pueden quedar en términos parciales, dada la poca profundidad con que a menudo se encuentran los

perfiles y las pocas herramientas visibles para obtener un alcance más certero y real de lo aparentemente público.

Por las razones anteriores, este trabajo pretende responder:

¿Es posible con herramienta propia, de bajo costo e inversión, mejorar el tiempo recolección, y la calidad de información apropiada de la red social Facebook, para las investigaciones que adelanta la FGN?

1.3. Objetivos.

1.3.1. Objetivo general.

Desarrollar e implementar una herramienta software que permita identificar, ordenar y presentar la información encontrada de un individuo o grupo, sus características y relaciones, en la fuente pública de la red social Facebook, a fin de conducir y facilitar los procesos investigativos acordes al direccionamiento estratégico en la FGN.

1.3.2. Objetivos específicos.

Los siguientes son los objetivos específicos planteados para el desarrollo del trabajo:

- Conocer las técnicas y fuentes de recolección para análisis de información de fuentes abiertas aplicables a Facebook.

- Identificar los datos e insumos que se pueden obtener de Facebook, y sean de utilidad a los procesos de acción penal desarrollada en la FGN.

- Diseñar, crear e implementar una plataforma que automatice la recolección estudio y análisis de información de la red social de mayor uso en Colombia, Facebook.

- Utilizar las herramientas tecnológicas, capacidades humanas y de infraestructura disponible que generen habilidades y conocimientos específicos en la búsqueda automatizada de información en la red social Facebook.

1.4. Motivación

Para la FGN, existe la necesidad latente de usar la información que aportan las fuentes abiertas, para conducir y alimentar a las investigaciones judiciales. La información obtenida por este medio permite crear mapas relacionales y generar escenarios completos de los individuos o grupos que puedan estar involucrados en conductas criminales punibles. En estos ambientes tecnológicos tan cambiantes, ya no basta con usar las herramientas comerciales; se hace necesario crear, modelar y apropiarse del conocimiento para enfrentar a los nuevos desafíos sociales y tecnológicos. Este proyecto brinda la posibilidad de explorar y comprender el conocimiento teórico y práctico del funcionamiento de la red social Facebook, para extraer información de interés y representar las relaciones de un usuario o un grupo de usuarios en la red social.

1.5. Contexto

La interacción social actual, a través de los medios tecnológicos modernos, ha traído consigo retos y nuevas posibilidades en las formas como se abordan las investigaciones. La información pública y el desconocimiento general de las restricciones provistas por los desarrolladores de redes sociales han dejado al alcance de quien quiera explotar en estas alternativas la posibilidad de búsqueda de relaciones y patrones que son de interés para cualquier tipo de investigación. Estas características de las redes sociales y de la interacción de sus usuarios, pueden ser identificadas a través de procesos manuales, de acuerdo al tipo de objetivo. Sin embargo, la construcción de relaciones a gran escala y de fuentes de

información que puedan ser relacionadas con otro tipo de orígenes, como bases de datos internas o públicas, requieren de procesos automatizados en los cuales, el analista tome su rol especializado en análisis y toma de decisiones y no gaste su tiempo en la búsqueda, recolección y organización de la información.

Los medios tecnológicos actuales que conforman los servicios de inteligencia de los países más avanzados del mundo constituyen una serie de acciones muy complejas y de distinta índole entre ellas, los sistemas de búsqueda y transmisión de señales electromagnéticas, la obtención y análisis de imágenes, la intercepción y el análisis de comunicaciones electrónicas, entre otras, y aquellas actividades tendientes a buscar información en fuentes abiertas.

OSINT es una de las clases de inteligencia que pretende crear nuevos productos de información a partir de datos tomados de las fuentes abiertas. Éstas están conformadas por todos aquellos materiales documentales de carácter público, gratuito o pagado, en cualquier tipo de formato. Los elementos que conforman a las fuentes abiertas son múltiples y de naturaleza muy variada. Estos pueden ser: "obras de referencia, bases de datos, monografías, publicaciones seriadas (tanto científicas como de información general o especializada), literatura gris, sitios y página web, colecciones de imágenes, emisiones radiofónicas o de televisión, grabaciones sonoras y audiovisuales." (Felip I Sarda, 2004)

Ahora bien, las "fuentes abiertas" se definen por la conformación y características de sus elementos constitutivos, aunque esto implique una definición muy elemental y sencilla: "todo documento impreso o electrónico de acceso y uso público en cualquier idioma, que contenga datos políticos, culturales, económicos, militares, científicos, técnicos,

sociológicos, geográficos, etc." (Felip I Sarda, 2004). Siendo esto de gran interés para la búsqueda de información.

Los expertos profesionales en la búsqueda y análisis de información están de acuerdo en que las fuentes abiertas, gestionadas correcta y adecuadamente, son un medio muy valioso y fundamental para la consecución de información pertinente a cualquier caso en proceso de investigación. Esta modalidad, con su gran cantidad, variedad y riqueza de sus fuentes y contenidos, ha adquirido una importancia muy significativa con el advenimiento de la multidimensionalidad de la seguridad digital, cibernética y nacional de los Estados. Esto ha posicionado a las fuentes abiertas como imprescindibles e incluso, se les ha dado mayor importancia que a otras modalidades de obtención de información.

La mayoría de las grandes invenciones tecnológicas de uso corriente, se originaron en la segunda guerra mundial e igualmente en el contexto de la guerra fría entre los Estados Unidos y la Unión Soviética. Hoy, las tecnologías heredadas de la guerra fría se han perfeccionado de tal manera que las amenazas convencionales de un enemigo identificado concretamente y con una táctica de batalla particular, se han trasladado hacia otros escenarios y grupos que actúan bélicamente en una estructura no convencional y de una volatilidad inimaginable que es utilizada como su principal medio de ocultación. A este fenómeno se denomina: "Amenazas asimétricas" (Martín, 2010). Esto se vio reflejado el 11 de septiembre de 2001, cuando a partir del ataque terrorista de las torres gemelas, los Estados Unidos han dado importancia significativa a las técnicas de inteligencia, especialmente a aquellas que administran la información obtenida en fuentes abiertas. La inteligencia militar de las fuerzas armadas estadunidenses incorporó a OSIF (Open Source

Información) y OSINT (Open Source Intelligence), en sus labores diarias de inteligencia y seguridad del país.

OSINT, en su base histórica, consistía en clasificar la información de interés, para una agencia específica, información de medios como prensa y radio. Para el siglo XXI, la sociedad se ha volcado a la era de la información digital y escenarios como internet ofrecen una gran disponibilidad de recursos para explotar OSINT. En su corto tiempo de evolución la Internet, ha cambiado radicalmente. En 1994, el servicio de inteligencia de los Estados Unidos afirmaba que "en las redes de distribución solamente había 450 direcciones útiles, y que un 99% de INTERNET no tenía valor desde el punto de vista de la inteligencia, sino que estaban dedicadas a la pornografía, opinión o anuncios comerciales." (Martín, 2010) Hoy la realidad es otra. El cambio ha sido radical. Internet es el medio fundamental de información y comunicación más voluminosa y de mayor uso en el mundo.

En los orígenes de Internet, el usuario era un elemento pasivo que solo recibía la información dispuesta por los diferentes portales de su interés. Posteriormente surge el primer gran cambio y nace la web 2.0, (Social Media), la cual brinda la posibilidad de que cualquier usuario de internet publique sus contenidos, y surgen las herramientas que facilitan a estos usuarios e internautas a publicar cualquier clase de contenido. "La Web 2.0 representaba una etapa nueva caracterizada por el surgimiento de aplicaciones que permiten la participación directa en Internet de la gente común". (Moya, 2012).

De esta manera nacen los blogs, wikis, redes p2p, y en especial, las redes sociales donde el internauta o navegante no es un elemento pasivo en la comunicación, y donde se presenta un crecimiento evidente. En la red social Facebook, el número de usuarios que acceden,

sigue creciendo exponencialmente convirtiéndose en la red social más usada del planeta con más 2.1 millones de usuarios activos, en comparación con Instagram que reporta 895 millones y de Twitter con 251 millones de usuarios, según el estudio, Digital 2019 Global Digital Overview, realizado en enero 2019, por las empresas "We are Social" y "Hootsuite". Colombia no es la excepción, La penetración de las redes sociales en la población colombiana es del 68%, y la población mayor de 13 años vinculados a la red social es del 83%. (We are social ; Hootsuite, 2019)

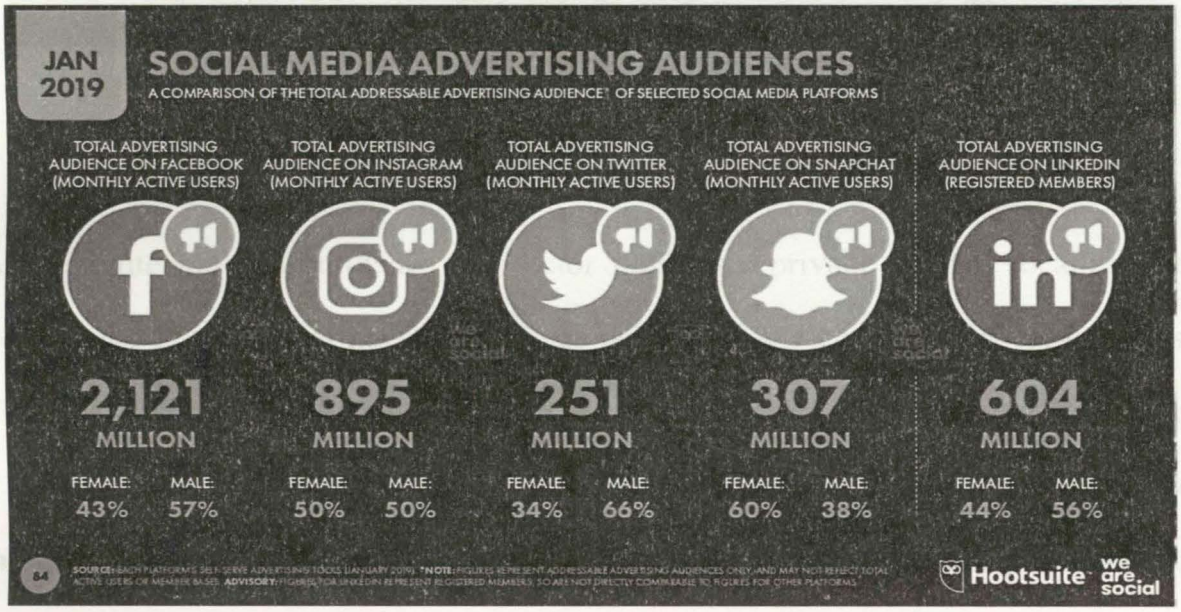


Ilustración 1 Comparativa audiencia Redes sociales, fuente Datareportal

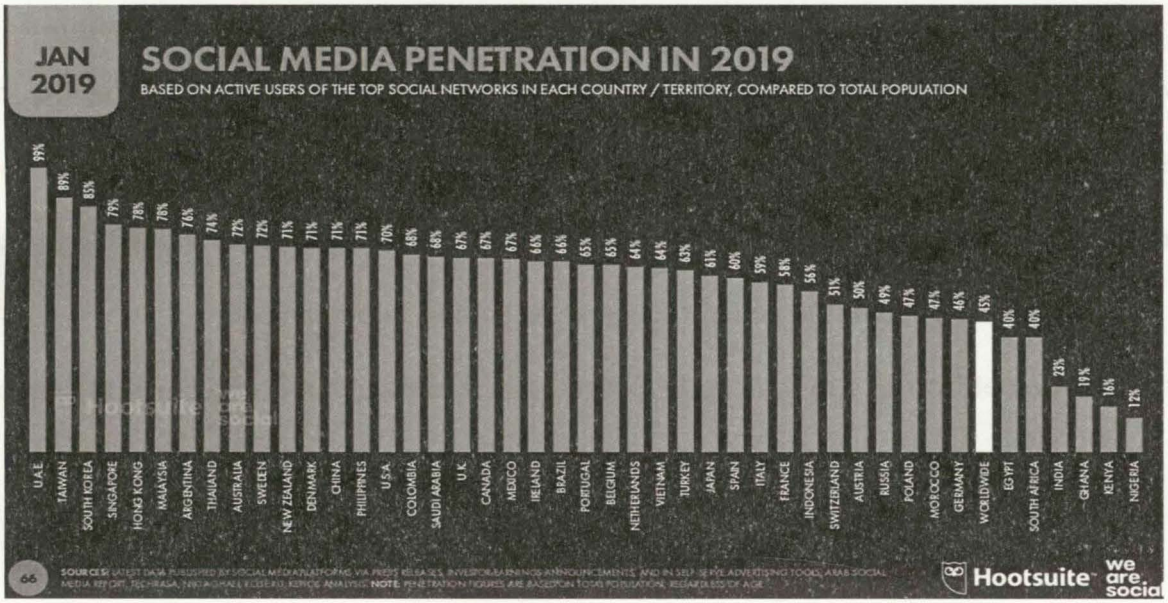


Ilustración 2 Penetración Social Media por países, Fuente Datareportal.

Actualmente muchas entidades del sector público y privado se han focalizado en el procesamiento de esta información con fines publicitarios y comerciales, como el caso de Netflix que realiza filtrado de su contenido según las tendencias visibles y las respuestas de sus seguidores en las redes sociales, (Fernadez & Quevedo, 2018) y en el caso de Cambridge Analytica de quien se presume que influencio las elecciones presidenciales de EEUU en 2016, explotando y perfilando millones de usuarios de la red social Facebook, para publicitar al candidato Donald Trump. (Rosenberg & J. X. Dance , 2018)

La FGN, es el ente encargado de la investigación judicial en Colombia, nace con la formulación de la constitución de 1991, y tiene el reto de enfrentar y realizar la acción penal de manera eficaz y eficiente. Para lograr este objetivo es necesario tener siempre un estudio del estado del arte de las técnicas y las herramientas que le permitan cumplir con su misión de la ejecución de la política criminal del Estado.

1.6. Estado del Arte.

En la realización de OSINT orientado en los recursos WEB de manera efectiva se pueden hallar múltiples publicaciones, herramientas y repositorios, que permiten agilizar o automatizar algún criterio de búsqueda, y reducen el espectro de información para su posterior análisis y procesamiento, entre estas herramientas algunas de las más importantes y que más información pueden aportar para brindar la solución a la problemática planteada son:

1.7. Herramientas o plataformas libres

En la categoría Open Source o de libre uso, se han identificado plataformas WEB o repositorios que permiten agilizar o automatizar algún criterio de búsqueda, estos productos reducen el espectro de información para su posterior análisis y procesamiento, en la búsqueda de una solución capaz recolectar datos de la red social Facebook los desarrollos más común mente usados son:

1.7.1.1. Inteltechniques.

Es la plataforma WEB de consulta, más utilizada por los investigadores adscritos a la FGN, funcional hasta junio 2019 y constituyo la base para muchas otras propuestas gracias a su compendio de herramientas, la capacidad específica para le red social Facebook, consistía en poder en crear links de consulta, relaciones entre perfiles e identificar la información pública de un perfil, hallada en perfiles de otros usuarios.

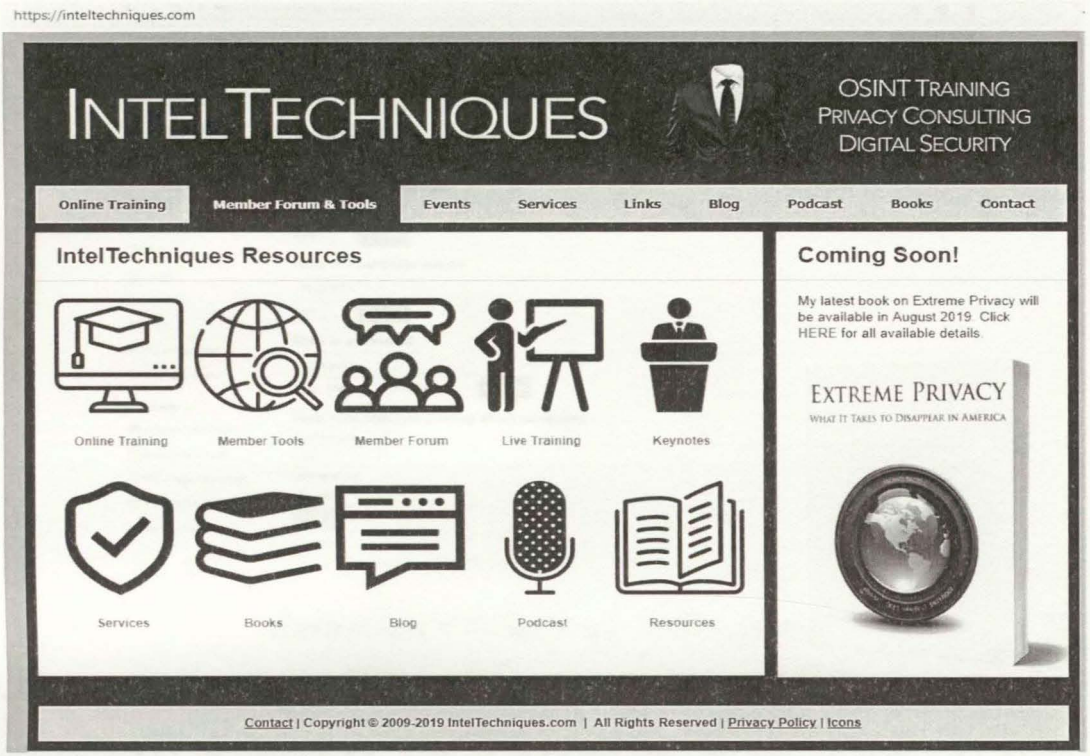


Ilustración 3 Plataforma IntelTecniques

1.7.1.2. IntellegenceX.

Plataforma web, con un interesante compendio de herramientas para realizar consultas e investigaciones en diferentes fuentes abiertas, en el caso de Facebook posee actualmente la única alternativa de búsqueda que permite profundizar un poco más sobre la red social.

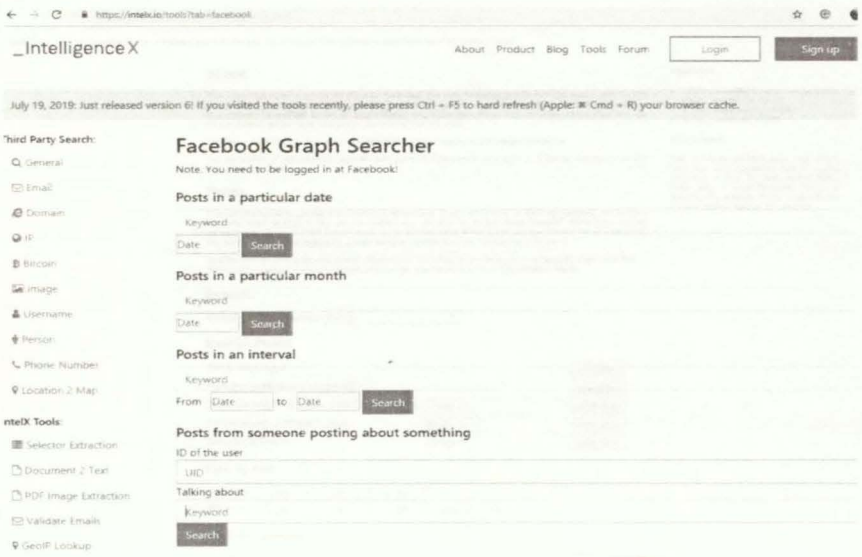


Ilustración 4 Plataforma IntelligenceX

1.7.1.3. Fb-Search.

Repositorio GIT[1], el funcionamiento es muy similar a la herramienta de IntelligenceX, y permite identificar publicaciones, fotos, páginas y eventos en la información pública de un perfil en Facebook, mediante la construcción de URL´s similares o iguales a las de la red social, (Ver ilustración 5).

1.7.1.4. OSINT Tools.

Plataforma web que posee un compendio de herramientas para la aplicación de OSINT, muy similar a INTELTECHNIQUES, y presenta el mismo problema, en el que para Facebook perdió funcionalidad en Junio 2019, debido a las políticas de seguridad implementadas por la red social para proteger sus usuarios.

[1] https://github.com/sowdust/searchbook

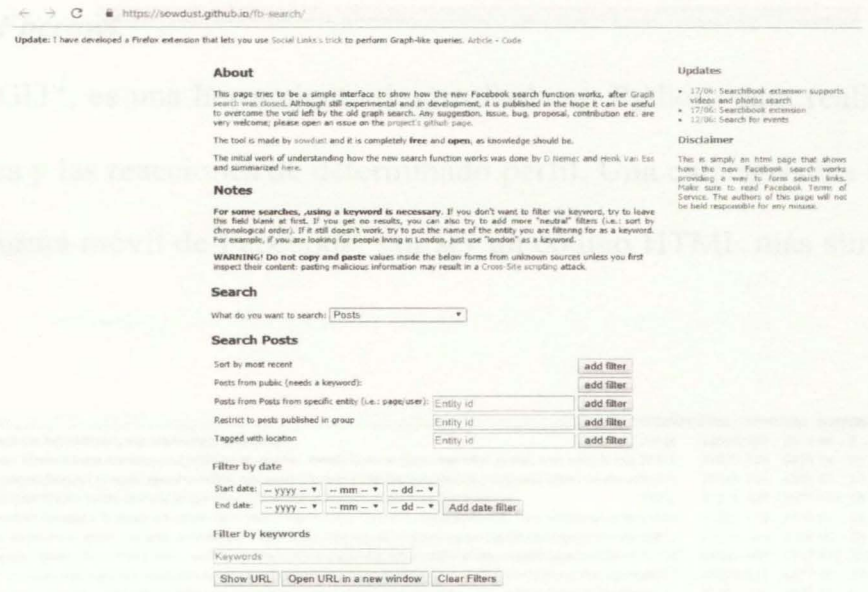


Ilustración 5 Plataforma FB-Search

Ilustración 6 Respuesta Facebook ante plataformas de búsqueda.

En la búsqueda de este tipo de proyectos, se identificaron múltiples repositorios que pretenden realizar, mediante la técnica de scraping y crawling, la recolección de información de un perfil de la red social Facebook, entre los más relevantes y más actualizados se identifican.

1.7.1.5. Fbcrawl.

Repositorio GIT[2], es una herramienta desarrollada en Python, para realizar descarga de las publicaciones y las reacciones de determinado perfil. Una característica interesante es el uso de la plataforma móvil de Facebook, por ser un código HTML más simple de entender y analizar.

source	date	text	reactions	likes	ahah	love	wow	sigh	grrr	comments	url
Donald J. Trump	2018-08-26	Thanks to Republicans, our economy is booming!	25930	22809	502	2513	49	6	51	2103	/story.php?st
Donald J. Trump	2018-08-25	Social Media Giants are silencing millions of people... People have to figure out what is real, and what is not,	25319	21873	624	2453	94	91	184	3605	/story.php?st
Donald J. Trump	2018-08-24	"President Donald Trump's massive rallies and primary endorsements are having a dramatic effect on the 2(	31474	26604	665	4100	47	17	41	2631	/story.php?st
Donald J. Trump	2018-08-26	President Trump holds meeting in the Oval Office this afternoon regarding North Korea.	24332	21271	637	2257	106	18	43	3295	/story.php?st
Donald J. Trump	2018-08-24	President Donald J. Trump's schedule for Friday, August 24th: · Travel to Columbus, OH · Visit Nationwide Ch	13188	11249	259	1595	41	11	33	1976	/story.php?st
Donald J. Trump	2018-08-24	Our economy is setting records on virtually every front. The world is respecting us again! Companies are mc	17871	14918	594	2266	39	13	41	2443	/story.php?st
Donald J. Trump	2018-08-24	The only place where we're doing badly is in the Fake News media, where 90% of coverage is negative! Every	76336	65197	2095	7629	510	392	513	7560	/story.php?st
Donald J. Trump	2018-08-23	We must change our very weak and ineffective immigration laws NOW. I have a plan that will put the intere	40027	34888	515	4457	56	17	94	3265	/story.php?st
Donald J. Trump	2018-08-23	President Donald J. Trump's schedule for Thursday, August 23rd: · Roundtable on the Foreign Investment Ris	9464	8129	236	1035	27	5	32	1457	/story.php?st
Donald J. Trump	2018-08-23	Looking forward to my next BIG RALLY in Evansville, Indiana on Thursday, August 30th at 7 PM CT!	18941	15897	377	2547	43	15	62	2376	/story.php?st
Donald J. Trump	2018-08-22	Longest bull run in the history of the stock market, congratulations America!	36610	31381	539	4478	158	11	43	2932	/story.php?st
Donald J. Trump	2018-08-22	After years of being ripped off by other countries, we're not going to be ripped off anymore.	23434	19915	616	2791	43	11	58	2609	/story.php?st
Donald J. Trump	2018-08-22	President Donald J. Trump's schedule for Wednesday, August 22nd: · Lunch with the Secretary of Defense · Pr	14521	12266	538	1625	48	8	36	2985	/story.php?st
Donald J. Trump	2018-08-21	Governor Jan Brewer is right! "President Trump will stay the course and the American people will ultimatel	15919	13584	495	1764	34	7	35	2119	/story.php?st
Donald J. Trump	2018-08-21	"West Virginia is being destroyed by one of the worst epidemics in American history." We need ACTION!	13317	11048	114	544	139	1351	121	2528	/story.php?st
Donald J. Trump	2018-08-21	Mike Braun is a great man and the right guy to represent Indiana in the U.S. Senate!	10437	9269	199	902	25	11	31	1209	/story.php?st
Donald J. Trump	2018-08-21		31601	27755	424	3198	64	45	115	3918	/DonaldTrun
Donald J. Trump	2018-08-21	A Blue Wave means Crime and Open Borders. A Red Wave means SAFETY and STRENGTH!	74386	64726	1751	7570	115	40	184	5767	/story.php?st
Donald J. Trump	2018-08-18	If the Democrats ever got back into power, they would immediately implement their socialist agenda at YOU	46411	36520	1740	885	630	673	5963	9742	/story.php?st
Donald J. Trump	2018-08-20	We're doing big things for the American workforce. This is perhaps one of the greatest economic revivals in	17632	15289	510	1728	40	10	55	1654	/story.php?st
Donald J. Trump	2018-08-20		35493	29814	684	4841	62	18	74	3489	/DonaldTrun
Donald J. Trump	2018-08-17	Looking forward to a big rally in Charleston, West Virginia this Tuesday, August 21st at 7 PM ET. Get your FR	17572	14839	331	2268	46	12	76	2554	/story.php?st
Donald J. Trump	2018-08-17		22143	18694	529	2798	48	10	64	2420	/DonaldTrun
Donald J. Trump	2018-08-17	There is nothing that I would want more for our country than true FREEDOM OF THE PRESS.	26577	22810	873	2715	52	36	91	3686	/story.php?st
Donald J. Trump	2018-08-17	We are being respected again, and we are not backing down.	45382	38180	1273	5764	74	11	80	4001	/story.php?st
Donald J. Trump	2018-08-17	President Donald J. Trump's schedule for Friday, August 17th: · Travel to Southampton, NY · Roundtable with	11556	9951	264	1272	24	4	41	1570	/story.php?st
Donald J. Trump	2018-08-16	Great Cabinet meeting today at The White House!	32348	28294	514	3358	107	13	62	3100	/story.php?st
Donald J. Trump	2018-08-17	Big Tech bias is very real! Social media platforms are "actively seeking to SILENCE and CENSOR conservative	16964	12570	420	405	337	308	2924	2745	/story.php?st

Ilustración 7 Resultados análisis FBCrawl

1.7.1.6. Facebookcrawler.

Repositorio GIT[3], desarrollado por Vinay Bharadwaj y Nishith Agarwal, pertenecientes a Georgia Institute of Technology, en el cual pretenden estudiar patrones en las redes sociales mediante el uso de la API Graph de Facebook, aunque este proyecto no utiliza las técnicas de Crawling o Scraping, obtiene y produce información de interés para los analistas a partir de la información pública de 30 grupos y 2000 perfiles en la red social.

