Universidad de Alcalá
Escuela Politécnica Superior

GRADO EN INGENIERÍA DE COMPUTADORES

**Trabajo Fin de Grado**

# Herramientas y Técnicas OSINT para la Extracción y el Análisis de la Información Procedente de Fuentes Abiertas



**Autor:** Adrián Cabello Gallego

**Tutor:** Manuel Sánchez Rubio

2021

UNIVERSIDAD DE ALCALÁ

Escuela Politécnica Superior

**GRADO EN INGENIERÍA DE COMPUTADORES**

Trabajo Fin de Grado

Herramientas y Técnicas OSINT para la Extracción y el Análisis de la Información Procedente de Fuentes Abiertas

**Autor:** Adrián Cabello Gallego

**Tutor:** Manuel Sánchez Rubio

**Tribunal:**

**Presidente: José Javier Martínez Herraiz**

**Vocal 1º: Carmen Pagés Arévalo**

**Vocal 2º: Manuel Sánchez Rubio**

**Calificación: _____**

Alcalá de Henares a, 12 de noviembre del 2021

# Agradecimientos

Quiero dar las gracias a mi familia por apoyarme durante todos estos años y ayudarme siempre en todo lo que han podido.

También quiero agradecer a mis compañeros y amigos que me han acompañado todos estos años y que siempre han estado ahí para todo.

Por último, quiero agradecer especialmente a mi tutor Manuel Sánchez Rubio, tanto por sus clases como por su ayuda siempre que lo he necesitado y por hacer este trabajo posible.

# Contenido

# Índice de Figuras

# 1. Sumario

Debido a todos los avances y cambios que ha habido en los últimos años, hoy en día, en internet existe gran cantidad de información sobre cualquier cosa, personas, instituciones, etc. Todo esto debido a la información que recopilan las páginas web, redes sociales, empresas, etc. y que está expuesta para que cualquier persona que sepa encontrarla la obtenga. Todo el proceso de investigación que se aprovecha de esto se denomina OSINT (Open Source Intelligence) y consiste en la recopilación de toda la información disponible en internet sobre una persona o institución para luego dotarla de inteligencia y darle validez. El objetivo de hacer el trabajo sobre OSINT es conseguir arrojar algo de luz sobre este tema, que mucha gente desconoce y explicar las técnicas y herramientas más utilizadas.

# 1. Sumary

Due to the advances and changes that have developed in the last few years, nowadays there is a great amount of information on the internet about any topic, people, institution… This is in part due to the information collected by Web pages, social networks, companies… All of which is exposed for anyone to see, provided you know how to access it. The process of getting and transforming data about a person or institution into useful intelligence is called OSINT (Open Source Intelligence). The aim of this document is to shine some light around this relatively unknown subject, and to explain the techniques and tools most widely used in this scope.

# 2. Palabras clave

Investigación, información, OSINT, inteligencia.

# 3. Resumen

OSINT (Open Source Intelligence) traducido al español como Inteligencia en fuentes abiertas es un importante ámbito dentro del mundo de la ciberseguridad y que cada vez se está extendiendo más y cobrando más importancia, debido al aumento exponencial de la cantidad de información de libre acceso que se almacena en internet.

OSINT consiste en un proceso llevado a cabo en varias etapas. Primero se recopila toda la información posible del objetivo obtenida de fuentes abiertas y posteriormente se dota de inteligencia a toda esa información.

Una de las principales características de este proceso de investigación es que se hace uso de información que está disponible en fuentes abiertas y que es de uso público, de modo que, no se hace uso de prácticas de hacking ilegales, manteniéndonos así siempre dentro del marco legal.

Existen numerosas fuentes disponibles donde podemos consultar y extraer información, pero sin duda la mayor fuente de información son los motores de búsqueda como Google o metabuscadores, que están continuamente indexando toda la información con los crawlers. Otra importante fuente de información son las redes sociales, estas además son peligrosas por el exceso de confianza o el desconocimiento de la gente, ya que al subir fotos o enviar twitts si se tiene activada la ubicación, un atacante podría averiguar donde ha estado su objetivo o incluso donde vive.

Vamos a analizar y estudiar las herramientas más utilizadas para obtener toda esta información en cada una de estas fuentes y aprender a sacar el máximo partido a la investigación de un objetivo.

# 4. Planteamiento

El TFG se va a dividir en varias secciones donde iremos explicando y desglosando cada etapa de una investigación OSINT y estudiaremos el proceso de obtención de información en fuentes abiertas desde diversas fuentes y herramientas de libre acceso.

Para ponernos en contexto, vamos a analizar el inicio de esta clase de investigaciones, que hasta hace relativamente poco no eran de gran relevancia ya que la información que había en internet a cerca de personas o instituciones era mucho más escasa. Hoy en día con el aumento exponencial del uso de internet por parte de las personas y con el auge de las redes sociales, las investigaciones OSINT están adquiriendo una gran importancia y sigue en aumento, ya que existe una enorme cantidad de información almacenada en internet sobre prácticamente todo el mundo. Y la mayoría de esta información es publica y abierta a cualquier persona.

Una vez en contexto se explicarán las distintas fuentes abiertas de las que vamos a hacer uso para realizar una investigación OSINT, ya que, como hemos comentado, las investigaciones OSINT se basan en la existencia de estas fuentes abiertas donde recopilar información y en la existencia de herramientas que se encarguen de automatizar el proceso y buscar en diferentes fuentes para luego mostrarnos el resultado. También estudiaremos varias de las herramientas más usadas de libre acceso para facilitar esta recopilación de información y frameworks de OSINT que nos ayudaran a estructurar la investigación y toda la

información obtenida ofreciéndonos un amplio abanico de herramientas y motores de búsqueda que podemos aplicar.

# 5. Internet y fuentes abiertas

Antes de empezar a analizar el proceso de la investigación OSINT y sus fases, vamos a hablar sobre internet y las distintas fuentes abiertas, y de su crecimiento a lo largo de los últimos años.

Existen multitud de fuentes a las que podemos recurrir para obtener información para nuestra investigación, pero sin duda la mayor y más importante fuente de información es internet. El crecimiento masivo que ha tenido internet en los últimos años ha hecho que sea la fuente de información por excelencia y ha contribuido para el aumento de la importancia de OSINT dentro del campo de la ciberseguridad.

Si hacemos una comparación de los usuarios de internet de hace unos años a la actualidad, podemos observar el gran crecimiento que estamos comentando, en 2010 según la Internet World Stats (IWS) había una población mundial de 6800 millones personas y un total de 1966 millones de usuarios en internet [1]. En cambio, este último año 2021 se estima que hay una población de 7830 millones de personas y un total de 4660 millones de usuarios en internet [2]. Si comparamos estos datos obtenemos que en 2010 un 28,9% de la población mundial usaba internet y que este año lo usa un 59,51% de la población.

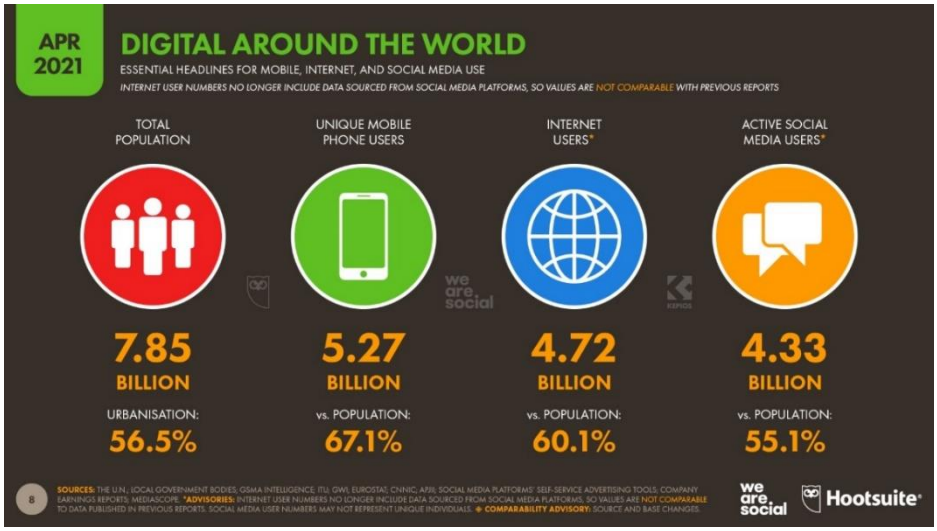



*Figura 1: Uso de medios digitales en el mundo*

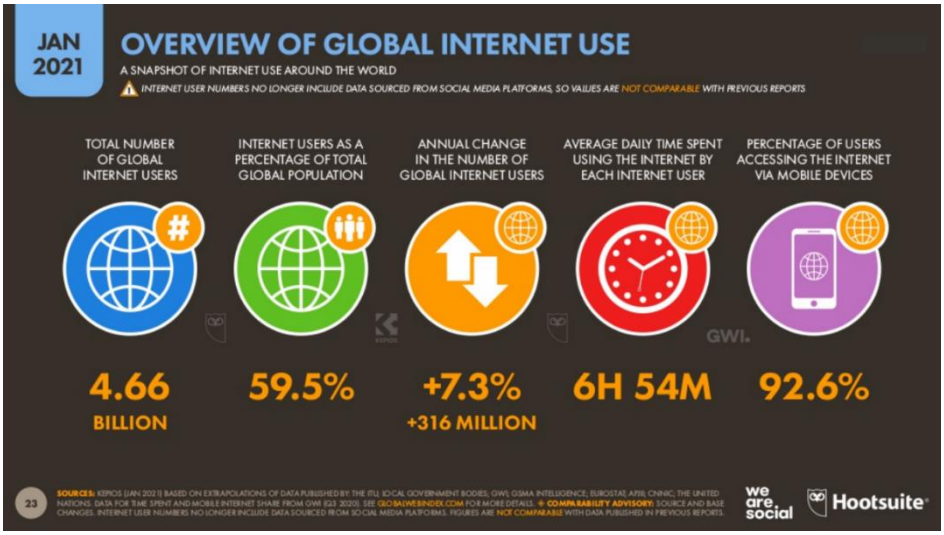



*Figura 2: Uso de internet en el mundo*

Este crecimiento del número de usuarios, sumado al aumento del tiempo que pasan navegando por internet, que tal y como podemos observar en la figura dos es una media de casi 7 horas, contribuye a que se vaya almacenando una gran cantidad de información de los usuarios en las distintas páginas web, a medida que los usuarios las visitan. Esto es debido a la huella digital que dejamos al navegar por internet cuando las páginas capturan información sobre nosotros y nuestros gustos.

# 6. OSINT (Open Source Intelligence)

## 6.1. Concepto de OSINT

OSINT consiste en un proceso de investigación, donde el objetivo es extraer toda la información a cerca de algo haciendo uso de las diferentes fuentes abiertas, y archivar toda esa información para posteriormente dotarla de inteligencia.

Este es un proceso que a pesar de estar en auge ahora con el crecimiento de internet, lleva realizándose desde cientos de años atrás. Usando otras fuentes como periódicos u otros medios de comunicación disponibles.

Toda esta información podemos recopilarla de numerosas fuentes:

- Internet: foros, blogs, redes sociales, páginas web.
- Medios de comunicación: Entrevistas, noticias, periódicos, radio, etc.
- Artículos: artículos de trabajo, artículos académicos, publicaciones académicas, conferencias.
- Comunicados: boletines oficiales del estado, comunicados del gobierno, agencias internacionales, etc.

Una vez terminada la fase de extracción de la información llega la parte realmente importante, dar inteligencia a toda esa información. De esta forma conseguiremos diferenciar la información válida que realmente pertenezca a la persona o institución que estamos analizando, de la información errónea que podamos obtener y que no nos sirva para nada [3][4].

# 6.2. Importancia y utilidad de OSINT

La importancia que ha adquirido OSINT en los últimos años es debida a su inmensa utilidad en cualquier ámbito.

Un ejemplo de la gran utilidad que ha adquirido es por ejemplo a nivel empresarial. Cuando una persona manda un CV a una empresa para conseguir trabajo, el OSINT se ha convertido en un gran aliado para estas empresas, que pueden realizar una investigación para estudiar y averiguar todo lo posible de esa persona y ver cuál es su imagen pública para saber si es una persona adecuada o no.

Otro ejemplo es en el ámbito de la ciberinteligencia y la delincuencia. Muchas veces se averigua un nombre, un username, una foto o cualquier mínimo de información sobre un delincuente, ya sea por un despiste suyo o por desconocimiento del potencial de OSINT, que puede servir como punto de partida para una investigación que arroje luz sobre su identidad o sobre su paradero y ayudar a encontrarlo, cosa que sin OSINT habría sido algo imposible de hacer.

*Figura 3: OSINT*

# 6.3. Fases de OSINT

OSINT se divide en dos partes, la parte de recopilación de toda la información disponible y posteriormente la parte donde se da inteligencia a toda esa información.

Estas a su vez están formadas por varias fases, a esto se le denomina el ciclo de la inteligencia.


*Figura 4: Fases OSINT*

Podríamos agrupar todas las fases de este ciclo en las dos partes principales que hemos mencionado. El establecimiento de los requisitos, búsqueda de las fuentes de información y adquisición formarían parte de la recopilación de la información y a su vez el procesamiento, análisis e inteligencia formarían parte de la dotación de inteligencia a toda la información recopilada.

# 7. Herramientas

A continuación, vamos a enumerar y explicar varias herramientas que podemos utilizar para realizar la investigación. Vamos a ver tanto herramientas propiamente desarrolladas para extraer información y algo más avanzadas (como podría ser iKy proyect) como herramientas de lo más comunes que usa cualquier persona en su día a día (como podría ser Google).

Para este trabajo solo se van a mostrar herramientas de software libre gratuitas a las que cualquiera pueda tener acceso y probar por su cuenta, aunque como en todo, también existen herramientas de pago para la extracción de información.

## 7.1. Motores de búsqueda

### 7.1.1. Introducción

Como no podía ser de otra manera, empezaremos explicando la utilidad de la mayor y principal fuente de información, que son los motores de búsqueda. Esto es debido a que cualquier persona usa los motores de búsqueda, ya sea Google, Firefox, Bing o cualquiera de los existentes.

Sea cual sea el motivo, por búsquedas personales, por trabajo, todos los usuarios de internet pasan una gran cantidad de horas navegando por estos buscadores y evidentemente es con la herramienta que más familiarizados están.

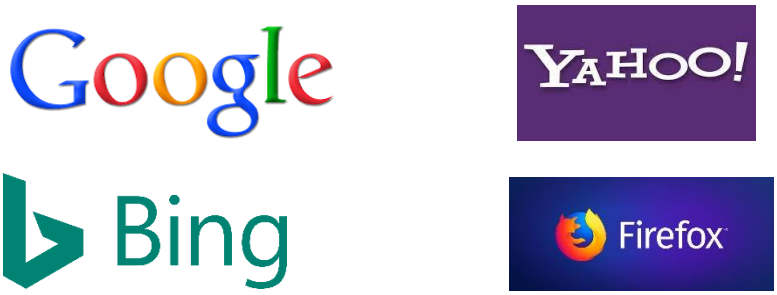



*Figura 5: Buscadores*

## 7.1.2. Concepto

Los motores de búsqueda son un software que se diseñaron para organizar y gestionar las páginas web y la información en internet. Ya que debido al gran incremento del número de páginas se necesitaba algo que nos permitiera, mediante palabras clave, navegar entre estas páginas y encontrar información precisa.

El funcionamiento de los motores de búsqueda es estar continuamente explorando internet e indexando todas las páginas nuevas que encuentran. Es tal la capacidad de indexación que tienen estos buscadores, que pueden tardar 10 minutos en indexar una página nueva.

Los encargados de realizar esta indexación son los crawlers, cualquier motor de búsqueda tiene crawlers indexando continuamente y recorriendo las páginas web. El funcionamiento de estos crawlers es empezar primero en unas páginas de referencia que tienen establecidas. A partir de esas páginas de referencia van entrando a todos los enlaces que encuentren y saltando a otras páginas, las cuales van registrando en la lista de páginas web, de esta manera van saltando por todas las páginas web en busca de nuevos enlaces que indexar [5].

