

# OSINT Tools and Techniques

# 6

---

## INFORMATION IN THIS CHAPTER

- OSINT Tools
- Geolocation
- Information Harvesting
- Shodan
- Search Diggity
- Recon-ng
- Yahoo Pipes
- Maltego

---

## INTRODUCTION

In the previous chapters we learned about the basics of the internet and effective ways to search it. We went to great depths of searching social media to unconventional search engines and further learned about effective techniques to use regular search engines. In this chapter we will move a step further and will discuss about some of the automated tools and web-based services which are used frequently to perform reconnaissance by professionals of various intelligence-related domains specially information security. We will start from the installation part to understanding their interface and will further learn about their functionality and usage. Some of these tools provide a rich graphic interface (GUI) and some of them are command line based (CLI), but don't judge them by their interface but by their functionality and relevance in our field of work.

Before moving any further we must install the dependencies for these tools so that we don't have to face any issues during their installation and usage. The packages we need are

- Java latest version
- Python 2.7
- Microsoft .NET Framework v4

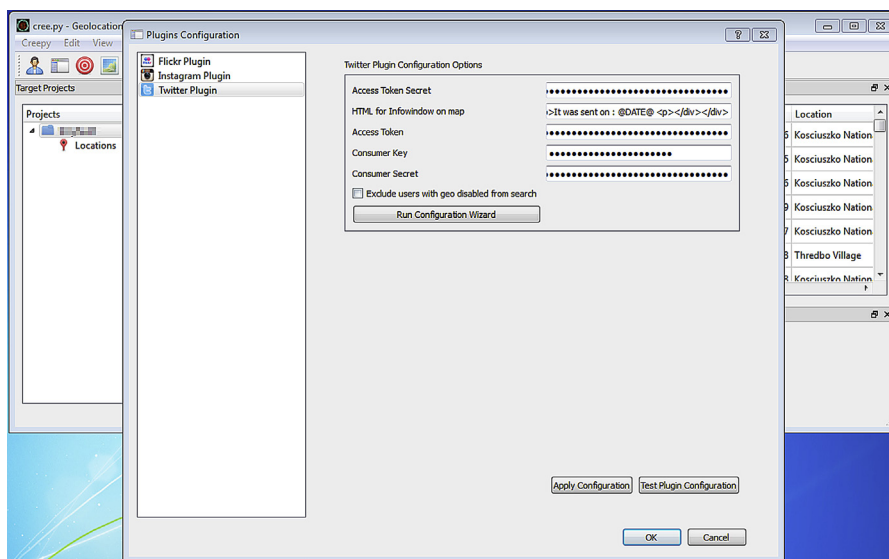
We simply need to download the relevant package depending upon our system configuration and we are good to go.

## CREEPY

Most of us are addicted to social networks, and image sharing is one of the most utilized features of these platforms. But sometimes when we share these pictures it's not just the image that we are sharing but might also the exact location where that picture was taken.

Creepy is a Python application which can extract out this information and display the geolocation on a map. Currently Creepy supports search for Twitter, Flickr, and Instagram. It extracts the geolocation based on EXIF information stored in images, geolocation information available through application programming interface (API), and some other techniques.

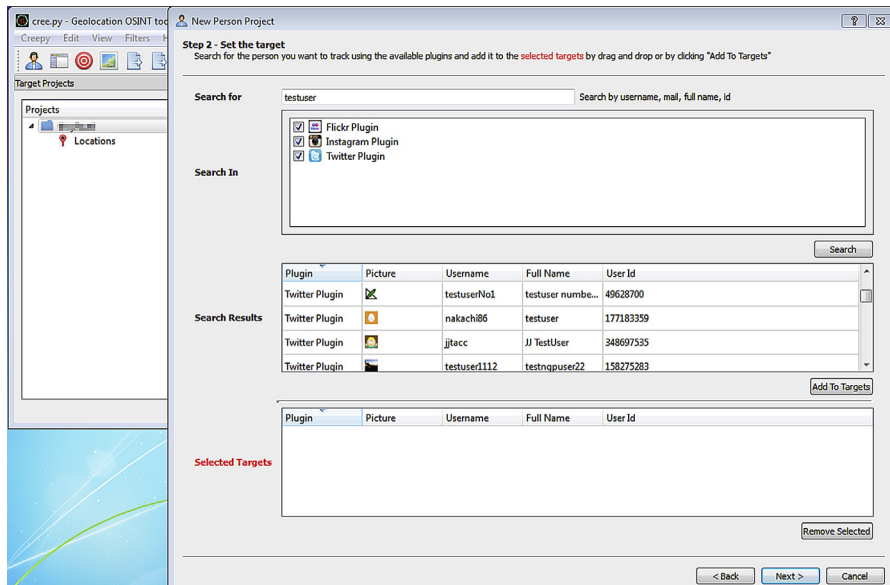
It can be downloaded from <http://ilektrojohn.github.io/creepy/>. We simply need to select the version according to our platform and install it. The next phase after installation of Creepy is to configure the plugins that are available in it, for which we simply need to click on the Plug-in Configuration button present under the edit tab. Here we can select the plugins and using their individual configuration wizard configure them accordingly. Once the configuration is done we can check whether it is working properly or not using the Test Plugin Configuration button.



**FIGURE 6.1**

Configure Creepy.

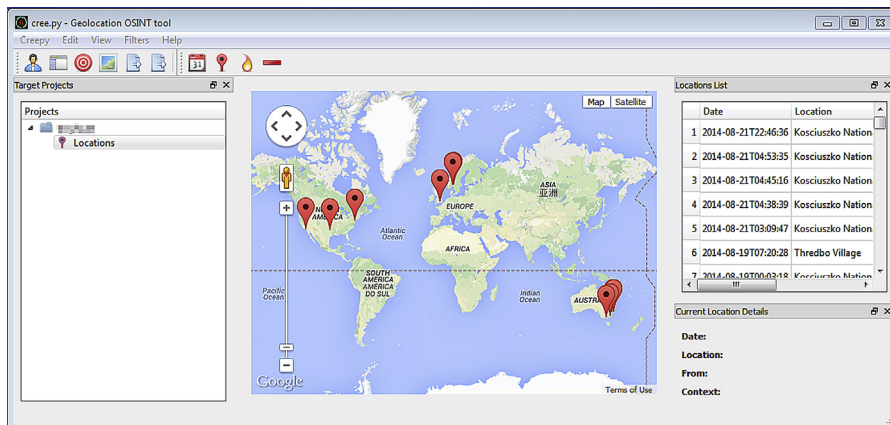
After the configuration phase is done, we can start a new project by clicking on the person icon on the top bar. Here we can name the project and search for people on different portals. From the search results we can select the person of interest and include him/her in the target list and finish the wizard. After this our project will be displayed under the project bar at the right-hand side.



**FIGURE 6.2**

Search users.

Now we simply need to select our project and click on the target icon or right click on the project and click Analyze Current Project. After this Creepy will start the analysis, which will take some time. Once the analysis is complete, Creepy will display the results on the map.

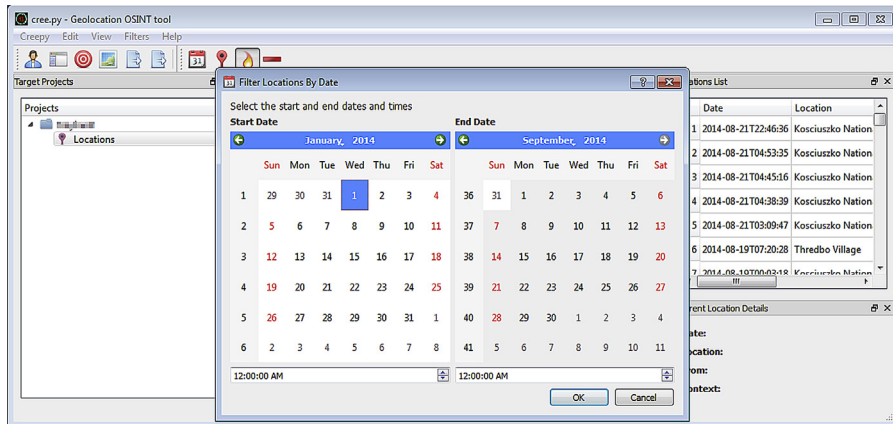


**FIGURE 6.3**

Creepy results.

Now we can see the results in which the map is populated with the markers according the identified geolocation. Now Creepy further allows us to narrow down these results based on various filters.

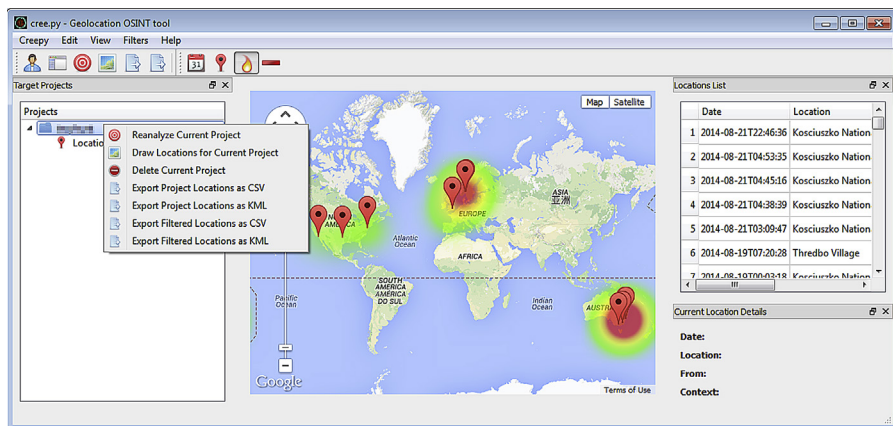
Clicking on the calendar button allows us to filter the results based on a time period. We can also filter the results based upon area, which we can define in the form of radius in kilometers from a point of our choice. We can also see the results in the form of a heat map instead of the markers. The negative sign (–) present at the end can be used to remove all the filters imposed on the results.