[2] https://github.com/rugantio/fbcrawl
[3] https://github.com/jedivind/Facebookcrawler

Al realizar un comparativo de las bondades de las herramientas Open Source, analizadas, se puede identificar que aunque el potencial de uso eran las búsquedas avanzadas, que ofrecían las plataformas WEB, estas ya no son tan útiles, debido a las restricciones implementadas por Facebook, por lo tanto el uso de todas las herramientas se fundamentan en la información pública de los perfiles de la red social.

Tabla 1

Atributos herramientas libres

Características	Inteltechniques	Intellegencex	Fb-search	Osint tools	Fbcrawl	Facebookcrawler
Búsquedas Información Púbica de perfil	Hasta Junio 2019	SI	SI	Hasta Junio 2019	NO	SI
Búsqueda de reacciones (Like, Love, Sad, etc…)	NO	NO	NO	NO	SI	SI
Búsqueda de Amigos	Hasta Junio 2019	NO	SI	Hasta Junio 2019	NO	SI
Búsqueda de Post o publicaciones	Hasta Junio 2019	SI	SI	Hasta Junio 2019	SI	SI
Identifica nivel de relación entre usuarios	NO	NO	NO	NO	NO	NO
Identifica factores comunes de educación y empleo entre usuarios	NO	NO	NO	NO	NO	NO

Podemos concluir de este análisis que las plataformas estudiadas en este caso, no realizan la identificación de patrones comunes de academia, ciudad o empleo entre los nodos amigos de un perfil. No se identifica el nivel de relación entre los perfiles en la red social.

Las herramientas analizadas no generan gráficos de interacción o no generan el insumo para crear los gráficos de relación entre los nodos y los perfiles.

La información que se puede recolectar mediante el uso de API Graph de Facebook, es limitada a los niveles de seguridad del perfil, pudiendo solo conseguir la información pública de un perfil o grupo. Además, se evidencia la urgencia de la plataforma de la red social en diseñar e implementar, mecanismos que resguarden la información consignada por sus usuarios.

1.8. Herramientas Licenciadas con Costo.

Actualmente se ha observado por parte de los directivos de la de la FGN, un directo interés por herramientas que permitan escanear redes sociales, que permitan obtener y procesar la información de ámbito público, se evidencia que existe un mercado naciente para ser explotado por las diferentes compañías con herramientas enfocadas a las agencias estatales de investigación e inteligencia. En los últimos dos años se realizaron pruebas a diferentes propuestas del mercado con herramientas como MNEMO (NERV), FUTURE SPACE, WEBINTPRO (HIWIRE) y CERTERIAN (ROGUEEYE). Los resultados de estas pruebas fueron tabulados en un cuadro comparativo con las diferentes capacidades evaluadas.

Tabla 2

Capacidades herramientas Pagas evaluadas durante licitación.

Capacidades	MNemo	Futurespace	Hiwire	Rouge Eye
Búsqueda en fuentes abiertas Google	si	si	parcialmente	Si
Búsqueda información redes sociales	parcialmente	si	parcialmente	si

Extracción contactos redes sociales	no	parcialmente	No	Si
Búsqueda por abonado celular	parcialmente	si	Si	Si
Búsqueda por correo electrónico	parcialmente	si	Si	Si
Búsqueda por nombres	si	si	si	No
Búsqueda por Url	no	si	si	Si
Copia de seguridad automática	si	si	no	Si
Manejo grandes volúmenes de información	si	si	si	Si
Análisis grafico de entidades y vínculos	parcialmente	si	parcialmente	Si
Integración bases de datos	si	parciamente	no	No
Ubicación geográfica	si	si	si	Si
Uso de avatares	no	si	si	Si
Internet profundo	no	no	si	Si
Soporte técnico local	si	no	no	No
Ambiente grafico amigable	parcialmente	si	parcialmente	Si

Adicionalmente se recibieron propuestas de herramientas como VOYAGER ANALYTICS y 4iQ (Telefónica), las cuales no se alcanzaron a probar y testear en el momento de este trabajo.

Es evidente que estos desarrollos presentan grandes capacidades de búsqueda, análisis y presentación de la información, pero es necesario analizar los riesgos y costos que presentan estas soluciones. Estas propuestas varían entre los 3.000 mil y los 6.000 mil millones de pesos en su costo y se limita su uso y soporte al tipo de licenciamiento adquirido, se desconocen sus acciones y comportamientos en el Backend y tienen grandes requerimientos de hardware y software.

1.9. Bondades y desventajas.

Las herramientas estudiadas presentan tanto como bondades y desventajas, y aunque todas permiten obtener datos de algún tipo, algunas en mayor medida, es el analista de la FGN, quien, durante las etapas de la investigación, agrega valor a estos datos y los convierte en información relevante, el objetivo de las herramientas es presentar los datos de la manera más clara y directa a los ojos de los analistas.

1.9.1. Bondades y desventajas Herramientas y plataformas Libres

- Aumento Capacidades Técnicas: El uso, la implementación y la explotación de herramientas y plataformas libres, constituyen una gran fuente de conocimientos para los investigadores que se perfilan en tareas orientadas a las redes sociales. Este aumento de capacidades se da porque no existe un equipo de soporte, o de instalación que solucione los problemas durante las diferentes etapas de la herramienta. Motivo que obliga al personal a investigar y explorar en las opciones que se encuentran en los diferentes medios y los diferentes repositorios.

- Poca o nula inversión. Las herramientas de uso libre, no representan costos para institución, desde que se cumplan las condiciones del licenciamiento, pero si pueden constituir una buena fuente de información y entrenamiento en la búsqueda de información en fuentes abiertas.

- Conocimiento del Backend: Dado que la mayora de proyectos libres y repositorios se encuentran sus códigos fuentes, es posible estudiar el comportamiento de la herramienta, y se pueden adaptar o extraer fragmentos de esta, que sean de utilidad a desarrollos propios en la FGN.

- Ciclo de vida de corta duración: El ciclo de vida de las herramientas orientas a fuentes abiertas es directamente proporcional a la fuente y a la cantidad de usuarios activos. Los entornos digitales son cambiantes y así como hoy en día en red social Facebook confluyen más de 2.1 millones de usuarios, (We are social ; Hootsuite, 2019) en pocos meses esa población podría migrar hacia otra plataforma, y las inversiones en herramientas para obtener información de esa fuente en particular, perderían su valor. Esa es la gran ventaja de usar plataformas libres, la inversión y el riesgo son menores que con una plataforma paga.

- Información Limitada: las herramientas y plataformas de libre uso, son básicamente proyectos que buscan acercase y apropiarse a nuevas técnicas y conocimientos, por lo tanto, su producción de información, y menos para un caso como este, puede considerarse como definitiva o concluyente, es más bien información de tipo conducente que permite guiar y orientar los caminos de la investigación.

1.9.2. Bondades y desventajas Herramientas y plataformas Pagas.

- Desconocimiento del Backend: Al estudiar rango de ofertas del mercado en soluciones y herramientas para realizar OSINT en fuentes abiertas y redes sociales, se evidencia que los principales proponentes son pertenecientes a países potencias y líderes en labores de inteligencia y espionaje. Al desconocer cuál es el funcionamiento y el comportamiento del Backend de las propuestas presentadas, se generan dudas, ¿Es seguro comprar una herramienta de la cual desconoces su funcionamiento interno? ¿Quiénes pueden visualizar la información, que de hecho para la FGN es de carácter reservada? ¿Es seguro vincular las bases de datos propias a herramientas diseñadas por las potencias en inteligencia y espionaje?

- Altos Costos: la característica común encontrada en las herramientas pagas, fue su alto costo, según la FGN en su reporte presupuestal del 2019, (Fiscalía General de la Nacón, 2019) en los ítems, FORTALECIMIENTO DE LA CAPACIDAD TÉCNICO-CIENTÍFICA DE LOS LABORATORIOS Y GRUPOS DE CRIMINALÍSTICA DE LA FISCALÍA A NIVEL NACIONAL. Posee una asignación presupuestal de $6.375.000.000 de pesos. Y en EFECTIVIDAD DE LA INVESTIGACIÓN PENAL Y TÉCNICO CIENTIFICA se observa una asignación de $13.250.000.000. Al comprar por la FGN alguna de las herramientas ofrecidas, se estaría comprometiendo aproximadamente entre el 20% y el 30% de la sumatoria de estos rubros. Por lo tanto, los altos costos de las herramientas evaluadas es su mayor impedimento para su consecución.

- Licenciamiento: Al posible valor de la compra de alguna herramienta propuesta, se debe sumar su costo de licenciamiento, el cual varía entre 10% y el 15% anual del costo de la solución. Sin mencionar que algunas de las plataformas evaluadas cobran adicional el número de consultas efectuadas, elevando notoriamente el costo total.

- Cantidad de Información recolectada: Es de especial reconocimiento la cantidad de información que se logra recuperar con las propuestas de pago, sus capacidades son muy distantes de las herramientas libres, o de un desarrollo in-house, esto es dado a la cantidad de recursos que se les han dedicado a los desarrollos y al tiempo y experticia que las potencias mundiales en inteligencia e investigación han logrado.

- Ciclo de vida: De igual manera que con las herramientas libres, el ciclo de vida y de funcionamiento de estas propuestas depende de agentes externos y de plataformas externas, Así como se han dado cambios en la estructura y la seguridad de Facebook, así se darán en

otras redes sociales, el surgimiento de nuevas plataformas, la evolución y la percepción de los usuarios cambia continuamente obligándolos a usar nuevas estructuras y a olvidar las actuales.

- Uso de Avatar: La tendencia para la investigación en fuentes abiertas, apunta hacia el uso de múltiples avatars, y se pudo observar que las herramientas pagas vinculan esta opción en su propuesta, esto reduce el tiempo transcurrido entre el inicio de la investigación y la obtención de la información. Los avatares en la FGN deben considerarse como agentes virtuales y con una asignación tal se le debe dedicar tiempo y preparación tanto como para parecer real en la red social, como también para estar lo más cerca posible del objetivo. El uso de avatares es una bondad que ya implementan la mayoría de las herramientas pagas.

1.9.3. Bondades y desventajas de un desarrollo In-house.

- Adquisición de conocimiento y capacidades: la gran ventaja que se puede obtener de un desarrollo in-house, son las capacidades adquiridas durante el ciclo de vida del proyecto. Estos conocimientos pueden ser aplicados a nuevas propuestas y nuevos desarrollos, además servir como materia prima en la profesionalización de los investigadores en fuentes abiertas y redes sociales.

- Diseño a la medida: Al diseñar un software a la medida, se optimizan recursos según las necesidades y las capacidades de la FGN, el desarrollo a la medida puede consumir más tiempo que la compra de una solución ya preparada, pero la información obtenida es dirigida puntualmente a las necesidades específicas de la entidad.

- Menor costo e inversión: Al estudiar las propuestas económicas de herramientas pagas, un desarrollo a la medida costaría entre el 0,5% y el 1% de la más barata de las soluciones ofrecidas.

- Conocimiento Comportamiento interno: Al realizar un desarrollo In-house, se conocen las funciones internas del programa y para una institución como la FGN es necesario mantener estricto control de la información de sus investigaciones por su carácter de reservada.

- Proyección y nuevos desarrollos: al desarrollar in-house, se incrementan las capacidades del grupo de desarrollo, incrementa el espectro de futuros proyectos y posibilidades, lo cual influye en el ciclo de vida del proyecto porque se han generado la capacidad de relacionarse con los entornos cambiantes de las redes sociales y de las fuentes abiertas.

- Unificación Bases de datos. Al poseer el control y flujo de la información en un desarrollo in-house, es posible vincular al proyecto las bases de datos propias de la FGN, minimizando el riesgo de fugas de información reservada.

2. Capitulo II

En el capítulo intermedio, y con la delimitación y propuesta tecnológica obtenida del capítulo anterior, se procederá a identificar las necesidades teóricas y técnicas para obtener información útil de la red social Facebook, se evalúa el escenario de implementación de la

solución presentada y se evalúan la percepción y capacidades de un segmento especifico de funcionarios investigadores de la FGN. Posteriormente se diseña, desarrolla y evalúa una herramienta llamada CRAWLERFB, la cual permitirá identificar patrones de ubicación, estudio y empleo, además de identificar los niveles de relación entre un perfil o un grupo de perfiles de la red social Facebook.

2.1. Metodología y referente Teórico.

Este proyecto se clasifica como un desarrollo tecnológico, y al culminarlo, aportara un producto tangible, aplicable en las labores inherentes de las investigaciones que adelanta la FGN, en la ejecución de la acción penal. Un producto que fortalecerá las capacidades de recolección y análisis de la información pública de un objetivo o de un grupo de objetivos que se interrelacionan de manera activa en la red social Facebook. Para cumplir los objetivos de este documento es necesario conocer los referentes teóricos sobre los que se construye esta propuesta.

2.1.1. Crawler - Spider.

Desde el surgimiento del internet, existió la necesidad de ordenar e indexar el contenido de las páginas web. En 1993 científico del MIT, Mathew Gray, (Parrilla, 2012) crea un programa robot que recopilaba e indexaba las direcciones web de los portales, surgiendo de este modo el primer buscador. En 1994 surge un programa de nombre RBSE, creado por Eichmann (Iglesias, 2014). El cual se podría considerar el primer spider o araña, que descargaba y almacenaba no solo la URL sino también el contenido del portal WEB. En 1998, cuatro años después, nace en manos de Brian y Page el más grande y rápido Crawler de páginas web, Google. (Iglesias, 2014)

Un "Crawler" es un programa robot que recorre y descarga el contenido de una página web, cuando la página contiene links o enlaces a otros sitios web, se forma una red de conexiones entre sitios, esto considera como "Spider".

2.1.2. Crawlers en redes sociales.

Desde la aparición de la WEB 2.0, 2006 cuando se hace público Facebook, red en la que los usuarios son quienes comparten material, contenido y publicaciones y donde se generan las primeras comunidades virtuales, se descubrió que este nicho información correspondía a un importante banco de datos para entender fenómenos comportamentales y tendencias de las diferentes comunidades vinculadas a la Red Social. Solo un par de años se necesitaron para empezar a surgir herramientas que realizan Crawlin a redes sociales, así como lo plateaba S. Ibrahim, (Siti, Selamat, & Selamat, 2008) quien propone la extracción de información de las redes sociales para la toma de decisiones en diferentes tipos de negocios. En el año 2009 (Fard & Ester), presenta un proyecto para identificar patrones criminales en grupos o redes sociales, por medio de minería de datos o "minería colectiva".

2.1.3. Crawler - Scrapy a páginas de búsquedas Facebook.

En 2016 mediante un Test de personalidad implementado en Facebook, la compañía Cambridge Analytica, (ROSENBERG & J. X. DANCE, 2018) recolectó datos personales de más de 50 millones de usuarios de la red social, con el propósito de influenciar a los votantes de las elecciones de EE. UU, donde Donald Trump logra la presidencia. Después de estos hechos Mark Zuckerberg, (fundador y director ejecutivo de Facebook), reconoce los errores en privacidad de la red social y se comienzan a implementar las medidas que limitan la consulta y recolección de información, así como la implementación métodos de

detección de robots que mediante la técnica de Crawlin descargan la información de los usuarios de su red.

2.1.4. Link de consulta Facebook.

El método de adquisición de información de las plataformas Inteltechniques e IntellegenceX consistió en la construcción de una URL con los parámetros de Facebook, esta URL, retornaba la información de un usuario que fue compartida por sus nodos cercanos o de sus contactos, este método fue útil hasta el mes de Julio 2019 y permitía identificar reacciones, fotos, publicaciones y otros elementos de un usuario mediante las reacciones y comentarios de sus contactos.

Tabla 3
Links de Consulta de Facebook antes de junio 2019.

Link	Respuesta
www.Facebook.com/search/ID_FACEBOOK/photos-by	Fotos publicadas por el Usuario
www.Facebook.com/search/ID_FACEBOOK/photos-liked	Fotos que le gustan a el usuario
www.Facebook.com/search/ID_FACEBOOK/photos-of	Fotos donde se ha etiquetado el usuario
www.Facebook.com/search/ID_FACEBOOK/stories-tagged	Post o publicaciones donde se ha etiquetado el usuario
www.Facebook.com/search/ID_FACEBOOK/stories-liked	Post o publicaciones que le gustan el usuario
www.Facebook.com/search/ID_FACEBOOK/stories-commented	Post o publicaciones que ha comentado el usuario
www.Facebook.com/search/ID_FACEBOOK/friends	Listado de amigos del usuario

En razón que la técnica de construir los LINK de búsqueda para Facebook dejo de ser útil en julio 2019 por bloqueo de la plataforma, y la API de Facebook entrega información muy limitada, la técnica de Crawling se convierte en la mejor opción para identificar las

interacciones de diferentes usuarios y una cuenta objetivo, reconstruir una red social y un nivel de interacción entre los usuarios, que pueda ser graficado.

2.1.5. Application Programming Interface - API

Las diferentes redes sociales, viendo las necesidades de atender los requerimientos de información de otros portales, con fines publicitarios y académicos, implementan en sus plataformas funciones para interactuar con estos aplicativos WEB, las API, aplicaciones para la programación de interfaces, permiten obtener de manera rápida y ordenada la información de la red social, en el caso de Facebook, su Api solo entrega la información pública y muy limitada de un usuario o grupo específico.

2.2. Descripción del Sistema.

Se plantea desarrollar un sistema que permita en la red social de Facebook, extraer la información de publica de perfiles de usuarios, puntos de convergencia y relaciones entre perfiles, Información recuperada de los datos públicos. Crear una aplicación que permita exportar los datos obtenidos y mediante el uso de Gephi, crear visualizaciones que sean intuitivas para el usuario y permitan una rápida interpretación de sus vínculos y redes por parte de los investigadores de la Fiscalía General de Nación.

2.2.1. Escenario

La solución propuesta es dirigida a los grupos de investigadores de la FGN, que utilizan las redes sociales como insumo en los procesos investigativos que adelantan. Durante la etapa de comprensión y para enriquecer las necesidades de este proyecto, mediante el método de encuestas, dirigidas a los usuarios investigadores de la FGN, se buscó modelar el estado del arte de la institución referente al conocimiento y uso de las técnicas y

herramientas OSINT. Se desarrolla en la plataforma Google Docs, y se envía a los funcionarios investigadores, pertenecientes a grupos de delitos informáticos de la FGN, estos grupos de funcionarios, poseen un mayor conocimiento, y mayores capacidades técnicas en la obtención, recolección y análisis de información almacenada en redes sociales y en fuentes abiertas, por el tipo de casos investigativos e información que administran.

2.2.2. Recolección de Información.

Se encuestaron un total de 34 funcionarios pertenecientes a los grupos de delitos informáticos de Caldas, Arauca, Bogotá, Valle, Risaralda, Villavicencio, Magdalena, Caquetá, Antioquia, Huila, Tolima, Nariño, Quindío, Bolívar Y Boyacá. Estos 34 funcionarios se clasificaron según el nivel y campo de acción en 3 grupos, Técnico, Investigación y profesional, obteniendo de la clasificación los siguientes resultados.

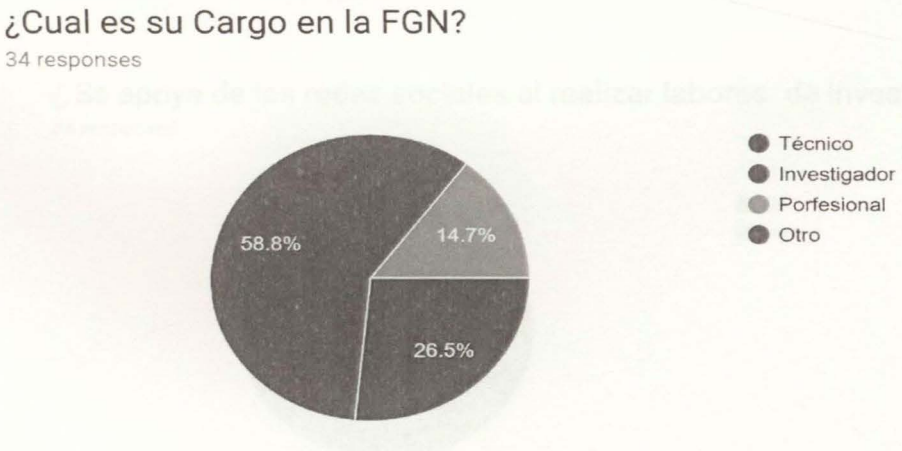


Ilustración 8 Clasificación personal encuestado

La población encuestada pertenece en un 100%, a grupos que realizan labores investigativas en la FGN, se indaga si conoce sobre OSINT y su terminología, se obtuvo como resultado un desconocimiento del término en un 17%.

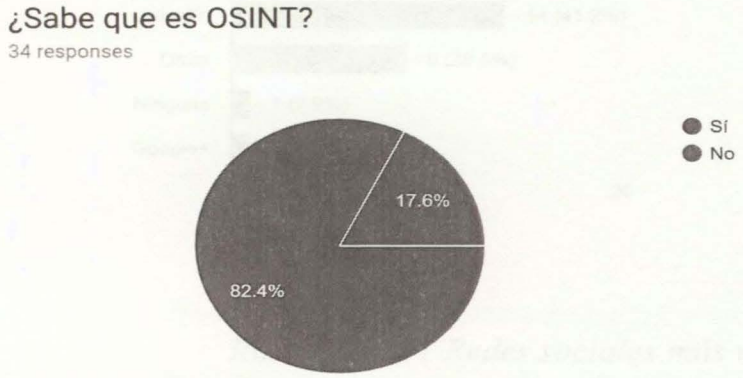


Ilustración 9 Conocimiento sobre OSINT

Al preguntar sobre el uso de redes sociales, como apoyo en las investigaciones, se concluyó que solo el 3% de la población no usa las redes sociales como fuente de información.

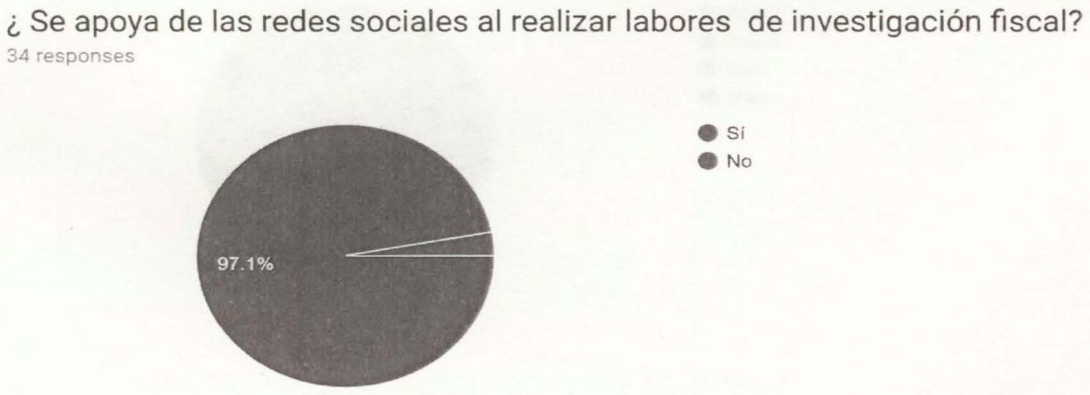


Ilustración 10 Uso de Redes Sociales al Investigar

Se identificaron las redes sociales más utilizadas para apoyar sus investigaciones donde se clasifica a FACEBOOK como la fuente de información de más uso.

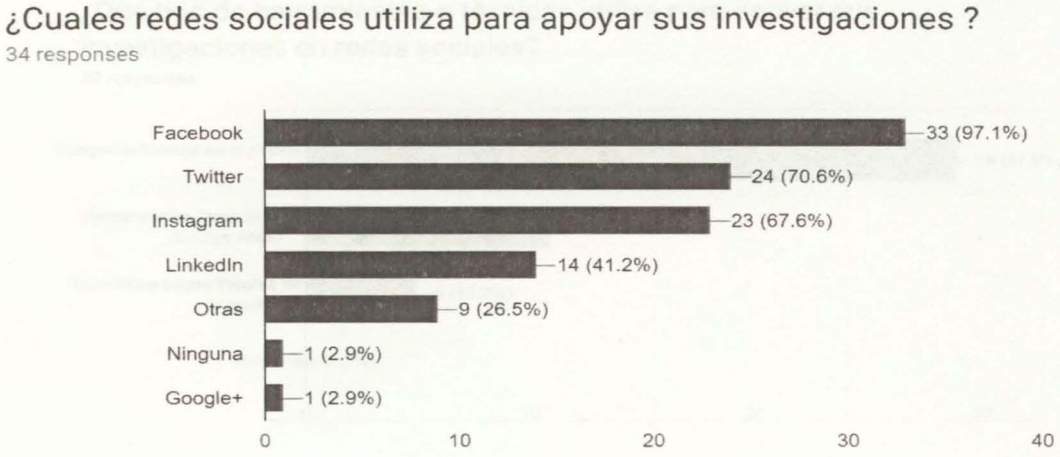


Ilustración 11 Redes sociales más usadas

Es de destacar que el 90% de los encuestados encuentra Alto o Medio, el nivel de beneficio de apoyar sus investigaciones con fuentes abiertas como redes sociales.