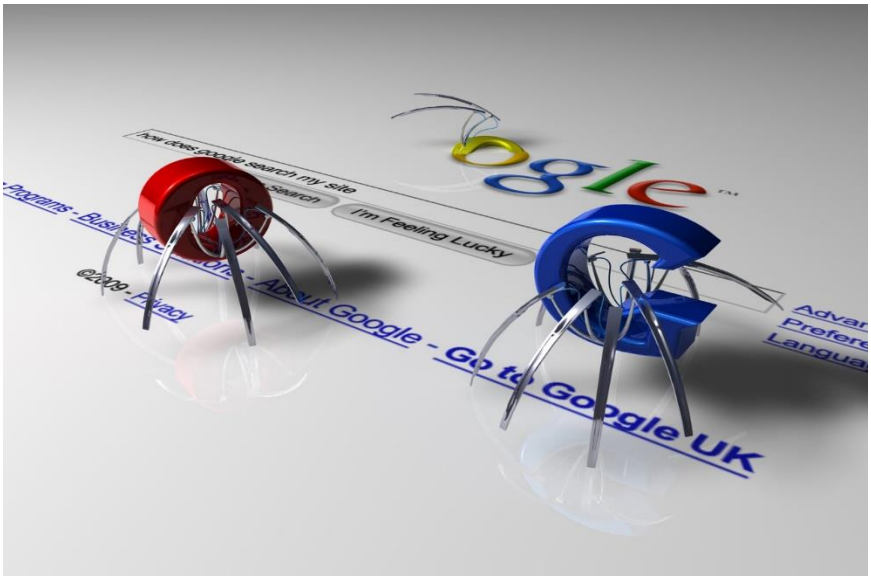



*Figura 6: Crawlers*

## 7.1.3. Extracción de información

Estos motores de búsqueda, tal y como hemos comentado, permiten obtener información haciendo búsquedas sencillas como palabras clave, páginas web, etc. Para ello podemos usar cualquier motor de búsqueda, pero existen operadores que nos permiten ir más allá y hacer búsquedas más complejas y aisladas.

Google y Bing poseen estos operadores conocidos como dorks, que como hemos comentado nos permiten hacer búsquedas complejas y nos permiten hacer búsquedas muy concretas. Búsquedas como palabras clave específicas en el contenido de texto de una página web o en su título, tipos de fichero, url y muchísimo más. En este caso nos vamos a centrar en los dorks de Google y a enumerar los más interesantes y más usados. También se darán ejemplos de búsqueda con combinaciones de estos dorks, ya que la mayoría pueden combinarse unos con otros para realizar búsquedas aún más específicas.

## 7.1.4. Google Dorks

Los Google dorks son diversos filtros que podemos aplicar en Google usando operadores, de manera que nos permitan buscar información muy específica dependiendo del operador utilizado, como por ejemplo tipos de fichero, palabras clave o cadenas de texto tanto en títulos como en las propias páginas o en las url, webcams, ficheros de contraseñas, etc.

Estos operadores los podemos clasificar en distintos tipos, operadores simples y operadores complejos o avanzados,

Dentro de los operadores simples podemos encontrar los operadores que realizan operaciones lógicas y también aquellos que nos sirven para crear

expresiones regulares, concatenar o excluir palabras. Los más utilizados son los siguientes:

- OR / '|': actúa como una operación or, se usa para buscar la palabra de la izquierda o la de la derecha, no las dos.
- NOT / '-': actúa como una operación not, se usa para excluir la siguiente palabra de la búsqueda.
- AND / ' ': actúa como una operación and, se usa para buscar ambas palabras, la de su izquierda y la de su derecha.
- " ": se usa para buscar una palabra o cadena de palabras por coincidencia exacta.
- '..': se usa para buscar dentro de un rango de números (25...100)
- '+': actúa inverso al operador '-', ya que busca resultados que incluyan la palabra de la derecha.
- '*': se usa para reemplazar cualquier palabra.
- '.': tiene la misma función que el '*', con la diferencia de que este puede reemplazar varias palabras.

Dentro de los operadores avanzados podemos encontrar todos ellos que nos permiten buscar en cualquier característica de una página (título, contenido, etc.), que permiten buscar por formato de archivo (pdf, xlsx), que permiten buscar por sitio web, etc. Los más utilizados son los siguientes.

- site: busca resultados dentro de un sitio especifico (site: uah.es)
- related: busca sitios relacionados (related: uah.es)
- allintitle: busca resultados cuyo título contenga la frase indicada (Allintitle: uah deportes)
- intitle: busca resultados cuyo título contenga solo el término indicado (intitle: ingeniería)

- inblogtittle: busca resultados de blogs cuyo título contenga la frase indicada (inblogtittle: programación en java)
- inposttittle: busca resultados de blogs cuyo título contenga el término indicado (inposttittle: universidad)
- cache: busca la versión del sitio web en cache (cache: uah.es)
- inurl: busca resultados cuya url contenga la palabra indicada (inurl: matrícula)
- intext: busca resultados cuyo contenido de texto de la página contenga lo indicado (intext: casa)
- filetype: busca resultados por tipo de archivo (filetype: pdf)
- location: busca artículos basados en una ubicación específica (location: Madrid)

A parte de tener operadores simples y complejos, también podemos hacer uso de estos combinándolos, de forma que podamos llegar a información muy específica, la cual no podríamos encontrar de no ser por estos, aunque no todos se pueden combinar. Los tipos de información buscada con más frecuencia son ficheros de usuarios y contraseñas, cámaras web sin restricciones de acceso, servidores FTP abiertos, ficheros de backup, páginas web con vulnerabilidades de versión y de SQL, ficheros log, ficheros de robots de páginas web, bases de datos accesibles, etc. [6][7][8].

- Ficheros de usuarios y contraseñas
    - **Intitle: "index of" "index of/" password.txt** (Este operador busca servidores con archivos password.txt)**.**

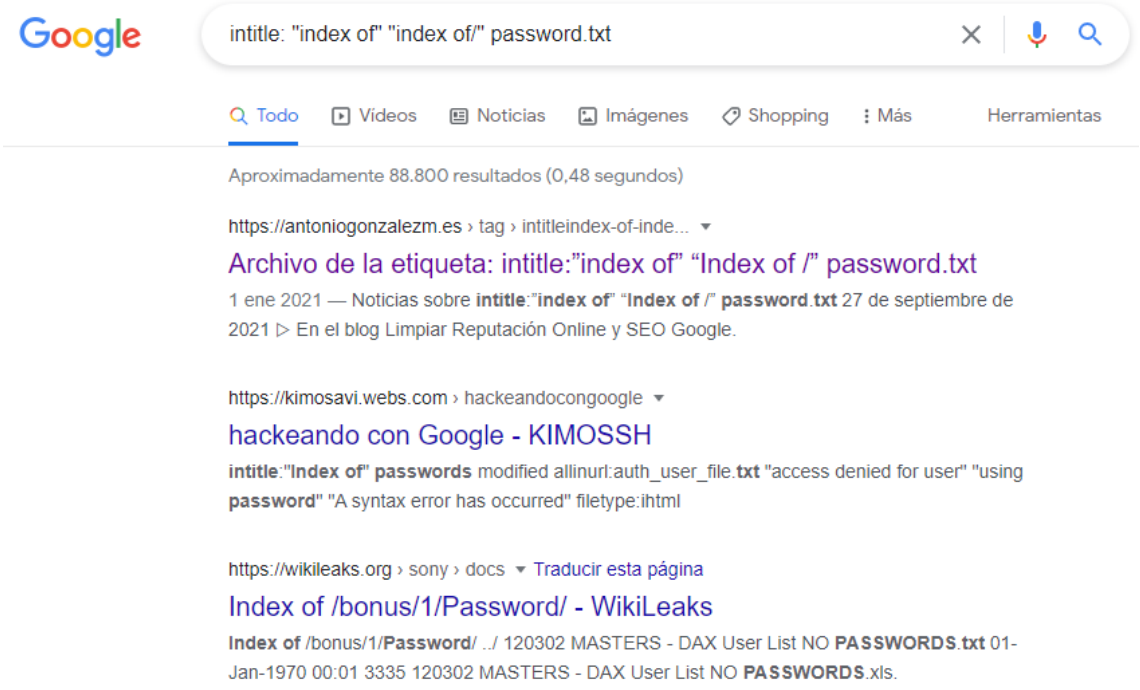



*Figura 7: Extracción de ficheros de contraseñas con Google Dorks*

Podemos observar que la tercera página que nos aparece en la búsqueda contiene un servidor con usuarios y contraseñas, si entramos a la página observamos lo siguiente y si entramos a uno de ellos podemos encontrar nombres de usuario con sus respectivas contraseñas e incluso con emails.

# Index of /bonus/1/Password/

| | | |
|---|---|---|
| ../ | | |
| 120302 MASTERS - DAX User List NO PASSWORDS.txt | 01-Jan-1970 00:01 | 3335 |
| 120302 MASTERS - DAX User List NO PASSWORDS.xls..> | 01-Jan-1970 00:01 | 51636 |
| 17 - Password History_09132007.pdf | 01-Jan-1970 00:01 | 32551 |
| 17 - Password History_09132007.txt | 01-Jan-1970 00:01 | 340 |
| 3N Notification password_userid.doc.pdf | 01-Jan-1970 00:01 | 14311 |
| 3N Notification password_userid.txt | 01-Jan-1970 00:01 | 1518 |
| 50 new user password.txt | 01-Jan-1970 00:01 | 4009 |
| 50 new user password.xls.pdf | 01-Jan-1970 00:01 | 33911 |
| 6 PWC password test.doc.pdf | 01-Jan-1970 00:01 | 50207 |
| 6 PWC password test.txt | 01-Jan-1970 00:01 | 125 |
| 90 Day Admin password.xlsx | 01-Jan-1970 00:01 | 8007 |
| ALL_SSL_Certs_2012.xlsx | 01-Jan-1970 00:01 | 58183 |
| Accounts Passwords.txt | 01-Jan-1970 00:01 | 376 |
| Accounts and Passwords.xlsx | 01-Jan-1970 00:01 | 19968 |
| All NT passwords.txt | 01-Jan-1970 00:01 | 28941 |
| All NT passwords.xls.pdf | 01-Jan-1970 00:01 | 48415 |
| Ariba User ID and Password memo.doc.pdf | 01-Jan-1970 00:01 | 92576 |

*Figura 8: Ficheros de contraseñas*

```
"User Id","Password","Lname","Fname","Phone","Email"
"MALBA","BLF5YJU","Alba","Maira","718-868-5971","Maira_Alba@spe.sony.com"
"MANDERSON","BKM6GWH","Anderson","Mark","310-482-4919","Manderson@sonypictures.com"
"AARVIZU-PERE","MPH9GLF","Arvizu-Perez","Agueda","310-571-3919","Agueda_Perez@spe.sony.com"
"RBAKER","WTW9WPW","Baker","Roy","212-833-6469","Roy_Baker@spe.sony.com"
"EBERGMAN","FYK3FEG","Bergman","Elizabeth","310-244-2610","the_dinner_party_accounting_office@spe.sony.com"
"TBIRKE-HAUEI","EMQ3GER","Birke-Haueisen","Tov",,"Tov_Birke-Haueisen@spe.sony.com"
"ABOBB","PFM1QJQ","Bobb","Andy","310-244-3589","King_of_Queens_Accounting_Office@spe.sony.com"
```

*Figura 9: Contenido de fichero de contraseñas*

- **inurl:/wp-content/uploads/ ext:txt "username" AND "password" | "pwd" | "pw"** (busca archivos de texto en páginas web que usen WordPress como gestor de contenidos y que contengan las palabras clave username y password o pw).

Si buscamos lo indicado anteriormente y entramos en la primera página que encontramos, observamos datos de usuario y contraseñas tanto de un usuario como del host.


*Figura 10: Exztracción de ficheros de contraseñas 2*

```
what is the username and password for http://www.sawchukwealth.com/wp-login.php?redirect_to=/clientt/private-client-section-home-page/

wp-login details

username: admin
pwd: 73osrx$V7

Ftp login details

host: ftp.ord1-1.websitesettings.com
username: terrysaw98xx
pwd: ef6&S6cU!
```

*Figura 11: Contenido de fichero de contraseñas 2*

- Cámaras web
    - **Intitle: "webcamXP 5"** (busca cámaras de modelo webcamXP5 que estén transmitiendo en vivo y que no requieran autenticación).

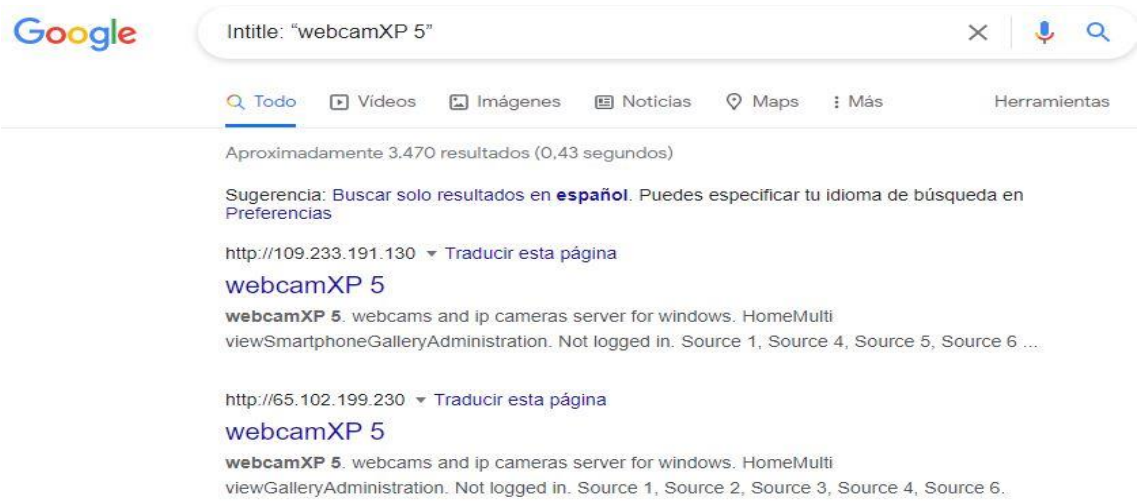



*Figura 12: Conexión a webcam con Google Dorks*


*Figura 13: Imagen webcam*

- Servidores FTP abiertos
    - **Intext: "indexo f" inurl: ftp** (busca servidores FTP abiertos que no requieran autenticación)

Con esta búsqueda podemos entrar en servidores FTP sin autenticación y obtener toda la información que tengan almacenada. Buscando la sentencia indicada obtenemos los siguientes resultados de servidores ftp, si entramos en el primer enlace podemos observar el contenido de dicho servidor.

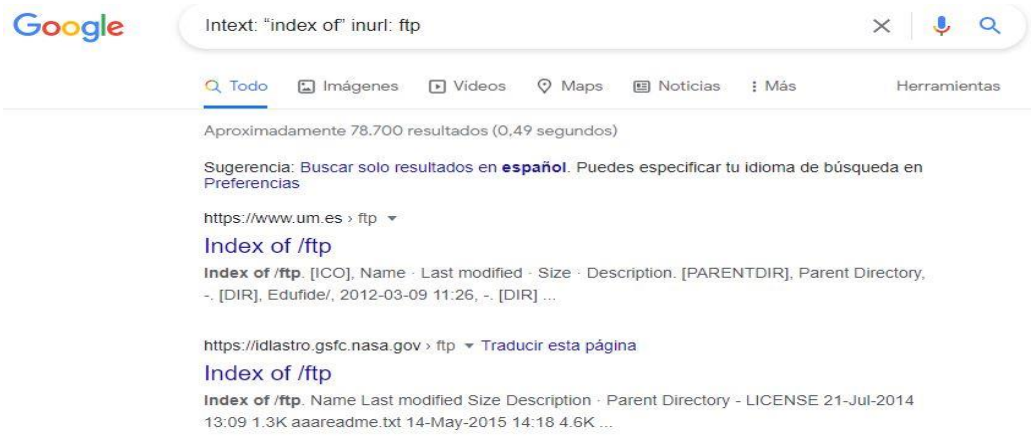



*Figura 14: Búsqueda de servidores FTP con Dorks*

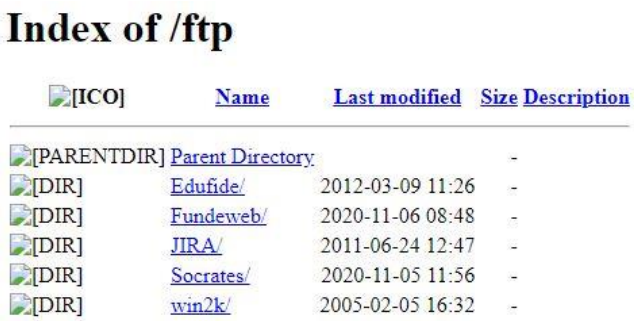



*Figura 15: Servidores FTP*

- Ficheros de backup
    - **Intitle: "index of" "dump.sql"** (archivos de backup o volcados en bases de datos)

Buscamos la sentencia dump ya que es la que se usa para hacer volcados en bases de datos, de forma que podemos encontrar información sensible de bases de datos o archivos de backup que se hayan guardado en una base de datos. Haciendo esta búsqueda podemos encontrar numerosos enlaces a archivos volcados en bases de datos, entrando en el primer enlace que nos aparece encontramos archivos .sql comprimidos a los que podemos acceder y obtener información sensible de estos.