**FIGURE 6.4**  
Applying filter.

The results that we get from Creepy can also be downloaded in the form of CSV file and also as KML, which can be used to display the markers in another map.

Creepy can be used for the information-gathering phase during a pentest (penetration test) and also as a proof-of-concept tool to demonstrate to users what information they are revealing about themselves.



**FIGURE 6.5**  
Download Creepy results.

## THEHARVESTER

TheHarvester is an open source intelligence tool (OSINT) for obtaining e-mail addresses, employee name, open ports, subdomains, hosts banners, etc. from public sources such as search engines like Google, Bing and other sites such as LinkedIn. It's a simple Python tool which is easy to use and contains different information-gathering functions. Being a Python tool it's quite understandable that to use this tool we must have Python installed in our system. This tool is created by Christian Martorella and one of the simple, popular, and widely used tools in terms of information gathering.

TheHarvester can be found here: <http://www.edge-security.com/theharvester.php>

Generally we need to input a domain name or company name to collect relevant information such as email addresses, subdomains, or the other details mentioned in the above paragraph. But we can use keywords also to collect related information.

We can specify our search, such as from which particular public source we want to use for the information gathering. There are lots of public source that Harvester use for information gathering but before moving to that let's understand how to use Harvester.

EX: theharvester -d **example.com** -l 500 -b Google

-d=Generally, domain name or company name

-l=Number of result limits to work with

-b=Specifying the data source such as in the above command its Google, but apart from that we can use LinkedIn and all (to use all the available public sources) as a source also to collect information.

```

root@kali: ~
File Edit View Search Terminal Help
root@kali:~# theharvester -d example.com -l 500 -b google
*****
*
* THE HARVESTER
*
* TheHarvester Ver. 2.2a
* Coded by Christian Martorella
* Edge-Security Research
* cmartorella@edge-security.com
*****

[-] Searching in Google:
    Searching 0 results...
    Searching 100 results...
    Searching 200 results...
    Searching 300 results...
    Searching 400 results...
    Searching 500 results...

[+] Emails found:
-----
st@example.com
del@example.com
com@example.com
au@example.com
tor@example.com
she@example.com
col@example.com

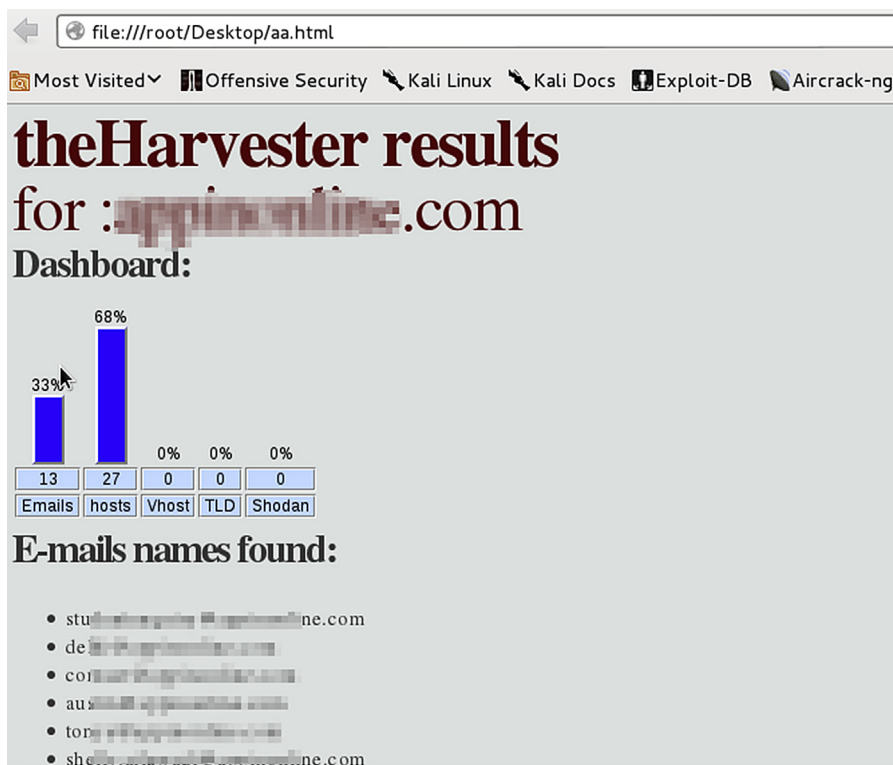
```

FIGURE 6.6

TheHarvester in action.

Apart from the above mentioned one harvester also has other options to specify, such as:

- s = to start with a particular result number (the default value is 0)
- v = to get virtual hosts by verifying hostnames via DNS resolution
- f = for saving the data. (formats available either html or xml)
- n = to perform DNS resolve query for all the discovered ranges
- c = to perform DNS bruteforce for all domain names
- t = to perform a DNS TLD expansion discovery
- e = to use a specific DNS server
- l = To limit the number of result to work with
- h = to use Shodan database to query discovered hosts.



**FIGURE 6.7**

TheHarvester HTML results.

The sources it uses are Google, Google profiles, Bing, pretty good privacy (PGP) servers, LinkedIn, Jigsaw, Shodan, Yandex, name servers, people123, and

Exalead, Google, Yandex, Bing, and Exalead are search engines that are used in backend as a source, while Shodan is also a search engine but not the conventional one and we already discussed a bit about it earlier and we will discuss in detail about the same in this chapter later. PGP servers are like key servers used for data security and those are also a good source to collect e-mail details. The people123 is for searching for a particular person and Jigsaw is the cloud-based solution for lead generation and other sales stuffs. From different sources harvester collects different information such as for e-mail harvesting it uses Google, Bing, PGP servers, and sometimes Exalead and run their specific queries in the background to get the desired result. Similarly for subdomains or host names it uses again Google, Bing, Yandex, Exalead, PGP servers, and Exalead. And finally for the list for employee names it uses LinkedIn, Google profiles, people123, and Jigsaw as a main source.

This is how theHarvester harvests all the information and gives us the desired result as per our query. So craft your query wisely to harvest all the required information.

---

## SHODAN

We have previously discussed about Shodan briefly in Chapter 4, but this unique search engine deserves much more than a paragraph to discuss its usage and impact. As discussed earlier Shodan is a computer search engine. The internet consists of various different types of devices connected online and available publicly. Most of these devices have a banner, which they send as a response to the application request send by a client. Many if not most of these banners contains information which can be called sensitive in nature, such as server version, device type, authentication mode, etc. Shodan allows us to search such devices over internet and also provides filters to narrow down the results.

It is highly recommended to create an account to utilize this great tool, as it removes some of the restrictions imposed on the free usage. So after logging into the application we will simply go to the dashboard at <http://www.shodanhq.com/home>. Here we can see some the recent searches as well as popular searches made on this platform. This page also shows a quick reference to the filters that we can use. Moving on let's see more popular searches listed under the URL <http://www.shodanhq.com/browse>. Here we can see there are various different search queries which look quite interesting, such as webcam, default password, SCADA, etc. Clicking on one of these directly takes us to the result page and lists details of machines on the internet with that specific keyword. The page <http://www.shodanhq.com/help/filters> shows the list of all the filters that we can use in Shodan to perform a more focused search, such as country, hostname, port, etc., including the usual filters "+,""-," and "l."

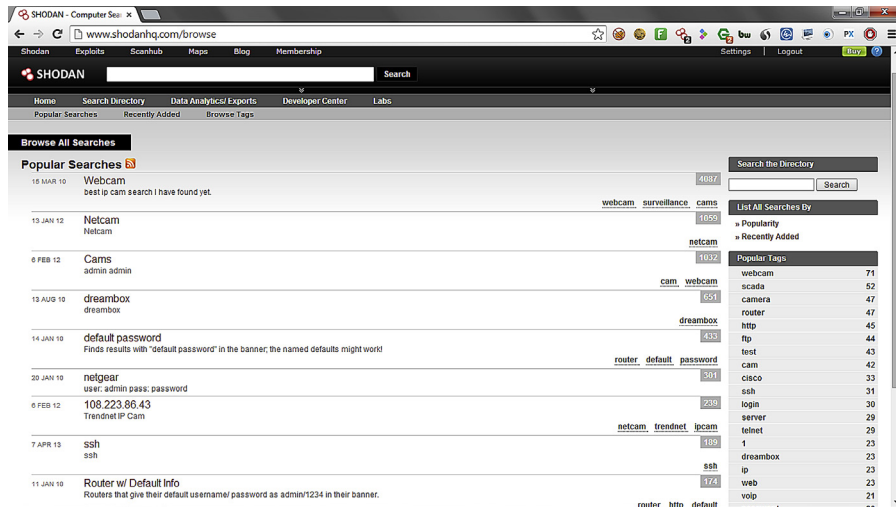


FIGURE 6.8

Shodan popular searches.