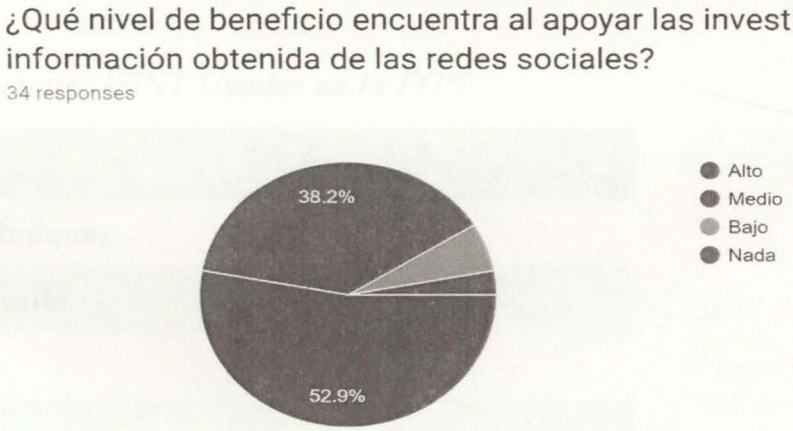


Ilustración 12 Nivel de beneficio

Al indagar Sobre el tipo de herramientas o técnicas utilizadas para apoyar las investigaciones en redes sociales, se identificó que solo el 15% de la población usa herramientas propias, y un 87% realiza búsquedas manuales sobre las plataformas.

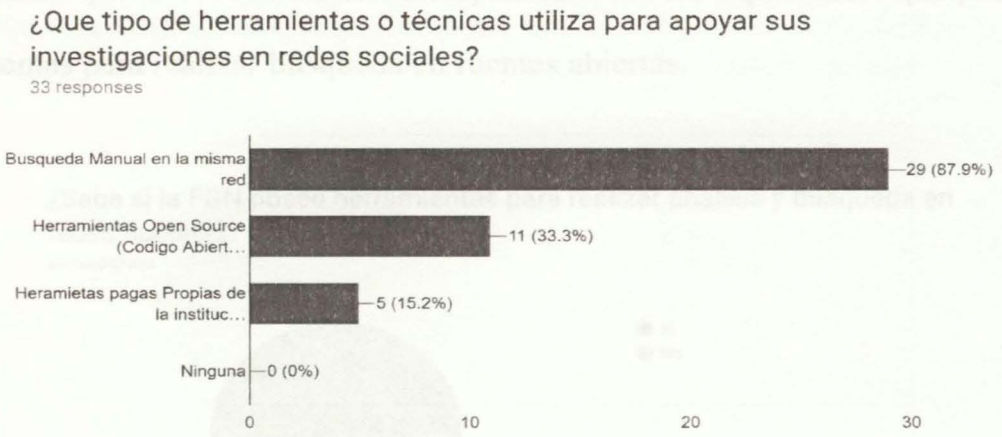


Ilustración 13 Técnicas y herramientas Utilizadas

Se encontraron similitudes en el uso de dos herramientas online, referentes en la búsqueda de información en fuentes abiertas, se identificaron dos herramientas licenciadas o pagas, pertenecientes al stock de la FGN.

Tabla 4

Herramientas OSINT Usadas en la FGN.

Herramienta	Cantidad de Encuestados
IntelTechniques	8
Ciberpatrulla	2
Ninguna	4
Google	2
Ufed Cloud	1
Axiom	1
Bases de datos publicas	2
No responde	10

Se identificó que el 79% de funcionarios, desconocen las capacidades que posee la FGN, en herramientas para realizar búsqueda en fuentes abiertas.

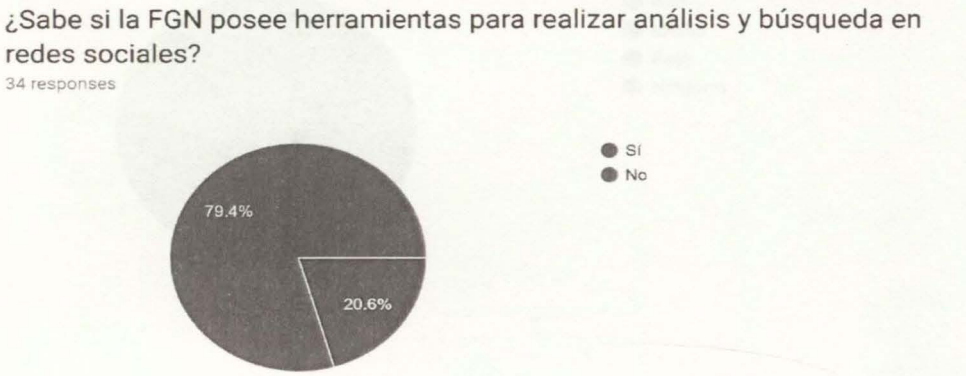


Ilustración 14 Conocimiento de herramientas licenciadas para la FGN

Y por lo tanto la percepción en la población encuestada es la poca existencia de capacidad e inversión por parte de la FGN, en técnicas y herramientas para realizar búsqueda y análisis en fuentes abiertas.

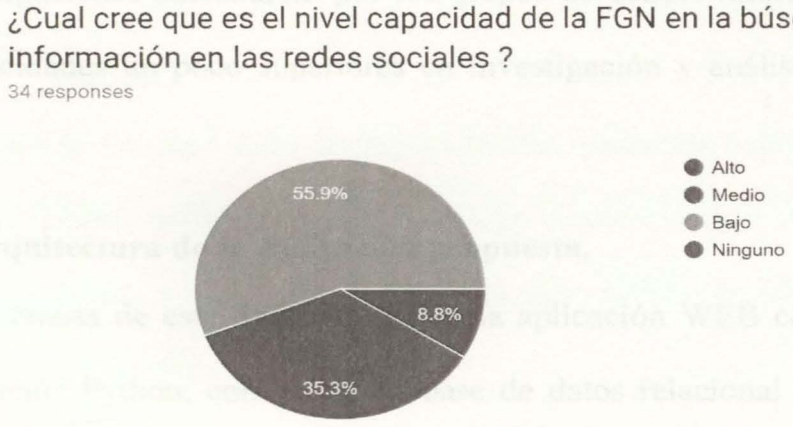


Ilustración 15 Percepción Capacidad en búsqueda y análisis en redes sociales

Ilustración 16 Percepción de Inversión

2.2.3. Resultados Encuestas.

Los resultados de las encuestas demostraron, la poca base en conocimiento, capacidades y herramientas para aplicar OSINT, la pobre percepción de los investigadores en inversión por parte de la FGN en herramientas y capacidades para reforzar las investigaciones con información de las redes sociales, y la alta percepción de beneficio de las redes sociales a las investigaciones adelantadas por los grupos de delitos informáticos de la FGN, Grupos con capacidades un poco superiores en investigación y análisis de información en redes sociales.

2.3. Arquitectura de la Aplicación propuesta.

La propuesta de este documento es una aplicación WEB construida en el lenguaje de programación Python, con motor de base de datos relacional en SQLite, una interfaz de usuario intuitiva y rápida, desarrollada en el framework Django, una aplicación que mediante el uso de la librería Selenium recorra la información pública de un perfil de Facebook y extraiga los contactos y relaciones, para posteriormente identificar

características de ubicación, de comportamiento y graficar sus relaciones y cercanía

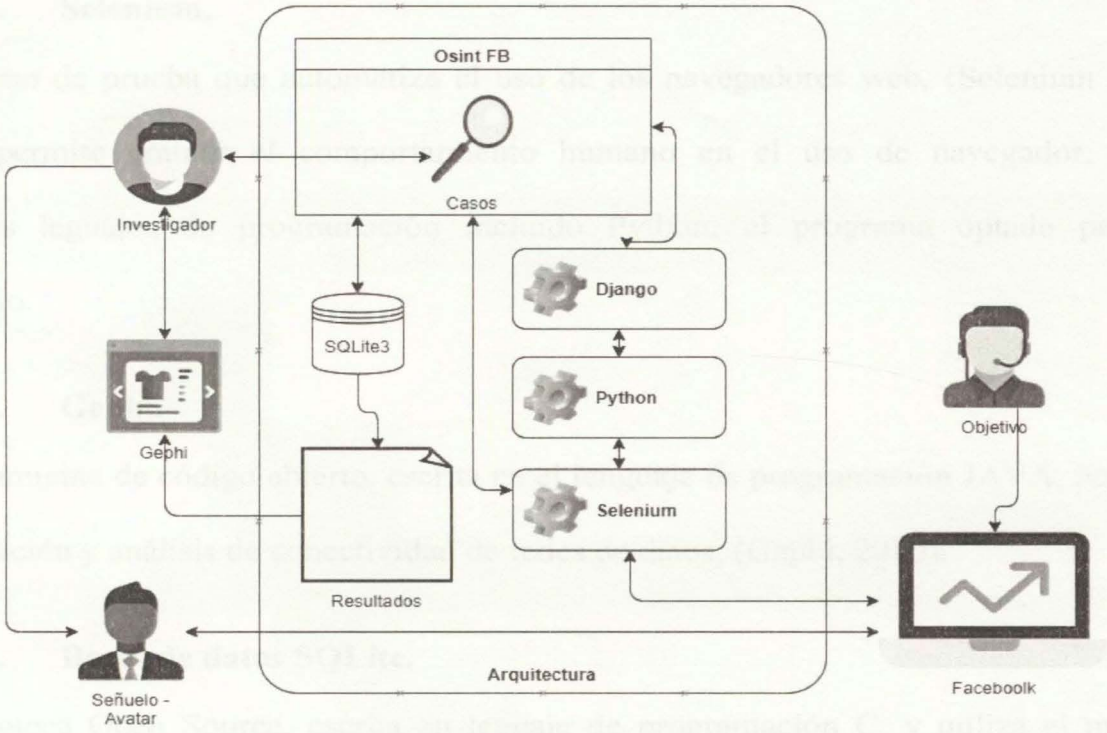


Ilustración 17 Arquitectura OSINTFB

mediante la herramienta Gephi.

2.3.1. Python.

Lenguaje de programación de código abierto, orientado a Objetos creado por Guido Van Rossum, (Python Software Fundation, 2001), reconocido por su potencia y flexibilidad, su fácil comprensión y una gran comunidad que realiza grandes aportes en desarrollo y soporte de sus librerías.

2.3.2. Django.

Es un Framework de desarrollo web de código abierto, escrito en Python, (Django Software Foundation, 2017) fomenta el desarrollo limpio y rápido de las aplicaciones web,

agiliza e implementa en su código tareas como las funciones de validación de entrada de datos y las interfaces de administración de usuarios.

2.3.3. Selenium.

Entorno de prueba que automatiza el uso de los navegadores web, (Selenium Project, 2019), permite emular el comportamiento humano en el uso de navegador, soporta múltiples leguajes de programación incluido Python, el programa optado para este desarrollo.

2.3.4. Gephi.

Herramienta de código abierto, escrita en el lenguaje de programación JAVA, permite la visualización y análisis de conectividad de redes de datos, (Gephi, 2016).

2.3.5. Bases de datos SQLite.

Biblioteca Open Source, escrita en leguaje de programación C, y utiliza el motor de bases de datos SQL, permite trabajar con bases de datos multiplataforma optimizando su tamaño y portabilidad. (SQLite Consortium, 2006).

2.4. Cronograma - Diagrama de GANT.

Se establecen 6 meses para completar el desarrollo, teniendo en cuenta los constantes cambios que se realizaron en los niveles de seguridad en Facebook, y la curva de aprendizaje necesaria en las técnicas, herramientas y leguajes que componen esta propuesta.

La actividad que más tiempo requiere, como se puede observar en la Tabla 5, es el análisis de la plataforma de Facebook, debido a que esta red social cambia constantemente

los id y nombres de sus maquetas HTML. Por lo tanto, se hizo necesario entender cuál es la estructura y comportamiento de la red social en la visualización en un navegador web.

Tabla 5
Cronograma de Actividades

Actividades /Semanas	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Estudio capacidades herramientas libres.																
Entendimiento Framework Django																
Entendimiento Facebook																
Pruebas Selenium																
Desarrollo del modelo																
Desarrollo de las vistas																
Desarrollo Templates																
Exportar y graficar																
Pruebas y caso de estudio																

2.5. Grupo de trabajo.

Inicialmente se propuso a la FGN, generar un grupo de trabajo constituido por tres ingenieros de sistemas, con conocimientos específicos en fuentes abiertas y con capacidades de programación, Esta propuesta no fue avalada por las directivas de la institución, por tal motivo el grupo de trabajo se conforma por un solo ingeniero e investigador, quien es el maestrante autor de este documento.

2.6. Presupuesto.

En las etapas iniciales del proyecto se planteó realizar este proyecto con recursos de FGN, Tales como personal, equipos de cómputos y servidores de prueba y desarrollo, pero dado el tiempo requerido para el inicio y aprobación del proyecto por parte de la Institución, se inició el desarrollo de esta propuesta con recursos propios del maestrante.

2.6.1. Costos de personal.

En los costos de personal se tiene en cuenta las horas dedicadas por el desarrollador del proyecto, como horas Ingeniero Senior. Tomando como base el Estudio de Salarios del sector de Software y TI de Colombia – 2015, (Fedesoft, 2015), el cual presenta el estudio de los rangos salariales de los diferentes niveles y rangos de los ingenieros en nuestro país. Según este estudio el valor hora promedio de desarrollo para un rango de ingeniero senior es de 86.522 pesos, y teniendo en cuenta los aumentos salariales decretados para los años 2016, 2017 y 2018 se establece que el valor hora de desarrollo para un ingeniero senior es de 102.942 pesos.

2.6.2. Costos materiales.

Estos costos son ciertamente menor dado que las etapas de desarrollo fueron asumidos por el propio ejecutor del proyecto.

2.6.2.1. Equipos de cómputo: Equipo Personal.

Software: Se utilizan para este desarrollo herramientas Open Source para el entorno de desarrollo y la creación de este proyecto.

Servidor desarrollo: como administrador de versiones se utiliza GitLab, el cual permite administrar el ciclo de vida del proyecto.

Tabla 6
Costos Desarrollo

Concepto	Cantidad	Valor	Tiempo	Total
Servidor producción	1	299000 Cop	6 meses	598000 Cop
Ingeniero senior	1	102942 Cop	96 horas	9882432 Cop
Total				14794957 Cop

2.7. Análisis de Riesgos.

Para el desarrollo de esta propuesta se tuvo en cuenta un modelo de gestión en la fase de diseño, este modelo implementado en el PMBOOK. Permite identificar y evaluar los eventos que se presentan en las diferentes etapas de análisis, desarrollo, implementación e implantación de software a la medida.

2.7.1. Identificación de Riesgos.

Se ha clasificado que existen tres tipos de riesgos en las diferentes etapas del proyecto, riesgos de tipo Técnico, estos corresponden a las capacidades y conocimientos del equipo de desarrollo. Riesgos Tecnológicos encontrados en los equipos físicos usados en las diferentes etapas del proyecto; Riegos Regulatorios y de control, los cuales son las medidas implementadas por las diferentes redes sociales para evitar la recolección de información de sus usuarios y por último, Riesgos sociales y legislativos, los cuales evidenciamos en la normativa y la aceptación de las herramientas OSINT en el marco de aplicación de la FGN.

En el desarrollo de esta propuesta hemos clasificado dos caracterizaciones del riesgo las cuales son la Probabilidad, clasificado en valores de 0 a 1 y el Impacto de la ocurrencia del riesgo clasificado en valores de 4, 8, 12, 16, 20.

Tabla 7
Clasificación de riesgo

	Nivel de Impacto				
Probabilidad	Menor -4	Moderado -8	Mayor -12	Critico -16	Catastrófico -20
Muy alta – 1	Transferir (4)	Transferir (8)	Transferir (12)	Evitar (16)	Evitar (20)
Alta 0.8	Transferir (3.2)	Transferir (6.4)	Transferir (9.6)	Evitar (12.8)	Evitar (16.8)
Media- 0.6	Mitigar (2.4)	Transferir (4.8)	Transferir (7.2)	Transferir (9.6)	Evitar (12)
Baja- 0.4	Mitigar (1.6)	Mitigar (3.2)	Transferir (4.8)	Transferir (6.4)	Evitar (8)
Muy baja-0.2	Mitigar (0.8)	Mitigar (1.6)	Mitigar (2.4)	Transferir (3.2)	Transferir (4)

Se han detectado los siguientes riegos para su valoración:

Tabla 8
Valoración de Riesgos

Clasificación	Riesgo	Probabilidad (0-1)	Impacto (4, 8, 12,16, 20)	PxI
Riesgo técnico RT-1	La curva de aprendizaje sobre Python, Django y Selenium es más larga de lo esperado	0.6	8	4.8
Riesgo técnico RT-2	El objetivo del proyecto es muy ambicioso.	0.4	8	3.2
Riesgo técnico RT-3	Desconocimiento del uso de las herramientas gráficas Gephi.	0.4	8	3.2
Riesgo técnico RT-4	Falta de entendimiento de la estructura de construcción de las páginas de Facebook.	0.4	16	9.6
Riesgo técnico RT-5	El equipo de trabajo no es suficiente.	0.8	12	9.6
Riesgo tecnológico RTg-1	Equipos y espacios de almacenamiento insuficiente.	0.2	16	3.2
Riesgo tecnológico RTg-2	Canales Vpn insuficientes para el ocultamiento de la Ip ante las redes sociales.	0.8	12	9.6
Clasificación	Riesgo	Probabilidad	Impacto	PxI

		(0-1)	(4, 8, 12,16, 20)	
Riesgo tecnológico RTg-3	Latencia en los procesos de Crawler y Scraping	0.8	8	6.4
Riesgo regulatorio y control RRC-1	bloqueo de direcciones ip por parte de la plataforma.	0.8	16	12.8
Riesgo regulatorio y control RRC-2	Bloqueo de navegadores y equipos por comportamiento automatizado.	0.8	16	12.8
Riesgo regulatorio y control RRC-3	Bloqueo de usuarios por comportamientos repetitivos.	0.8	8	6.4
Riesgo regulatorio y control RRC-4	No recolección de datos por Cambios en las presentaciones Web de las redes sociales.	1	20	20
Riesgo social y legislativo RSL-1	No aceptación por parte de los usuarios.	0.4	4	1.6
Riesgo social y legislativo RSL-2	Migración de usuarios de las redes sociales a otras plataformas.	0.6	12	7.2

2.7.2. Plan de Mitigación del Riesgo.

Es importante entender que los riesgos de mayor impacto en el desarrollo de este proyecto son los riegos ajenos al entorno de desarrollo e implementación.

La técnica de Crawler a redes sociales presenta inconvenientes que dificultan su ejecución, uno de ellos es la latencia humana, ya es necesario que los servicios de Facebook identifiquen un comportamiento Humano en la visita a sus páginas, por lo tanto, se bebe adicionar una latencia aleatoria en cada consulta para emular un comportamiento humano en cada visita a un link específico.

Tabla 9
Plan de Mitigación.

CODIGO	ACCION	RESPUESTA
RT-1	Transferir	Para estos riesgos es necesario aceptar las limitaciones del grupo de trabajo, y su plan de respuesta incluye ampliación de tiempo en la curva de aprendizaje y en las capacidades técnicas. Las implicaciones de ocurrencia de estos riegos son el consumo de tiempo y recursos en la etapa de desarrollo.
RT-2	Transferir	
RT-3	Transferir	
RT-4	Transferir	
RT-5	Transferir	
RTG-1	Transferir	Riesgos asociados a los equipos y recurso tecnológicos. Gracias al plan de modernidad y arquitectura en la FGN, implementado en 2018, se dispone de servidores virtuales y capacidad de almacenamiento suficiente para dar marcha a la implementación de este desarrollo.
RTG-2	Transferir	
RTG-3	Transferir	
RRC-1	Evitar	En los riesgos de tipo Regulatorio y control se encuentra la mayor probabilidad de ocurrencia e impacto. Es necesario evitar el control por parte de las plataformas de las redes sociales, esto se logra utilizando VPN´s, que enmascaren la dirección IP del origen de las consultas. Se logra también generando comportamientos humanos en las etapas de Crawling y generando perfiles o avatares bien diseñados y con interacción en las redes sociales. En el caso específico del RRC-4 es el riesgo de mayor impacto, el cual obliga a estudiar los cambios en las redes sociales y adaptar el desarrollo a estas modificaciones.
RRC-2	Evitar	
RRC-3	Evitar	
RRC-4	Evitar	
RSL-1	Mitigar	Este tipo de riesgo se debe a los usuarios que se resisten al cambio, para mitigar el impacto de este riesgo es necesario mostrar los beneficios y casos de éxito en la herramienta.
RSL-2	Mitigar	El plan de respuesta es la aceptación, ya que el ciclo de vida y la evolución de las redes sociales tienen su inicio y su fin, el nacimiento de nuevas redes sociales obliga a continuar generando proyectos de este tipo que fortalezcan las capacidades de recolección y estudio de información en fuentes abiertas.

Un segundo inconveniente de este método son las credenciales de usuario, ya que para poder visitar los enlaces o links de búsqueda se necesita un inicio de sesión o un perfil autorizado en la plataforma de Facebook, el perfil debe presentar comportamientos e interacción en la plataforma para no ser desconectado de esta.

2.8. Diseño Y Desarrollo.

En las fases de diseño y desarrollo de la herramienta se toma la notación UML(Unified Modeling Lenguage) como base para la construcción de esta propuesta. Aplicando la metodología RUP(Rational Unified Process), herramientas de modelado que permiten, identificar, delimitar y modelar los artefactos que constituirían este proyecto.

Con el fin de obtener un panorama completo de los requerimientos, actividades y actores del sistema, y basados en la metodología RUP, se realizan acciones de modelado que permiten comprender el Core del negocio, el diseño de la plataforma y de la base de datos.

2.8.1. Diseño del Sistema - Diagramas de comportamiento.

2.8.1.1. Casos de Uso

Mediante la construcción de los modelos de casos de uso, es posible identificar los roles y comportamientos de los actores en el sistema propuesto.

2.8.1.1.1. Actores.

- Administrador: Usuario que se encarga de crear y administrar los usuarios del sistema, sus permisos son totales.

- User: es el investigador que crea, edita y elimina casos, inicia las búsquedas de nodos y sus relaciones y exporta los resultados para la creación de gráficos.

2.8.1.1.2. Caso de Uso No1: Creación de Usuarios.

Dado que es un sistema que maneja información reservada de las investigaciones que adelanta la FGN, se hace necesario que un funcionario, sea quien valide, cree y autorice el acceso de los usuarios en la plataforma.

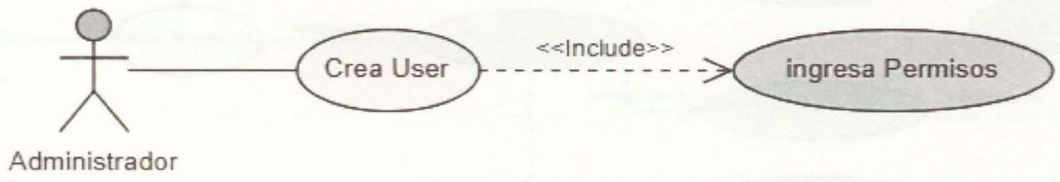


Ilustración 18 Caso de uso crea User

2.8.1.1.3. Caso de Uso No2. Usuario crea Caso

Un usuario validado en el sistema crea un caso o investigación en la cual debe ingresar los datos del objetivo investigar y los datos del señuelo a usar.

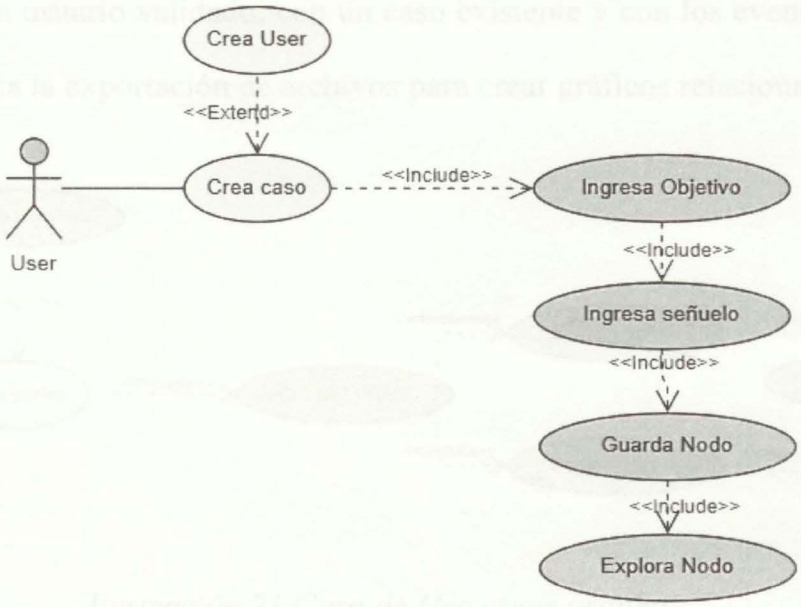


Ilustración 19 Caso de Uso Crea caso

2.8.1.1.4. Caso de Uso No3. Inicia Búsqueda

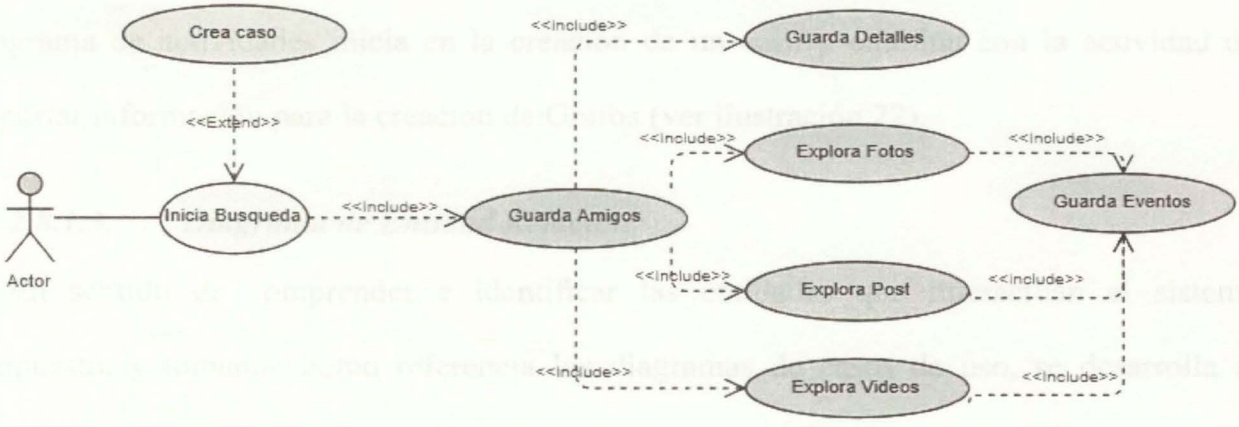


Ilustración 20 Caso de uso Inicio Búsqueda

Un usuario validado y con un caso existente, inicia la búsqueda de nodos o amigos en el perfil del objetivo e identificando la cantidad de relaciones que existen entre estos.