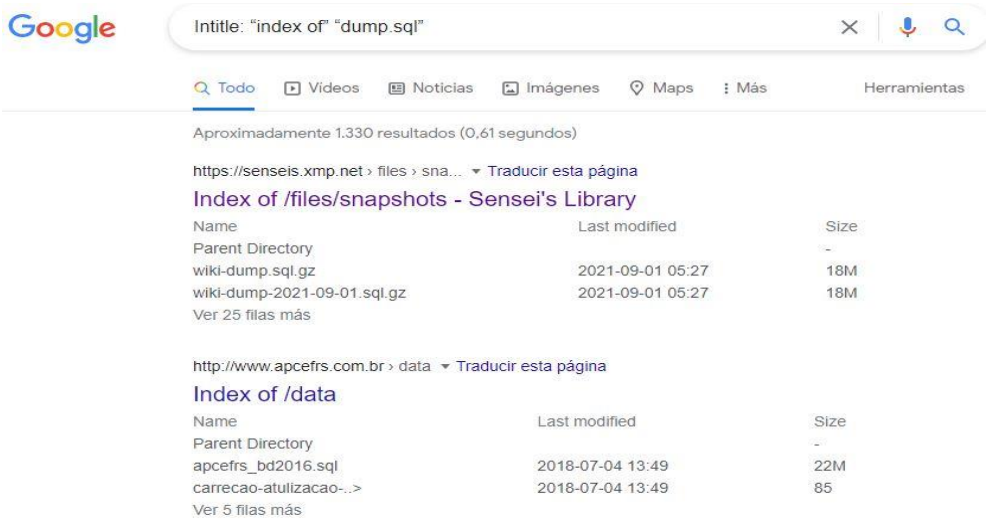



*Figura 16: Búsqueda de servidores SQL con Dorks*


*Figura 17: Servidores SQL*

- Páginas con vulnerabilidades
    - **Inurl: shop.php? id=6** (buscamos páginas con id=6)

    Esta búsqueda nos permite encontrar páginas que, debido a su versión, sean vulnerables a sql (mediante Kali, por ejemplo).

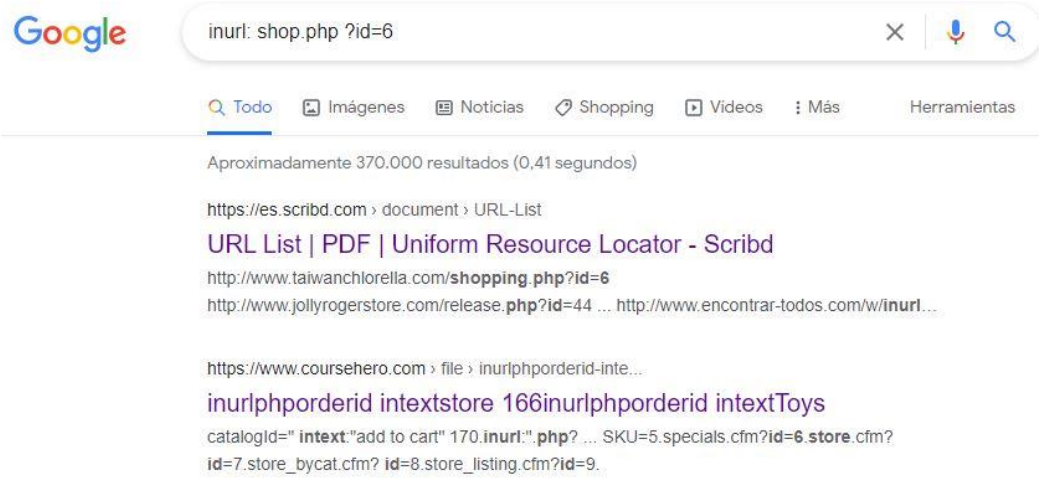



*Figura 18: Búsqueda de páginas vulnerables a SQL*

- Ficheros de log
    - **allintext: password filetype: log after:2020**

    Buscamos ficheros de log que además posean el término password y que estén actualizados (a partir de 2020), si miramos los resultados obtenidos podemos observar que el primer resultado que obtenemos es un .log que tal y como queríamos, contiene usuario y contraseña y en este caso ni siquiera nos hace falta entrar al enlace para verlos.

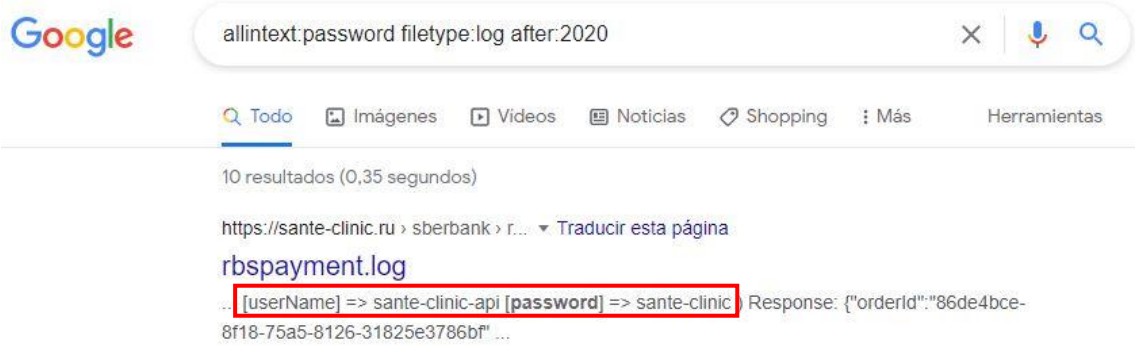



*Figura 19: Búqueda de ficheros de log con contraseñas con Dorks*

- Ficheros de robots
    - Site: uah.es inurl: robots.txt

    Realizando esta búsqueda obtendremos la estructura de ficheros permitidos y no permitidos para que los crawlers de Google puedan obtenerlos. Lo interesante de esta búsqueda no es acceder a la información a priori, sino saber cuál es la información delicada de una determinada página, la cual los atacantes querrán conseguir


*Figura 20: Búsqueda de robots con Dorks*

```
Sitemap: http://www.uah.es/es/sitemap.xml
Sitemap: http://www.uah.es/sitemap.xml

User-agent: *
Disallow: /es/accesibilidad/page/#
Disallow: /es/conoce-la-uah/organizacion-y-gobierno/organos-de-representacion/actas-de-la-junta-de-personal-de-administracion-y-servicios/
Disallow: /es/conoce-la-uah/organizacion-y-gobierno/organos-de-representacion/actas-del-comite-de-empresa/
Disallow: /es/buscador-general/
Disallow: /es/estudios/profesor/Lidia-Ruiz-Llorente/
Disallow: /en/estudios/profesor/Lidia-Ruiz-Llorente/

Crawl-delay: 30
```

*Figura 21: Fichero de robots de la uah*

- Bases de datos
    - "Index of" inurl: phpmyadmin

    Buscamos listados de bases de datos phpmyadmin, como se puede observar en el resultado obtenido, Google nos muestra varios listados. Si entramos al primer enlace, tendremos acceso a varios logins de bases de datos phpmyadmin.

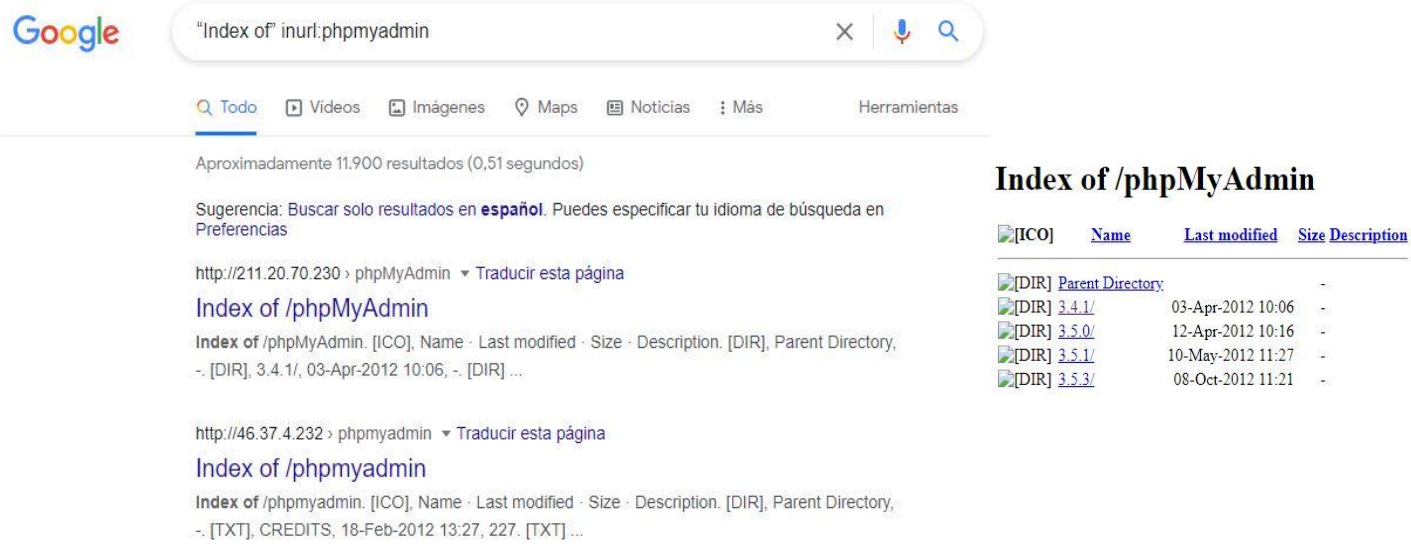



*Figura 22: Búsqueda de bases de datos phpmyadmin con Dorks*

## 7.1.5. Webs de interés

Dentro de la búsqueda y extracción de información a través de los motores de búsqueda me gustaría destacar ciertas páginas que pueden ser de gran utilidad y aportarnos información valiosa y sensible de un objetivo. Estas páginas en concreto pueden aportarnos información sobre el domicilio de una posible persona a la que estemos analizando, podríamos obtener la dirección de este, además de información sobre el propio domicilio.

### 7.1.5.1. Catastro

https://www1.sedecatastro.gob.es/Cartografia/mapa.aspx?buscar=S

Haciendo uso de la página web oficial del catastro en España podemos buscar cualquier dirección o incluso abrir el mapa para buscar manualmente una propiedad. Esta página nos permitirá ver las características de la vivienda, ya sean metros cuadrados totales, distribución de los metros en las diferentes

plantas de la vivienda, instalaciones deportivas (piscina, pistas de tenis, pistas de fútbol, etc.) [9].

Esto además nos permite encontrar irregularidades en la vivienda del objetivo, por ejemplo, podemos ver en el Google maps a través de la vista satélite que tenga una piscina y que en el catastro no aparezca declarada, lo que nos aportaría más información sensible sobre el objetivo.

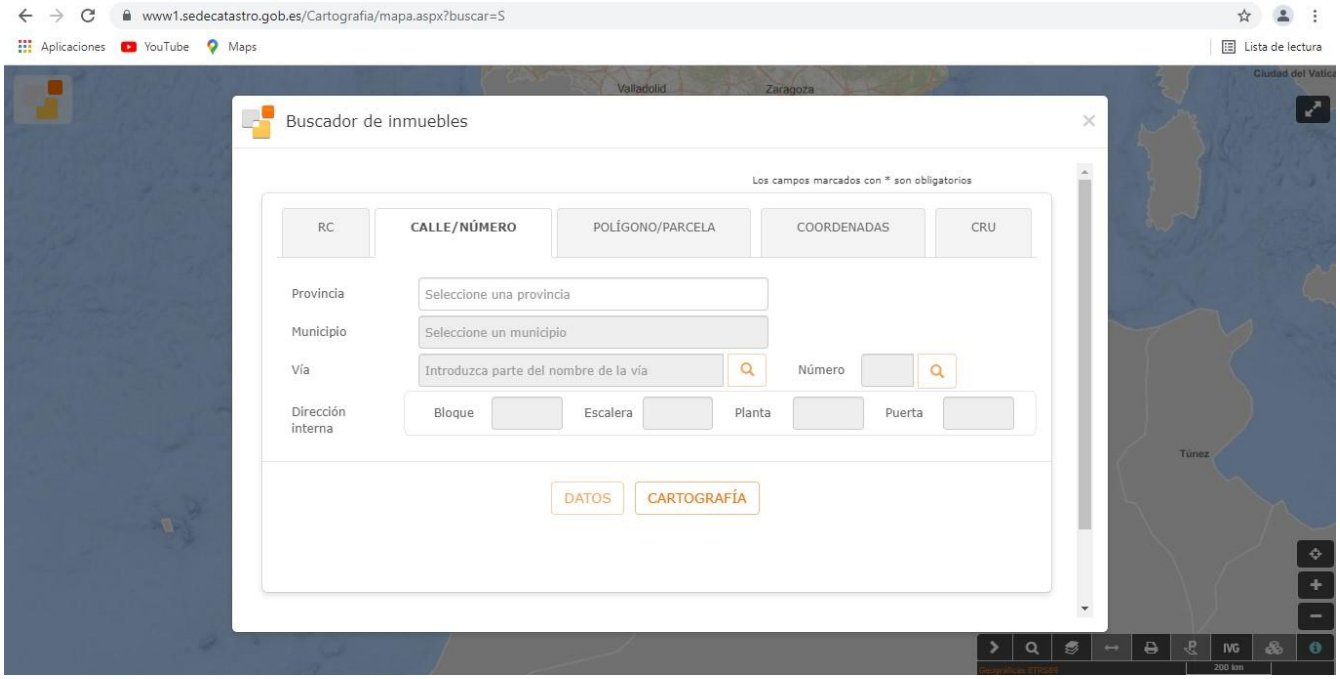



*Figura 23: Buscador de sedecatastro*

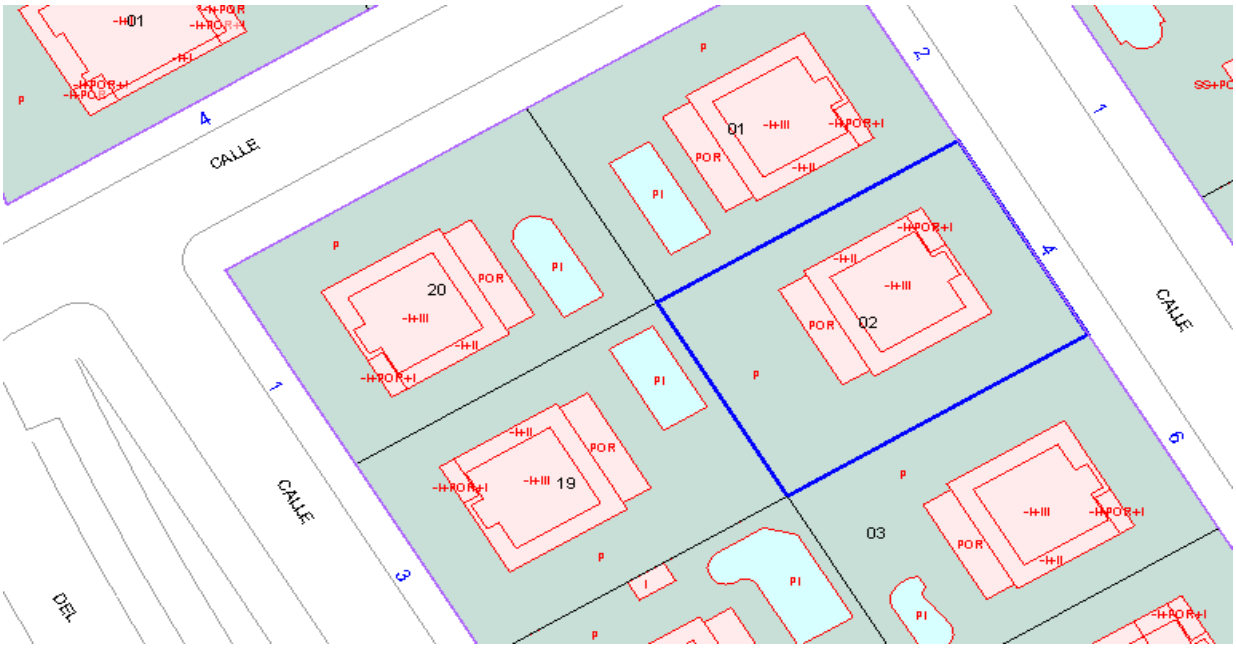



*Figura 24: Cartografía sedecatastro*

PARCELA CATASTRAL

Parcela construida sin división horizontal

Localización: CL NEPTUNO 4
PARACUELLOS DE JARAMA (MADRID)

Superficie gráfica: 450 m²

CONSTRUCCIÓN

| Uso principal | Escalera | Planta | Puerta | Superficie m² | Tipo Reforma | Fecha Reforma |
|---|---|---|---|---|---|---|
| APARCAMIENTO | | -1 | 01 | 56 | | |
| ALMACEN | | -1 | 02 | 39 | | |
| SOPORT. 50% | | 00 | 01 | 12 | | |
| VIVIENDA | | 00 | 02 | 91 | | |
| VIVIENDA | | 01 | 01 | 95 | | |
| ALMACEN | | 02 | 01 | 51 | | |

*Figura 25: Información inmueble sedecatastro*

# 7.1.5.2. Abctelefonos

https://www.abctelefonos.com/

La página abctelefonos nos permite usar las páginas amarillas para encontrar la ubicación del domicilio de una persona [10].