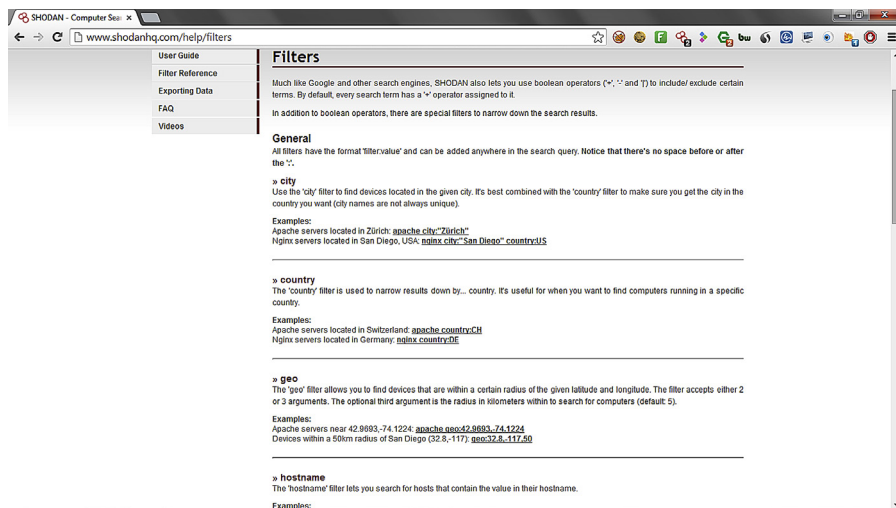


FIGURE 6.9

Shodan filters.

Let's perform a simple search on Shodan for the keyword "webcam." Shodan has simply found more than 15,000 results for this keyword; though we cannot view all the results under the free package, yet what we get is enough to understand its reach and availability of such devices on the internet. Some of these might be protected by some kind of authentication mechanism such as username and password, but some might be publicly accessible without any such mechanism. We can simply find out by opening



their listed IP address in our browsers (Warning: It might be illegal to do so depending upon the laws of the country, etc.). We can further narrow down these results to a country by using the “country” filter. So our new query is “webcams country:us” which gives us a list of webcams in the United States of America.

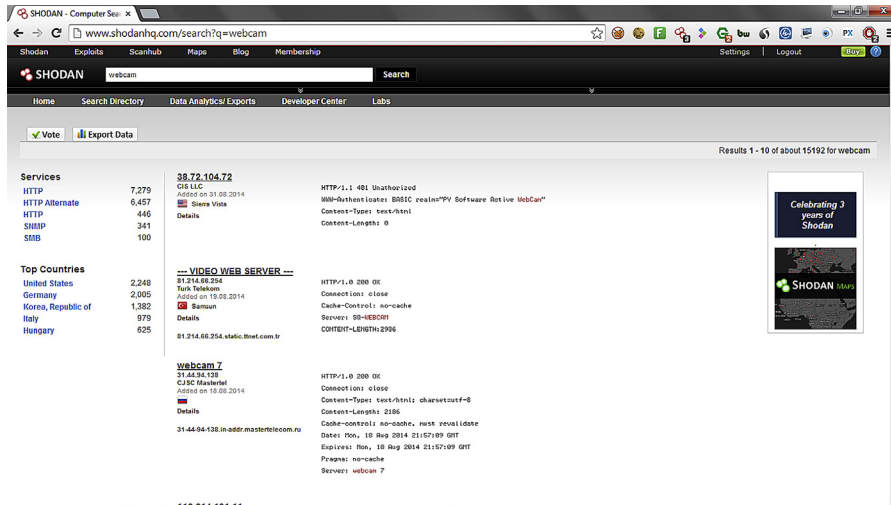


FIGURE 6.10

Shodan results for query “webcam”

To get a list of machines with file transfer protocol (FTP) service, residing in India, we can use the query “port:21 country:in”. We can also perform search for specific IP address or range of it using the filter “net.” Shodan is providing a great deal of relevant information and its application is only limited by the creativity of its users.

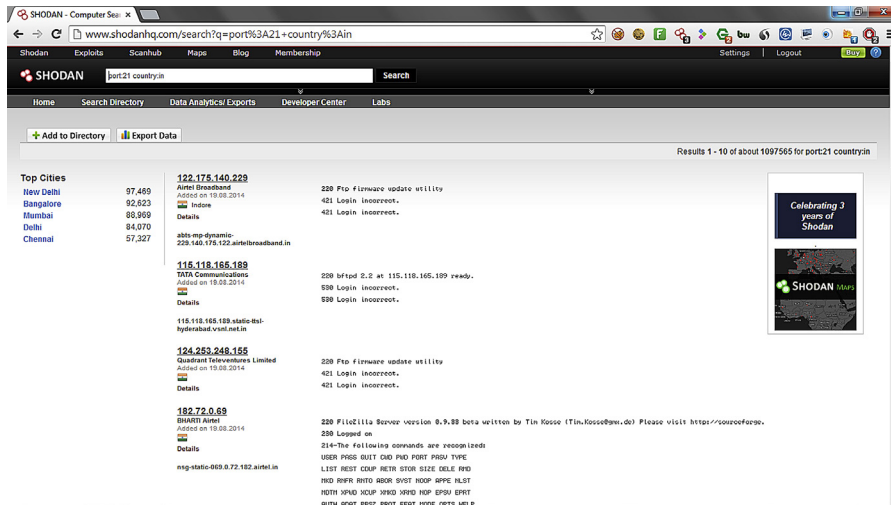


FIGURE 6.11

Shodan results for query “port:21 country:in.”

Apart from this Shodan also offers an API to integrate its data into our own application. There are also some other services provided by it at a price and are worth a try for anyone working in the information security domain. Recently there has been a lot of development in Shodan and its associated services which makes this product a must try for information security enthusiasts.

## SEARCH DIGGITY

In the last chapter we learned a lot about using advanced search features of various search engines and also briefly discussed about the term “Google Hacking.” To perform such functions we need to have the list of operations that we can use and will have to type each query to see if anything is vulnerable, but what if there was a tool which has a database of such queries and we can simply run it. Here enters the Search Diggity. Search Diggity is tool by Bishop Fox which has a huge set of options and a large database of queries for various search engines which allow us to gather compromising information related to our target. It can be downloaded from <http://www.bishopfox.com/resources/tools/google-hacking-diggity/attack-tools/>. The basic requirement for its installation is Microsoft .NET framework v4

Once we have downloaded and installed the application, the things we need are the search ids and API keys. These search ids/API keys are required so that we can perform more number of searcher without too many restrictions. We can find how to get and use these keys in the contents section under the Help tab and also from a some simple Google searches. Once all the keys (Google, Bing, Shodan, etc.) are at their place we can move forward with the usage of the tool.

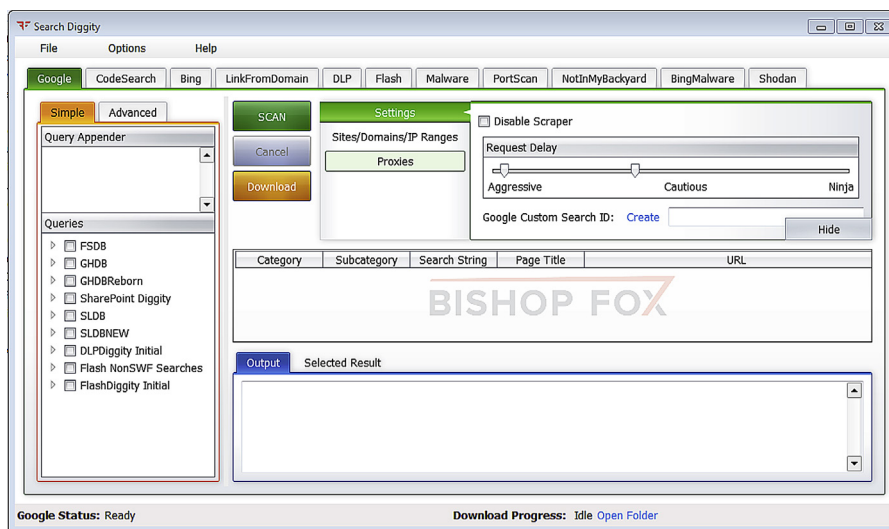


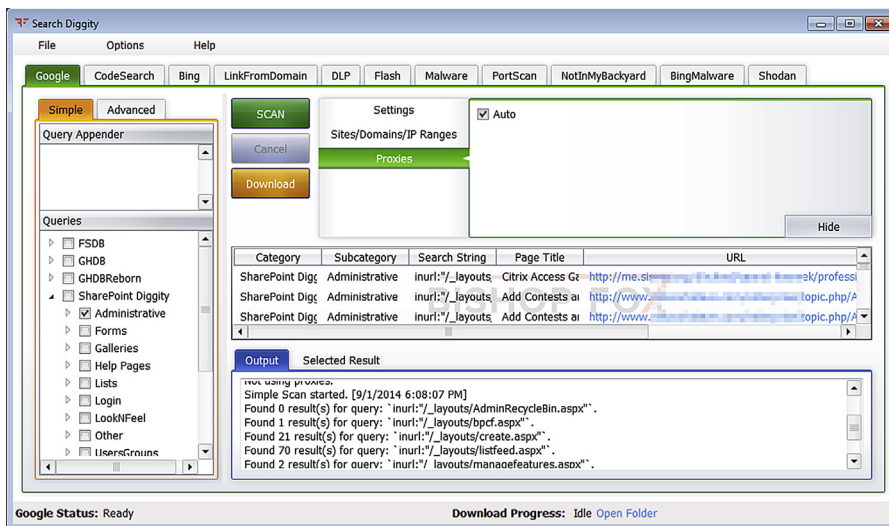
FIGURE 6.12

Search Diggity interface.