2.8.1.1.5. Caso de Uso No5. Crear gráficos.

En el sistema un usuario validado, con un caso existente y con los eventos recolectados de un objetivo inicia la exportación de archivos para crear gráficos relacionales y de red.

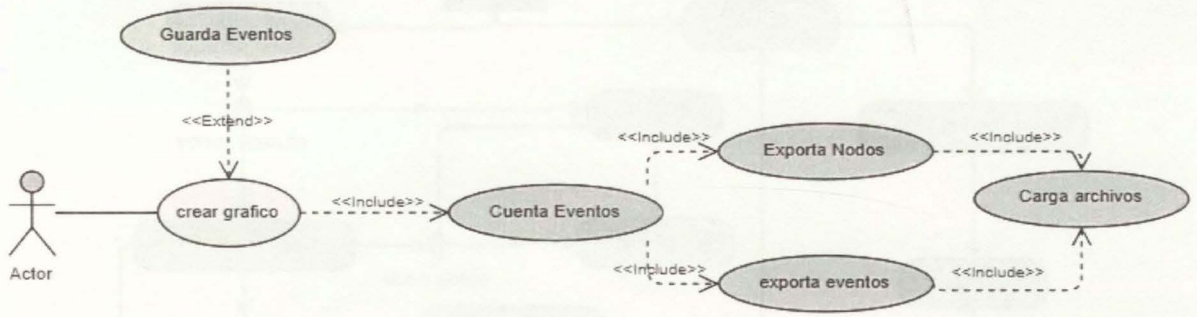


Ilustración 21 Caso de Uso crear grapho.

2.8.1.2. Diagrama de actividades

Con el fin de comprender los posibles flujos del sistema y la interacción de los casos de uso modelados, se utiliza el diagrama de comportamiento diagrama de actividades, en el cual se pude observar el ciclo de operación de la herramienta de software propuesta. El diagrama de actividades inicia en la creación de un caso y culmina con la actividad de exportar información para la creación de Grafos (ver ilustración 22).

2.8.1.3. Diagrama de Entidad Relación.

En sentido de comprender e identificar las entidades que interactúan el sistema propuesto, y tomando como referencia los diagramas de casos de uso, se desarrolla el diagrama de objetos (ver ilustración 23). Modelo que delimita las relaciones y entidades el sistema completo.

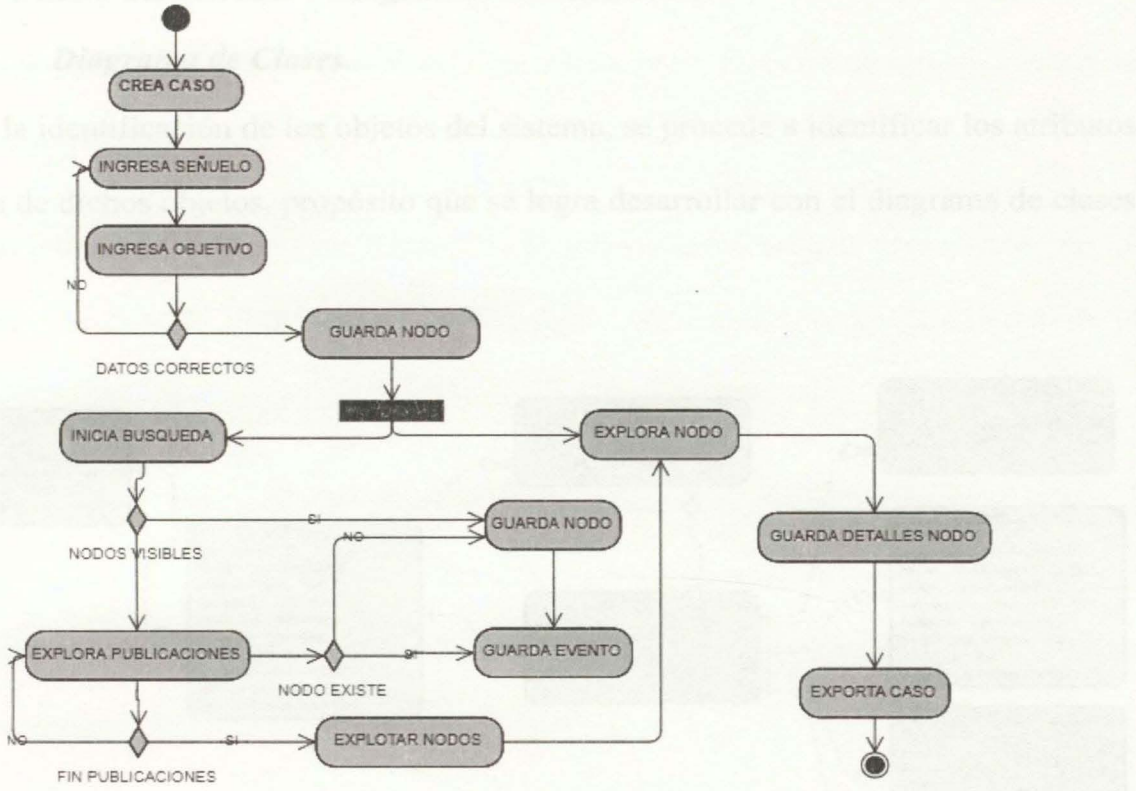


Ilustración 22 Diagrama de Actividades

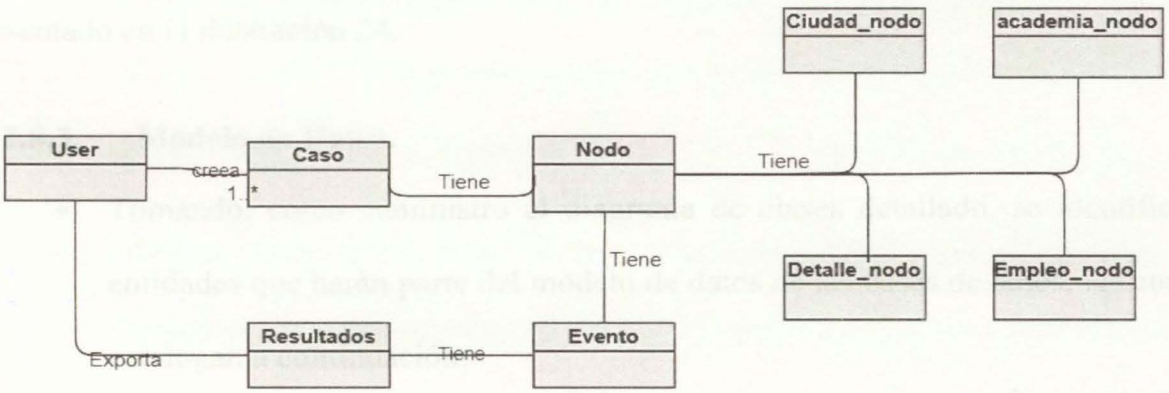


Ilustración 23 Diagrama de Entidades

2.8.2. Diseño del Sistema – Diagramas Estructurales.

2.8.2.1. Diagrama de Clases.

Lograda la identificación de los objetos del sistema, se procede a identificar los atributos y relaciones de dichos objetos, propósito que se logra desarrollar con el diagrama de clases

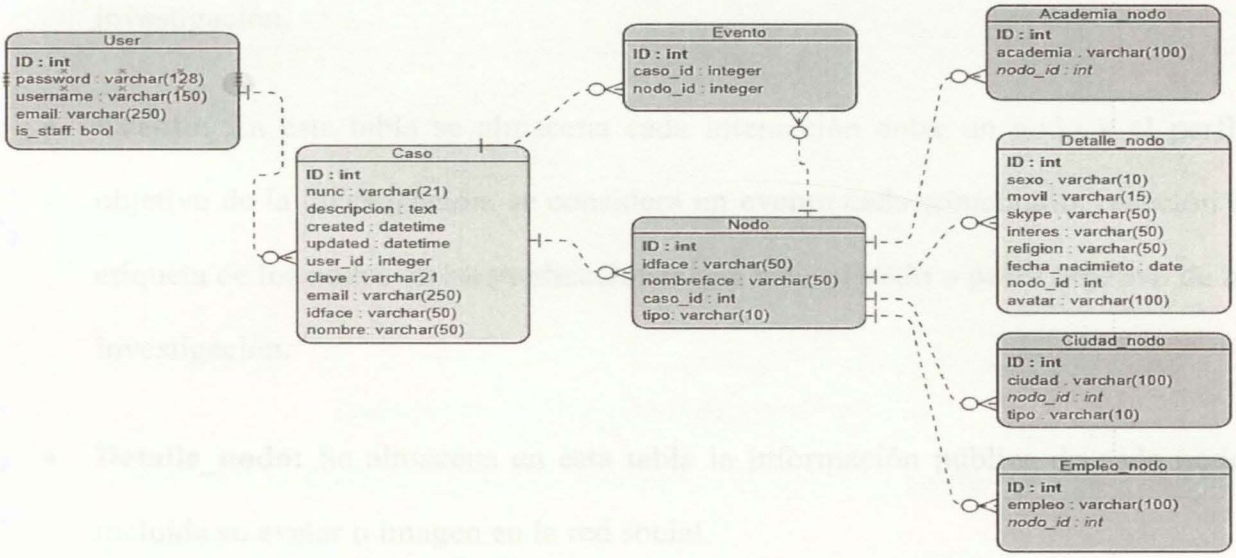


Ilustración 24 Diagrama de Clases

presentado en la ilustración 24.

2.8.3. Modelo de Datos.

- Tomando, como suministro el diagrama de clases detallado, se identifican las entidades que harán parte del modelo de datos de las bases de datos, las cuales de catalogan a continuación:

- **User:** Tabla que contiene los usuarios del sistema, se requiere que el usuario es su campo "is_staff" = TRUE, para poder acceder a las funciones del software.

- **Caso:** Tabla donde se almacenan los casos de los investigadores, esta tabla registra la noticia criminal, el señuelo a usar para la investigación, su clave y el objetivo.

- **Detalle_Nodo:** En esta tabla se almacenan todos los amigos encontrados de un caso, si el nodo es de tipo = "perfil origen" corresponde al perfil objetivo de la investigación.

- **Evento:** En esta tabla se almacena cada interacción entre un nodo y el perfil objetivo de la investigación, se considera un evento cada comentario, reacción o etiqueta de los nodos en las publicaciones que hace el nodo o perfil objetivo de la investigación.

- **Detalle_nodo:** Se almacena en esta tabla la información pública de cada nodo incluida su avatar o imagen en la red social.

- **Academia_nodo:** en esta tabla se almacenan los nombres de los centros de estudio, universidades y colegios que se encuentran en la información pública de los nodos en un caso o investigación.

- **Ciudad_nodo:** en esta tabla se almacenan las ciudades natales o actuales que se identifican en la información pública de los nodos en un caso o investigación.

- **Empleo_nodo:** en esta tabla se almacenan los lugares de trabajo que publican los nodos en su perfil de la red social en un caso o investigación.

2.8.4. Diagrama de Paquetes.

Mediante esta herramienta de modelado se identifican los paquetes y dependencias que interactúan en sistema.

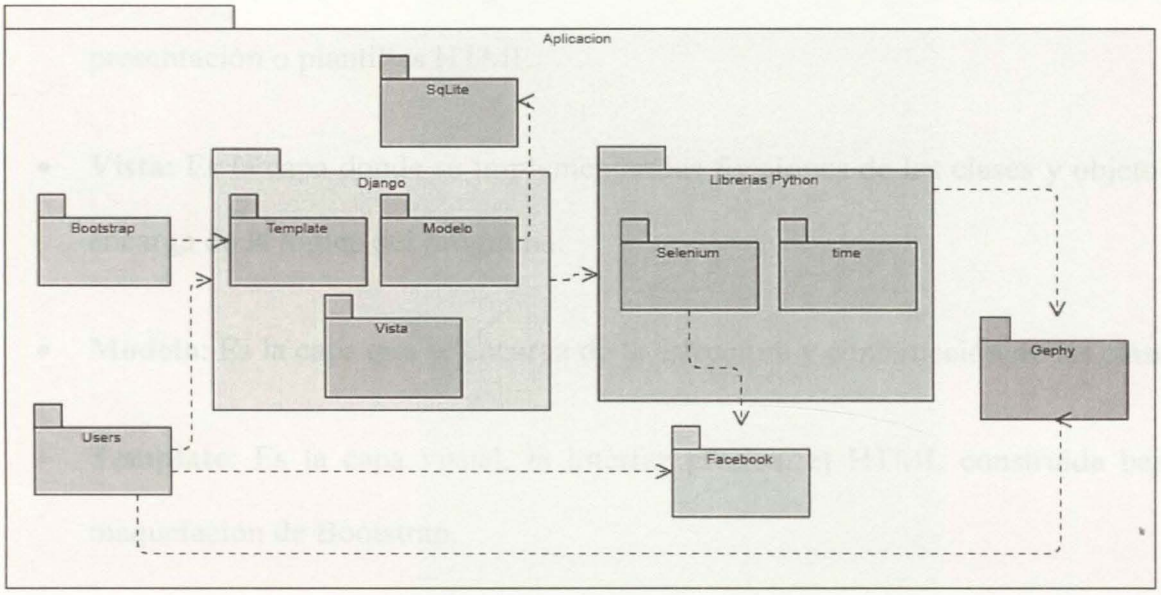


Ilustración 25 Diagrama de paquetes

- **Bootstrap**: Framework de maquetación de aplicaciones y páginas WEB basadas en HTML, CSS y JS.

- **Selenium:** Es una librería escrita en Python, que permite realizar pruebas sobre navegadores, y en el caso específico de este proyecto, permite emular los comportamientos humados sobre un navegador WEB y sobre la red social de la cual queremos obtener información.

- **Gephi**: Herramienta open-source que permite visualizar y analizar de manera gráfica datos de redes. Para generar hipótesis y descubrir patrones.

- **SQLite:** Motor o gestor de bases de datos relacional de tipo cliente – servidor, basado en SQL (Structured Query Language). Son bases de datos que ocupan poco tamaño y fáciles de portar.

- **Django** (Referencia): Framework de desarrollo ágil, orientado a objetos y escrito en el lenguaje de programación Python, Basado en el esquema MVT (modelo, Vista, Témplate) el cual separa la estructura de datos, de las funciones y la presentación o plantillas HTML.

- **Vista**: Es la capa donde se implementan las funciones de las clases y objetos, se encarga de la lógica del programa.

- **Modelo**: Es la capa que se encarga de la estructura y construcción de las clases.

- **Template**: Es la capa visual, la interfaz gráfica, el HTML construida bajo la maquetación de Bootstrap.

2.9. Estructura.

Como se ha explicado, este desarrollo se plantea bajo estructura del Framework Django V2.0, y por ende la arquitectura de archivos, se regula por el modelo (MVT) Modelo, Vista, Template, el cual divide el código del programa en los siguientes carpetas y archivos (ver ilustración 26).

Entre las carpetas y archivos principales podemos encontrar:

- **Templates:** Carpeta donde se almacenan los archivos HTML con los que interactúa el usuario del sistema (ver ilustración 27).

- **Models.py:** en este archivo se encuentra el modelo del sistema, la estructura de la base de datos se construye a partir de este modelo, también se encuentran validaciones de integridad de los datos (ver ilustración 28).

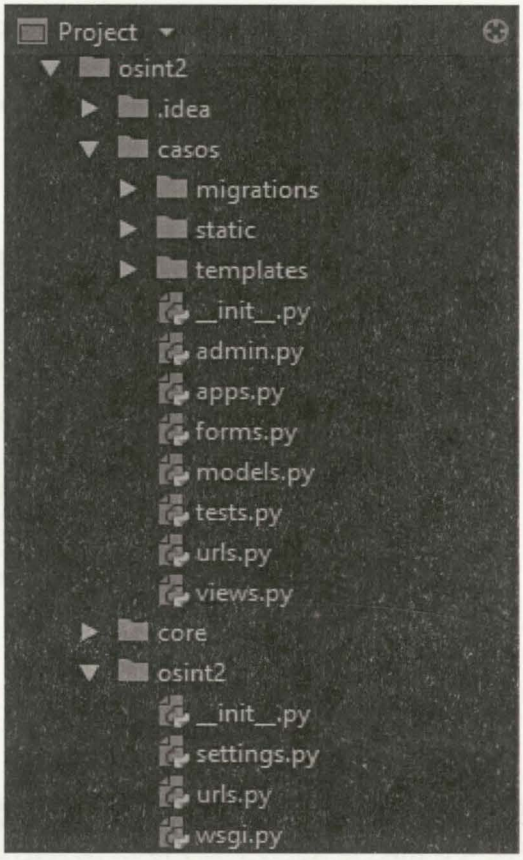


Ilustración 26 Estructura proyecto

- **Views.py** archivo que contiene las funciones y acciones propias de la aplicación, es la lógica del sistema que envía a los Template los resultados de las funciones.


```
from django.db import models
from ckeditor.fields import RichTextField
from django.contrib.auth.models import User

class Caso(models.Model):
    user = models.ForeignKey(User, on_delete=models.CASCADE)
    nunc = models.CharField(verbose_name="Noticia Criminal", max_length=21)
    descripcion = models.TextField(verbose_name="descripcion Caso", null=True)
    created = models.DateTimeField(auto_now_add=True, verbose_name="fecha de creacion")
    updated = models.DateTimeField(auto_now=True, verbose_name="Fecha de edición")
    idface = models.CharField(verbose_name="ID facebook Objetivo", max_length=50, null=True)
    nombre = models.CharField(verbose_name="Nombre Objetivo", max_length=50, null=True)
    email = models.EmailField(verbose_name="señuelo Facebook", null=True)
    clave = models.CharField(verbose_name="clave señuelo", null=True, max_length=21)

    class Meta:
        verbose_name = "Caso"
        verbose_name_plural = "casos"
        ordering = ['nunc']

    def __str__(self):
        return self.nunc

class Nodo(models.Model):
    caso = models.ForeignKey(Caso, null=True, blank=True, on_delete=models.CASCADE)
    idface = models.CharField(verbose_name="ID facebook Nodo", max_length=50, null=True)
    nombreface = models.CharField(verbose_name="Nombre Nodo", max_length=50, null=True)
    tipo = models.CharField(verbose_name="Tipo de Nodo", max_length=10, null=True)
```

Ilustración 28 Fragmento Archivo models.py

2.10. Flujo y funcionamiento del Sistema.

El desarrollo propuesto es una aplicación Web, que permita descargar y almacenar los datos de un usuario encontrados en Facebook, almacenarlos y procesarlos para identificar

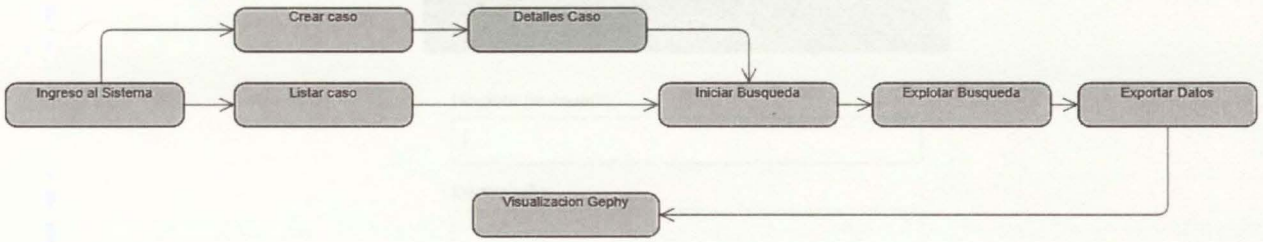


Ilustración 29 Flujo del Sistema

las relaciones y los niveles de relaciones que este usuario presenta con su comunidad.

2.10.1. Ingreso al Sistema.

Se ingresa a la aplicación vía navegador web, y se proporcionan las credenciales de usuario.

Gracias a las facultades de Django, es posible utilizar el entorno de administración. Este entorno ya tiene construido todo un panel administrativo y de manejo de las clases del sistema.

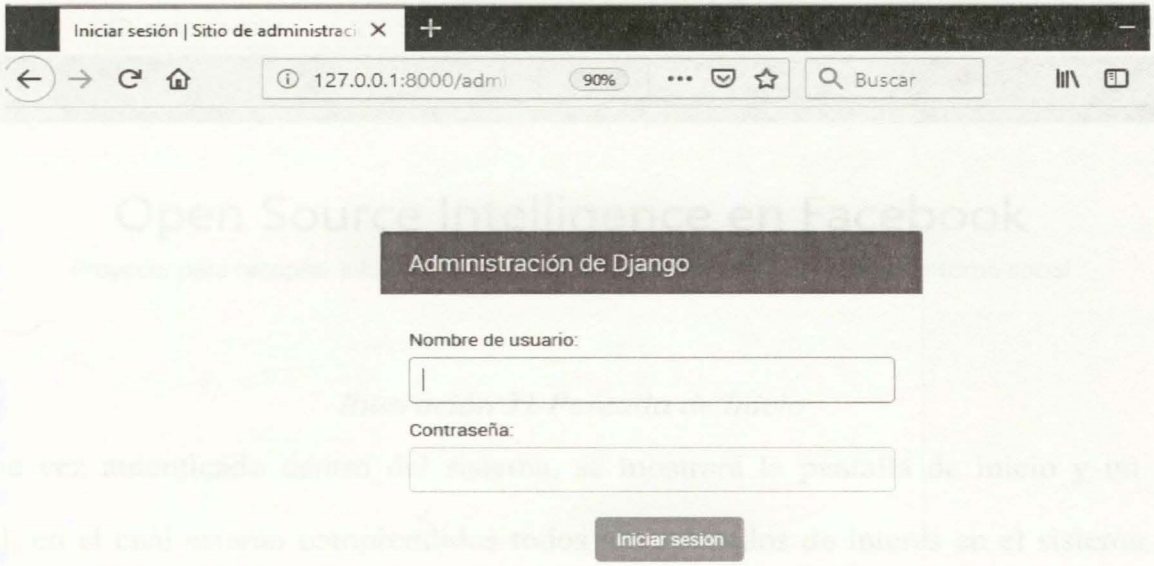


Ilustración 30 Inicio Sesión usuario

2.10.2. Pantalla de inicio

Open Source Intelligence en Facebook

Proyecto para recopilar informacion de perfiles de facebook y construir su entorno social

Ilustración 31 Pantalla de Inicio

Una vez autenticado dentro del sistema, se mostrará la pantalla de inicio y un menú lateral, en el cuál estarán comprendidos todos los apartados de interés en el sistema. A su vez, en el centro de la pantalla está el título y un pequeño menú comprendido por algunos botones, para acceder a los apartados de la cuenta, configuración, ayuda y salir.

2.10.3. Pantalla Lista de casos.

En esta sección se encuentra la lista de casos que cada usuario ha creado, desde esta sección pude ver los detalles, iniciar la creación de caso, iniciar la búsqueda del caso e iniciar la exportación de datos.

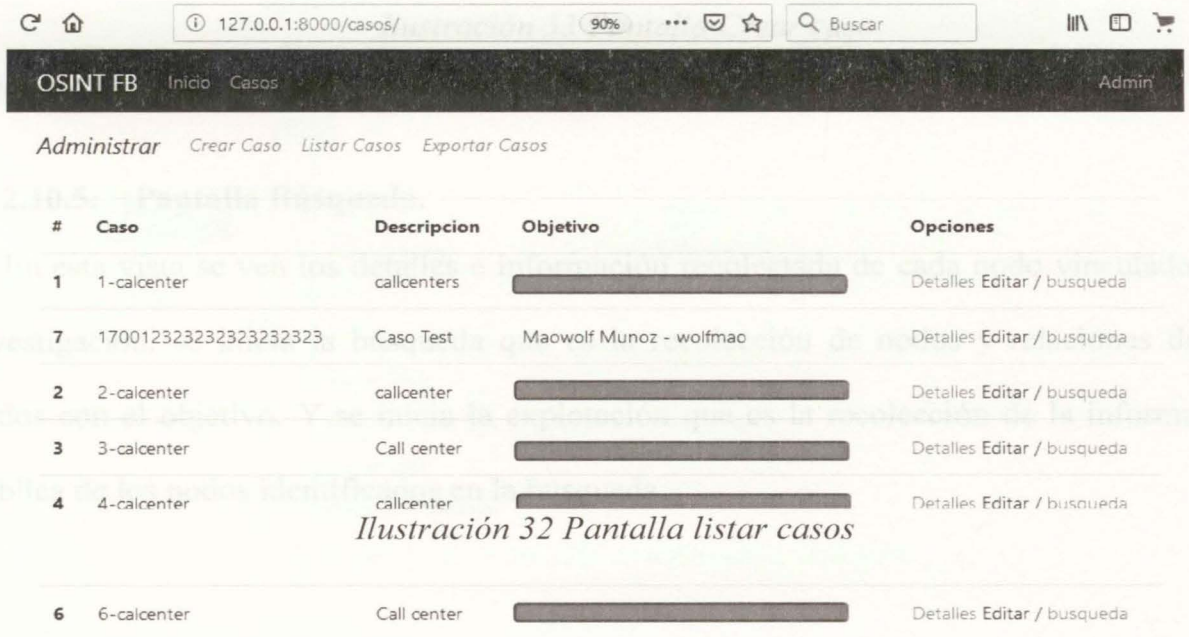


Ilustración 32 Pantalla listar casos

Ilustración 33 Pantalla Listar casos

2.10.4. Pantalla Crear Caso.

En esta interface se inicia la creación de un caso, se registran los datos del caso, los

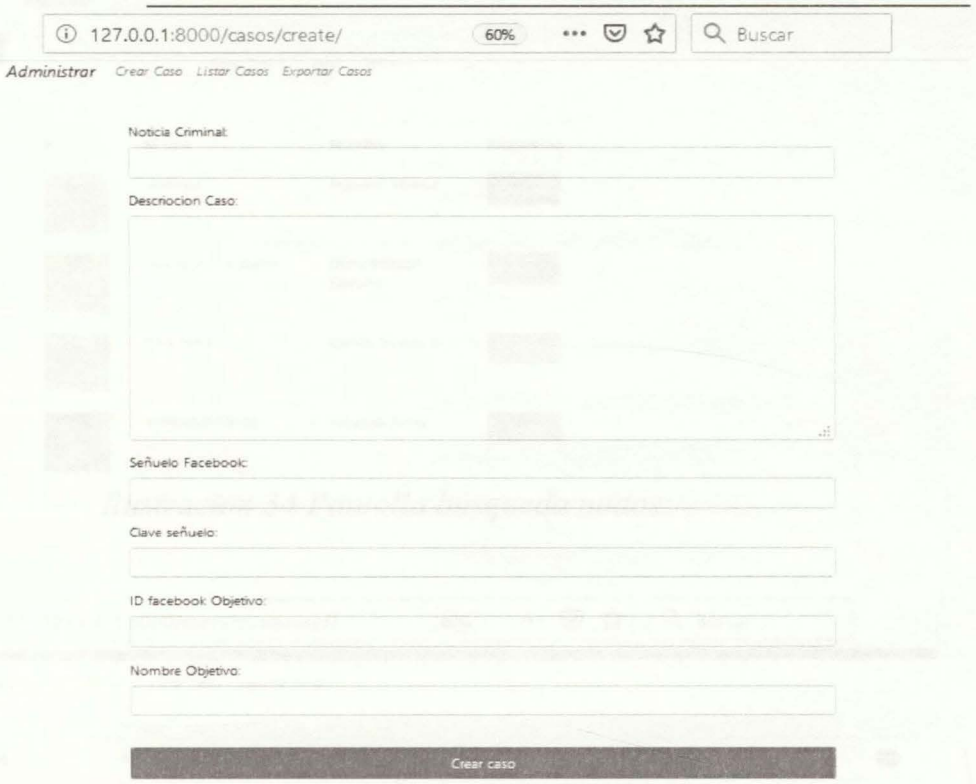


Ilustración 33 Pantalla Crear caso

datos del señuelo a utilizar y del objetivo a investigar.