Esta página nos permite buscar por nombre y ubicación, donde podemos introducir y buscar el nombre de nuestro objetivo dentro de un municipio concreto. Una vez realizada la búsqueda, nos mostrara un listado de todas las personas con un teléfono registrado a ese nombre en el municipio indicado y además nos indicara su número fijo y la dirección de su casa. Por lo tanto, podrían aparecernos más de una posibilidad, dependiendo del nombre y del municipio indicado ya que no es lo mismo buscar un “Esteban Galindo” en “Mejorada del campo” que un “Carlos García” en “Madrid”. Probablemente la primera opción encontráramos muy pocos o solo el objetivo en el caso de que existiera y en la

segunda opción, al ser un nombre tan común en una ciudad de millones de habitantes encontraríamos miles de coincidencias.

En cualquier caso, es una página muy recomendable para la fase de recolección de información a la que luego dar inteligencia.


*Figura 26: abctelefonos*

## 7.1.5.3. Permutador de correos

http://metricsparrow.com/toolkit/email-permutator/

Esta página puede ser de gran utilidad para obtener posibles correos electrónicos del objetivo. Nos permite introducir su nombre y apellidos y nombre de usuario, en el caso de que los conozcamos, y generara combinaciones entre ellos para formar distintas posibilidades de correos electrónicos [11].

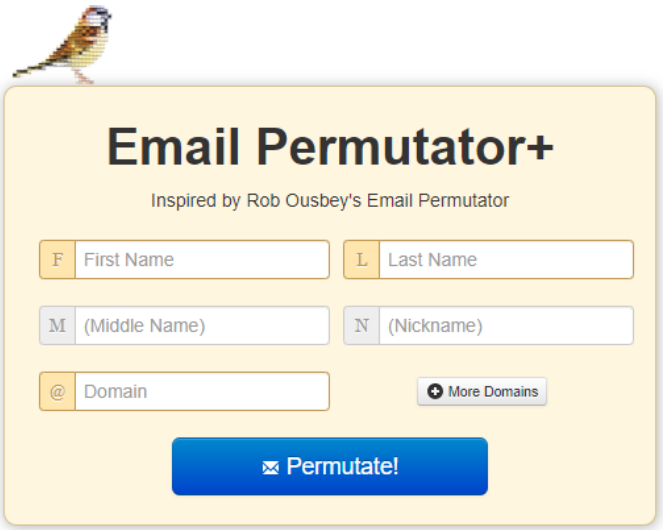



*Figura 27: Permutador de email*

Una vez introducida la información y ejecutado el permutador, obtendremos algo de este estilo.

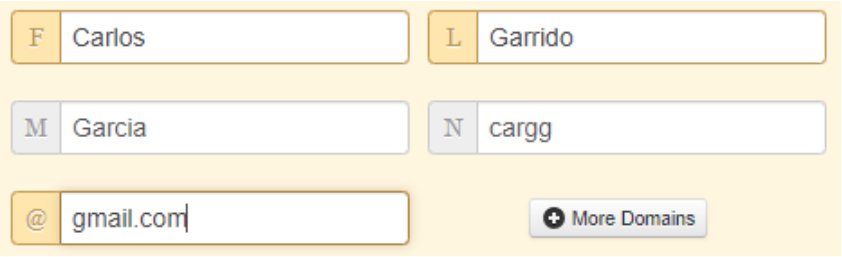



*Figura 28: Introducción de datos Permutador de email*

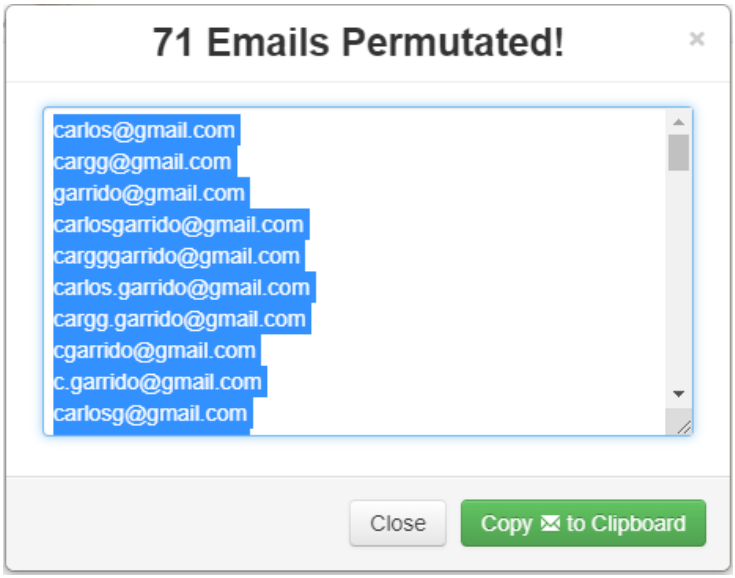



*Figura 29: Resultados Permutador de email*

### 7.1.6. Carrot2

https://search.carrot2.org/#/search/web

Carrot2 es un buscador menos convencional que los mencionados anteriormente, (Google, Bing, Yahoo!, etc.), sin embargo, es un motor de búsqueda muy potente y útil para OSINT. Este es un motor de búsqueda que se basa en un clúster de máquinas basado en el proyecto de código abierto de Stanislaw Osinski y Dawid Weiss [12].

*Figura 30: Carrot2*

La característica principal de este buscador que lo diferencia del resto de motores de búsqueda convencionales que ya hemos nombrado es que carrot2 separa los resultados de la búsqueda por categorías. Crea categorías como nombres, tipos de archivo, ámbitos, etc. Y lo desglosa para que podamos ir accediendo a la información según nuestros intereses, además carrot2 ofrece varios tipos de interfaces con menús esquematizados, para que cada usuario pueda adaptarlo a sus gustos o para su mayor comodidad y facilidad al extraer la información.

Cuando entramos a carrot2, encontramos una interfaz sencilla con una barra de búsqueda donde introducir lo que queremos buscar.

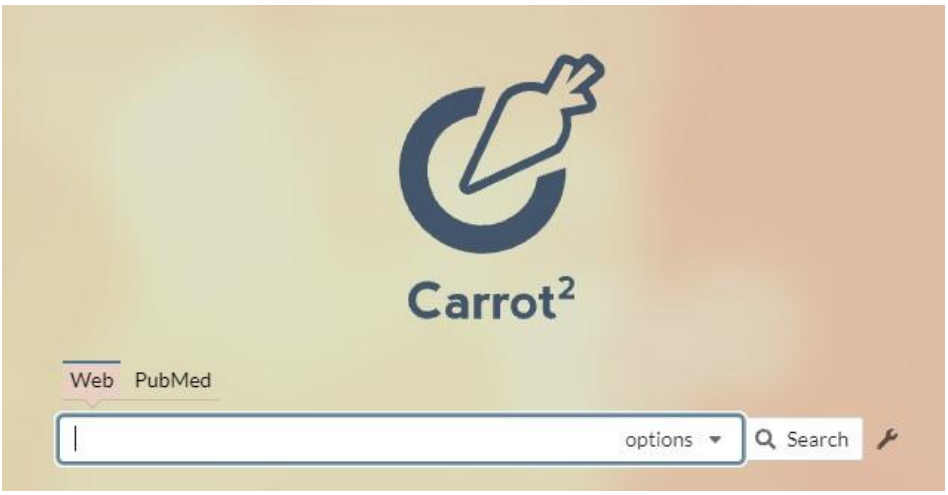



*Figura 31: Buscador carrot2*

Una vez introducimos un nombre y buscamos nos mostrará los resultados en 3 posibles paneles que podemos elegir, el primero es un panel tipo lista, el segundo es un panel tipo árbol y el tercero es un panel en forma de rueda de contenidos. Cada uno de los paneles muestra la misma información, pero organizada y clasificada de distinta forma.

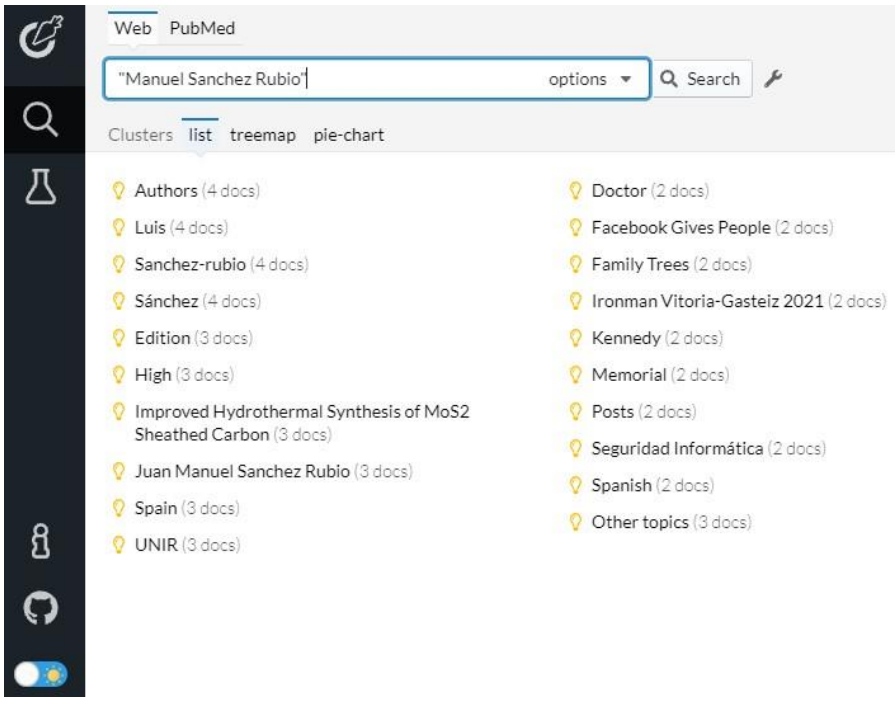



*Figura 32: Resultados modo lista en carrot2*


*Figura 33: Resultados modo mapa en carrot2*

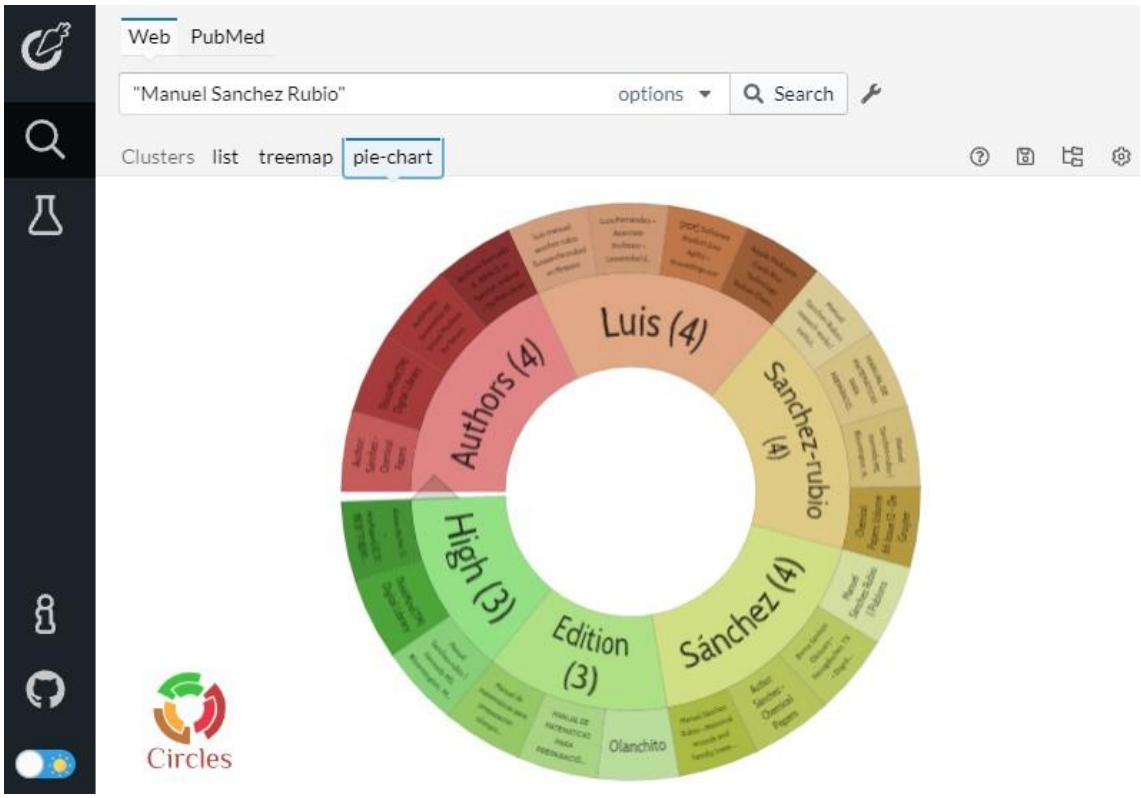



*Figura 34: Resultados modo rueda en carrot2*

## 7.2. Redes Sociales

Las redes sociales son probablemente, después de los motores de búsqueda, la mayor fuente de información que podemos consultar, ya que las redes sociales conectan todo, personas, instituciones, negocios, ubicaciones, etc.

Cuando una persona tiene gente cercana como amigos, familiares, compañeros de trabajo o incluso sus propias empresas. Es muy probable que, si esa persona usa redes sociales, siga o le sigan las cuentas de todas las personas o instituciones que se acaban de mencionar.

Además, normalmente la gente tiene un gran desconocimiento sobre el peligro que conllevan estas redes y suelen tener sus perfiles públicos, de manera que cualquier persona puede acceder a toda esa información. Incluso pueden subir fotos, tweets y más información mostrando su ubicación actual, aportando así información sensible sobre ellos mismos y su círculo de gente cercana.

Estas plataformas surgieron en la década de los 2000, sin embargo, la gran revolución de las redes sociales llego con la aparición de MySpace en 2003, que fue la primera red social del estilo de las que conocemos hoy en día y que permitía la creación de un perfil completo.


*Figura 35: MySpace*

Hoy en día las principales redes sociales son Facebook, Twitter, LinkedIn, Instagram, TikTok y Snapchat.


*Figura 36: Redes Sociales*

Todas estas redes permiten hacer lo que se ha comentado anteriormente. Podemos crearnos un perfil personal completo donde plasmar nuestros gustos, información personal como nombre y apellidos, dirección, zonas por las que nos movemos, trabajo, estudios.

Estas redes también nos permiten publicar fotos, videos, música, hablar con gente tanto de forma privada como de forma pública y exponer nuestras conversaciones a que las vea cualquier persona. Existe tanta cantidad de información en las redes sociales y es tan sencillo acceder a ella, ya que no hacen falta ningún tipo de herramienta ni búsqueda compleja para extraer toda esta información, que se han convertido en una de las principales fuentes de obtención de información y herramientas para extraerla en sí mismas.

### 7.2.1. Facebook

Facebook es la red social más extendida y que más usuarios activos posee hoy en día, con más de 2700 millones de usuarios [13].


*Figura 37: Crecimiento de Facebook*

Esta red social apareció en 2004, creada por Mark Zuckerberg junto a otro grupo de alumnos de Harvard, en un primer momento fue lanzada para los alumnos de Harvard. Sin embargo, fue adquiriendo mucha popularidad y extendiéndose a otras universidades, hasta que se hizo conocida mundialmente y se extendió a otras instituciones y a cualquier persona.

Facebook, al contrario que otras muchas redes sociales, ofrece una gran cantidad de servicios a sus usuarios y no solo servicios de mensajería y publicación de fotos y mensajes.

En esta red social podemos encontrar servicios como aplicaciones y juegos donde los usuarios pueden jugar y competir y mostrar sus resultados en la plataforma, también se pueden crear grupos de personas, listas de amigos, posee servicio de mensajería tanto privada como en forma de publicación y al igual que otras redes sociales también permite subir fotos, comentarlas, valorarlas con me gusta, etc.

Todos estos servicios que ofrece Facebook han conseguido una gran interacción del usuario con la plataforma, ya que prácticamente se puede hacer cualquier cosa, puedes hablar con amigos, hablar con gente por temas laborales, crear grupos tanto de amigos como de trabajo, subir y comentar tus fotos, incluso jugar a juegos, por lo tanto, engloba en una sola red social las características de muchas. Por ello Facebook cada día almacena una enorme cantidad de información de todo tipo de cada usuario.

Es tal la cantidad de información que Facebook genera que, según la compañía, generan 4 petabytes de datos por día.