There are many tabs in the tool such as Google, Bing, DLP, Flash, Shodan etc. Each of these tabs provides specialized functions to perform targeted search to identify information which can be critical from an information security point of view.

To use the tool we simply need to select one of the tabs at the top and further select the type of queries that we want to use. We can also specify the domain that we want to target and simply perform the scan. Depending upon what is available online the tool will provide us the results for various different queries related to the query type we have selected. It is highly recommended to select only the query types that we are really interested into, as it will help us to narrow down the total number of queries. The queries present are categorized properly to identify and make choice accordingly.

Let's use the queries to identify SharePoint Administrative pages. For this we simply need to select the Google tab and from the left-hand menu, check the Administrative checkbox under SharePoint Diggity, and run the scan.



**FIGURE 6.13**

Search Diggity scan—Google tab.

To make this scan more targeted we can specify a list of targets under the option Sites/Domains/IP Ranges. As soon as we start the scan we can see the results coming up with various information like category, page title, URL, etc. Similarly we can also use the Bing scan which has its own set of search queries.

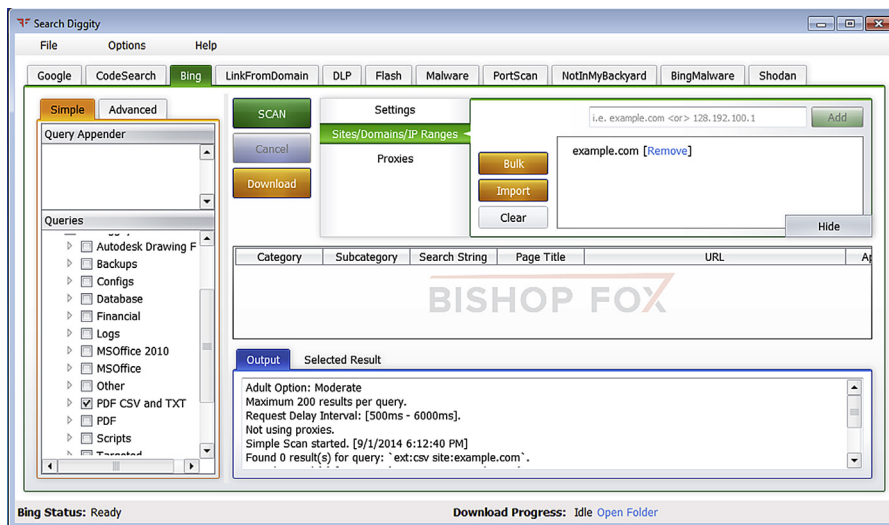


FIGURE 6.14

Search Diggity scan—Bing tab.

One of the interesting options is NotInMyBackyard which allows us to specify various options such as locations, filetypes, and keyword to get interesting data. Similarly we can also access information from Shodan using its API key.

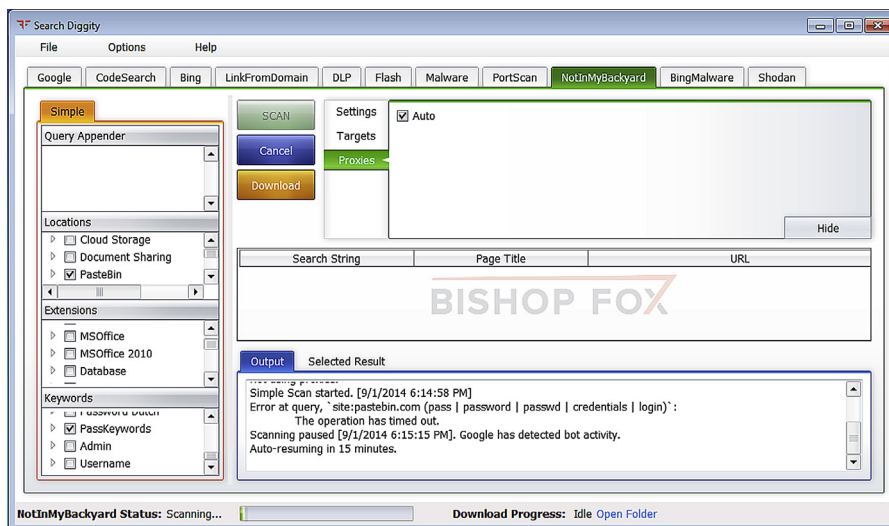
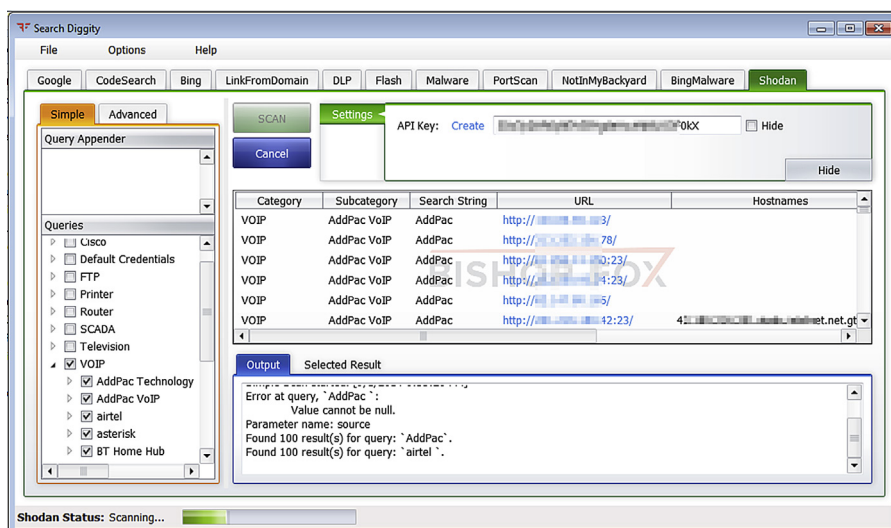


FIGURE 6.15

Search Diggity—NotInMyBackyard.



**FIGURE 6.16**

Search Diggity—Shodan scan.

## Recon-ng

There are many tools for reconnaissance but a special mention should be given to Recon-ng. This is an open source tool written in Python majorly by Tim Tomes (@Lanmaster53). There are many other researchers, coders, and developers who have contributed to this project. This project is one of its kind in terms of complete OSINT framework. The authors might have different opinion on my previous statement but still this framework helps all the OSINT enthusiast to perform various stages of reconnaissance in automated way.

It mainly focus on web-based open-source reconnaissance and provides its users with unique independent modules, elaborated and much required command based help, database interaction and command completion facility to perform reconnaissance deeply and fast paced. Apart from that it's made in a fashion that if a newbie into the field of security wants to contribute to it, he/she can easily do it with a little Python knowledge. It's just possible only because of well-structured modules, fully fledged documentation, and the uses of only-native Python functions that a new user or contributor will not face problem to download and install third party modules of Python for a specific task.

The tool can be downloaded from: <https://bitbucket.org/LaNMaSteR53/recon-ng>  
 The user guide: [https://bitbucket.org/LaNMaSteR53/recon-ng/wiki/Usage\\_Guide](https://bitbucket.org/LaNMaSteR53/recon-ng/wiki/Usage_Guide)  
 The development guide: [https://bitbucket.org/LaNMaSteR53/recon-ng/wiki/Development\\_Guide](https://bitbucket.org/LaNMaSteR53/recon-ng/wiki/Development_Guide)

Apart from the perspective of developer or contributor the author also focused on the ease of use for the users. The framework looks quite same as Metasploit which is a quite popular tool for exploitation in information security community. If you are from information security community or you have prior experience of using Metasploit, it's quite the same to use Recon-ng.

Recon-ng is quite easy to install. And to run the same we just need Python 2.7.x installed in our system. Just call recon-ng.py file from a terminal and you will get a fancy banner of the tool with credits and all along with that a recon-ng prompt.

To check all the available commands we can use command help. It will show all the available commands

```
> help
```

add	Adds records to the database
back	Exits the current context
del	Deletes records from the database
exit	Exits the framework
help	Displays this menu
keys	Manages framework API keys
load	Loads specified module
pdb	Starts a Python Debugger session
query	Queries the database
record	Records commands to a resource file
reload	Reloads all modules
resource	Executes commands from a resource file
search	Searches available modules
set	Sets module options
shell	Executes shell commands
show	Shows various framework items
spool	Spools output to a file
unset	Unsets module options
use	Loads specified module
workspaces	Manages workspaces

Here in this framework some fine features are provided such as workspaces. It consists of different settings, database, etc., and a self-independent place for a single project.