2.10.5. Pantalla Búsqueda.

En esta vista se ven los detalles e información recolectada de cada nodo vinculado a la investigación, se inicia la búsqueda que es la recolección de nodos y relaciones de los nodos con el objetivo. Y se inicia la explotación que es la recolección de la información pública de los nodos identificados en la búsqueda.

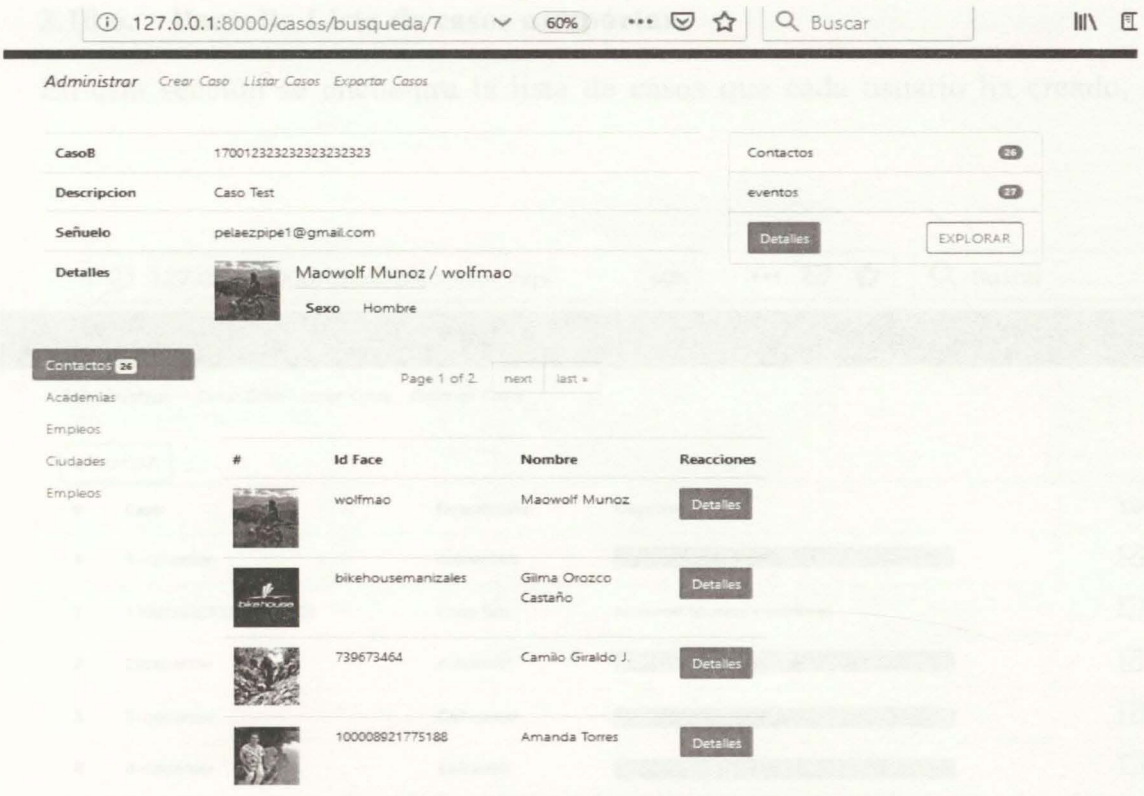


Ilustración 34 Pantalla búsqueda nodos

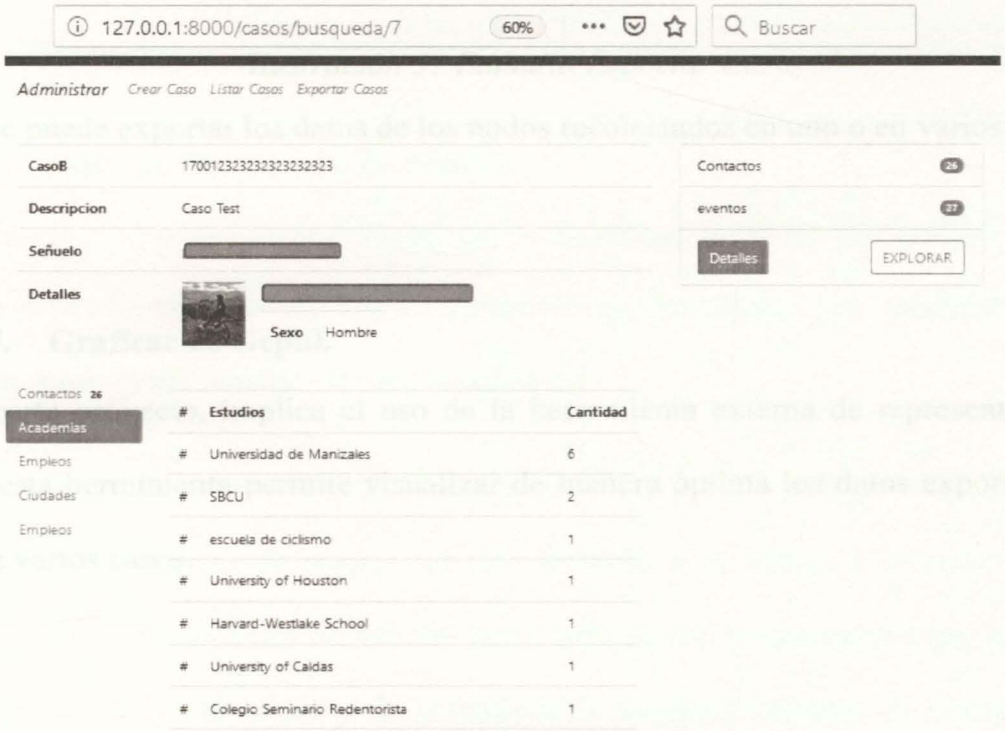


Ilustración 35 Pantalla compilado academias

2.10.6. Pantalla Lista de casos a exportar.

En esta sección se encuentra la lista de casos que cada usuario ha creado, desde esta

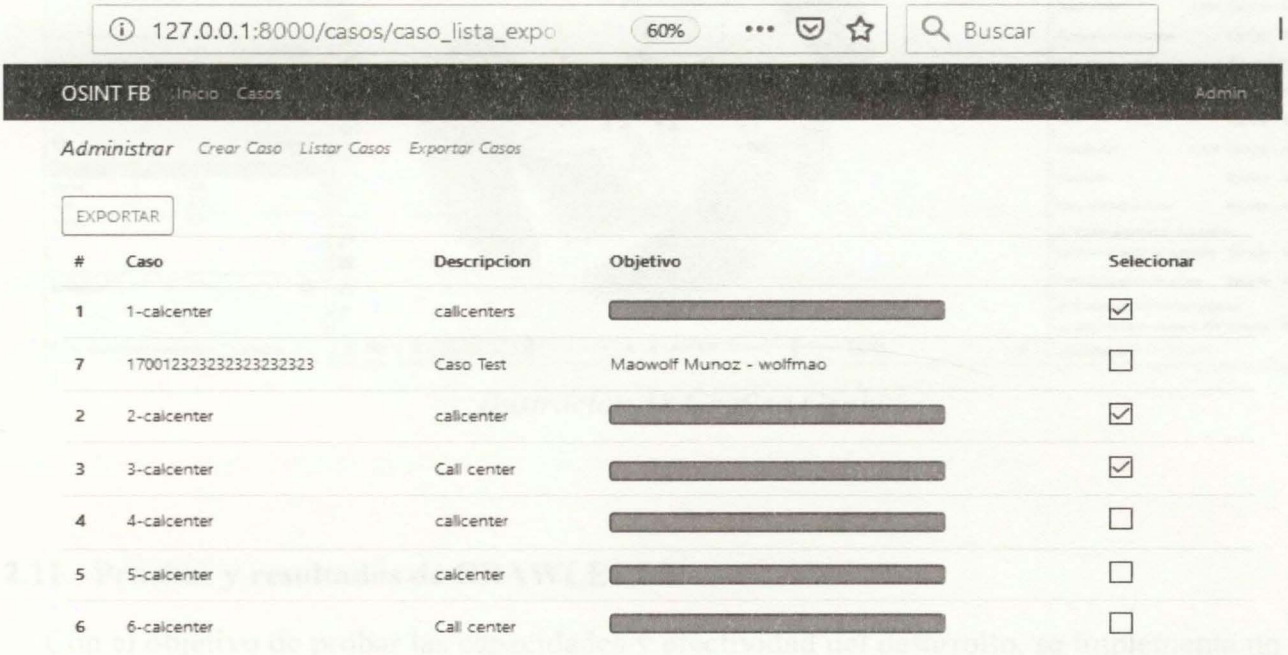


Ilustración 37 Pantalla Exportar Caso

sección se puede exportar los datos de los nodos recolectados en uno o en varios casos.

2.10.7. Graficar en Gephi.

Esta parte proyecto, implica el uso de la herramienta externa de representa de redes, GEPHI, esta herramienta permite visualizar de manera óptima los datos exportados de un caso o de varios casos.

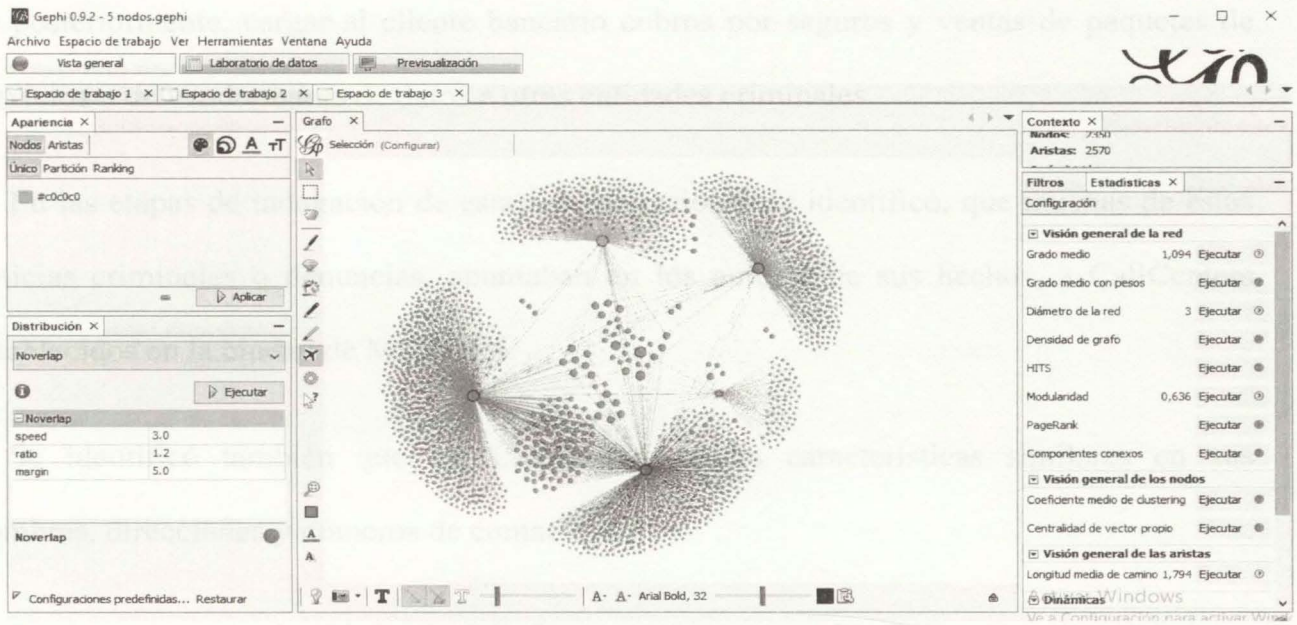


Ilustración 38 Grafica Gephi

2.11. Pruebas y resultados de CRAWLERFB.

Con el objetivo de probar las capacidades y efectividad del desarrollo, se implementa un laboratorio de pruebas y un entorno de trabajo para un caso investigativo real y activo.

2.11.1. Descripción del caso de prueba.

Por labores investigativas en el marco de las funciones como policía judicial de la FGN, se ha identificado, dado el aumento en número de denuncias, una modalidad delictiva etiquetada como "CallCenters", la cual consiste en:

La víctima, un cliente bancario, recibe vía telefónica una llamada por parte de un asesor de franquicias de tarjetas de crédito, en esta llamada se le ofrece a la víctima servicios especiales de seguros o cuotas de manejo gratis solo por ser propietarios o por hacer uso de sus tarjetas de crédito. El objetivo de la llamada es, mediante técnicas de ingeniería social, extraer los datos privados y propios de la tarjeta de crédito y del tarjetahabiente, con el fin

de posteriormente, cargar al cliente bancario cobros por seguros y ventas de paquetes de servicios o de vender bases de datos a otras entidades criminales.

En las etapas de indagación de estas investigaciones se identificó, que muchas de estas noticias criminales o denuncias, apuntaban en los autores de sus hechos, a CallCenters establecidos en la ciudad de Manizales.

Se identificó también que los CallCenters tenían características similares en sus nombres, direcciones y números de contacto.

Por lo tanto, entre las hipótesis delictivas se presume que los Callcenters denunciados y sus representantes legales poseen alguna relación o interacción entre ellos.

2.11.2. Metodología de caso Clásica – Búsqueda Manual.

En las etapas investigativas en la FGN, es útil realizar labores de búsqueda en fuentes abiertas, para este caso específico Facebook, labores que se realizan sin herramientas ni conocimientos y requieren de muchas horas hombre máquina, labores muy difíciles de medir y de calcular. Además, son labores limitadas por las capacidades y conocimientos de los investigadores en las redes sociales, los niveles de seguridad y restricción de Facebook, y las restricciones y bloqueos que los mismos usuarios configuran en la red social.

2.11.3. Metodología de caso propuesta – Búsqueda CRAWLERFB.

Para este caso de estudio se identificaron 5 perfiles de la red social Facebook, estos perfiles pertenecen a los representantes legales o propietarios de 5 Callcenters que se han visto vinculados en alguna noticia criminal. Para los objetos de este documento y con el fin de proteger la información que goza de carácter de reservada, por ser una investigación en

la Fiscalía, se ha decidido cambiar o cubrir la información sensible que pueda revelar detalles de la investigación.

Utilizando un perfil real en la red social Facebook, el cual no tiene fuerza en relación y cercanía a los objetivos, y de manera manual, se identifican y visitan los perfiles objetos de la investigación, obtenido la información pública de cada perfil, posteriormente se realiza la misma búsqueda con la herramienta CRAWLERFB.

Para los efectos de la búsqueda se desarrolla un perfil en la red social Facebook, este perfil el cual se llamará "señuelo", se alimenta, por una semana, con noticias y publicaciones sobre la ciudad de origen o actual de los objetos (Manizales- Caldas). Teniendo en cuenta que el objetivo4, posee pública su información de amigos, se envían invitaciones a nodos de este perfil, con el fin de obtener algunos amigos en común.

Después de adquirir 5 amigos en común con el objetivo4, se envían solicitudes de amistad a los perfiles objetivos de esta investigación. Se obtiene aceptación de solicitud de amistad de los objetivos 1 a 4.

En la plataforma propuesta y desarrollada por este maestrante se crean 5 casos o investigaciones, donde se ingresa los objetivos y el señuelo creado para este escenario. se lanza el proceso de exploración de los objetivos y se logra identificar sus nodos y principales relaciones.

Identificados los nodos de cada perfil, se lanza la explotación de nodos, proceso en el que se verifica la información pública de cada nodo. Este proceso permite identificar las

ciudades, empleo y Academias de mayor ocurrencia entre los nodos amigos del objetivo de la investigación.

Tabla 10
Comparativa Métodos de búsqueda manual y CRAWLERFB

Objetivos	Búsqueda Manual		Búsqueda CRAWLERFB	
	Nodos	Relaciones	Nodos	Relaciones
Objetivo1 / M***S*****z	1	0	434	733
Objetivo2 / L***C*******a	1	0	890	3903
Objetivo3 / L***P****C***P***a	0	0	719	1938
Objetivo4 / G******E*******a	545	0	423	2209
Objetivo5 / D****A****a	0	0	99	185

Tabla 11
Academias, Empresas y ciudades de Red Objetivo 1

	#	Academias de Estudio	Cantidad
Objetivo1 / M***S*****z	1	Universidad de Caldas	27
	2	colegio Gerardo arias Ramirez	19
	3	Universidad de Manizales	15
	4	SENA	14
	5	Universidad nacional de Colombia	10
	#	Empresas O Empleos	Cantidad
	1	Digitex Internacional	12
	2	Facebook	9
	3	Emergia Contact Center	7
	4	Community Service	3
	5	TOP LINE CALL S.A.S	3
	#	Ciudad	Cantidad
	1	Manizales / Ciudad actual	158
	2	Manizales / Localidad natal	123
	3	Villamaría / Ciudad actual	34
	4	Villamaría / Localidad natal	16
	5	Medellín / Localidad natal	14

Tabla 12
Academias, Empresas y ciudades de Red Objetivo 2

	#	Academias de Estudio	Cantidad
Objetivo2 / L***C*******a	1	Universidad de Caldas	80
	2	Universidad de Manizales	33
	3	SENA	22
	4	Universidad Nacional de Colombia	21
	5	Universidad Católica Luis Amigó - Manizales	15
	#	**Empresas**	**Cantidad**
	1	Autónomo	34
	2	Facebook	18
	3	DIGITEX MANIZALES	11
	4	Emergia Contact Center	10
	5	Universidad de Caldas	6
	#	**Ciudad**	**Cantidad**
	1	Manizales / Ciudad actual	347
	2	Manizales / Localidad natal	331
	3	Bogotá / Ciudad actual	87
	4	Bogotá / Localidad natal	65
	5	Medellín / Localidad natal	13

Tabla 13
Academias, Empresas y ciudades de Red Objetivo 3

	#	Academias de Estudio	Cantidad
Objetivo3 / L***P****C***P***a	1	Universidad de Caldas	7
	2	SENA	5
	3	Inem Baldomero Sanin Cano Manizales	4
	4	Universidad de Manizales	3
	5	Instituto Mariscal Sucre Manizales	3
	#	**Empresas**	**Cantidad**
	1	Colombia RED 365	1
	2	Emergia Contact Center	1
	3	Digitex	1

4	P& G CONTAC CENTER	1
5	PetService	1
#	**Ciudad**	**Cantidad**
1	Manizales / Localidad natal	60
2	Manizales / Ciudad actual	53
3	Villamaría / Ciudad actual	5
4	Bogotá / Localidad natal	3

Tabla 14

Academias, Empresas y ciudades de Red Objetivo 4

	#	Academias de Estudio	Cantidad
Objetivo4 / G******E*******a	1	Universidad de Manizales	45
	2	Universidad de Caldas	36
	3	Universidad Autónoma de Manizales - UAM	21
	4	Universidad Nacional de Colombia sede Manizales	12
	5	Instituto Universitario De Caldas	7
	#	**Empresas**	**Cantidad**
	1	Autónomo	10
	2	Universidad de Caldas	4
	3	Digitex	3
	4	P& G CONTAC CENTER	3
	5	Teleperformance Colombia	2
	#	**Ciudad**	**Cantidad**
	1	Manizales / Ciudad actual	165
	2	Manizales / Localidad natal	157
	3	Bogotá / Ciudad actual	17
	4	Medellín / Ciudad actual	9
	5	Medellín / Localidad natal	7

Tabla 15
Academias, Empresas y ciudades de Red Objetivo 5

	#	Academias de Estudio	Cantidad
Objetivo5 / D****A****a	1	Universidad de Caldas	7
	2	Universidad Autónoma de Manizales - UAM	4
	3	Escuela Normal Superior de Caldas	4
	4	fe y alegria	3
	5	Universidad de Manizales	3
	#	**Empresas**	**Cantidad**
	1	Facebook	5
	2	DIGITEX MANIZALES	3
	3	Dynamic Center S.A.S	2
	4	Emergia Contact Center	2
	5	P& G CONTAC CENTER	1
	#	**Ciudad**	**Cantidad**
	1	Manizales / Localidad natal	33
	2	Manizales, Antioquia, Colombia / Ciudad actual	5
	3	Medellín / Localidad natal	3
	4	Medellín / Ciudad actual	3
	5	Marquetalia (Caldas) / Ciudad actual	1

2.11.4. Graficas en Gephi

Se exportan los resultados en dos archivos nombre Nodos.csv y Relaciones.csv, estos archivos contienen los resultados de la herramienta CRAWLERFB, y para su mejor análisis y comprensión se cargan en Gephi, Herramienta que permite generar gráficos relacionales.

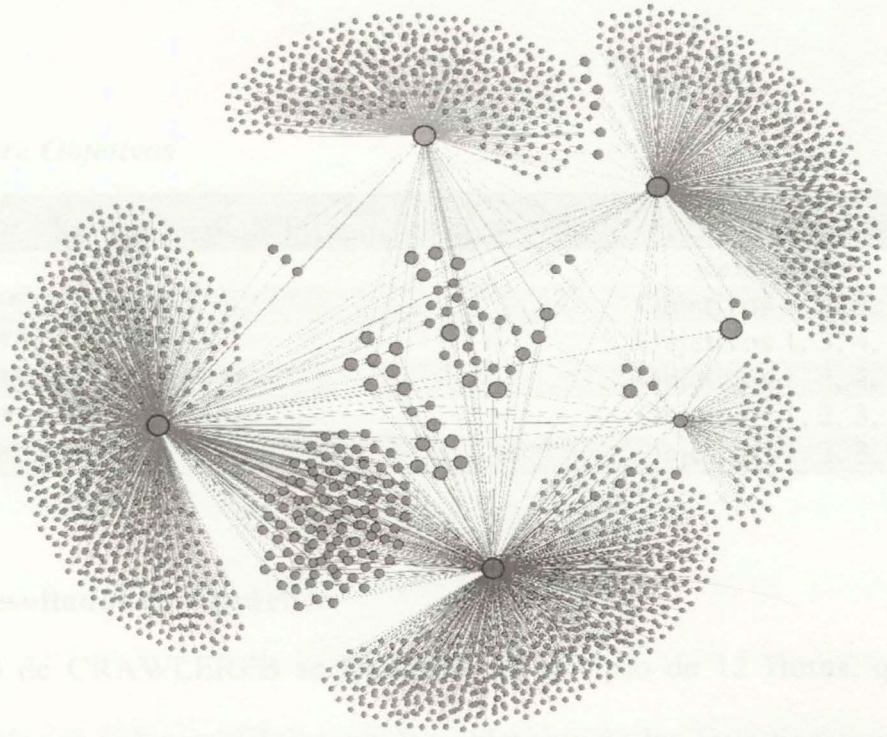


Ilustración 39 Resultado nodos Gephi

2.11.4.1. Resultados Análisis en Gephi.

Tabla 16
Relaciones por nodo

Nodos totales	2349
Nodos con 1 relación	2172
Nodos con 2 Relaciones	148
Nodos con 3 relaciones	23
Nodos con 4 Relaciones	6

Se puede observar en la gráfica y resultados que todos los nodos objetivos se conocen entre sí y poseen relaciones e interacciones a sus publicaciones. Es evidente que existe una fuerte relación de cercanía entre los nodos objetivos, Existe un nodo que se relaciona directamente a los nodos objetivos.

Tabla 17
Relaciones entre Objetivos

Relaciones Entre Nodos Objetivos Investigación	
Objetivo	Relaciones
Objetivo1 / M***S*****z	Objetivos 2, 3, 4, 5
Objetivo2 / L***C*******a	Objetivos 1, 3, 4, 5
Objetivo3 / L***P****C***P***a	Objetivos 1, 2, 4, 5
Objetivo4 / G******E*******a	Objetivos 1, 2, 3, 5
Objetivo5 / D****A****a	Objetivos 1, 2, 3, 4

2.11.5. Resultados caso práctico

Con el Uso de CRAWLERFB se identificó en el lapso de 12 Horas, que existe una evidente relación en redes sociales entre los objetivos de las investigaciones propuestas. Estos objetivos poseen coincidencia en su ciudades natales y actuales, en sus lugares de trabajo y de estudio.

Se identifican 23 nuevos perfiles relacionados directamente con nuevos Callcenters que no se encontraban en el radar de la investigación, entre estos 23 perfiles se identificaron parejas sentimentales, hermanos y otros familiares que son representantes legales de nuevos Callcenters que también se encuentran vinculados en noticias criminales en la FGN.

Como resultado la herramienta CRAWLERFB entrega a los analistas, más capacidades visuales y agilidad que le permiten generar un mejor mapa conceptual de las investigaciones donde exista uso de la red social Facebook, obtener solo la información de los nodos amigos, de manera manual involucraría más de 36 horas de trabajo hombre máquina, tal como se demuestra en el siguiente capítulo de este documento.

3. Capítulo III

En el capítulo final del documento, se evalúan los resultados obtenidos, se identifican las posibilidades de expansión y propuestas a futuro y finalmente, se presentan las conclusiones logradas durante las diferentes etapas del proyecto.