Facebook nos facilita el acceso a todos estos datos ya que nos permite buscar y filtrar la búsqueda por tipo, podemos introducir un nombre y filtrar por publicaciones, personas, fotos, videos, sitios de compras, páginas, lugares, grupos y eventos.

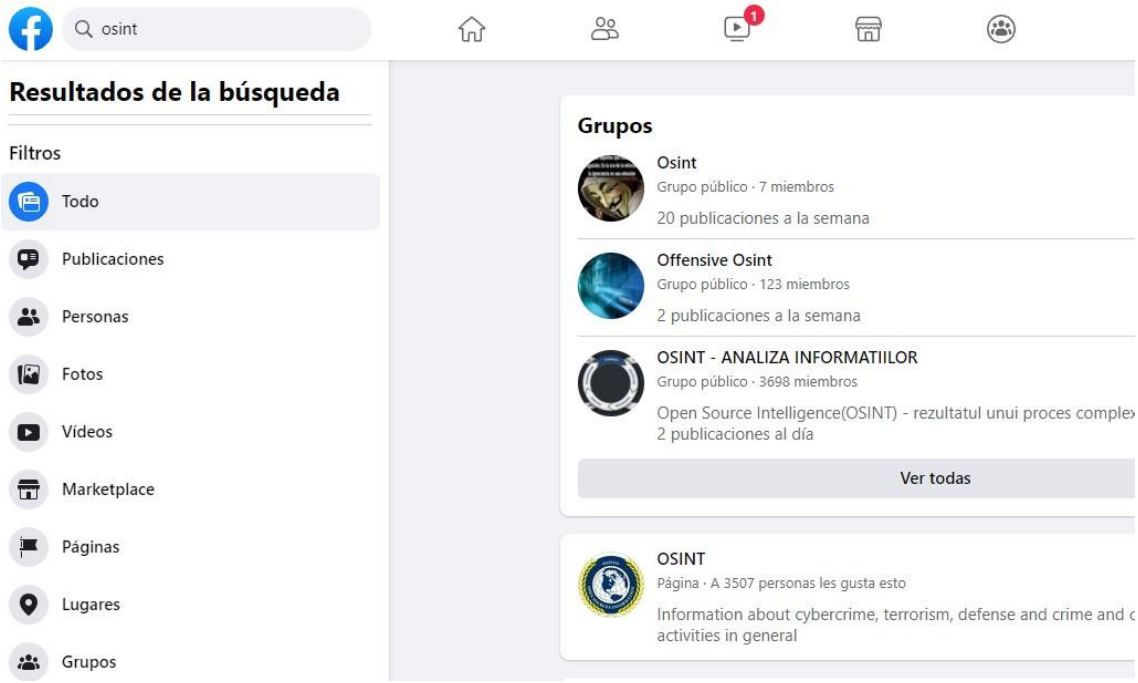



*Figura 38: Búsqueda Facebook*

También existen páginas web que se encargan de hacer búsquedas personalizadas en Facebook y que nos ahorraran tiempo en muchos casos, una de ellas es https://whopostedwhat.com/ [14].

Esta página nos permitirá:

- Obtener la ID de un usuario de Facebook.

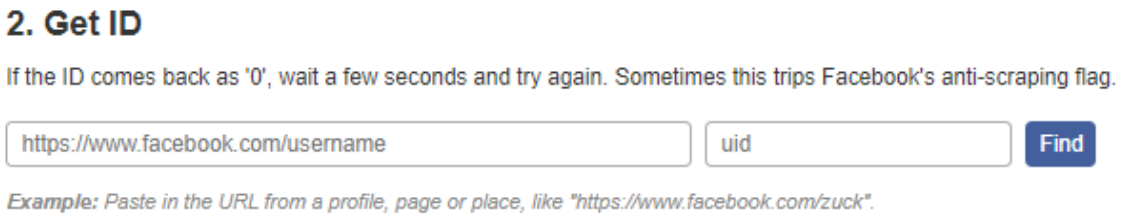



*Figura 39: Obtener ID Facebook*

- Buscar publicaciones por fecha (día, mes, año).


*Figura 40: Buscar publicaciones por fecha en Facebook*

- Buscar publicaciones por rango de tiempo (entre dos fechas).


*Figura 41: Buscar publicaciones por rango de tiempo en Facebook*

- Buscar publicaciones por lugar.

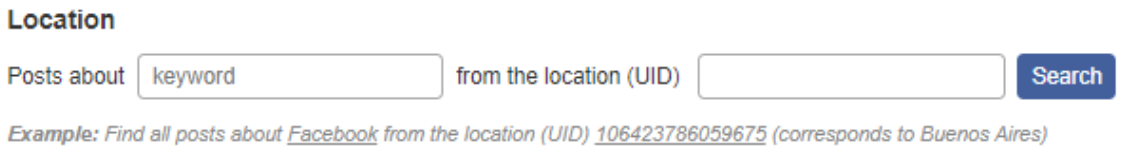



*Figura 42: Buscar publicaciones por lugar en Facebook*

- Buscar publicaciones por ID.

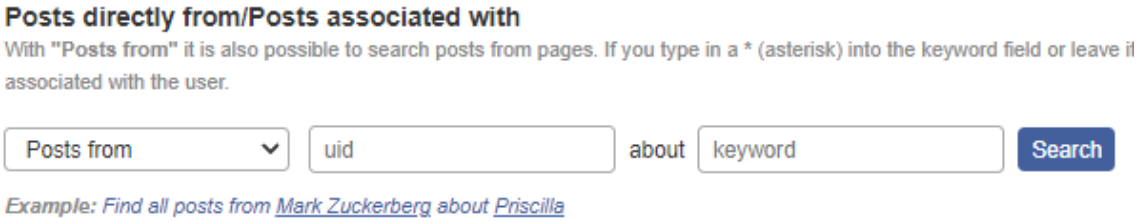



*Figura 43: Buscar publicaciones por ID en Facebook*

## 7.2.2. Twitter

Twitter es otra de las grandes redes sociales que se usan en la actualidad, quizás la más relevante en el ámbito del micro blogueo, donde la gente sube pequeños posts a la plataforma a los que se llama tweets, con un máximo de 280 caracteres, por esto también se le conoce como el SMS de internet [15][16].

Esta plataforma fue lanzada en 2006 por Jack Dorsey y desde entonces ha tenido un crecimiento masivo, hasta el punto de que hoy en día posee un total de 330 millones de usuarios activos.

Twitter gracias a su formato, es una de las mejores redes sociales de las que recopilar información, ya que la gente suele twittear y retwittear una gran cantidad de mensajes de forma pública. Además, el desconocimiento del peligro de esto, provoca que mucha gente tenga conversaciones públicas, de forma que nos desvelan su gente cercana, amigos, familia, etc. A parte de lo acabado de

mencionar, los usuarios también exponen una gran cantidad de información personal en sus perfiles, donde nombran los lugares donde han estudiado, lugar donde viven, incluso enlaces o sus nombres de otras cuentas de redes sociales. Todo esto ha provocado que aparezcan multitud de herramientas y páginas web dedicadas a la recopilación de todo tipo de información de los usuarios de Twitter. A continuación, se enseñarán algunas de las más relevantes.

- Foller.me

  https://foller.me/

  Esta página web nos permite introducir un nombre de usuario de Twitter, una vez introducido, la página buscara la cuenta de Twitter y empezara a extraer información de todo tipo. Información como nombre completo, fecha de inicio en Twitter, información del perfil (followers, tweets, etc.), palabras clave de temas que suele tratar el objetivo, hashtags, menciones [17][18].

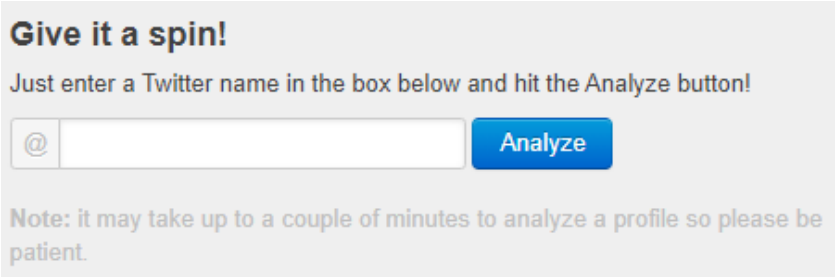
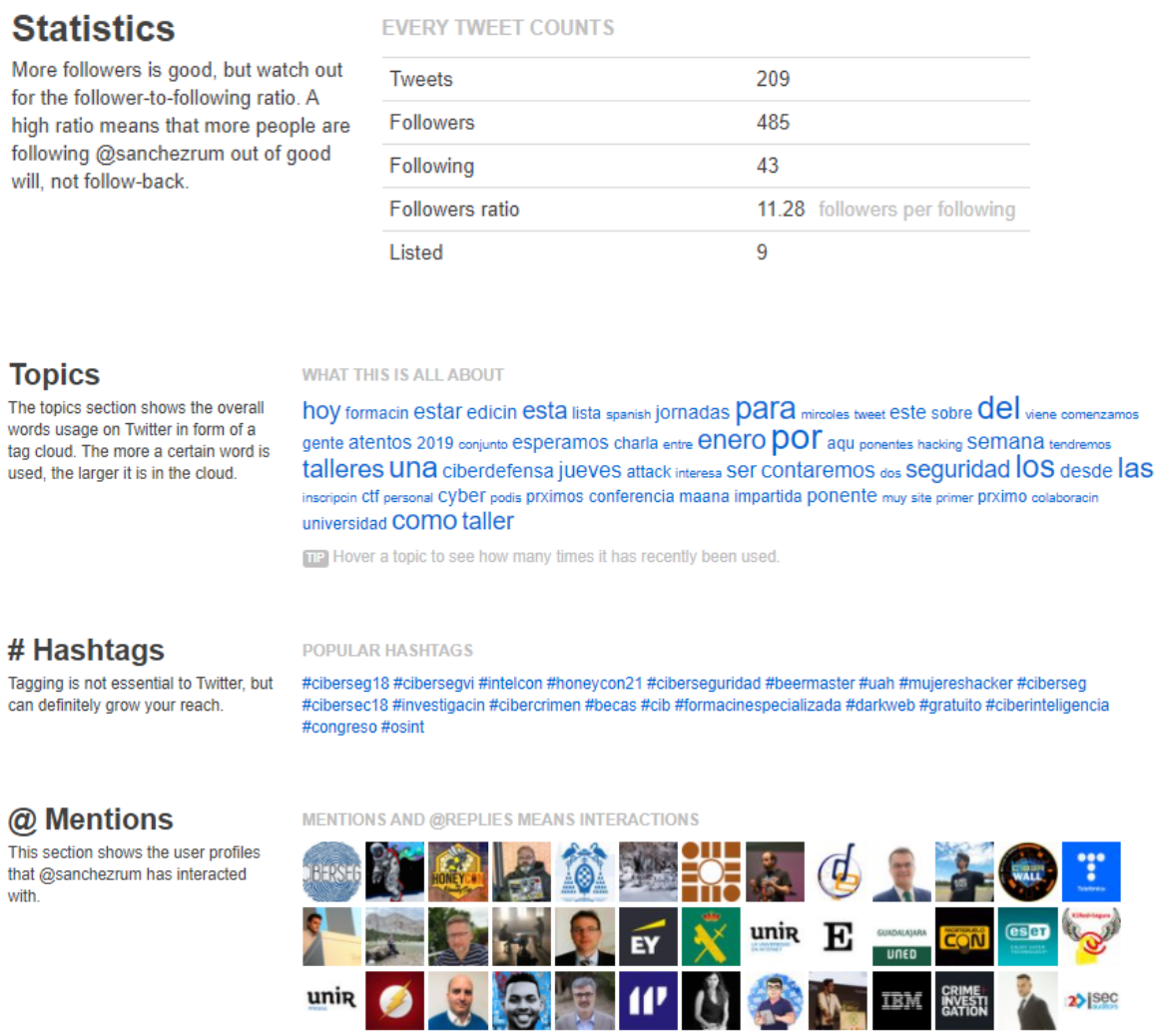



*Figura 44: Buscador Foller.me*

Resultado obtenido:


*Figura 45: Resultado Foller.me 1*



## Statistics

More followers is good, but watch out for the follower-to-following ratio. A high ratio means that more people are following @sanchezrum out of good will, not follow-back.

EVERY TWEET COUNTS

| | |
|---|---|
| Tweets | 209 |
| Followers | 485 |
| Following | 43 |
| Followers ratio | 11.28 followers per following |
| Listed | 9 |

## Topics

The topics section shows the overall words usage on Twitter in form of a tag cloud. The more a certain word is used, the larger it is in the cloud.

WHAT THIS IS ALL ABOUT

hoy formacin estar edicin esta lista spanish jornadas para mircoles tweet este sobre del viene comenzamos gente atentos 2019 conjunto esperamos charla entre enero por aqu ponentes hacking semana tendremos talleres una ciberdefensa jueves attack interesa ser contaremos dos seguridad los desde las inscripcin ctf personal cyber podis prximos conferencia maana impartida ponente muy site primer prximo colaboracin universidad como taller

TIP Hover a topic to see how many times it has recently been used.

## # Hashtags

Tagging is not essential to Twitter, but can definitely grow your reach.

POPULAR HASHTAGS

#ciberseg18 #cibersegvi #intelcon #honeycon21 #ciberseguridad #beermaster #uah #mujereshacker #ciberseg #cibersec18 #investigacin #cibercrimen #becas #cib #formacinespecializada #darkweb #gratuito #ciberinteligencia #congreso #osint

## @ Mentions

This section shows the user profiles that @sanchezrum has interacted with.

MENTIONS AND @REPLIES MEANS INTERACTIONS

*Figura 46: Resultado Foller.me 2*

# 7.2.3. Instagram

Instagram es una red social de origen estadounidense que inicialmente apareció como una aplicación para iPhone en 2010 y que lanzó su versión para Android en 2012. Después de ver su gran crecimiento y potencial esta fue comprada por Facebook [19].

Instagram consiste en una red social orientada al contenido multimedia como fotos o videos, aunque también tiene servicio de mensajería privada a lo que se

denomina direct. Esta además ofrece multitud de posibilidades, ya que cualquier usuario puede subir fotos o videos. En el caso de subir fotos con más gente, esta gente se puede etiquetar para identificarles, pueden comentarse las fotos o darles me gusta, pueden hacerse directos donde la gente puede comentar y hablar en tiempo real y opinar sobre el directo. Esto ha ocasionado que Instagram cobre una gran importancia entre los adolescentes.

Instagram tiene un gran peligro, como se acaba de comentar, ya que principalmente la usan adolescentes, jóvenes, influencers. Gente que buscan tener el mayor número de seguidores posible, sin importar los riesgos que esto conlleva y el nivel de exposición a la gente que obtienen y que para ello mantienen sus perfiles en público y con una gran cantidad de fotos y de información sobre ellos.

Esto es el principal motivo por el que Instagram es otra de las grandes y más fiables fuentes de información para una investigación OSINT.

### 7.2.4. TikTok

TikTok es una red social relativamente nueva lanzada en 2016 por la empresa tecnológica china Bytedance y que ha alcanzado una popularidad y números abrumadores [20].

Esta red social tiene un gran parecido a Instagram (mencionada anteriormente), ya que se centra en el contenido multimedia. Se suelen usar videos de corta duración, videos de pocos segundos a un minuto generalmente. También tiene un servicio de mensajería directa y privada donde puedes comunicarte con distintos usuarios de la plataforma.

Al igual que Instagram, este contenido tiene un gran peligro, debido a que normalmente los videos que se suben suelen mostrar a la propia persona o gente cercana, a los que se está exponiendo a millones de usuarios.

### 7.2.5. LinkedIn

LinkedIn es una red social orientada al mundo laboral que se lanzó en 2002. Esta red social, a diferencia del resto de redes de las que hemos hablado, es una red social destinada a dar acceso y dar a conocer a profesionales y empresas [21].

LinkedIn permite crear un perfil a modo de currículum, donde se puede indicar los estudios que se han realizado y donde se han completado, además de la experiencia laboral, destrezas, etc. Además, permite conectar con personas o empresas y en el caso de que acepten dicha conexión nos permite acceder también a sus contactos. De esta manera LinkedIn permite crear una red enorme, lo que nos aporta una gran versatilidad y facilidad para encontrar información sobre personas o instituciones.