To know more about workspaces, we can use the command

```
> help workspaces
```

This command is used to manage workspaces such as providing user freedom to list down, add, select, delete workspaces. If a user does not set a workspace externally, then he/she will be under default workspace. If we want to check in which workspace we are exactly then the command is

```
> show workspaces
```

```
+-----+
| Workspaces |
+-----+
| default   |
+-----+
```

And we will get something similar to this showing that we are under default workspace.

Let's say we want to change the workspace to something that we want, let's say osint, then the command would be

```
> workspaces add osint
```

The prompt itself shows the workspace so the default prompt we will get in fresh installation is

```
[recon-ng] [default] >
```

After the above command the prompt will change in to

```
[recon-ng] [osint] >
```

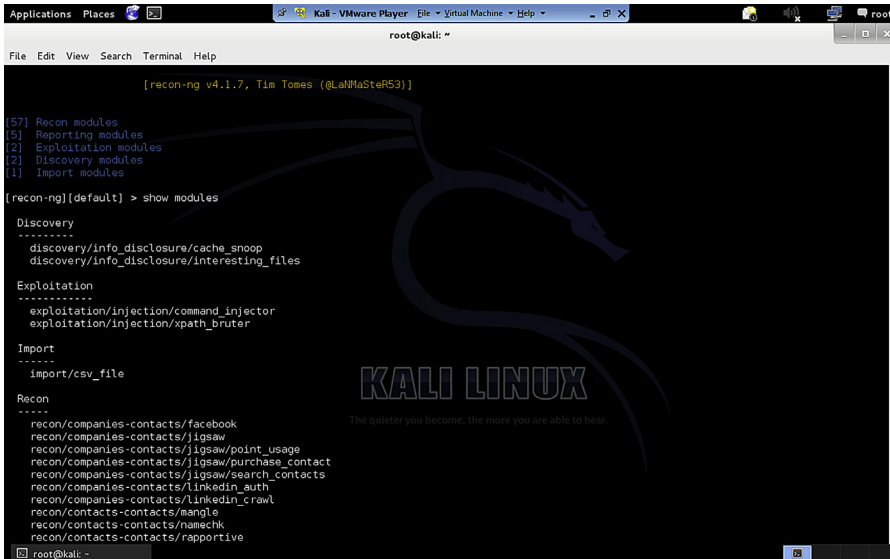
Now it's time to explore the commands and its capabilities. If you are using this tool for the first time the most needed command after "help" is "show."

```
[recon-ng] [osint] > show
```

Using this command we can see available details of banner, companies, contacts, credentials, dashboard, domains, hosts, keys, leaks, locations, modules, netblocks, options, ports, pushpins, schema, vulnerabilities, and workspaces details but here we want to explore the modules section to see what all are possibilities available.

Basically recon-ng consists of five different sections of modules.

1. Discovery
2. Exploitation
3. Import
4. Recon
5. Reporting



```

root@kali: ~
File Edit View Search Terminal Help
[recon-ng v4.1.7, Tim Tomes (@LulMaSteR53)]

[57] Recon modules
[5] Reporting modules
[2] Exploitation modules
[2] Discovery modules
[1] Import modules

[recon-ng][default] > show modules

Discovery
-----
discovery/info_disclosure/cache_snoop
discovery/info_disclosure/interesting_files

Exploitation
-----
exploitation/injection/command_injector
exploitation/injection/xpath_bruter

Import
-----
import/csv_file

Recon
-----
recon/companies-contacts/facebook
recon/companies-contacts/jigsaw
recon/companies-contacts/jigsaw/point_usage
recon/companies-contacts/jigsaw/purchase_contact
recon/companies-contacts/jigsaw/search_contacts
recon/companies-contacts/linkedin_auth
recon/companies-contacts/linkedin_crawl
recon/contacts-contacts/mangle
recon/contacts-contacts/namechk
recon/contacts-contacts/rapportive

root@kali: ~

```

FIGURE 6.17

Recon-ng modules.

And by using following commands we will be able to see more details as well as off the available options about these five sections

```
[recon-ng] [osint] > show modules
```

such as under-discovery interesting files, under-exploitation command injection, under-import CSV files, under-recon company contacts, credentials, host details, location information and many more and last but not the least under-reporting CSV, HTML, XML, etc.

Now we can use these modules based on our requirements. To use any of these modules, first, we need to load that module using following command but before that we must know that this framework has a unique capability to load a module by auto completing it or if more modules are available with a single keyword then giving all the module list. Let's say we want to check the pwnedlist and we are so lazy to type the absolute command. Nothing to worry just do as shown below

```
[recon-ng] [osint] > load pwnedlist
```

Now recon-ng will check whether this string is associated with a single module or multiple modules. If it is associated with a single module then it will load that or else it will give the user all the available modules that contain this keyword.



As in our case pwnedlist keyword is associated with multiple functions, thus it will show all. Let's say we want to use the module recon/contacts-creds/pwnedlist the command we will use is

```
[recon-ng] [osint] > load recon/contacts-creds/pwnedlist
```

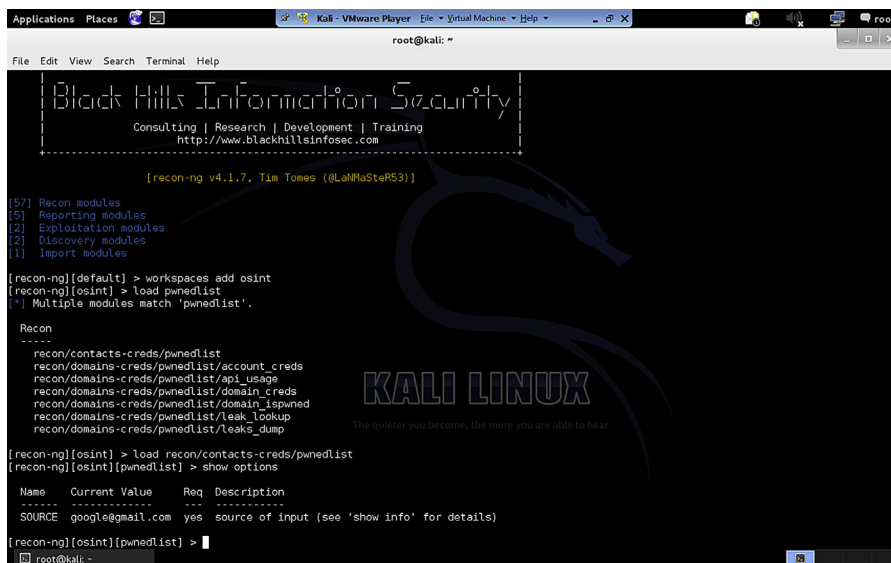
Once a module get loaded, for easy understanding it gets added in the prompt so the prompt will change to

```
[recon-ng] [osint] [pwnedlist] >
```

As we wanted to use this module first we need to check the options available. To check that the command is as follows

```
[recon-ng] [osint] [pwnedlist] > show options
```

This command will show all the required details in a tabular manner such as name of the required field, its current status or value, whether it is mandatory field or not, description in a bit.



```

Applications  Places  [Kali - VMware Player] File Virtual Machine Help  root
root@kali: ~
File Edit View Search Terminal Help
-----
Black Hills Information Security
Consulting | Research | Development | Training
http://www.blackhillsinfosec.com
-----
[recon-ng v4.1.7, Tim Tomes (@LaMaSteR53)]

[57] Recon modules
[5] Reporting modules
[2] Exploitation modules
[2] Discovery modules
[1] Import modules

[recon-ng][default] > workspaces add osint
[recon-ng][osint] > load pwnedlist
(*) Multiple modules match 'pwnedlist'.

Recon
-----
recon/contacts-creds/pwnedlist
recon/domains-creds/pwnedlist/account_creds
recon/domains-creds/pwnedlist/api_usage
recon/domains-creds/pwnedlist/domain_creds
recon/domains-creds/pwnedlist/domain_ispwned
recon/domains-creds/pwnedlist/leak_lookup
recon/domains-creds/pwnedlist/leaks_dump

[recon-ng][osint] > load recon/contacts-creds/pwnedlist
[recon-ng][osint][pwnedlist] > show options

Name Current Value Req Description
-----
SOURCE google@gmail.com yes source of input (see 'show info' for details)

[recon-ng][osint][pwnedlist] >
root@kali: ~

```

**FIGURE 6.18**

Recon-ng module options.

If we are still in confusion then this framework has other command to elaborate in more detailed fashion about a module:

```
[recon-ng] [osint] [pwnedlist] > show info
```



Voilà! The above e-mail id has been pwned somewhere. If we want to use some other modules we can simply use the “load” command along with the module name to load and use the same.

This is how we can easily use this recon-ng. The commands and approaches will remain quite same. First look for the modules. Choose the required module, load it, check for its options, provide values to the required fields, and then run. If required repeat the same process to extend the reconnaissance.

Now let’s discuss some of the scenarios and the modules that can be handy for the same.

## CASE 1

If we are into sales and desperately wanted to collect database to gather prospective clients then there are certain modules available here that will be pretty helpful. If we want to gather these information from social networking sites, LinkedIn is the only place where we can get exact names and other details as compared to other sites which generally consists of fancy aliases. And if we are in core sales then we might have heard of portals like Sales Force or Jigsaw, where we can get certain details either by free or by paying reasonable amount of money. And mostly nowadays in IT sector sales teams focus less on cold calling and more on spreading details on e-mail. So getting valid e-mails from a target organization is always like half work done for sales team. So here we will discuss the sources available to get these information and its associated modules in recon-ng.