3.1. Resultados

Si bien es cierto que no se poseen las capacidades técnicas y el nivel de desarrollo de cualquiera de los proponentes de herramientas licenciadas, se ha demostrado en este documento que es posible con poca inversión y en poco tiempo, desarrollar, para la FGN una herramienta que realice labores de análisis y recolección de datos sobre la plataforma de la red social Facebook.

Durante las etapas de desarrollo de esta propuesta, se pudo identificar que existen tres reconocidas técnicas, que permiten recolectar información en las redes sociales. La utilización de las API´s propias de cada red social, la extracción manual y la técnica de Crawlin automatizado, cada una de estas técnicas se limitan a la información pública del objetivo a investigar y también a la respuesta, control y verificación de la red social.

Con el propósito de determinar como la herramienta CRAWLERFB, mejora las técnicas de recolección de datos en la red social Facebook, se estable un comparativo entre la aplicación y la recolección de datos manual. Para realizar esta comparación se elige como objetivo un perfil identificado como "wolfmao", en la red social, este perfil es real y pertenece a un funcionario adscrito al grupo de delitos informáticos de la FGN, además el perfil seleccionado cuenta con una configuración de seguridad óptima dentro de los

parámetros establecidos en la red social, donde no son visibles, para perfiles fuera de su

Configuración y herramientas de privacidad

Tu actividad	¿Quién puede ver las publicaciones que hagas a partir de ahora?	Amigos	Editar
	Revisa todas tus publicaciones y los contenidos en los que se te etiquetó		Usar registro de actividad
	¿Quieres limitar los destinatarios de las publicaciones que compartiste con los amigos de tus amigos o que hiciste públicas?		Limitar el público de publicaciones anteriores
Cómo pueden encontrarte y ponerse en contacto contigo	¿Quién puede enviarte solicitudes de amistad?	Amigos de amigos	Editar
	¿Quién puede ver tu lista de amigos? Recuerda que tus amigos controlan quién puede ver sus amistades en sus propias biografías. Si alguien puede ver tu amistad en la biografía de otra persona, podrá verla en la sección de noticias, en la búsqueda y en otros lugares de Facebook. Si cambias la privacidad a Solo yo, solo tú podrás ver tu lista de amigos completa en tu biografía. Las demás personas solo podrán ver los amigos que tienen en común.	Solo yo	Editar
	¿Quién puede buscarte con la dirección de correo electrónico que proporcionaste?	Amigos	Editar
	¿Quién puede buscarte con el número de teléfono que proporcionaste?	Solo yo	Editar
	¿Quieres que los motores de búsqueda fuera de Facebook enlacen a tu perfil?	No	Editar

Ilustración 40 Configuración seguridad del objetivo a investigar.

círculo de red, las publicaciones, los amigos y la información privada.

Se realiza un procedimiento de búsqueda manual de los datos públicos del perfil objetivo, y dado que el señuelo utilizado no pertenece al círculo social, no es posible identificar sus amigos y nodos cercanos. Por tal motivo se procedió a revisar las publicaciones públicas del perfil y de manera manual, visitando cada una de las publicaciones y de estas, utilizando una hoja de cálculo, se toman los datos de los perfiles que comentan, o reaccionan a las imágenes y o publicaciones visibles en la biografía del objetivo.

Posteriormente se realiza la búsqueda del mismo perfil con el uso de la herramienta CRAWLERFB, y adicional se realiza la explotación de la información pública de los perfiles nodos amigos identificados, etapa que no es posible realizar de modo manual. A continuación, se presentan los resultados obtenidos.

Tabla 18

Comparativo CRAWLERFB y extracción manual.

Objetivo	Búsqueda Manual			Búsqueda CRAWLERFB		
	Nodos	Relaciones	Tiempo	Nodos	Relaciones	Tiempo
Wolfmao	3	28	32 Min 10 Seg	26	28	8 Min 42 Seg

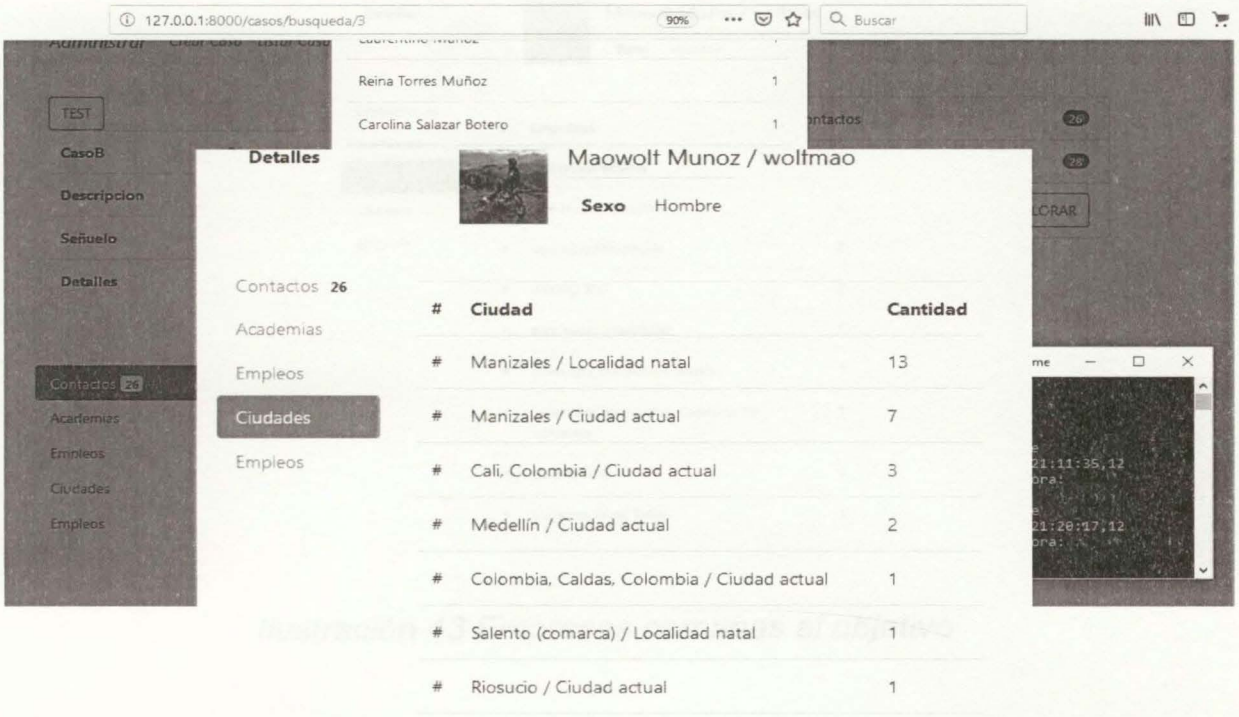


Ilustración 42 Recopilación de Ciudades Actuales y Natales, cercanas al objetivo

Es posible observar en la Tabla 18, como la herramienta CRAWLERFB, emplea menos de un tercio del tiempo que se utiliza la búsqueda manual. Adicionalmente la herramienta vista cada uno de los perfiles o nodos encontrados, extrayendo de estos los datos públicos en la red social y así poder determinar las relaciones de ciudades natales o actuales, las instituciones educativas y las empresas donde trabajan los nodos amigos identificados. Esta información es de gran importancia para los investigadores de la FGN, porque permite perfilar un poco mejor el objetivo de la investigación.

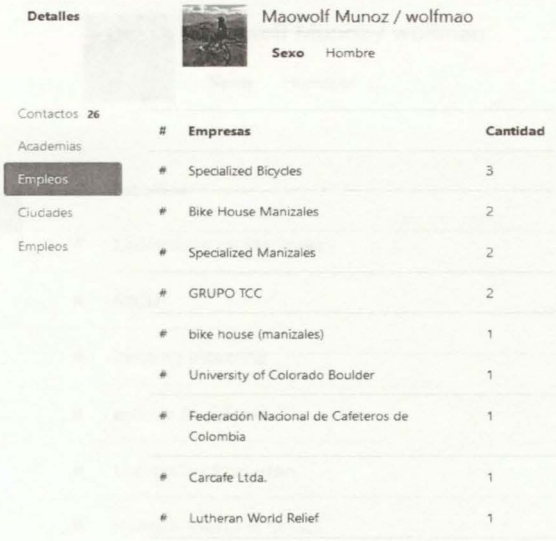


Ilustración 43 Empresas cercanas al objetivo

Detalles

Maowolf Munoz / wolfmao

Sexo Hombre

Contactos 26

Academias

Empleos

Ciudades

Empleos

#	Estudios	Cantidad
#	Universidad de Manizales	6
#	SBCU	2
#	instituto villamaria	1
#	escuela de ciclismo	1
#	University of Houston	1
#	Harvard-Westlake School	1
#	University of Caldas	1
#	Colegio Seminario Redentorista	1
#	sena regional valle	1

Ilustración 44 Recopilación de Academias cercanas al objetivo.

Se puede constatar que el tiempo, la cantidad y la información recolectada son mucho mejor que los obtenidos con las técnicas manuales, y adicionalmente durante el tiempo de ejecución del programa, no se requiere presencia humana que interactúe con el sistema, caso muy diferente comparado con la búsqueda manual donde el tempo se contabiliza en horas hombre – máquina.

A continuación, se realiza un comparativo cualitativo entre las herramientas licenciadas, las herramientas libres y el desarrollo CRAWLERFB, en este análisis se observa que, a pesar de no poseer las capacidades de las herramientas licenciadas, la relación costo beneficio de CRAWLERFB es mejor, para las necesidades de la FGN, además de brindar capacidades de análisis e información específica en las investigaciones apoyadas en fuentes de información como la red social Facebook.

Tabla 19

Comparativo entre herramientas libres, pagas y CRAWLERFB

Características	Herramientas Libres	Herramientas Pagas	CRAWLERFB
Costo	Ninguno	Entre 3000 y 6000 millones	15 millones
Licenciamiento Anual	Ninguno	Entre el 10% y 15% de valor inicial	Ninguno
Nivel Conocimientos adquiridos	Medio	Bajo	Alto
Nivel de posibilidad de Futuros desarrollos	Medio	Bajo	Alto
Adaptación a los cambios en las redes sociales	Ninguno	Solo contratado	Alto
Nivel de información recolectada	Bajo	Alto	Medio
Nivel de Riesgo al vincular con bases de datos propias de la FGN	Medio	Alto	Bajo

3.2. Propuestas a futuro.

Existen múltiples posibilidades de expansión en este tipo de desarrollos, dado el exponencial crecimiento y uso de las redes sociales, y dada también a la necesidad de la FGN de implementar medidas contra los ciberdelitos, las ciberamenazas. Con base en este desarrollo se pueden generar proyectos futuros como:

- Sistema de alertas para redes sociales: resulta interesante, con los conocimientos adquiridos, generar y aplicar las técnicas y las estructuras de Crawling a otras redes sociales que sean de interés, y de este modo poder efectuar sistemas de alarmas y seguimientos antes hechos delictivos como ofrecimiento de productos

ilícitos y amenazas, temas que se ha identificado como necesarios para la detección y prevención del delito por parte de las autoridades colombianas.

- Crawling a la DEEP WEB: Por la estructura y constitución de la red profunda, donde las direcciones u onions son volátiles y cambiantes, resulta un producto interesante realizar mediante técnicas de Crawling y Scraping, la identificacion actividades delictivas de páginas que se publicitan a través de foros. Actividad que se puede realizar de manera automática y almacenar los sitios de la DEPP WEB que puedan interesar a nuestra jurisdicción para la prevención e identificación del delito.

- Descarga y Firma digital de perfiles: teniendo en cuenta la volatilidad de la información en las redes sociales, sería importante poder realizar descargas completas de la información pública perfiles de usuarios y generar una marca temporal y una firma HASH que permita legalizar esta información ante los estrados y los procesos judiciales.

- Búsqueda facial en redes sociales: Adquirida la posibilidad de recorrer y descargar un perfil completo de las redes sociales, sería un interesante proyecto detectar los rostros de las fotografías adquiridas y realizar búsquedas de personas con características morfológicas específicas. existen compendios de herramientas de desarrollo libre que realizan una evaluación cuantitativa de las características morfológicas brinda un grado de certeza e identificación.

- Perfilación virtual: Dado la poca formación en seguridad de la información que posee nuestra base poblacional, la comunidad ha reflejado en las redes sociales,

sus gustos, comportamientos, odios, problemáticas y afinidades. Toda esta información es un material importante que se puede obtener de con las técnicas presentadas en este documento, y de este modo generar una perfilación del mundo virtual de un individuo, útil para procesos de selección laboral, o para procesos de análisis y perfilación criminal, identificando patrones de comportamiento que pueden resultar factores de riesgo para una organización o para el estado social de derecho.

- Analizar tendencias comportamentales: Es un proyecto interesante que se puede generar a partir de este desarrollo, generar sistemas de alertas y alarmas ante palabras de peligro, criterios y comportamientos que identifican potenciales actos delictivos.

3.3. Conclusiones.

La obtención y análisis de información de las redes sociales brinda a los investigadores de la FGN, grandes capacidades que les permite segregar e identificar los niveles de relaciones y los grupos de personas vinculados a una investigación, esta nueva información de carácter orientativo, es de gran importancia en la conducción de las hipótesis delictivas construidas en las características de ubicación, cercanía y relación.

La alta reducción en consumo de tiempo y la posibilidad de delimitar el foco de la investigación, es la mayor ventaja que se obtiene con el uso de herramientas orientadas a la investigación en fuentes abiertas, para el caso de estudio tratado en este documento, el proceso de recolección de información y la presentación al analista, tomo 12 horas de trabajo máquina, durante este tiempo no es necesario la participación directa del usuario,

gracias a la automatización que nos brinda la herramienta planteada, calcular este tiempo en función de una investigación en la FGN, resulta imposible por las continuas y variadas actividades que los investigadores deben realizar en el marco de la investigación.

Sin ser la programación, una gran fortaleza, del equipo de desarrollo de este proyecto, ha logrado crear una aplicación, CRAWLERFB, con altos niveles de seguridad y efectividad en los procesos de recolección de información de la plataforma de Facebook. Se ha logrado comprender la técnica y desarrollar la metodología que permite mejorar las capacidades de análisis de información enfocada a una investigación criminal y recogida en la red social.

Se evidencio, durante las etapas de recolección de información mediante encuestas, que además del uso de herramientas tecnológicas, se necesita también mejorar las capacidades de técnicas en el uso de las redes sociales, se necesitan avatares o perfiles bien construidos que cumplan las funciones de agentes encubiertos virtuales que se acercar a los objetivos de la investigación.

Los costos de las herramientas OSINT licenciadas, como se logró dejar evidenciado, superan a un desarrollo In-house, teniendo en cuenta que el ciclo de vida de estas propuestas depende de las restricciones y controles que implementan las fuentes de información, las capacidades de obtención y análisis de información de las fuentes abiertas es mucho más que la compra de herramientas, el cambiante entorno de las redes sociales obligan a instituciones como la FGN a generar capacidades y habilidades específicas de obtención, análisis, desarrollo y producción de información a partir de los medios abiertos y las redes sociales.

Resulta necesario que las entidades del estado, como la FGN, se vinculen a los grupos investigativos de los centros educativos superiores, quienes pueden aportar las capacidades técnicas para desarrollar herramientas a la medida, aumentando las capacidades investigativas.

Se concluye que es necesario, en entidades como la FGN, implementar grupos de investigación enfocados a los medios virtuales y a las redes sociales, estos grupos deben estar constantemente en la búsqueda de conceptos y estructuras, que apunten a nuevas metodologías de obtención de información, enriquecidas con capacitaciones, nuevas tecnologías y técnicas, que permitan afrontar las necesidades de investigación del siglo XXI.

3.4. Referencias

Brin, S., & Page, L. (2012). *http://infolab.stanford.edu.* Obtenido de http://infolab.stanford.edu/~backrub/google.html

Django Software Foundation. (1 de Diciembre de 2017). *Django*. Recuperado el 05 de febrero de 2019, de https://docs.djangoproject.com/en/2.2/

Eichmann, D. (1994). *http://citeseerx.ist.psu.edu.* Obtenido de http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.589.3925&rep=rep1&type=pdf

Fard, M., & Ester, M. (s.f.). Published in: 2009 International Conference on Computational Science and Engineering. *Collaborative Mining in Multiple Social Networks Data for Criminal Group Discovery.* Vancouver. Obtenido de https://ieeexplore.ieee.org/document/5283849

Fedesoft. (2015). *cenisoft.* Recuperado el 22 de junio de 2018, de http://cenisoft.org/estudios/EstudiodeSalarios2015.pdf

Felip I Sarda, J. M. (2004). La gestión de fuentes abiertas por los servicios de Inteligencia y equipos de investigacion. *Cuadernos constitucionales de la Cátedra Fadrique Furió Ceriol*, 41-50. Recuperado el Agosto de 2018, de https://dialnet.unirioja.es/servlet/articulo?codigo=2270934

Fernadez, E., & Quevedo, J. M. (Noviembre de 2018). El profesional de la informacion. *27*(6). Obtenido de https://recyt.fecyt.es/index.php/EPI/article/view/epi.2018.nov.12

Fiscalía General de la Nacón. (01 de Febrero de 2019). *Fiscalia*. Recuperado el 15 de Julio de 2019, de https://www.fiscalia.gov.co/colombia/la-entidad/presupuesto-general-asignado/

Gephi. (01 de Febreo de 2016). *gephi*. Recuperado el 18 de Junio de 2018, de https://gephi.org/

Iglesias, A. (20 de noviembre de 2014). *Blogthinkbig*. Recuperado el 22 de Marzo de 2014, de https://blogthinkbig.com/buscadores-de-internet

Martín, J. R. (15 de Mayo de 2010). *Como Explotar Osint Eficazmente.* Recuperado el 22 de Octubre de 2018, de Ministerio de Defensa España: http://www.defensa.gob.es/ceseden/Galerias/esfas/destacados/en_portada/COMOx20EXPLOTARx20OSINTx20EFICAZMENTE.pdf

Moya, E. (Octubre de 2012). *Las Redes Sociales como fuentes de información (OSINT).* Recuperado el 22 de Octubre de 2018, de https://intranet.bibliotecasgc.bage.es/intranet-tmpl/prog/local_repository/documents/5149.pdf

Parrilla, M. (2012). *Universidad Carlos III de Madrid.* Recuperado el 22 de marzo de 2019, de https://e-archivo.uc3m.es/bitstream/handle/10016/16759/TFG%20-%20Marta%20Parrilla%20Monrocle.pdf?sequence=1&isAllowed=y

Python Software Fundation. (2001). *python*. Recuperado el 01 de Junio de 2018, de https://www.python.org/

ROSENBERG, M., & J. X. DANCE, G. (10 de 4 de 2018). *The New York Times*. Obtenido de https://www.nytimes.com/es/2018/04/10/facebook-cambridge-analytica/

Rosenberg, M., & J. X. Dance , G. (10 de Abril de 2018). Así funcionaba la recolección de datos de Cambridge Analytica. *The New York Times*. Recuperado el 15 de Noviembre de 2018, de https://www.nytimes.com/es/2018/04/10/facebook-cambridge-analytica/

Selenium Project. (15 de Mayo de 2019). *seleniumhq*. Recuperado el 05 de 06 de 2019, de https://www.seleniumhq.org/docs/

Siti, I., Selamat, A., & Selamat, H. (2008). International Conference on Computer and Communication Engineering. *Scalable e-business social network using MultiCrawler agent.* Kuala Lumpur. Obtenido de https://ieeexplore.ieee.org/document/4580695

SQLite Consortium. (06 de Febrero de 2006). *sqlite.org*. Recuperado el 05 de Febreo de 2019, de https://www.sqlite.org

We are social ; Hootsuite. (Enero de 2019). *Datereportal*. Recuperado el 14 de mayo de 2019, de https://datareportal.com/reports/digital-2019-global-digital-overview.

3.5. Listado de Tablas

3.6. Listado de Ilustraciones

3.7. Palabras Clave

3.8. Anexos

- Especificación de Requisitos Estándar IEEE 830.
- Manual Técnico y de Instalación.
- Manual de Usuario.

OBTENCION DE REQUERIMIENTOS SEGÚN ESTÁNDAR IEEE 830

1. Introducción.

En el presente documento se describen las Especificaciones de Requisitos Software (ERS), para el desarrollo plateado como CRAWLERFB, teniendo en cuenta las necesidades de información los investigadores de la FGN sobre objetivos identificados en las redes sociales.

Durante las etapas investigativas en la FGN, los investigadores se ven evocados a utilizar las redes sociales, en especial Facebook, para identificar relaciones y algunos patrones que enriquezcan sus hipótesis delictivas, dado que no se encuentra al alcance una herramienta que les facilite esta tarea se plantea CRAWLERFB.

1.1. Propósito del Sistema

El propósito es plasmar claras, las funcionalidades y requisitos del sistema CRAWLERFB, este documento va dirigido tanto como para el departamento de Tecnologías de la Fiscalía General Nación, como para El departamento de seguridad ciudadana, quienes serían los indicados para impulsar la propuesta en el ente fiscal.

1.2. Alcance

CRAWLERFB, es un sistema que permitirá en la rede social de Facebook, extraer la información de publica de perfiles de usuarios, puntos de convergencia y relaciones entre perfiles, Información recuperada de los datos públicos. Es un sistema que genera la materia prima para representar grupos relaciones entre perfiles o comunidades de Facebook, mediante el uso de herramientas externas como GEPHI o I2, además de ser un sistema que identifica puntos de convergencia entre redes o comunidades de perfiles de la red social.

1.3. Definiciones Siglas y Abreviaturas.

- FGN: Fiscalía General de la Nación, ente encargado de ejecutar la acción penal en la república de Colombia.

- GEPHI: Plataforma Open Graph Viz, software para visualización y exploración para de gráficos y redes.

- I2: Software desarrollado por IBM para identificar relaciones y niveles de convergencia de una red, mediante gráficos interconectados.

- SELENIUM: Entorno de pruebas de software para aplicaciones basadas en la web, librería escrita en Python que permite la automatizacion y control de navegadores WEB.

- SQLITE: Sistema de bases de datos relaciones con grandes facilidades de compresión y portabilidad.

- NUNC: Numero Único de Noticia Criminal, código de 21 dígitos que identifica e individualiza cada uno de los casos investigativos en la FGN.

- PYTHON: Lenguaje de programación Interpretado, caracterizado por su legibilidad en su código fuente.

- DJANGO: Plataforma de desarrollo orientado a la WEB, escrito en Python, busca fomentar el desarrollo rápido y seguro de aplicaciones.

- SEÑUELO: Perfil de Facebook diseñado para interactuar y realizar las búsqueda en la red social.

- OBJETIVO: Perfil de Facebook objeto del caso investigativo.

- NODO: Perfiles de Facebook identificados como amigos de un perfil Objetivo.

- WEB: World Wide Web (WWW) o Red informática mundial.

- RF: Requerimiento Funcional

- RNF: Requerimiento No funcional

1.4. Referencias

Título del Documento	Referencia
Standard IEEE 830 - 1998	IEEE

1.5. Resumen

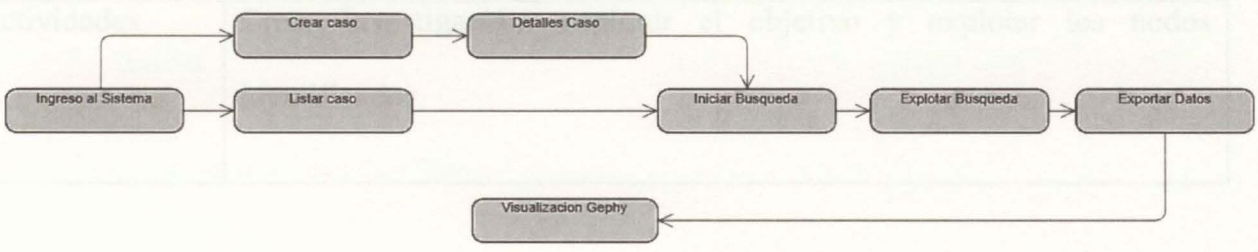
Este proyecto reducirá el tiempo de horas hombre máquina que utiliza el investigador de la FGN en realizar búsqueda de relaciones y patrones de convergencia de academia, ubicación y empleo de un perfil o un grupo de perfiles identificados en la red social Facebook.

2. Descripción General

2.1. Perspectiva del producto

Esta herramienta está diseñada para trabajar en entornos WEB, y que sea de fácil acceso desde la red interna de la FGN donde el sistema mediante el uso de la librería SELENIUM recorrerá todas las publicaciones y la información pública de un perfil, identificando los Nodos y los niveles de relación de estos con el objetivo de la investigación.

2.2. Funcionalidades de CRAWLERFB



El sistema CRAWLERFB se plantea para crear casos investigativos en los cuales existen tres funciones principales.

- Iniciar búsqueda: El investigador busca la información pública del Objetivo de la investigación.
- Explotar búsqueda: Se inicia el proceso de recolección de nodos y niveles de relación entre el objetivo y los nodos que han reaccionado a las publicaciones del objetivo de la investigación.
- Exporta Datos: Se generan los archivos necesarios para crear los gráficos de relación.

2.3. Características de usuarios.

Tipo de usuario	Investigador
Formación	Perfil Investigador o analista de la FGN, Nivel básico.
Habilidades	Identificar el objetivo de la investigación, crear y alimentar un perfil señuelo que se pueda acercar lo más posible al objetivo de la investigación. Debe poder crear y manipular grafico de relaciones en Gephi o en I2.
Actividades	Crear Investigación, explorar el objetivo y explotar los nodos identificados

Tipo de usuario	Administrador
Formación	Perfil Investigador o analista de la FGN, Nivel medio.
Habilidades	Identificar los perfiles de los investigadores asociados al sistema y a la FGN.
Actividades	Crear y administrar los perfiles de los Investigadores en la plataforma.

2.4. Restricciones

- Al ser una red social la plataforma de investigación, se requiere conexión a internet.
- Leguaje de programación Python y la librería Selenium.
- Se requieren perfiles de la red social o señuelos los cuales deben ser validados y activos en la red social, además de ser perfilados de acuerdo al objetivo de la investigación.
- Es necesario emular comportamiento humano al recorrer el perfil Objetivo, porque la red social Facebook detecta los robots que recorren su red y bloquea su dirección IP de acceso.
- Es necesario conocer la estructura de las secciones y publicaciones de la red social.

2.5. Suposiciones dependencias.

- Depende de internet y sus servidores.
- Depende de Facebook, su construcción y sus políticas regulatorias.
- Depende de la cantidad de información pública y la configuración de seguridad del perfil objetivo de la investigación.
- Depende de herramientas de representación gráfica para visualizar las relaciones y los niveles de relación entre los nodos y el objetivo de la investigación.