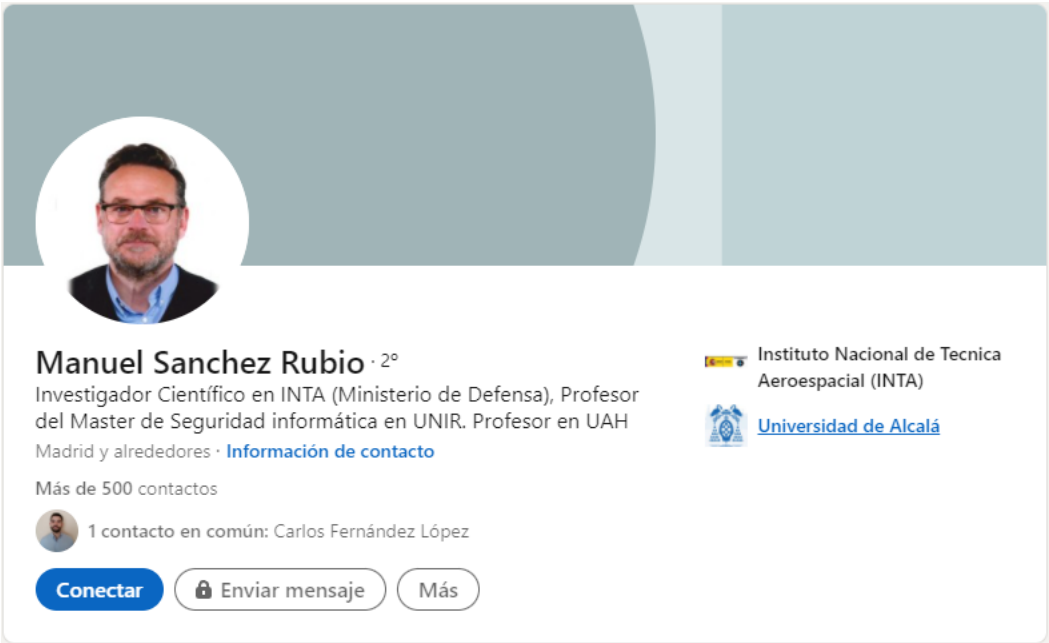



*Figura 47: Perfil LinkedIn*

## Datos destacados

**Estudiasteis en Universidad de Alcalá**
Manuel empezó en Universidad de Alcalá antes que tú

Saludar

## Experiencia

**Investigador Científico de OPI (INTA) Ministerio de Defensa**
Instituto Nacional de Tecnica Aeroespacial (INTA)
oct 1991 – actualidad · 30 años y 1 mes
http://www.inta.es

Jefe del Laboratorio de Telemedida e Instrumentación del Área de Ensayos en Vuelo

**Director del Master de Seguridad Informática en UNIR**
UNIR Universidad Internacional de la Rioja
mar 2016 – actualidad · 5 años y 8 meses
UNIR

Director del Master de Seguridad Informática
Investigador Principal del Grupo de investigación "cybersecuritics"
Docencia en el Master de Seguridad Informática
Docencia en el Máster en Ingeniería de Software y Sistemas Informáticos. ...ver más

**Profesor**
Universidad de Alcalá
feb 2000 – actualidad · 21 años y 9 meses
http://www.cc.uah.es/msanchez

Director Cátedra DARS "Ciberinteligencia"
Docencia en el Máster de Ingeniería del Software para la Web
Docencia en el Máster de Ciberdefensa
Docencia en Grado (Sistemas de Información e Ingeniería de Computadores) ...ver más

## Educación

**Universidad de Alcalá**
Doctor Ingeniero en Informática (cum laude)
2008 – 2013

**Universidad de Alcalá**
Ingeniero, Informatica
2003 – 2008

**UAH**
Diplomado en Informática de Sistemas
1993 – 1998

*Figura 48: Perfil LinkedIn 2*

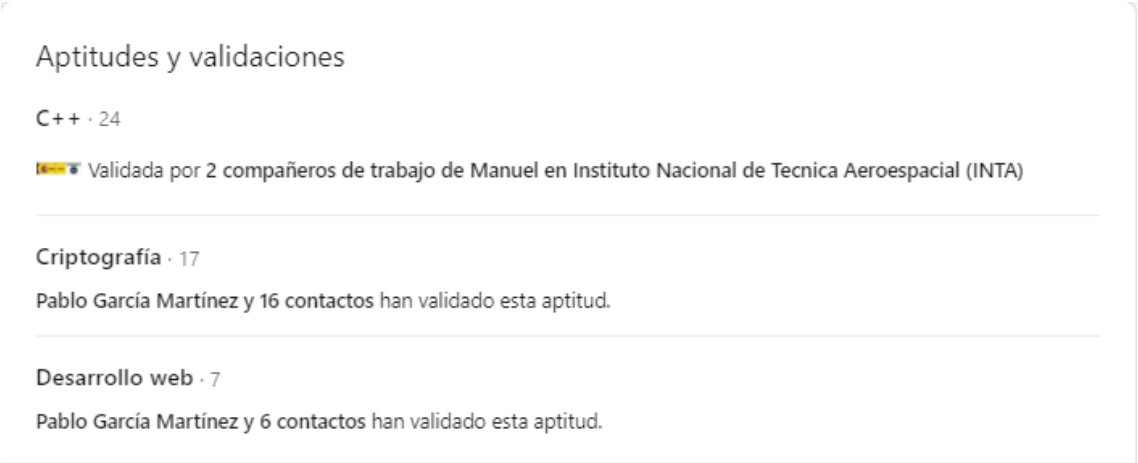



*Figura 49: Perfil LinkedIn 3*

Este perfil mostrado es un ejemplo de perfil de LinkedIn. Como podemos observar, a partir de LinkedIn podemos obtener una gran cantidad de información importante a la que puede acceder cualquier persona.

Por ejemplo, viendo donde ha completado sus estudios una persona, podemos hacernos una idea aproximada de donde viva posiblemente, ya que normalmente se intenta estudiar lo más cerca de casa posible. También podemos ver el mundo en el que se mueve el objetivo, incluso ver contactos que tenga que puedan ayudarnos a obtener información extra.

## 7.3. Username

Todas las redes sociales normalmente identifican a las personas con un username o Nick y no por el nombre completo real y es muy frecuente que cuando una persona en una red social tenga un username, use también ese mismo username en distintas redes sociales que tenga. Esto es muy frecuente ya que si no habría que tener distintos nombres de usuario para cada red social en vez de tener solo uno.

Gracias a esto podemos beneficiarnos a la hora de extraer información para la investigación, ya que, si supiéramos el nombre de usuario del objetivo en una red social o lo averiguáramos, podríamos buscar con ese mismo username en las distintas redes sociales con el objetivo de encontrar más información accesible.

Pero, tal y como hemos comentado antes, existen una gran multitud de redes sociales y sería una tarea muy tediosa buscar una a una. Para eso existen determinadas páginas que sirven para identificar usernames en las distintas redes. De esta manera podemos introducir un username y al hacer la búsqueda, la página nos indicara que redes sociales tienen ya una coincidencia y por lo tanto alguien tiene ya registrado ese username o, al contrario.

Tres de las páginas más usadas para esta tarea son knowem, checkusernames, namecheck.

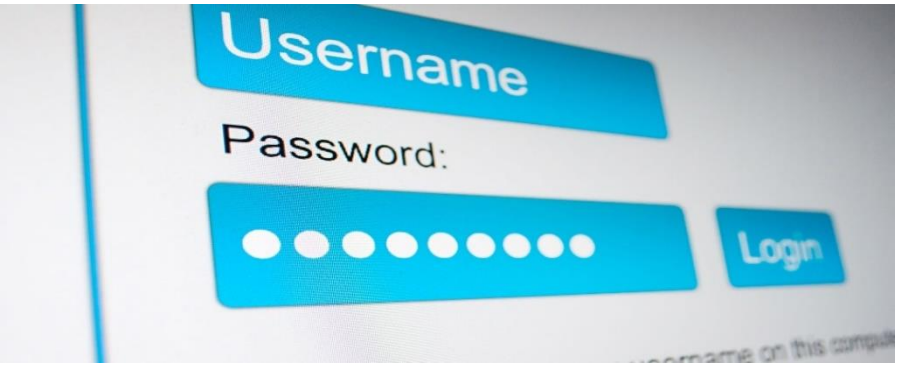



*Figura 50: Username*

*Figura 51: Knowem*

*Figura 52: Checkusernames*

*Figura 53: Namecheck*

## 7.3.1. Knowem

https://knowem.com/

Cuando entramos en la página, lo primero que nos aparece es un buscador donde nos permite introducir un nombre de usuario.


*Figura 54: Buscador Knowem*

Una vez buscado el nombre de usuario, la página nos mostrara las 25 redes sociales más populares indicándonos si el nombre está disponible para cada una de ellas o no.


*Figura 55: Resultado Knowem 2*

También nos mostrara un listado de los dominios más conocidos, indicando una vez más si está disponible o si está ya en uso.


*Figura 56: Resultado Knowem 2*

## 7.3.2. Checkusernames

https://checkusernames.com/

Esta página es muy similar a la anterior, al entrar nos aparecerá arriba a la izquierda un cuadro de texto donde podemos introducir un nombre de usuario y buscarlo.


*Figura 57: Buscador Checkusernames*

Una vez buscado, se mostrarán en color grisáceo apagado las redes sociales que no tengan ese username disponible porque ya este usado y en color normal aquellas que si este disponible.

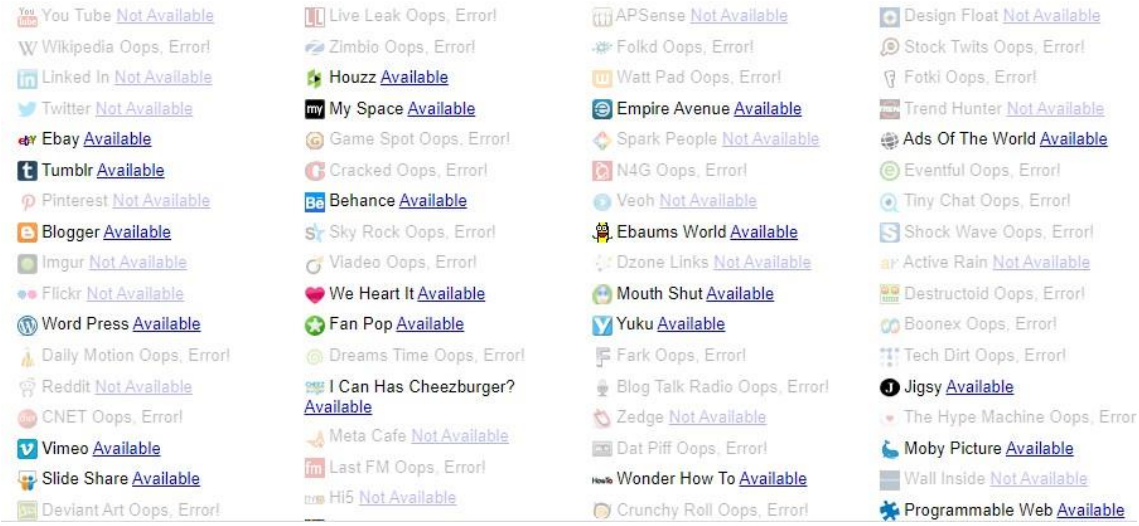



*Figura 58: Resultado Checkusernames*

## 7.3.3. Namecheck

https://www.namecheck.com/en/

Al entrar en la página de Namecheck encontramos el cuadro de texto donde podemos introducir el nombre de usuario a buscar, solo aparece eso, por lo tanto, no hay opción a equivocarse.

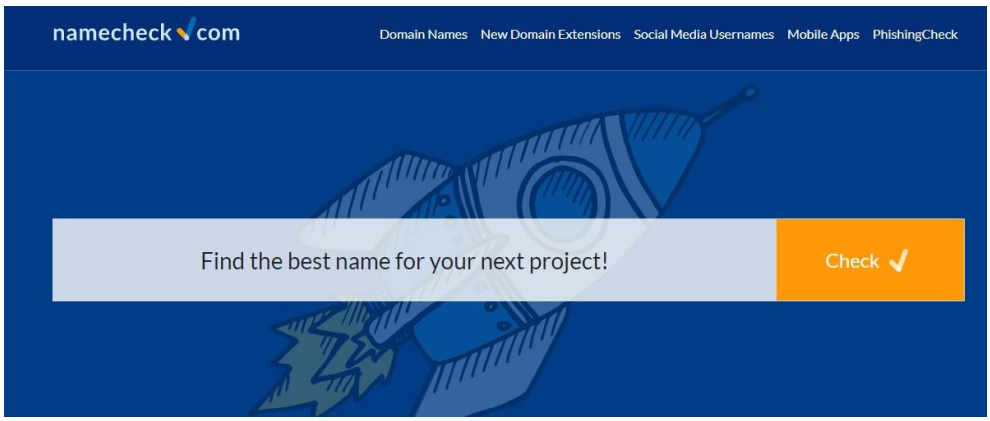



*Figura 59: Buscador Namecheck*

Una vez introducido el nombre, la página nos mostrara un listado de dominios y redes sociales donde también nos va a indicar si ese nombre ya está usado o está disponible.

Domain Names learn more

| | | |
|---|---|---|
| sanchezrum.com | available | Register |
| sanchezrum.co.uk | available | Register |
| sanchezrum.org | available | Register |
| sanchezrum.us | available | Register |
| sanchezrum.net | available | Register |
| sanchezrum.de | available | Register |

*Figura 60: Resultado Namecheck 1*

Social Media Usernames learn more

| | | |
|---|---|---|
| facebook | already taken | View |
| twitter | already taken | View |
| linkedin | already taken | View |
| instagram | already taken | View |
| tumblr | available | Register |
| pinterest | available | Register |
| youtube | already taken | View |

*Figura 61: Resultado Namecheck 2*

En general estas tres páginas mostradas tienen un funcionamiento muy similar, una interfaz muy sencilla donde prácticamente solo encontramos el cuadro de texto donde introducir el nombre a buscar y que una vez buscado nos mostraran un listado de las diferentes redes sociales indicando cuales están disponibles con ese respectivo nombre y cuáles no.

# 7.4. Metabuscadores

Con la aparición de tantos buscadores y navegadores puede hacerse una tarea laboriosa ir buscando uno a uno que información podemos obtener sobre el objetivo.

Para ello entran en juego los metabuscadores, los cuales recopilan información de varios buscadores y motores de búsqueda a la vez. Estos actúan como si fueran un buscador de buscadores, son buscadores que no tienen base de datos propia, sino que se nutren de otros buscadores para obtener la información. Los metabuscadores suelen acceder a los buscadores más conocidos o importantes y muestran toda la información que extraen de estos buscadores a los que consultan.

Son una muy buena opción teniendo en cuenta que, al buscar en muchos motores de búsqueda, nos mostraran un resultado de búsqueda más completo y con más información, ya que es la información combinada de varios buscadores. Sin embargo, también pueden mostrar la información peor indexada y menos relevante a lo que estamos buscando y en general son más difíciles de usar a la hora de manejar la información.

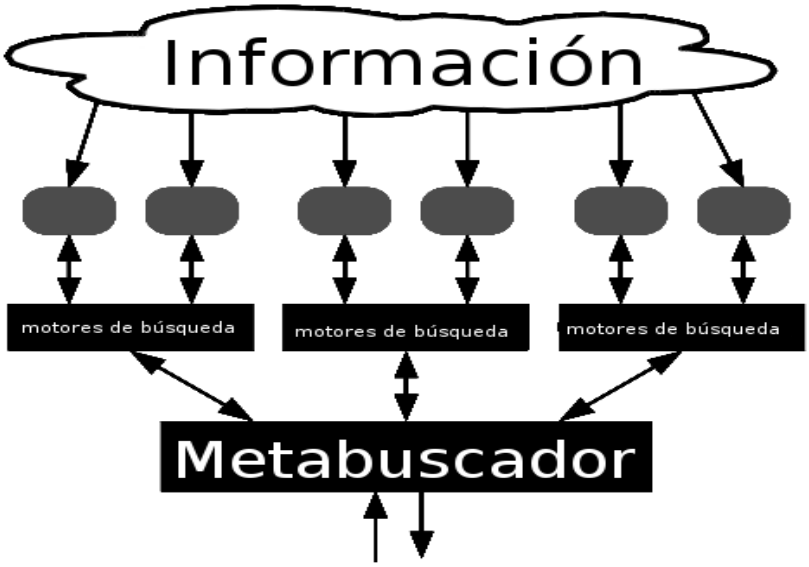



*Figura 62: Metabuscadores*

## 7.4.1. Webcrawler

https://www.webcrawler.com/

Webcrawler es un metabuscador que combina búsquedas de Google, Yahoo!, Bing, Ask.com, About.com, MIVA, Look Smart.

Este metabuscador además permite buscar por imágenes, videos y noticias. La interfaz es muy sencilla, una vez entramos al metabuscador encontramos una barra de búsqueda y una serie de botones para seleccionar si queremos que el resultado de la búsqueda sea todo, imágenes, videos o fotos.