Available modules:

```
recon/companies-contacts/facebook
recon/companies-contacts/jigsaw
recon/companies-contacts/linkedin_auth
```

These are some of the modules but not all that can be helpful to gather information such as name, position, address, etc.

But e-mail addresses are the key to contact. So let’s look into some options to collect e-mail addresses. We can collect some e-mail id details from Whois database. Search engines also sometimes play a vital role in collecting e-mail address, using PGP servers.

Available modules:

```
recon/domains-contacts/pgp_search
recon/domains-contacts/whois_pocs
```

## CASE 2

Physical tracking. The use of smart phones intentionally or unintentionally allowed users to add their geolocation with data that they upload to different public sites

such as YouTube, Picasa, etc. In that case we can collect information by the help of geotagged media. This can be used for behavioral analysis, understanding a person's likes and dislikes, etc.

Available modules:

```
recon/locations-pushpins/flickr
recon/locations-pushpins/picasa
recon/locations-pushpins/shodan
recon/locations-pushpins/twitter
recon/locations-pushpins/youtube
```

### CASE 3

If some organization or person wants to check whether he/she or any of a company's e-mail id has been hacked, then there are certain modules that can be helpful. Similar to what we already discussed above, i.e., pwnedlist, there are other modules that can give similar results:

```
recon/contacts-creds/pwnedlist
recon/contacts-creds/havebeenpwned
recon/contacts-creds/should_change_password
```

### CASE 4

For penetration testers it is also like hidden treasure because they can perform penetration testing without sending a single packet from their environment. The first approach to do any penetration testing is information gathering. Let's say we want to perform a web application penetration testing then the first thing we want to enumerate is what technology or server the site is running on. So that we can manually search later for the publicly available exploits to exploit the same. In this case reconng has a module to find the technology details for us.

Available module:

```
recon/domains-contacts/builtwith
```

Now after getting these details, generally, we look into the vulnerabilities available in the net associated with that technology. But we can also look at the vulnerabilities associated with that domain. And it is possible by the use of punkspider module. Punkspider uses a web scanner to scan the entire web and collect detailed vulnerabilities and store it in its database, which can be used to directly search for the available exposed vulnerabilities in a site.

Available modules:

```
recon/domains-vulnerabilities/punkspider
recon/domains-vulnerabilities/xssed
```



As we all are OSINT enthusiasts, the only thing that matters to us is valid required information. There are information available in different parts of the web. And there are different sources to get information regularly. The problem is that how to differentiate the information we want from the numerous information provided by a particular source. If we need to filter the required information manually from a set of information then it requires a lot of manual effort. So to ease the process this application will help us a lot.

Requirements:

- A web browser
- Internet connectivity
- Yahoo Id

As it's a web application we can access it from anywhere, and the lesser dependency make it more usable along with its user friendly GUI. We can access the application from below mentioned URL.

<https://pipes.yahoo.com/>

Visit this URL, login with your yahoo id and we are all set to use this application. Another major plus point of this application is its well-formed documentation. Apart from that we can find links to different tutorials (text as well as video) in the application site itself describing how to start and other advance stuffs. Along with that for reference purpose there are also links to popular pipes available. Let's create our own pipe.

To create an own pipe we need to click on Create pipe button on the application. It will redirect to <http://pipes.yahoo.com/pipes/pipe.edit>

In the right top corner we can find tabs like new, save, and properties. By default there is no necessity to do anything with these tabs. As we are about to start creating a new pipe, the things to be noted are that in the left side of the application we will find different tabs and subtabs such as sources, user inputs, operators, URL, etc.

These are the tabs from where we can drag the modules to design the pipe. Basically a pipe starts with a source or multiple sources. Then we need to create some filters as per our requirements using operators, date, location, etc., and then finally need to add an output to get the desired filtered information.

So to start with lets drag a source from the source sub tab, there are different options available such as Fetch CSV, Fetch data, Fetch feed, etc. Let's fetch from feeds as it's a very good source of information. Drag the Fetch Feed sub tab to the center of the application. When we drag anything to the center it will generate an output box for us, where it will ask us to add the feed URL. Add any feed URL in my case I am using <http://feedads.bbc.co.uk/news/rss.xml?edition=int>.

For the demo purpose I'll show only single source example but we can also add multiple sources for one pipe. Now it's very important to create a proper filter, which will give us the proper output. Now drag filter sub tab from Operators tab. By default we will see "block," "all," "contains" keywords there and some blank spaces to fill. Change that to "Permit," keep the "all" as it is and add item description in first blank space following with "contains" following with US. So our filter will provide us data which contains only keyword "US" in its item description.

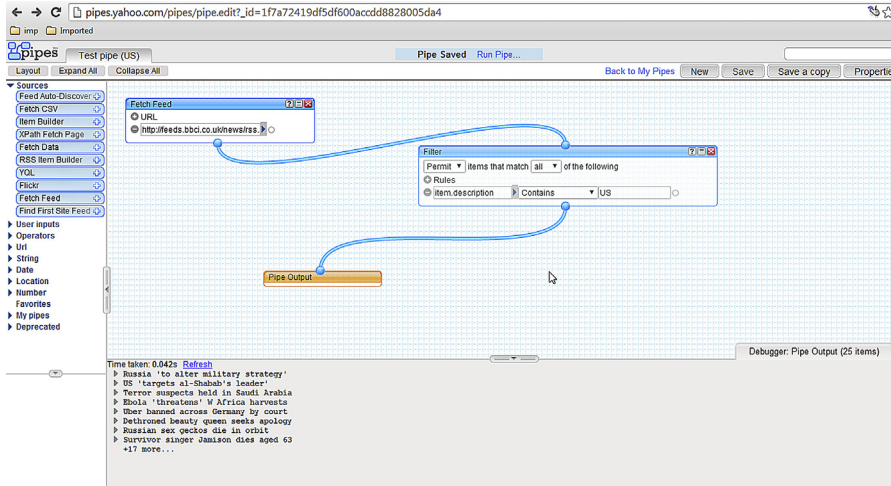


FIGURE 6.22

Creating a Yahoo Pipe.

Now connect all the pipe points from sources box (Fetch Feed) to Filters box and from Filter box to Pipe Output box. First save the pipes and then run the pipes to get the output in a new tab.

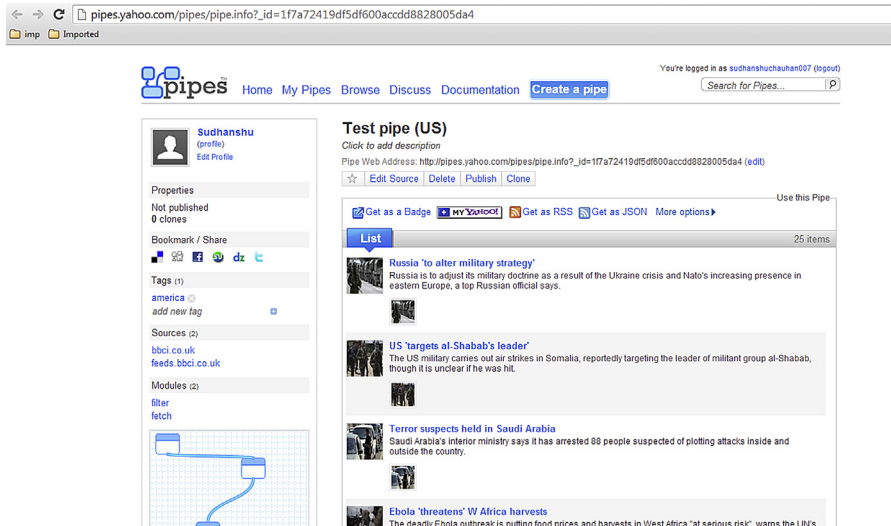


FIGURE 6.23

Yahoo Pipe result.

We can use it in many other scenarios like collecting images of a specific person from flicker, filtering information by URL, date- and location-based and many others. Explore this to create as customized pipes as possible. This tool provides freedom to create pipes way beyond our imagination.

---

## MALTEGO

There are many OSINT tools available in the market, but one tool stands out because of its unique capabilities, Maltego.

Maltego is an OSINT application which provides a platform to not only extract data but also represent that data in a format which is easy to understand as well as analyze. It's a one stop shop for most of the recon requirements during a pentest, what adds to its already great functionalities is the feature which allows users to create custom add-ons for the platform (which we will discuss later) depending upon the requirement.

Currently Maltego is available in two versions: commercial and community. Commercial version is paid and we need a license key for it. The community version however is free and we only need to register at the site of Pateva (creator of Maltego) at this page: <https://www.paterva.com/web6/community/maltego/index.php>. Though community version has some limitations in comparison to commercial version, like limited amount of data extraction, no user support, etc., still it is good enough to feel the power of this great tool. During this chapter we will be using the community version for the demo purpose.

Let's see how this tool works and what we can utilize it for.

First of all unlike most of the application software used for recon, Maltego provides a GUI, which not only makes it easier to use but is a feature in itself, as the data representation is what makes it stand out of the crowd. It basically works on client-server architecture, which means that what we as a user get is a Maltego client which interacts with a server to perform its operations.