2.6. Evolución previsible del sistema.

- Descarga y firma digital de las publicaciones de un objetivo de investigación.
- Búsqueda por reconocimiento facial en las fotos de un objetivo.
- Sistema de alertas ante palabras claves y o comportamientos de un objetivo de investigación en la red social de Facebook.

3. Requisitos Específicos

3.1. Requerimientos funcionales

Identificación del requerimiento	RF01
Nombre del requerimiento	Autenticación de usuario
Características	Los dos tipos de usuarios Administrador o investigador deberán identificarse en el sistema

Identificación del requerimiento	RF02
Nombre del requerimiento	Asignación de Staff al perfil de investigador
Características	El usuario administrador valida la información de los usuarios investigadores y les asigna el rol de staff.

Identificación del requerimiento	RF03
Nombre del requerimiento	Registrar usuarios
Características	El usuario administrador registra la información de los usuarios investigadores.

Identificación del requerimiento	RF04
Nombre del	Crear Caso

requerimiento	
Características	El usuario investigador crea los casos a investigar insertando las características y valores del señuelo y del objetivo de la investigación

Identificación del requerimiento	RF05
Nombre del requerimiento	Buscar datos
Características	El usuario investigador inicia la investigación consultado la información pública del objetivo, como son academias, ciudades actuales y natales y las empresas donde ha laborado.

Identificación del requerimiento	RF06
Nombre del requerimiento	Buscar
Características	El usuario investigador continúa la investigación recorriendo cada una de las publicaciones del objetivo, extrayendo y contabilizando

	los nodos y las reacciones de estos contra el objetivo de la investigación.

Identificación del requerimiento	RF07
Nombre del requerimiento	Explotar
Características	El usuario investigador continua con la explotación o con la búsqueda de la información pública de todos los nodos identificados en la fase de buscar, identificando patrones de coincidencia en ciudades, empleo y academias

Identificación del requerimiento	RF08
Nombre del requerimiento	Exportar
Características	El usuario investigador genera los archivos necesarios para graficar los niveles de relación entre los nodos y los objetivos de la Investigación

3.2. Requisitos No Funcionales

Identificación del requerimiento	RNF01
Nombre del requerimiento	Interfaz
Características	El sistema debe presentar una interfaz intuitiva y de fácil uso.

Identificación del requerimiento	RNF02
Nombre del requerimiento	Administración
Características	El sistema se podrá administrar completamente desde la interfaz WEB

Identificación del requerimiento	RNF03
Nombre del	Nivel de Usuario

requerimiento	
Características	El sistema solo mostrara a un usuario la información pertinente a sus casos investigativos.

Identificación del requerimiento	RNF04
Nombre del requerimiento	Seguridad
Características	El sistema debe garantizar la información consignada por cada usuario sobre sus investigaciones.

Identificación del requerimiento	RNF05
Nombre del requerimiento	Centralidad
Características	El sistema debe instalado en un servidor central por procesos de administración y verificación de la información y los perfiles de usuario

MANUAL TECNICO Y DE INSTALACION DE CRAWLERFB

1. INTRODUCCION

El sistema de información CRAWLERFB, es una herramienta diseñada para apoyar las investigaciones de la FGN, mediante la recolección de la información pública de un perfil de la red social Facebook, permitiendo identificar patrones de ubicación, educación y empleo de un individuo o un grupo de individuos que se relacionan entre sí. La herramienta crea los insumos para generar gráficos de nivel de relación entre un objetivo y sus nodos asociados.

2. REQUERIMIENTOS TECNICOS.

2.1. Requerimientos de Hardware

La instalación se recomienda realizarla sobre un equipo que cuente mínimo con las siguientes especificaciones, que se deben aumentar los requerimientos a medida que aumentan los usuarios en el sistema. Estas especificaciones se plantearon para un grupo de trabajo de 20 usuarios concurrentes.

- Procesador Celeron
- Memoria RAM 8 GB

- Disco duro 500 GB

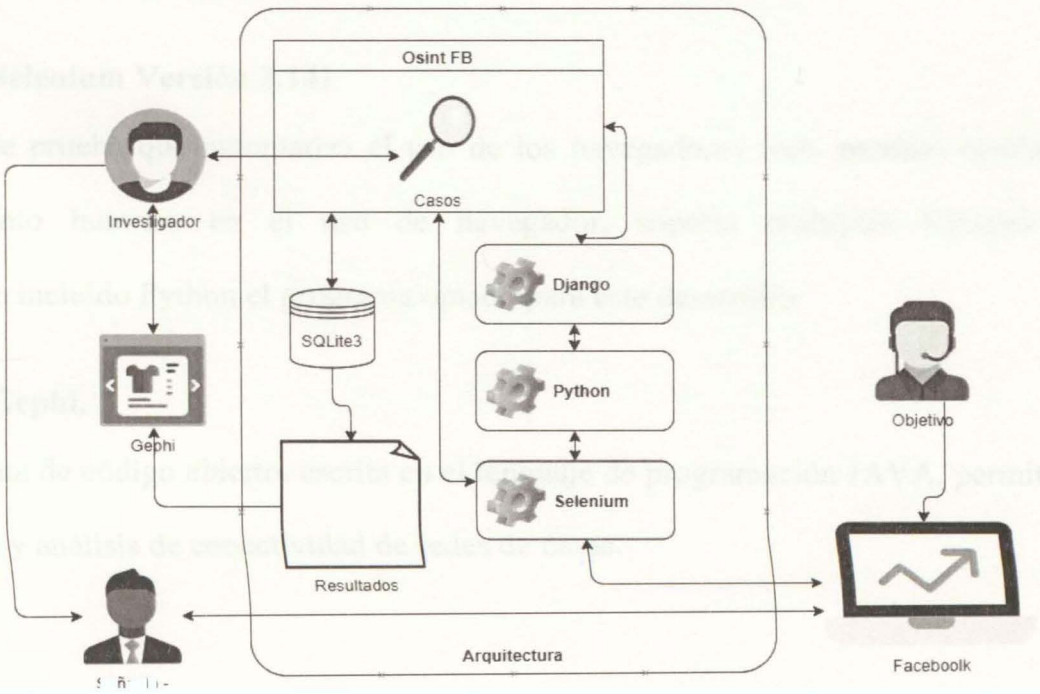
- Conexión Internet 5 conexiones VPN y un canal de 10 megas.

Es necesario tener en cuenta que el aplicativo emula el comportamiento humano en la conexión y exploración de la red social Facebook, por lo tanto requiere de acceso ilimitado a la red y depende de la disponibilidad del servicio de la plataforma WEB.

Es necesario poseer un rango de conexiones VPN, para evitar que la dirección IP utilizada en os procesos de exploración, se bloqueada por la red social Facebook.

2.2. Requerimientos de Software

El aplicativo se encuentra desarrollado en el lenguaje de programación Python, con motor de base de datos relacional en SQLite, una interfaz de usuario intuitiva y rápida desarrollada en el framework Django, la cual mediante el uso de la librería Selenium recorra la información pública de un perfil de Facebook y extraiga los contactos y relaciones, para posteriormente identificar características de ubicación, de comportamiento



y graficar sus relaciones y cercanía mediante la herramienta Gephi.

2.2.1. Python Versión 3.7

Lenguaje de programación de código abierto, orientado a Objetos creado por Guido Van Rossum, reconocido por su potencia y flexibilidad, su fácil comprensión y una gran comunidad que realiza grandes aportes en desarrollo y soporte de sus librerías.

2.2.2. Django Versión 2.2.3

Ilustración 45 Arquitectura de OSINTFB

Es un Framework de desarrollo web de código abierto, escrito en Python, fomenta el desarrollo limpio y rápido de las aplicaciones web, agiliza e implementa en su código tareas como las funciones de validación de entrada de datos y las interfaces de administración y login.

2.2.3. Selenium Versión 3.141

Entorno de prueba que automatiza el uso de los navegadores web, permite emular el comportamiento humano en el uso de navegador, soporta múltiples leguajes de programación incluido Python el programa optado para este desarrollo.

2.2.4. Gephi.

Herramienta de código abierto, escrita en el lenguaje de programación JAVA, permite la visualización y análisis de conectividad de redes de datos.

2.2.5. Bases de datos SQLite.

Biblioteca Open Source, escrita en leguaje de programación C, y utiliza el motor de bases de datos SQL, permite trabajar con bases de datos multiplataforma optimizando su tamaño y portabilidad.

3. INSTALACION

Para el proceso de instalación sobre la plataforma de Windows, se hace necesario configurar el entorno de trabajo instalando Python como variable de entorno.

3.1. Instalación de Python 3.7

Acceder a la página de Python sección de descargas y obtener el paquete de instalación de acuerdo a las especificaciones del sistema, el paquete se puede descargar en la web, sitio oficial, https://www.python.org/downloads/windows/. Una vez obtenido el paquete de instalación, y utilizando un usuario con perfil de administrador, se ejecuta la instalación siguiendo los pasos recomendados.

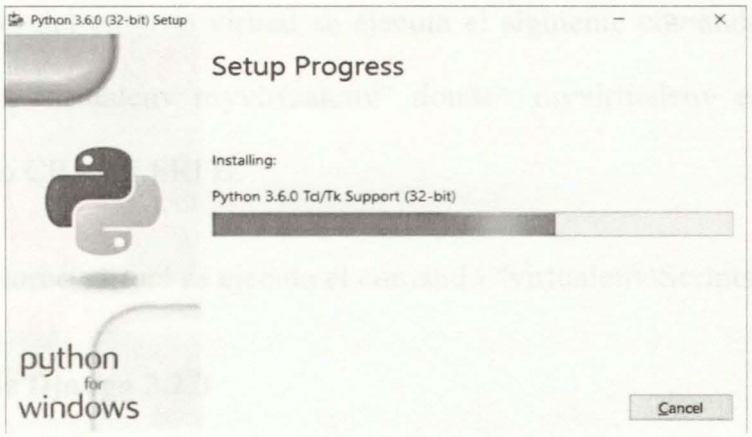


Ilustración 46 Instalación de Python

Una vez instalado el paquete seleccionado de Python, se recomienda agregar Python al PATH, mediante las configuraciones de variables de entorno de Windows. O mediante la consola CMD utilizando el comando "set PATH=%PATH%;c:\python36\"

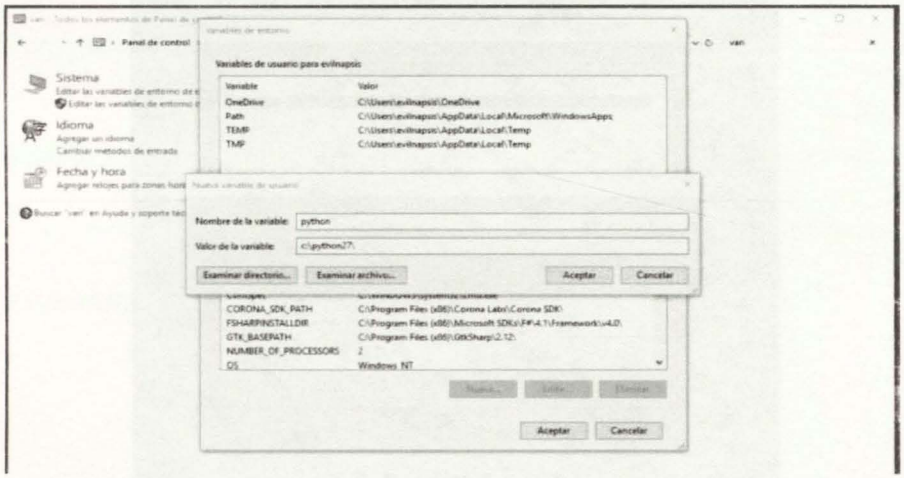


Ilustración 47 Variable Python en el PATH

3.2. Creación del Entorno Virtual

El entorno virtual o virtualenv es una herramienta que encapsula la configuración del proyecto, permitiendo movilidad entre servidores y evitando que la configuración de proyectos ubicados en el mismo servidor interfiera o modifique la configuración de nuestro desarrollo.

Para la instalación del entorno virtual se ejecuta el siguiente comando en una consola CMD: "python3 –m virtualenv myvirtualenv" donde myvirtualenv es el nombre del proyecto en este caso CRAWLERFB.

Para activar el entorno virtual se ejecuta el comando "virtualenv\Scripts\activate"

3.3. Instalación de Django 2.2.3

Una vez instalado y activo el entorno virtual se procede a instalar DJANGO, mediante el paquete de instalación PIP, con el comando "Python3 –m pip install django==2.2.3"

Logrado e instalado DJANGO, se procede a instalar las dependencias del proyecto, estas son las librerías utilizadas que facilitan las acciones y funcionalidades del proyecto.

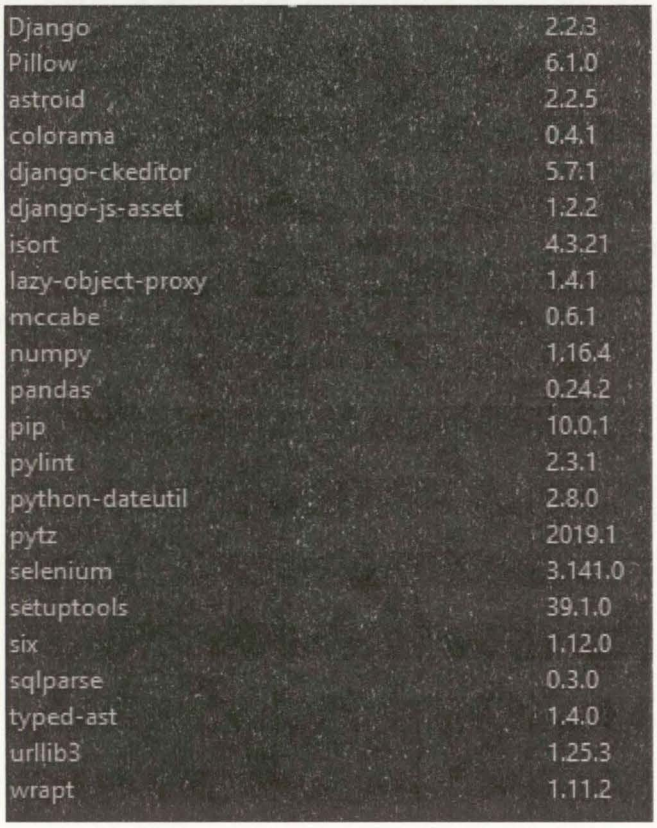


Ilustración 48 Librerías y/o dependencias del proyecto

Las dependencias se encuentra descritas en un archivo de nombre *requirements.txt,* y para su instalación solo es necesario ejecutar el comando "*python3 –m pip install –r requeriments.txt*".

4. ESTRUCTURA DEL PROYECTO

Como se ha explicado, este desarrollo se plantea bajo estructura del Framework Django V2.2.3, y por ende la arquitectura de archivos, se regula por el modelo (MVT) Modelo, Vista, Template, en la carpeta "osint2" se encuentra los archivos de la aplicación. La cual debe ser copiada al entorno virtual creado.

Entre las carpetas y archivos principales podemos encontrar:

- **Templates:** Carpeta donde se almacenan los archivos HTML con los que interactúa el usuario del sistema.
- **Models.py:** en este archivo se encuentra el modelo del sistema, la estructura de la base de datos se construye a partir de este modelo, también se encuentran validaciones de integridad de los datos.

- **Views.py** archivo que contiene las funciones y acciones propias de la aplicación, es la lógica del sistema que envía a los Template los resultados de

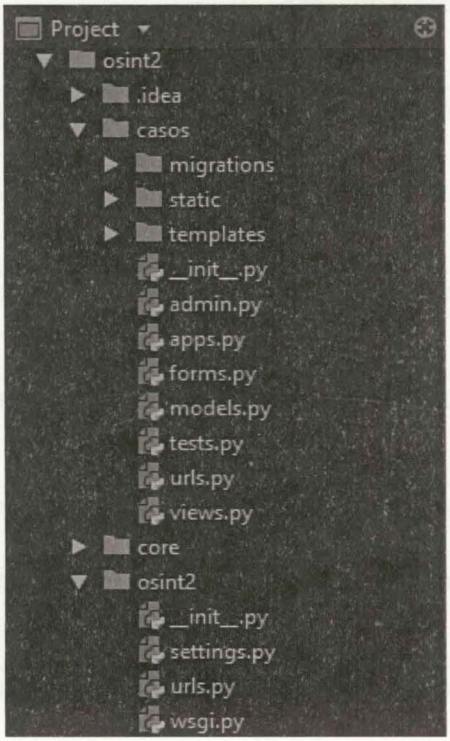


Ilustración 49 Estructura del código fuente

las funciones.

4.1.1. Views o Vistas del proyecto.

Función	Descripción
CasoListView	Página principal de casos, lista los casos de el usuario registrado en el sistema
CasoDetailView	Muestra los detalles de un caso especifico
CasoCreate	Función para crear un caso por un usuario especifico
encontrar_idface	Función para identificar el ID de un usuario de Facebook, obtenido de un link o dirección web.
encontrar_idface_reaccion	Función para identificar los usuarios que reacciona ante las publicaciones de un objetivo.
guarda_nodo	Función para almacenar en la base de datos un nodo o un amigo identificado de un objetivo
guarda_nodo_origen	Función para almacenar el objetivo de la investigación
guarda_reaccion	Función para registrar en la base de datos las reacciones identificadas de las publicaciones, pueden ser comentarios, like, love etc.
guarda_amigos_si_accede	Función para guardar los nodos si son públicos en el perfil

	objetivo.
Exportar	Función que crea archivo de tipo csv, con los nodos y con las reacciones a graficar
caso_lista_exportar	Función que muestra en pantalla la lista de casos, para que el usuario elija de cuales exportar información para ser graficada
busqueda	Función que inicial la búsqueda de perfiles o nodos vinculados con el objetivo de la investigación.
Lanzar	Función que toma cada uno de os nodos identificados en la investigación y los recorre en búsqueda de la información pública.
dato_nodo	Función que identifica la información publica de academias, ciudades y empleos de los nodos ubicados en la investigación

4.1.2. Modelos y bases de datos

CRAWLERFB, utiliza el modelo por defecto de bases de datos implementado por DJANGO, sqlite3, por lo tanto no se requieren instalaciones adicionales de motor de bases de datos, solo es necesario configurar el archivo "*settings.py*"

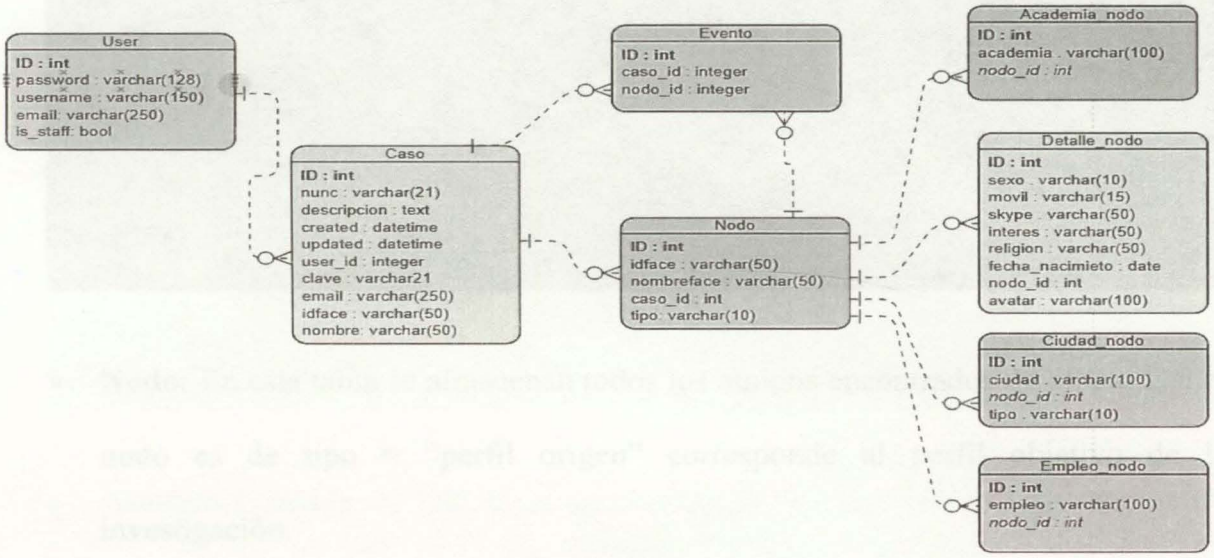
DATABASES = {

'default': {

'ENGINE': 'django.db.backends.sqlite3',

'NAME': os.path.join(BASE_DIR, 'db.sqlite3'),

}



}

El archivo *'db.sqlite3'* contiene la base de datos completa y se encuentra construida mediante los comandos "*makemigrations*" y "*migrate*" de Django, donde toma el archivo "*models.py*" que contiene el diccionario de datos del sistema, el cual se encuentra

Ilustración 50 Modelo Entidad relación

construido con las siguientes clases:

- **User:** Tabla que contiene los usuarios del sistema, se requiere que el usuario es su campo "is_staff" = TRUE, para poder acceder a las funciones del software. Se utiliza por defecto el modelo de usuarios incorporado por DJANGO

- **Caso:** Tabla donde se almacenan los casos de los investigadores, esta tabla registra la noticia criminal, el señuelo a usar para la investigación, su clave y el objetivo.

```
class Caso(models.Model):
    user = models.ForeignKey(User, on_delete=models.CASCADE)
    nunc = models.CharField(verbose_name="Noticia Criminal", max_length=21)
    descripcion = models.TextField(verbose_name="descripcion Caso", null=True)
    created = models.DateTimeField(auto_now_add=True, verbose_name="Fecha de creación")
    updated = models.DateTimeField(auto_now=True, verbose_name="Fecha de edición")
    idface = models.CharField(verbose_name="ID facebook Objetivo", max_length=50, null=True)
    nombre = models.CharField(verbose_name="Nombre Objetivo", max_length=50, null=True)
    email = models.EmailField(verbose_name="señuelo Facebook", null=True)
    clave = models.CharField(verbose_name="clave señuelo", null=True, max_length=21)
```

- **Nodo:** En esta tabla se almacenan todos los amigos encontrados de un caso, si el nodo es de tipo = "perfil origen" corresponde al perfil objetivo de la investigación.

```
class Nodo(models.Model):
    caso = models.ForeignKey(Caso, null=True, blank=True, on_delete=models.CASCADE)
    idface = models.CharField(verbose_name="ID facebook Nodo", max_length=50, null=True)
    nombreface = models.CharField(verbose_name="Nombre Nodo", max_length=50, null=True)
    tipo = models.CharField(verbose_name="Tipo de Nodo", max_length=10, null=True)
```

- **Evento:** En esta tabla se almacena cada interacción entre un nodo y el perfil objetivo de la investigación, se considera un evento cada comentario, reacción o etiqueta de los nodos en las publicaciones que hace el nodo o perfil objetivo de la investigación.

```
class Evento(models.Model):
    caso = models.ForeignKey(Caso, null=True, blank=True, on_delete=models.CASCADE)
    nodo = models.ForeignKey(Nodo, null=True, blank=True, on_delete=models.CASCADE)
```

- **Detalle_nodo:** Se almacena en esta tabla la información pública de cada nodo incluida su avatar o imagen en la red social.

```
class Detalle_nodo (models.Model):
    nodo = models.ForeignKey(Nodo, null=True, blank=True, on_delete=models.CASCADE)
    sexo = models.CharField(verbose_name="Sexo Nodo", max_length=20, null=True)
    movil = models.CharField(verbose_name="Movil Nodo", max_length=15, null=True)
    skype = models.CharField(verbose_name="Skype Nodo", max_length=50, null=True)
    interes = models.CharField(verbose_name="Interes Nodo", max_length=50, null=True)
    religion = models.CharField(verbose_name="Religion Nodo", max_length=50, null=True)
    fecha_nacimiento = models.CharField(verbose_name="Fecha Nacimiento", max_length=30, null=True)
    avatar = models.ImageField(upload_to='avatares', null=True)
    caso = models.ForeignKey(Caso, null=True, blank=True, on_delete=models.CASCADE)
```

- **Academia_nodo:** en esta tabla se almacenan los nombres de los centros de estudio, universidades y colegios que se encuentran en la información pública de los nodos en un caso o investigación.

```
class Academia_nodo(models.Model):
    nodo = models.ForeignKey(Nodo, null=True, blank=True, on_delete=models.CASCADE)
    academia = models.CharField(verbose_name="Formacion Academica Nodo", max_length=100, null=True)
    caso = models.ForeignKey(Caso, null=True, blank=True, on_delete=models.CASCADE)
```

- **Ciudad_nodo:** en esta tabla se almacenan las ciudades natales o actuales que se identifican en la información pública de los nodos en un caso o investigación.

```
class Ciudad_nodo(models.Model):
    nodo = models.ForeignKey(Nodo, null=True, blank=True, on_delete=models.CASCADE)
    ciudad = models.CharField(verbose_name="Ciudad Nodo", max_length=100, null=True)
    tipo = models.CharField(verbose_name="Tipo de Ciudad", max_length=10, null=True)
    caso = models.ForeignKey(Caso, null=True, blank=True, on_delete=models.CASCADE)
```

- **Empleo_nodo:** en esta tabla se almacenan los lugares de trabajo que publican los nodos en su perfil de la red social en un caso o investigación.

```
class Empleo_nodo(models.Model):
    nodo = models.ForeignKey(Nodo, null=True, blank=True, on_delete=models.CASCADE)
    empleo = models.CharField(verbose_name="Empleo Nodo", max_length=100, null=True)
    caso = models.ForeignKey(Caso, null=True, blank=True, on_delete=models.CASCADE)
```

5. LICIENCIAMINETO

En el entendimiento de las características de este desarrollo, definido como un prototipo que representa las bases creativas, de futuros productos de ámbito publico y privado, donde se pueden ver relacionadas entidades educativas y entidades de gobierno, las cuales vincularan a sus propias creaciones y productos, sus bases de datos y conocimientos específicos, se elige como tipo de licenciamiento a APACHE v 2.0, una licencia permisiva, creada por "Apache Software Foundation", que no requiere que los proyectos derivados de este trabajo, se publiquen bajo el mismo tipo de licencia, y tampoco exige la liberación del código fuente.

Este tipo de licenciamiento, APACHE 2.0, brinda la autonomía necesaria a las entidades estatales de para continuar en sus propios proyectos, y conectar sus recursos, y mantener la información de carácter reservada, protegida.

5.1. Licencia

Los términos y condiciones de la licencia escogida, APACHE v2.0 pueden ser encontrados en el enlace http://www.apache.org/licenses/LICENSE-2.0.