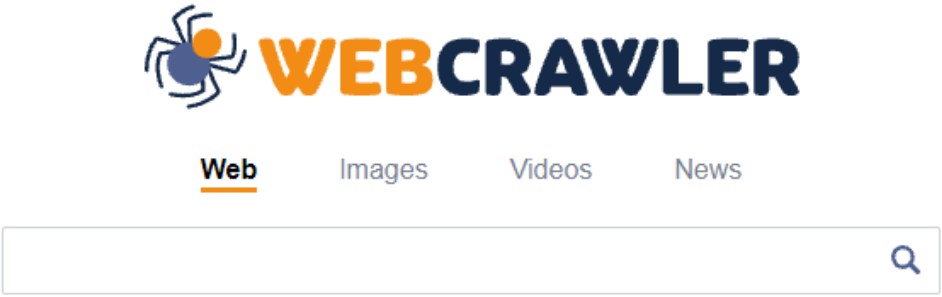



*Figura 63: Webcrawler*

## 7.4.2. Metacrawler

https://www.metacrawler.com/

Metacrawler es otro de los metabuscadores más conocidos y utilizados, este metabuscador empezó alimentándose de las bases de datos de los distintos

motores de búsqueda que usaba para obtener la información. En la actualidad tiene su propia base de datos, de donde obtiene información, aparte del resto de motores de búsqueda a los que accede.

La interfaz es similar a la de cualquier buscador y prácticamente idéntica a la del metabuscador mencionado anteriormente.


*Figura 64: Metacrawler*

## 7.4.3. Yasni

http://www.yasni.com/

Yasni es un metabuscador que a su vez está orientado a la búsqueda de personas y perfiles online.

Este metabuscador permite buscar empresas y empleos dependiendo del ámbito que selecciones, permite buscar personas en un ámbito laboral y permite buscar personas introduciendo nombre completo, apodo conocido, etc.

Yasni obtiene los resultados de varios motores de búsqueda como Google y Bing entre otros.

En la interfaz encontramos tres barras de búsqueda, cada una para buscar uno de los tipos de búsqueda que se acaban de nombrar.

## 7.5. Software de búsqueda

Aparte de las fuentes más accesibles, conocidas y usadas por la gente mencionadas anteriormente como motores de búsqueda, redes sociales, etc. Existen diversas herramientas software diseñadas para la búsqueda y recopilación de información, que, con la facilidad de un solo clic, ellas mismas se encargan de buscar en una gran cantidad de sitios a la vez y de mostrarnos todos los resultados obtenidos.

Estas herramientas nos brindan una facilidad abrumadora, ya que evitan que tengamos que ir de página en página buscando y recopilando la información o que tengamos que estar logueandonos y entrando en cada una de las redes sociales para buscar posibles cuentas de un objetivo. Para ello existen herramientas que buscan información de todo tipo, tanto información contenida en internet y en motores de búsqueda como información relacionada con las redes sociales (nombres de usuario, mensajes, tweets, cuentas, etc.).

### 7.5.1. IKy Proyect

Iky proyect es una herramienta ejecutada en Linux que funciona a través de una página web localhost alojada en un servidor que tenemos que hostear en la propia máquina Linux. Esta herramienta nos permite hacer una búsqueda sencilla introduciendo el correo electrónico del objetivo, o también hacer una búsqueda avanzada donde nos permite introducir más información, en el caso de que la tengamos, en esta búsqueda avanzada podemos introducir un email, nombre de usuario, nombres de usuario de redes sociales concretas, etc.

Esta herramienta en concreto tiene una instalación algo tediosa, en la siguiente página está indicado como instalar esta herramienta paso a paso [22].

https://esgeeks.com/iky-osint-recopilar-informacion-email/

Una vez instalado todo lo necesario, tenemos que abrir cuatro terminales e iniciar las cuatro tareas necesarias para funcionar, iniciamos el servidor redis, la aplicación y celery en el backend y el frontend y finalmente abrir la aplicación en la web.

1. Iniciamos el servidor redis:

```
adri@Adri-Ubuntu:~$ redis-server
```

*Figura 65: Iniciar servidor redis*

Una vez iniciamos el servidor redis debe aparecer algo como esto

```
adri@Adri-Ubuntu: ~
adri@Adri-Ubu...   adri@Adri-Ubu...   adri@Adri-Ubu...   adri@Adri-Ubu...
2270:M 07 Oct 2021 12:48:49.597 * monotonic clock: POSIX clock_gettime

                Redis 6.2.5 (00000000/0) 64 bit

                Running in standalone mode
                Port: 6379
                PID: 2270

                      https://redis.io

2270:M 07 Oct 2021 12:48:49.599 # Server initialized
2270:M 07 Oct 2021 12:48:49.599 # WARNING overcommit_memory is set to 0! Background s
ave may fail under low memory condition. To fix this issue add 'vm.overcommit_memory
= 1' to /etc/sysctl.conf and then reboot or run the command 'sysctl vm.overcommit_mem
ory=1' for this to take effect.
2270:M 07 Oct 2021 12:48:49.612 * Loading RDB produced by version 6.2.5
2270:M 07 Oct 2021 12:48:49.612 * RDB age 789555 seconds
2270:M 07 Oct 2021 12:48:49.612 * RDB memory usage when created 2.29 Mb
2270:M 07 Oct 2021 12:48:49.613 * DB loaded from disk: 0.014 seconds
2270:M 07 Oct 2021 12:48:49.613 * Ready to accept connections
```

*Figura 66: Servidor redis iniciado*

2. Accedemos al directorio de backend dentro de iKy (en una nueva terminal):

```
adri@Adri-Ubuntu:~$ cd iKy
adri@Adri-Ubuntu:~/iKy$ cd backend/
adri@Adri-Ubuntu:~/iKy/backend$
```

*Figura 67: Directorio backend*

3. Iniciamos celery:

```
adri@Adri-Ubuntu: ~/iKy/backend
adri@Adri-Ubu...  adri@Adri-Ubu...  adri@Adri-Ubu...  adri@Adri-Ubu...
adri@Adri-Ubuntu:~$ cd iKy
adri@Adri-Ubuntu:~/iKy$ cd backend/
adri@Adri-Ubuntu:~/iKy/backend$ ./celery.sh
/home/adri/.local/lib/python3.8/site-packages/fuzzywuzzy/fuzz.py:11: UserWarning: Usi
ng slow pure-python SequenceMatcher. Install python-Levenshtein to remove this warnin
g
  warnings.warn('Using slow pure-python SequenceMatcher. Install python-Levenshtein t
o remove this warning')
[2021-10-07 12:57:21,126: DEBUG/MainProcess] | Worker: Preparing bootsteps.
[2021-10-07 12:57:21,134: DEBUG/MainProcess] | Worker: Building graph...
[2021-10-07 12:57:21,135: DEBUG/MainProcess] | Worker: New boot order: {Beat, Timer,
Hub, Pool, Autoscaler, StateDB, Consumer}
[2021-10-07 12:57:21,212: DEBUG/MainProcess] | Consumer: Preparing bootsteps.
[2021-10-07 12:57:21,213: DEBUG/MainProcess] | Consumer: Building graph...
[2021-10-07 12:57:21,320: DEBUG/MainProcess] | Consumer: New boot order: {Connection,
 Events, Heart, Mingle, Tasks, Control, Agent, Gossip, event loop}

 -------------- celery@Adri-Ubuntu v4.4.5 (cliffs)
--- ***** -----
-- ******* ---- Linux-5.8.0-48-generic-x86_64-with-glibc2.29 2021-10-07 12:57:21
- *** --- * ---
```

*Figura 68: Iniciar Celery*

4. Iniciamos aplicación app.py (en una nueva terminal, en el directorio backend):

```
adri@Adri-Ubuntu: ~/iKy/backend
adri@Adri-Ubu...  adri@Adri-Ubu...  adri@Adri-Ubu...  adri@Adri-Ubu...
adri@Adri-Ubuntu:~$ cd iKy/
adri@Adri-Ubuntu:~/iKy$ cd backend/
adri@Adri-Ubuntu:~/iKy/backend$ python3 app.py
 * Serving Flask app "factories.application" (lazy loading)
 * Environment: production
   WARNING: This is a development server. Do not use it in a production deployment.
   Use a production WSGI server instead.
 * Debug mode: on
 * Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
 * Restarting with stat
 * Debugger is active!
```

*Figura 69: Iniciar Aplicación*

5. Iniciamos el servidor Frontend (en una nueva terminal, en el directorio frontend):

```
adri@Adri-Ubuntu: ~/iKy/frontend
adri@Adri-Ub...   adri@Adri-Ub...   adri@Adri-Ub...   adri@Adri-Ub...

adri@Adri-Ubuntu:~$ cd iKy/
adri@Adri-Ubuntu:~/iKy$ cd frontend/
adri@Adri-Ubuntu:~/iKy/frontend$ npm start

> ngx-admin-iKy@2.0.0 start /home/adri/iKy/frontend
> ng serve

Browserslist: caniuse-lite is outdated. Please run the following command: `npm u
pdate`
10% building 3/3 modules 0 activei ｢wds｣: Project is running at http://localhost
:4200/webpack-dev-server/
i ｢wds｣: webpack output is served from /
i ｢wds｣: 404s will fallback to //index.html

chunk {app-pages-pages-module} app-pages-pages-module.js, app-pages-pages-module
.js.map (app-pages-pages-module) 36.3 MB  [rendered]
chunk {main} main.js, main.js.map (main) 3.13 MB [initial] [rendered]
chunk {polyfills} polyfills.js, polyfills.js.map (polyfills) 480 kB [initial] [r
endered]
chunk {runtime} runtime.js, runtime.js.map (runtime) 9.01 kB [entry] [rendered]
chunk {scripts} scripts.js, scripts.js.map (scripts) 1.81 MB [entry] [rendered]
chunk {styles} styles.js, styles.js.map (styles) 4.87 MB [initial] [rendered]
chunk {vendor} vendor.js, vendor.js.map (vendor) 7.82 MB [initial] [rendered]
Date: 2021-10-13T07:25:51.646Z - Hash: 962c860c9dc345d4c23e - Time: 414706ms
** Angular Live Development Server is listening on localhost:4200, open your bro
wser on http://localhost:4200/ **
i ｢wdm｣: Compiled successfully.
```

*Figura 70: Iniciar servidor Frontend*

Una vez iniciadas todas las tareas necesarias para el funcionamiento de iKy, procedemos a iniciar la aplicación. Para ello abrimos un buscador e introducimos la siguiente dirección http://127.0.0.1:4200/. Una vez buscado esto, nos llevara directamente a la página web donde tendremos acceso a la herramienta y donde podremos empezar la búsqueda.

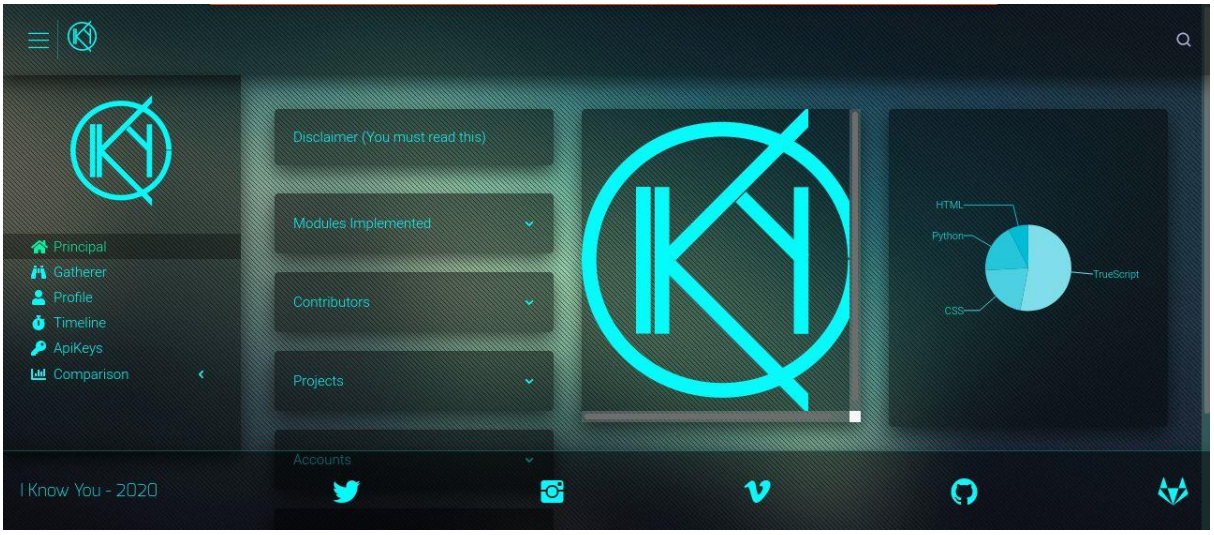



*Figura 71: IKy Proyect*

Para iniciar la búsqueda de una persona hay que entrar en el apartado "Gatherer", donde tenemos varias opciones de búsqueda. Hay una búsqueda sencilla donde nos permite introducir el correo electrónico del objetivo y una búsqueda avanzada donde permite introducir además del correo, nombres de usuario de distintas redes sociales en el caso de que los sepamos, para evitar posibles errores.

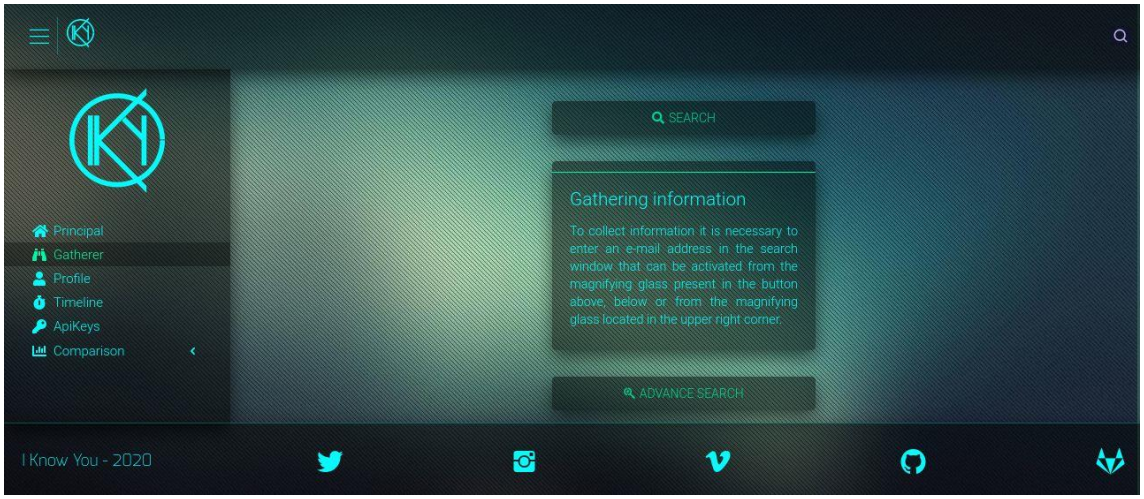



*Figura 72: Buscar en IKy Proyect*

Esta herramienta comenzara a buscar usuarios con el correo electrónico introducido y con el nombre de usuario, e ira ordenando toda esa información en tablas y grafos. Entre la información que buscará se encuentran redes sociales con ese correo o nombre de usuario, tweets (en el caso de que tenga Twitter), menciones, posibles nombres, etc.

## 7.5.2. Maltego

Maltego es una herramienta software desarrollada por paterva y diseñada para la búsqueda e inteligencia en fuentes abiertas [23]. Es una de las herramientas más potentes que existen para la inteligencia en fuentes abiertas.

Maltego nos permite buscar a una persona, página web, etc. y buscara en fuentes abiertas todo lo que encuentre sobre esta y nos lo mostrara en forma de grafo, de manera que tendremos un fácil acceso a toda la información.

Esta herramienta está disponible tanto en Windows como Linux o Mac y para poder usarla solo es necesario registrarse en paterva para tener acceso a sus servidores comunitarios de forma gratuita e instalar la herramienta en sí, lo cual es un proceso de instalación normal y corriente como cualquier programa.