Before going any further let's understand the building blocks of Maltego as listed below.

## ENTITY

An entity is a piece of data which is taken as an input to extract further information. Maltego is capable of taking a single entity or a group of entities as an input to extract information. They are represented by icons over entity names. E.g. domain name [xyz.com](http://xyz.com) represented by a globe-like icon

## TRANSFORM

A transform is a piece of code which takes an entity (or a group of entities) as an input and extracts data in the form of entity (or entities) based upon the relationship. E.g. DomainToDNSNameSchema: this transform will try to test various name schemas against a domain (entity).

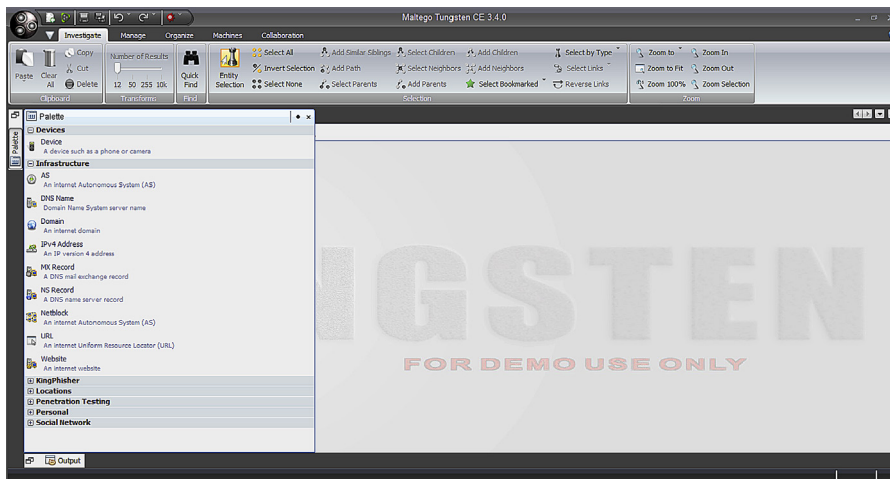


## MACHINE

A machine is basically a set of transforms linked programmatically. A machine is very useful in cases where the starting data (in form of an entity) and the desired output data are not directly linked through a single transform but can be reached through a series of transforms in a custom fashion. E.g. Footprint L1: a transform which takes a domain as an input and generates various types of information related to the organization such as e-mails, Autonomous System AS number, etc.

First of all as mentioned above, we need to create an account for the community version. Once we have an account we need to download it from <https://www.pater.va.com/web6/products/download3.php>. The installation of the application is pretty straightforward and the only requirement is Java. Once the installation is complete we simply need to open the application and login using the credentials created during the registration process.

Now as the installation and login processes are complete, let's move on to the interface of Maltego and understand how it works. Once we are logged into the application it will provide us with some options to start with; we will be starting with a blank graph so that we can understand the application from scratch. Now Maltego will present a blank page with different options on top bar and a palette bar on the left. This is the final interface we will be working on.



**FIGURE 6.24**

Maltego interface.

On the top left corner of the interface is the Maltego logo, clicking on which will list down the options to create a new graph, save the graph, import/export configurations/entities, etc. The top bar in the interface presents five options, let's discuss them in detail:

## INVESTIGATE

This is the first option in the top bar which provides basic functions such as cut, copy, paste, search, link/entity selection, as well as addition. One important option provided is Select by Type, this options comes in handy when there is a huge amount of data present in the graph after running a different set of transforms or machines and we are seeking a specific data type.

## MANAGE

The Manage option basically deals with entity and transform management with some other minor functions such as notes and different panel arrangements. Under the Entities tab we get the options to create new entities, manage existing ones, and their import/export; similarly the Transforms tab presents the options to discover new transforms, manage existing ones, and create new local transforms (we will discuss creating local transforms in later chapter.

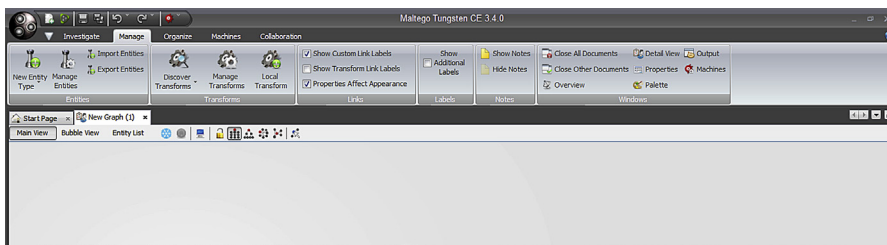


FIGURE 6.25

Maltego Manage tab.

## ORGANIZE

Once we are done with extracting the data, we need to set the arrangement of the graph to make a better understanding of it, this is where the Organize option comes in. Using the underlying options we can set the layout of the complete graph or selected entities into different forms, such as Hierarchical, Circular, Block, etc. We can also set the alignment of entities using the functions under “Align Selection” tab.

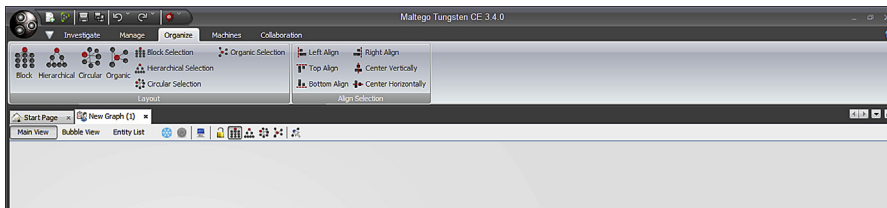


FIGURE 6.26

Maltego Organize tab.

## MACHINES

As described before machines are an integral part of the application. Machines tab provides the options to run a machine, stop all machines at once, create new machines (which we will discuss in later chapter) and to manage existing ones.

## COLLABORATION

This tab is used to utilize the feature introduced in late version of Maltego which allows different users to work as a team. Using the underlying options users can share their graphs with other users in real time as well as communicate through the chat feature. This feature can be very helpful in Red Team environments.

The palette bar on the left is used to list all the different types of entities present in Maltego. The listed entities are categorized according to their domain. Currently Maltego provides 20+ entities by default.

Now as we are familiar with the interface we can move on to the working of Maltego.

First of all to start with Maltego we need a base entity. To bring an entity into the graph we simply need to drag and drop the entity type we need to start with, from the palette bar on the left. Once we have the entity in the graph, we can either double click on the name of the entity to change its value to the value of our desire or double click on the entity icon which pops up the details window where we can change data, create note about that entity, attach an image, etc. One thing that we need to keep in mind before going any further is to provide the entity value correctly depending upon the entity type e.g. don't provide a URL for an entity type "domain."

Once we have set the value of an entity we need to right click on that entity and check the transforms listed for that specific entity type. Under the "Run Transform" tab we can see the "All Transforms" tab at the top, which will list all the transforms available for the specific entity type; below that tab we can see different tabs which contains the same transforms classified under different categories. The last tab is again "All Transforms," but use this one carefully as it will execute all the listed transforms at once. This will take up a lot of time and resources and might result into a huge amount of data that we don't desire.

Now let's take up the example of a domain and run some transforms. To do this simply drag and drop the domain entity under infrastructure from the palette bar to the graph screen. Now double click on the label of the entity and change it to let's say [google.com](http://google.com). Now right click on it and go to "All Transforms" and select the "To DNS Name - NS (name server)." This transforms will find the name server records of a domain. Once we select the transform we can see that results start to populate on the graph screen. The progress bar at the bottom of the interface shows if the transform is complete or is still running. Now we can see that Maltego has found some name server (NS) records for the domain. We can further select all the listed NS records and run a single transform on them. To do this simply, select the region containing all the records and right click to select a transform. Let's run the transform "To Netblock [Blocks delegated to

this NS],” this transform will check if the NS record have any (reverse) DNS netblocks delegated to them. In the graph window itself we can see at the top that there are some options to try like Bubble View, which shows the graph as a social network diagram with the entity size depending upon the number of inbound and outbound edges; the Entity List as the name suggests lists down all the entities in the graph and some others like freeze view, change layout to Block, Hierarchical, Circular, etc.

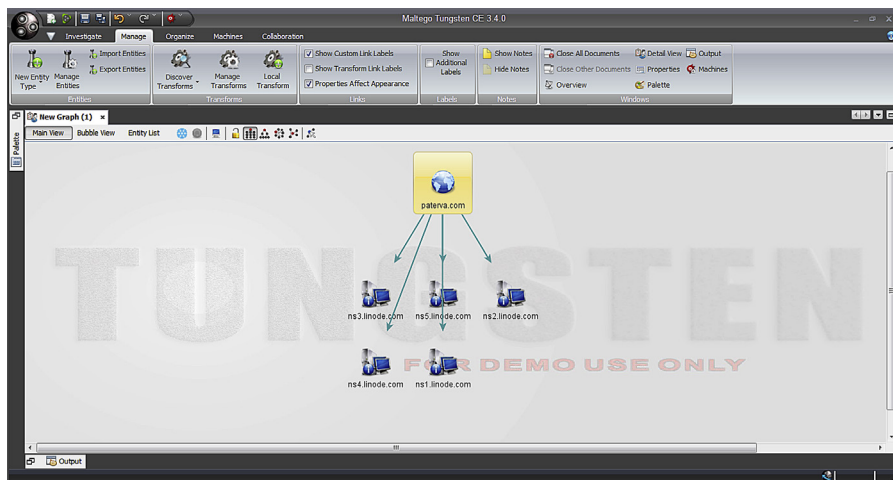


FIGURE 6.27

Maltego Transform result (Domain to DNS Name - NS (name server)).