Licencia Apache 2.0

Apache License

Version 2.0, January 2004

http://www.apache.org/licenses/

TERMS AND CONDITIONS FOR USE, REPRODUCTION, AND DISTRIBUTION

1. Definitions.

"**License**" shall mean the terms and conditions for use, reproduction, and distribution as defined by Sections 1 through 9 of this document.

"**Licensor**" shall mean the copyright owner or entity authorized by the copyright owner that is granting the License.

"**Legal Entity**" shall mean the union of the acting entity and all other entities that control, are controlled by, or are under common control with that entity. For the purposes of this definition, "**control**" means (i) the power, direct or indirect, to cause the direction or management of such entity, whether by contract or otherwise, or (ii) ownership of fifty percent (50%) or more of the outstanding shares, or (iii) beneficial ownership of such entity.

"**You**" (or "**Your**") shall mean an individual or Legal Entity exercising permissions granted by this License.

"**Source**" form shall mean the preferred form for making modifications, including but not limited to software source code, documentation source, and configuration files.

"**Object**" form shall mean any form resulting from mechanical transformation or translation of a Source form, including but not limited to compiled object code, generated documentation, and conversions to other media types.

"**Work**" shall mean the work of authorship, whether in Source or Object form, made available under the License, as indicated by a copyright notice that is included in or attached to the work (an example is provided in the Appendix below).

"**Derivative Works**" shall mean any work, whether in Source or Object form, that is based on (or derived from) the Work and for which the editorial revisions, annotations, elaborations, or other modifications represent, as a whole, an original work of authorship. For the purposes of this License, Derivative Works shall not include works that remain separable from, or merely link (or bind by name) to the interfaces of, the Work and Derivative Works thereof.

"**Contribution**" shall mean any work of authorship, including the original version of the Work and any modifications or additions to that Work or Derivative Works thereof, that is intentionally submitted to Licensor for inclusion in the Work by the copyright owner or by an individual or Legal Entity authorized to submit on behalf of the copyright owner. For the purposes of this definition, "**submitted**" means any form of electronic, verbal, or written communication sent to the Licensor or its representatives, including but not limited to communication on electronic mailing lists, source code control systems, and issue tracking systems that are managed by, or on behalf of, the Licensor for the purpose of discussing and improving the Work, but excluding communication that is conspicuously marked or otherwise designated in writing by the copyright owner as "**Not a Contribution.**"

"**Contributor**" shall mean Licensor and any individual or Legal Entity on behalf of whom a Contribution has been received by Licensor and subsequently incorporated within the Work.

2. Grant of Copyright License. Subject to the terms and conditions of this License, each Contributor

hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable copyright license to reproduce, prepare Derivative Works of, publicly display, publicly perform, sublicense, and distribute the Work and such Derivative Works in Source or Object form.

3. Grant of Patent License. Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable (except as stated in this section) patent license to make, have made, use, offer to sell, sell, import, and otherwise transfer the Work, where such license applies only to those patent claims licensable by such Contributor that are necessarily infringed by their Contribution(s) alone or by combination of their Contribution(s) with the Work to which such Contribution(s) was submitted. If You institute patent litigation against any entity (including a cross-claim or counterclaim in a lawsuit) alleging that the Work or a Contribution incorporated within the Work constitutes direct or contributory patent infringement, then any patent licenses granted to You under this License for that Work shall terminate as of the date such litigation is filed.

4. Redistribution. You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

1. You must give any other recipients of the Work or Derivative Works a copy of this License; and
2. You must cause any modified files to carry prominent notices stating that You changed the files; and
3. You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
4. If the Work includes a "**NOTICE**" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.

   You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

5. Submission of Contributions. Unless You explicitly state otherwise, any Contribution intentionally submitted for inclusion in the Work by You to the Licensor shall be under the terms and conditions of this License, without any additional terms or conditions. Notwithstanding the above, nothing herein shall supersede or modify the terms of any separate license agreement you may have executed with Licensor regarding such Contributions.

6. Trademarks. This License does not grant permission to use the trade names, trademarks, service marks, or product names of the Licensor, except as required for reasonable and customary use in

describing the origin of the Work and reproducing the content of the NOTICE file.

7. Disclaimer of Warranty. Unless required by applicable law or agreed to in writing, Licensor provides the Work (and each Contributor provides its Contributions) on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied, including, without limitation, any warranties or conditions of TITLE, NON-INFRINGEMENT, MERCHANTABILITY, or FITNESS FOR A PARTICULAR PURPOSE. You are solely responsible for determining the appropriateness of using or redistributing the Work and assume any risks associated with Your exercise of permissions under this License.

8. Limitation of Liability. In no event and under no legal theory, whether in tort (including negligence), contract, or otherwise, unless required by applicable law (such as deliberate and grossly negligent acts) or agreed to in writing, shall any Contributor be liable to You for damages, including any direct, indirect, special, incidental, or consequential damages of any character arising as a result of this License or out of the use or inability to use the Work (including but not limited to damages for loss of goodwill, work stoppage, computer failure or malfunction, or any and all other commercial damages or losses), even if such Contributor has been advised of the possibility of such damages.

9. Accepting Warranty or Additional Liability. While redistributing the Work or Derivative Works thereof, You may choose to offer, and charge a fee for, acceptance of support, warranty, indemnity, or other liability obligations and/or rights consistent with this License. However, in accepting such obligations, You may act only on Your own behalf and on Your sole responsibility, not on behalf of any other Contributor, and only if You agree to indemnify, defend, and hold each Contributor harmless for any liability incurred by, or claims asserted against, such Contributor by reason of your accepting any such warranty or additional liability.

END OF TERMS AND CONDITIONS

Dado que el desarrollo esta pensado en la estructura del Framework DJANGO 2, en el cual el usuario solo tiene acceso a los HTML de las vistas, se ha insertado en el archivo base.html del proyecto, el fragmento de comentario donde se especifica la declaración de la licencia, el tipo de licenciamiento seleccionado y la relación pertinente a la maestría en Ciberseguridad y Ciberdefensa de la Escuela Superior de Guerra, adicionalmente se ubica un archivo de nombre LICENSE.txt en la raíz del código fuente con los detalles y términos de la licencia APACHE 2.0.

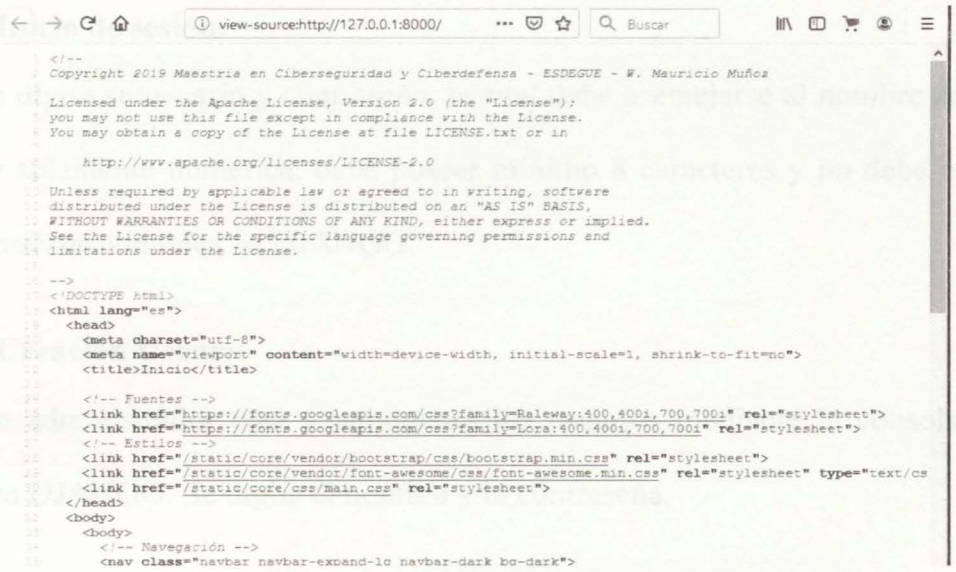


Ilustración 51 Declaración de la Licencia APACHE 2.0

1. MANUAL DE USUARIO CRAWLERFB

CRAWLERFB Es un aplicativo WEB, que recolecta la información pública de un perfil de la red social Facebook, esta recolección se realiza mediante el uso de la técnica de CRAWLER implementada con la librería Selenium. CRAWLERFB fue desarrollado utilizando el Framework de desarrollo Django en su versión 2. Entre las ventajas de Django están el desarrollo rápido y limpio, además de su ya establecida interfaz o panel de administración, el cual ya trae implementado un sistema de administración de usuarios con todas las validaciones necesarias en seguridad e inicios de sesión.

1.1. Creación de usuario e Inicio de sesión

Dado el nivel de seguridad que requiere la plataforma, por el manejo de información de carácter reservada, solo el usuario administrador y de tipo "STAFF" puede crear, validar y dar de alta en el sistema de información.

1.1.1. Inicio de sesión:

El usuario digita su usuario y contraseña, la cual debe asemejarse al nombre de usuario, no puede ser solamente numérica, debe poseer mínimo 8 caracteres y no debe estar en la lista de contraseñas comunes de DJANGO.

1.1.2. Creación Usuario

El usuario administrador crea y administra los usuarios mediante la consola que por defecto aporta DJANGO. Se digita el nombre y la contraseña.

Ilustración 1 Pantalla Inicio de sesión

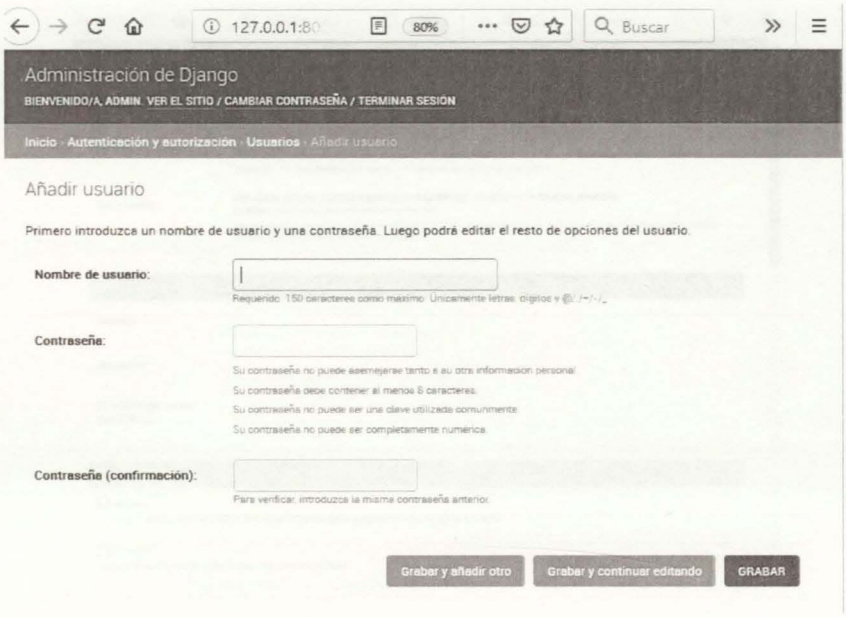


Ilustración 2 Pantalla Creación usuario Django

1.1.3. Usuario Es Staff

Logueado en el sistema como usuario administrador, y ubicado en panel de administración de Django, se procede a listar todos los usuarios del sistema, en la lista se identifican los usuarios y si es o no "STAFF", solo usuarios que hacen parte del grupo "STAFF" pueden realizar las acciones en el sistema. Para migrar los usuarios a el nivel de "STAFF" es necesario completar el formulario desplegado bajo el nombre de cada usuario donde debe completar los campos de nombre, apellidos y correo electrónico, además de seleccionar la opción "STAFF" en la sección de permisos.

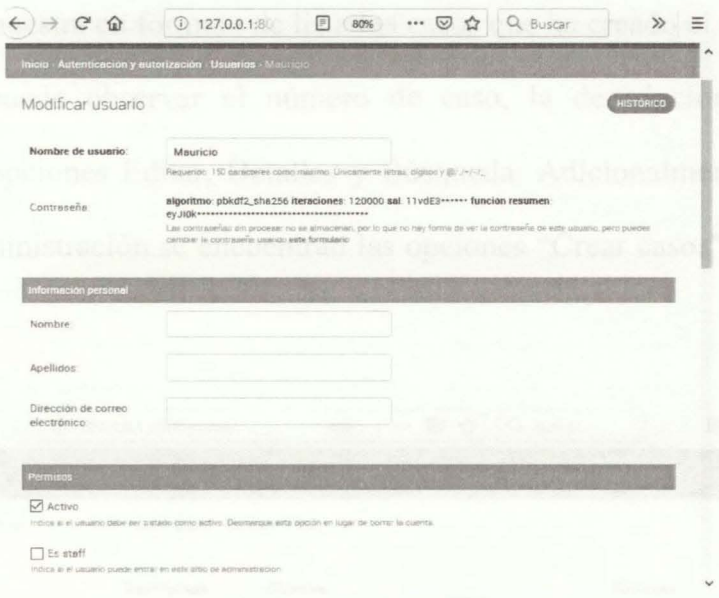


Ilustración 3 pantalla modificación usuario django

1.2. Inicio del sistema CRAWLERFB

Después de estar logueado y con los permisos necesarios, el sistema presenta al usuario la pantalla de inicio del sistema del sistema, en esta pantalla se observa la opción de casos y la opción de usuario donde se muestra el usuario logueado o la opción acceder cuando la sesión ha terminado.

Ilustración 4 Pantalla Inicio CRAWLERFB

1.3. Casos

1.3.1. Lista de casos

En este panel se muestra en formato de lista los casos que ha creado el usuario registrado en el sistema, se puede observar el número de caso, la descripción, el objetivo de investigación y las opciones Editar, Detalles y Búsqueda. Adicionalmente en la barra de navegación o de administración se encuentran las opciones "Crear casos", "Listar casos" y "exportar casos"

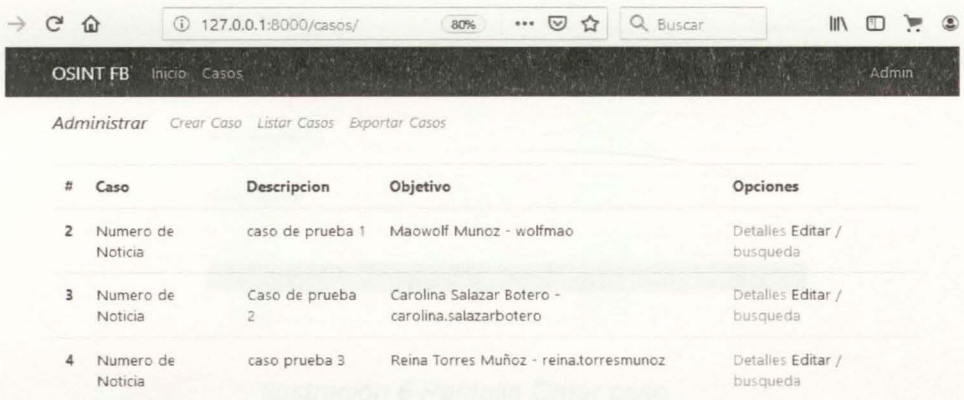


Ilustración 5 Pantalla listado de casos

1.3.2. Crear Caso

En esta opción se despliega el formulario para la creación del caso e investigación, se debe registrar el Numero de noticia criminal, propio de cada investigación de la fiscalía, una descripción del caso que le permita al investigador recordar de que se trata, insertar el señuelo que es usuario de la red de Facebook que se utilizara par alanzar la exploración o búsqueda, la contraseña de Facebook del señuelo, los datos del objetivo que son su ID en la red social Facebook y el nombre que posee el objetivo. El ID el objetivo deben ser tomados de manera exacta como registran en Facebook y se encuentra al visitar la página principal del objetivo.

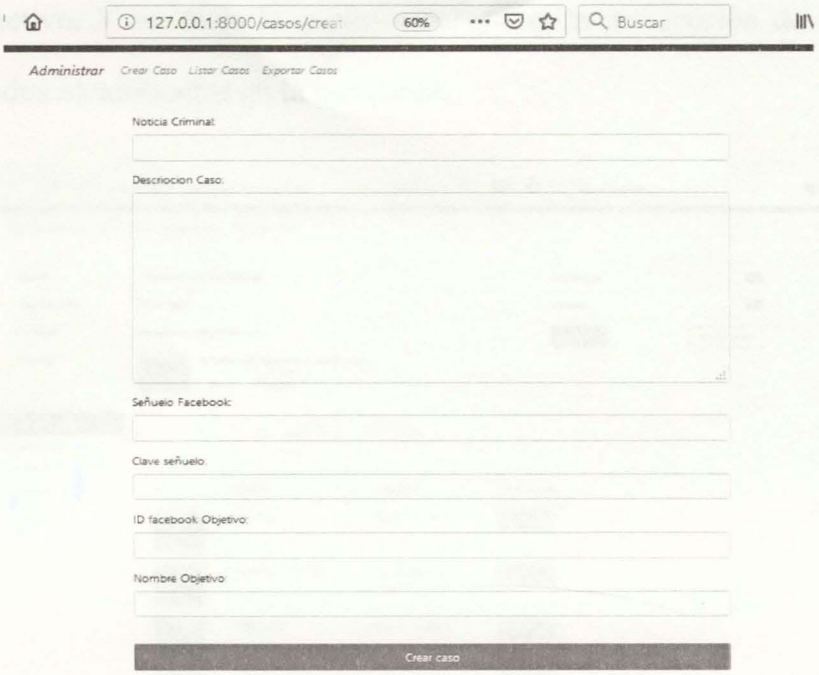


Ilustración 6 Pantalla Crear caso

Ilustración 7 Identificación de ID y nombre de objetivo

1.3.3. Búsqueda

En esta vista se ven los detalles e información recolectada de cada nodo vinculado a la investigación, se inicia la búsqueda que es la recolección de nodos y relaciones de los

nodos con el objetivo. Y se inicia la explotación que es la recolección de la información pública de los nodos identificados en la búsqueda.

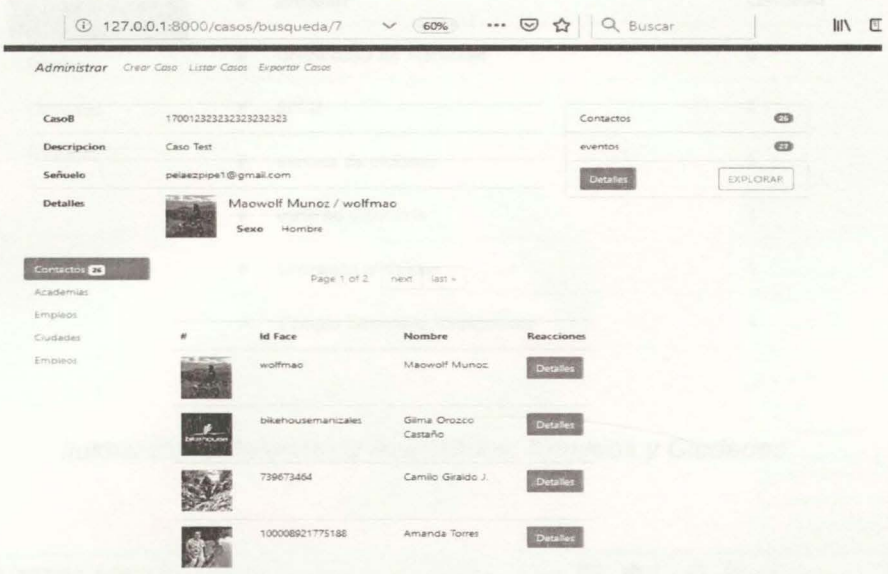


Ilustración 8 Pantalla de Búsqueda

En esta opción existen tres nivele de búsqueda representado en os botones "Datos", "Buscar" y "explotar". En la opción datos el sistema busca los datos públicos del objetivo de la investigación, en la opción Buscar el sistema recorre cada publicación del objetivo identificando los nodos amigos y sus relaciones. El nivel Explotar recorre cada uno de los nodos identificados y busca en estos la información pública de ubicación, empleo y academia.

1.3.4. Pantalla Lista de casos a exportar.

En esta sección se encuentra la lista de casos que cada usuario ha creado, desde esta sección se puede exportar los datos de los nodos recolectados en uno o en varios casos.

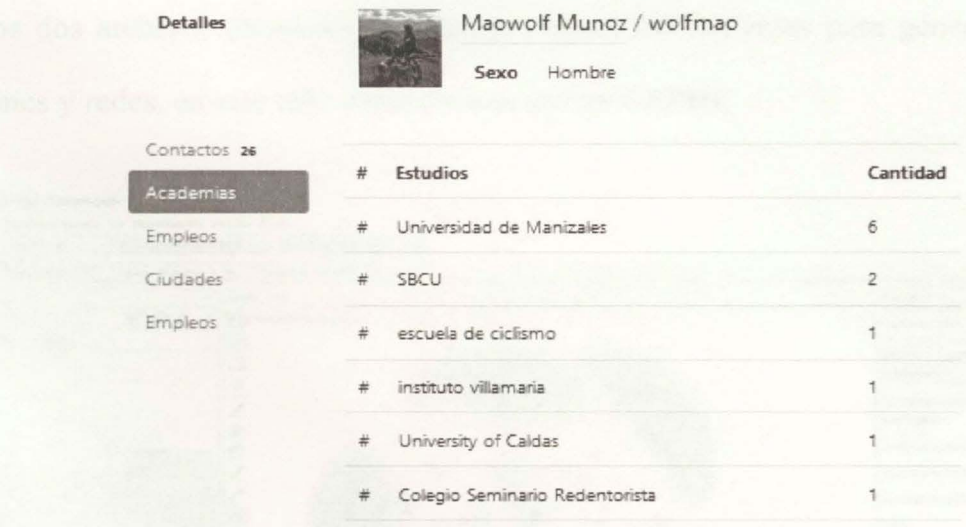


Ilustración 9 Detalles de Academias, Empleos y Ciudades

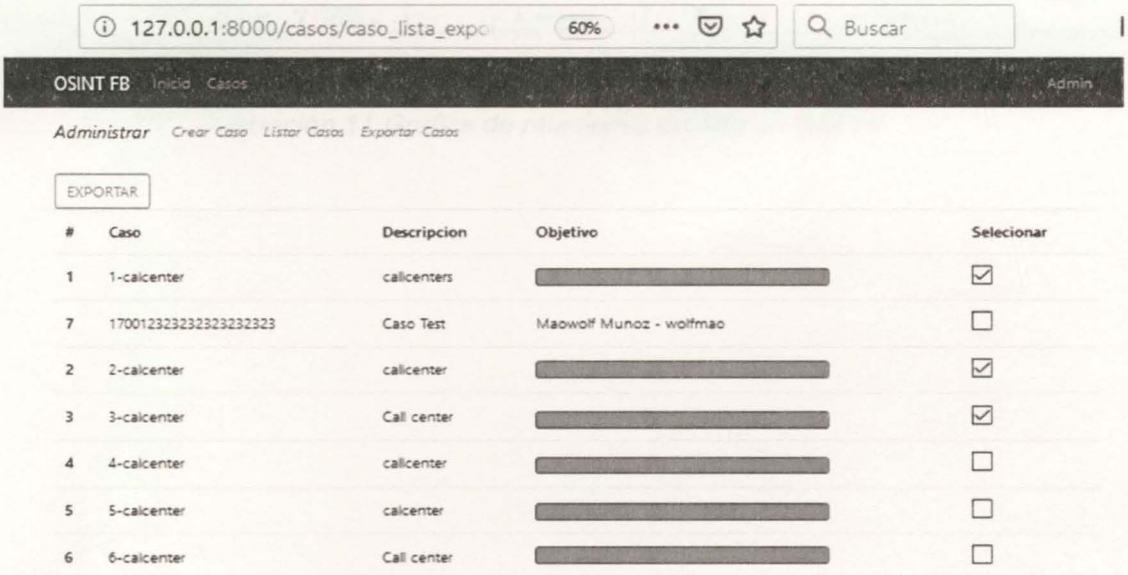


Ilustración 10 Pantalla Exportar casos

Al seleccionar la opción exportar se crean dos archivos de tipo CSV, los cuales contienen los nodos identificados y las relaciones entre los diferentes nodos y el objetivo.

Con los dos archivos obtenidos se pueden utilizar herramientas para generar gráficos de relaciones y redes, en este caso específico se utiliza GEPHI.

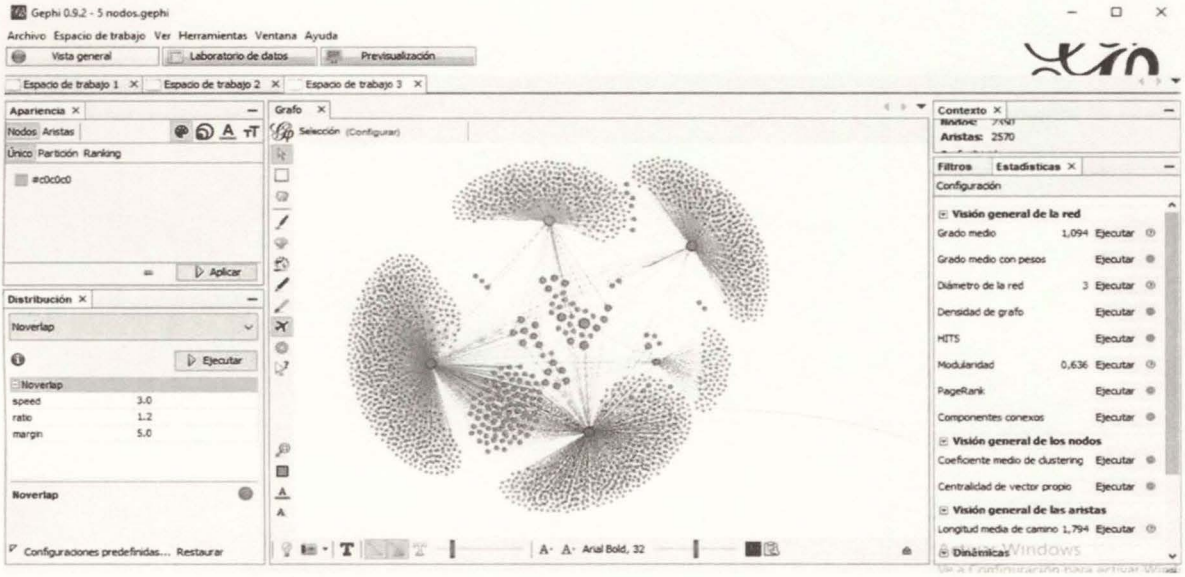


Ilustración 11 Grafica de relaciones creada en GEPHI