Una vez abierta la aplicación nos encontramos una interfaz como esta:

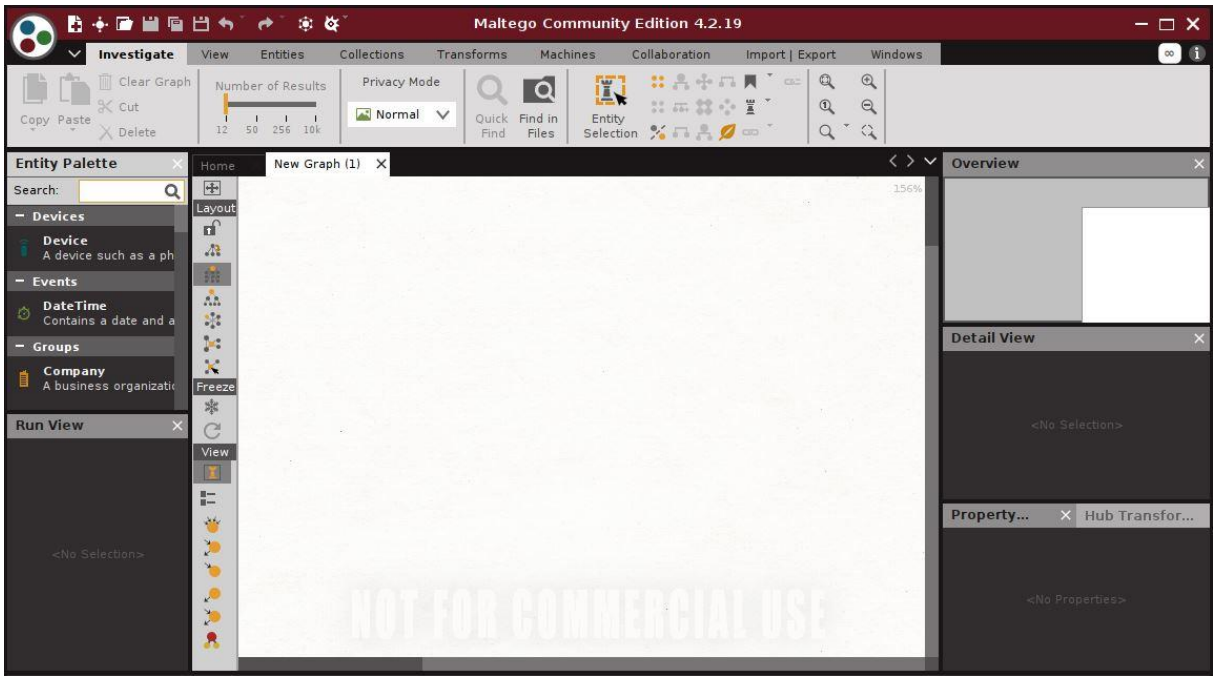



*Figura 73: Maltego*

Una vez en la pantalla principal de Maltego podemos seleccionar que tipo de búsqueda queremos realizar. Podemos buscar dominios, personas, alias y una gran cantidad de elementos más. Por ejemplo, si buscamos por dominio "uah.es", Maltego buscara cualquier información que encuentre en internet sobre el dominio, ya sean enlaces a páginas web, pdf, redes, perfiles, etc.

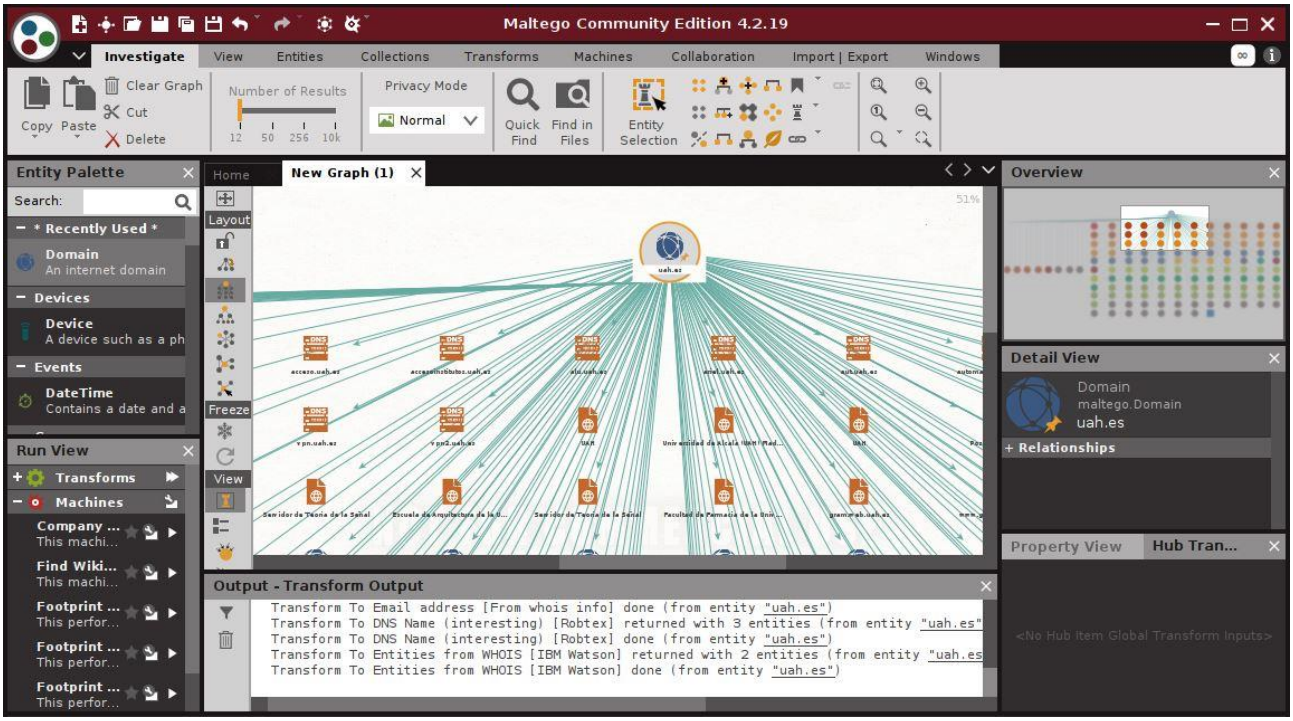



*Figura 74: Resultado Maltego*

## 7.5.3. FOCA

FOCA es otra de las principales y más usadas herramientas para OSINT. Esta herramienta se encarga principalmente de encontrar metadatos ubicados en documentos [24].

FOCA usa Google, Bing y DuckDuckGo, para explorar páginas web y encontrar documentos como ficheros pdf, Word, Excell, open office, etc. Y analizarlos en busca de metadatos del objetivo, de manera que va recopilando todos los metadatos que encuentra para posteriormente mostrárnoslos en forma de lista.

# 7.6. Osint Framework

Osint framework es una página web o repositorio que agrupa un gran número de herramientas OSINT donde poder hacer búsquedas siempre en fuentes abiertas.

Esta página nos muestra una agrupación de herramientas en forma de árbol desplegable. Podemos ir accediendo a cada ramificación del árbol y cada una muestra una categoría o ámbito para la búsqueda de información.

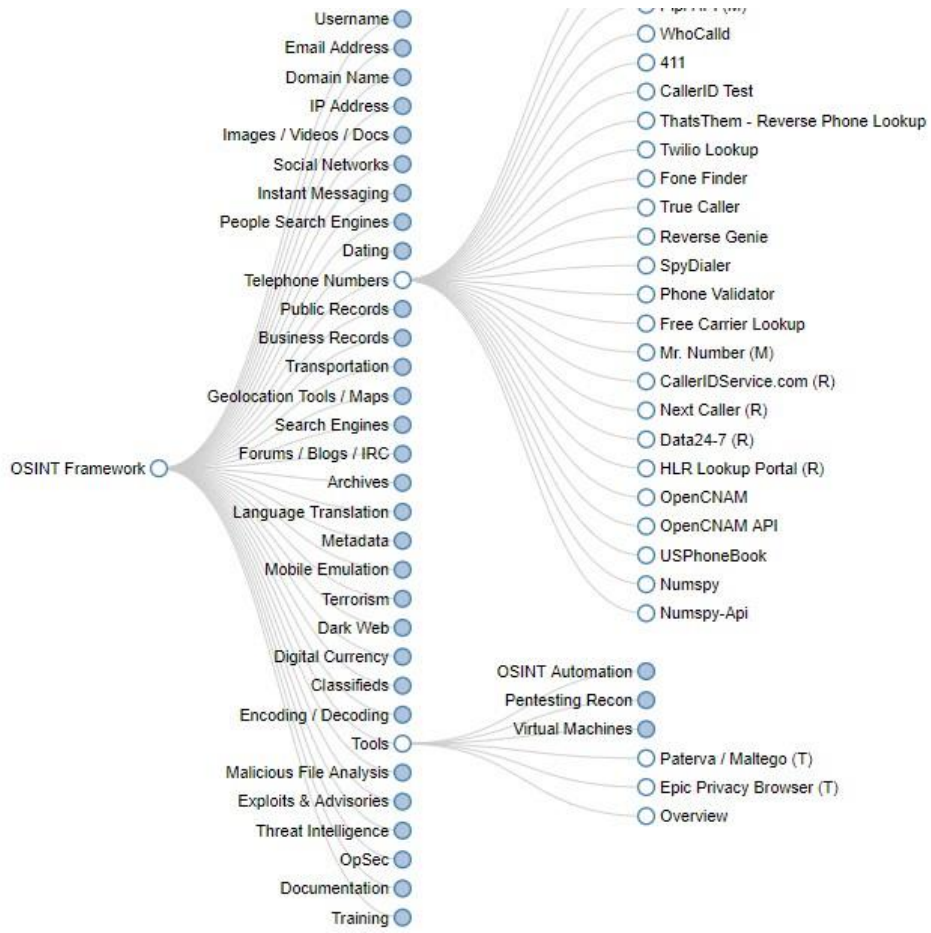



*Figura 75: Osint Framework*

# 8. Conclusiones

Como principal conclusión sobre OSINT y el mundo de la investigación y extracción de información en fuentes abiertas podemos destacar el tema de la privacidad en internet y el riesgo al que estamos constantemente expuestos.

Creo que es de los temas más importantes y con los que debemos tener mucho cuidado, ya que después de estar investigando sobre este sector de la ciberinteligencia y probando numerosas herramientas. Me he dado cuenta de que estamos realmente muy expuestos, subimos fotos, videos, posts y una gran cantidad de contenido en redes sociales, todo esto sumado a las filtraciones de datos de empresas o metadatos en documentos oficiales donde podemos encontrar DNI y una gran cantidad de información sensible.

Todo esto ha llevado además a la aparición de multitud de herramientas dedicadas a esta tarea de recolectar información sobre personas o empresas. Sin embargo, estas no suelen ser del todo precisas, ya que suelen estar desarrolladas por pequeños equipos o por aficionados del sector, que no pueden llegar a conseguir como resultado herramientas con una precisión del 100% y que muestran fallos.

Aun así, estas herramientas resultan de una gran utilidad, debido a que, aunque no nos muestren información 100% verídica y de utilidad, nosotros a partir de toda la información obtenida y haciendo un proceso de inteligencia podemos distinguir y clasificar la información según si validez.

Una vez mencionado todo esto es totalmente normal el enorme crecimiento que ha tenido la inteligencia en fuentes abiertas y seguirá creciendo mucho más con las nuevas generaciones, donde cada vez más personas y cada vez desde más pequeños empiezan a usar internet.

# 9. Trabajos futuros

OSINT ha tenido un crecimiento histórico en los últimos años y el mundo ha experimentado un crecimiento masivo de la cantidad de información en internet sobre cualquier cosa.

Sin embargo, no hemos observado un crecimiento tan grande a nivel de herramientas de OSINT, ya sean páginas web o programas o aplicaciones software.

Muchas de las herramientas que se usan actualmente para extraer información de las fuentes abiertas son herramientas de hace años que llevan funcionando desde los inicios del crecimiento de OSINT, y aunque estas herramientas se van actualizando, no llegan a aportarnos un nivel y calidad de información realmente buena. Además, todas estas herramientas, como ya se ha comentado, suelen estar desarrolladas por pequeños equipos o por personas aficionadas de manera semiprofesional.

Por ello se debería centrar el trabajo de OSINT en esta parte de la creación de nuevo software o dar cobertura y actualizar páginas y herramientas que están ahora en funcionamiento. Crear grupos o equipos para el desarrollo de estas, con el respaldo de grandes compañías.

Se deberá también adentrarse más en el mundo del análisis de fotografías o videos. Actualmente las herramientas o páginas web de las que disponemos de forma libre y gratuita para analizar fotografías o videos de alguien dejan mucho que desear. Es un campo que está muy atrasado y que si queremos tener una funcionalidad decente debemos invertir miles de euros en aplicaciones de pago. Para ello podrían combinarse distintas herramientas y haciendo uso de inteligencia artificial, para poder buscar una imagen, que se busquen coincidencias y una vez encontrado un nombre, username o correo, proceder a hacer una búsqueda y extracción de información en las fuentes abiertas.

# 10. Bibliografía

[1] Octavio Islas Carmona (20 de enero de 2011), *"Principales estadísticas sociodemográficas de internet y Facebook"*
https://dialnet.unirioja.es/descarga/articulo/5896188.pdf

[2] EUROPA PRESS (22 de abril de 2021), *"¿Sabes cuántas personas en el mundo usan internet?"*
https://www.excelsior.com.mx/hacker/sabes-cuantas-personas-en-el-mundo-usan-internet/1444773

[3] Alberto Fonte (8 de marzo de 2021), *"OSINT, ¿Qué es? ¿Para qué sirve?"*
https://derechodelared.com/osint/

[4] Julián Gutiérrez (10 de junio de 2019), *"¿Qué es OSINT? Usos y beneficios de aplicar este sistema para recopilar información"*
https://ciberpatrulla.com/que-es-osint/

[5] Cliento.mx (27 de octubre de 2020), *"Motores de búsqueda: qué son, como funcionan y cuantos tipos existen"*
https://blog.cliento.mx/que-son-motores-de-busqueda-como-funcionan-tipos

[6] Tibor Kopca (16 de abril de 2021), *"Google Dorks Como Interesantes y Buscar Como Un Hacker"*
https://www.ma-no.org/es/seguridad/google-dorks-como-encontrar-datos-interesantes-y-buscar-como-un-hacker

[7] Nicolás Raggi (29 de julio de 2021), *"Google hacking: averigua cuanta información sobre ti o tu empresa aparece en los resultados"*
https://www.welivesecurity.com/la-es/2021/07/29/google-hacking-averigua-que-informacion-sobre-ti-o-empresa-aparece-resultados/

[8] Antonio González, *"Google hacking & Dorks (46 ejemplos): cómo consigue un hacker contraseñas usando solo Google. Google puede ser tu peor enemigo."*
https://antoniogonzalezm.es/google-hacking-46-ejemplos-hacker-contrasenas-usando-google-enemigo-peor/

[9] Sede electrónica del catastro, *"Buscador de inmuebles"*
https://www1.sedecatastro.gob.es/Cartografia/mapa.aspx?buscar=S

[10] Abctelefonos, *"Buscar dirección por nombre o empresa en una localidad"*
https://www.abctelefonos.com/

[11] Permutador de correo, *"Generador de direcciones de correo a partir de nombre, apellidos, apodo y dominio"*
http://metricsparrow.com/toolkit/email-permutator/

[12] Carrot2, *"Buscador que ordena toda la información encontrada en categorías"*
https://search.carrot2.org/#/search/web

[13] Lucía Berlanga (4 de octubre de 2021), *"Que es Facebook, cómo funciona y qué te puede aportar esta red social"*
https://www.ciudadano2cero.com/que-es-facebook/

[14] Whopostedwhat, *"Recolectar información de id y posts de Facebook"*
https://whopostedwhat.com/

[15] Webempresa (1 de marzo de 2018), *"¿Qué es Twitter? ¿Cómo funciona? ¿Cómo puedo usarlo para mi organización?"*
https://www.webempresa.com/blog/que-es-twitter-como-funciona-2.html

[16] Hotmart (20 de marzo de 2021), *"¿Qué es Twitter, cómo funciona y para qué sirve esta red social?"*
https://blog.hotmart.com/es/que-es-twitter/

[17] Héctor Russo (23 de agosto de 2014), *"Foller.me, analiza gratis y a fondo cualquier cuenta de Twitter"*
https://geeksroom.com/2014/08/foller-me-analiza-gratis-y-a-fondo-cualquier-cuenta-de-twitter/88032/#:~:text=Foller.me%20es%20una%20aplicaci%C3%B3n,dejar%20su%20direcci%C3%B3n%20de%20email.

[18] Foller.me, *"Analizador de cuentas de Twitter"*
https://foller.me/

[19] Juan Antonio Soto (16 de agosto de 2020), *"¿Qué es Instagram y para qué sirve?"*
https://www.geeknetic.es/Instagram/que-es-y-para-que-sirve

[20] Yúbal Fernández (7 de abril de 2021), *"Qué es TikTok, de donde viene y que ofrece la red social de videos"*
https://www.xataka.com/basics/que-tiktok-donde-viene-que-ofrece-red-social-videos

[21] Berto López (3 de septiembre de 2021), *"Qué es LinkedIn, para qué sirve y cómo funciona"*
https://www.ciudadano2cero.com/linkedin-que-es-como-funciona/

[22] Esgeeks, *"IKy Proyect: recopilar información correo electrónico gui"*
https://esgeeks.com/iky-osint-recopilar-informacion-email/

[23] Maltego, *"Buscador de información OSINT"*
https://www.maltego.com/

[24] FOCA, *"Buscador de metadatos en documentos de internet"*
https://github.com/ElevenPaths/FOCA

[25] Osint Framework, *"Framework con herramientas de OSINT"*
https://osintframework.com/

Universidad de Alcalá
Escuela Politécnica Superior

ESCUELA POLITECNICA
SUPERIOR

Universidad
de Alcalá