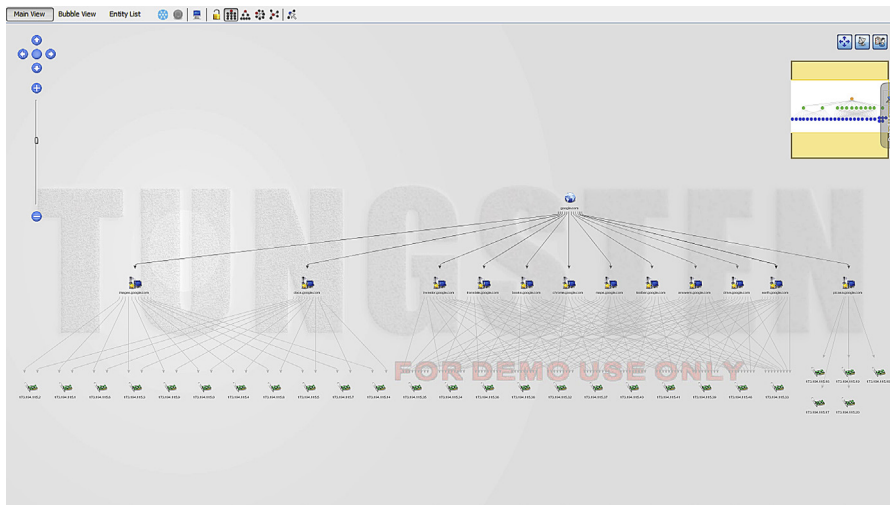
Similar to running a transform on an entity we can also run a machine. Let’s stick to our example and take a domain entity with value [google.com](http://google.com). Now we simply need to right click on the entity, go to “Run Machines” tab and select a machine. For this example let’s simply run the machine “Footprint L1.” This machine will perform a basic footprint of the domain provided. Once this machine is executed completely we can see that it displays a graph with different entities such as name servers, IP addresses, websites, AS number, etc. Let’s move forward and see some specific scenarios for data extraction.

## DOMAIN TO WEBSITE IP ADDRESSES

Simply take a domain entity. Run the transform “To Website DNS [using Search Engine].” It queries a search engine for websites and returns the response as website entities. Now select all the website entities we got after running the transform and run the transform “To IP Address [DNS].” This will simply run a DNS query and get us the IP addresses for the websites. This sequence of

transforms can help us to get a fair understanding of the IP range owned by the organization (owning the domain). We can also see which websites have multiple IP addresses allocated to them. Simply changing the layout of the graph, to say circular, can be helpful in getting a better understanding of this particular infrastructure. Information like this is crucial for an in-depth pentest and can play a game changing role.

E.g.: Domain = [google.com](https://www.google.com)



**FIGURE 6.28**

Maltego Transform result (Domain to Website IP).

## DOMAIN TO E-MAIL ADDRESS

There is a set of transforms for extracting e-mail address directly from a domain, but for this example we will be following a different approach using metadata. Let's again take a domain entity and run all the transforms in the set "Files and Documents from Domain." As the name itself says, it will look for files listed in search engine for the domain. Once we get a bunch of files, we can select them and run the transform "Parse meta information." It will extract the metadata from the listed files. Now let's run all the transforms in the set "Email addresses from person" on the entities of type entity and provide the appropriate domain (domain we are looking for in the e-mail address) and a blank for additional terms. We can see the result from this final transform and compare it with the result of running the transform set for e-mail extraction running directly on the domain and see how the results are different.

E.g.: Domain = [paterva.com](https://www.paterva.com)

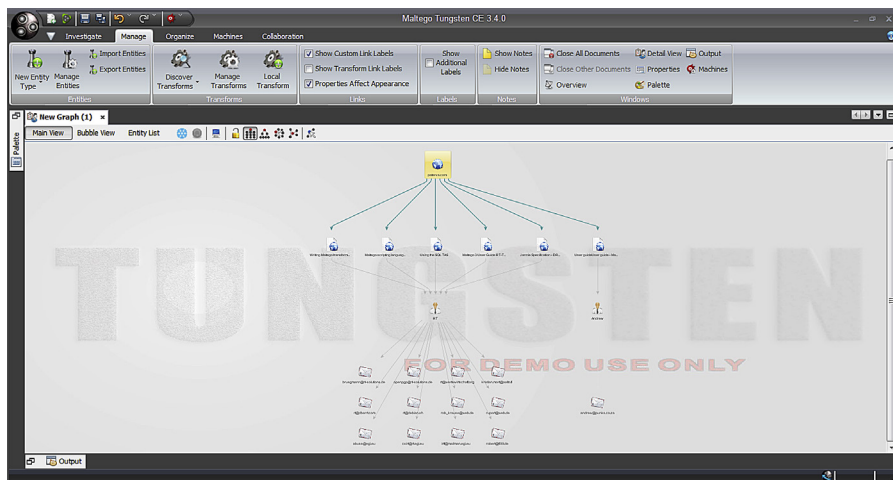


FIGURE 6.29

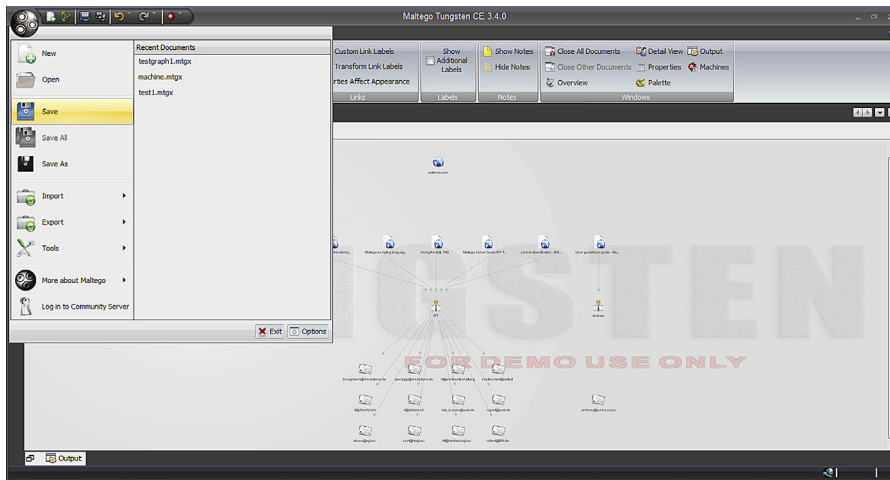
Maltego Transform result (Domain to Email address).

## PERSON TO WEBSITE

For this example we will be using a machine “Person - Email address.” Let’s take an entity of type person and assign it a value “Andrew MacPherson” and run the machine on this entity. The machine will start to enumerate associated e-mail IDs using different transforms. Once it has completed running one set of transforms it will provide us the option to move forward with selected entities, enumerated till now. From the above example we know “andrew@punks.co.za” is a valid e-mail address so we will go ahead with this specific entity only. What we get as an end result is websites where this specific e-mail address occurs, by running the transform “To Website [using Search Engine]” (as a part of the machine).

The examples shown clearly demonstrate the power of this sophisticated tool. Running a series of transforms or a machine can enumerate a lot of data which can be very helpful during a pentest or a threat-modeling exercise. Extracting a specific type of data from another data type can be done in different ways (using different series of transforms). The best way to achieve what we want is to run a series of transforms, eliminate the data we don’t need, then parallelly run another sequence of transforms to verify the data we have got. This exercise not only helps to verify the credibility of the data we have got but sometimes also produce unique revelation.

Maltego even allows to save the graph we have generated into a single file in “mtgx” format for later usage or sharing. We can even import and export entities as well as configuration. This feature allows us to carry our custom environment with us and use it even on different machines.



**FIGURE 6.30**

Saving Maltego results.

Apart from the prebuilt transforms Maltego allows us to create our own transforms. This feature allows us to customize the tool to extract data from various other sources that we find useful for specific purpose, for example an API which allows to get the company name from its phone number.

For custom transforms we have got two options:

**Local transforms:** These transforms are stored locally in the machine on which the client is running. These type of transforms are very useful when we don't need/want others to run the transform or execute a task locally. They are simple to create and deploy. Major drawback is that if we need to run it on multiple machines we need install them separately on each one of them, and same is the case for updates.

**TDS transforms:** TDS stands for transform distribution server. It is a web application which allows the distribution as well as management of transforms. The client simply probes the TDS, which calls the transform scripts and presents the data back to the client. Compared to local transforms they are easy to setup and update.

We will learn how to create transforms in later chapter.

So these are some of the tools which can play a very crucial part in an information-gathering exercise. Some of these are more focused on information security and some are generic. The main takeaway here is that there are a bunch of tools out there which can help us to extract relevant information within minutes and if used in a proper and efficient manner these tools can play a game changing role in our data extraction process. There is something for everyone, it's just a matter of knowing how data is interconnected and hence how one tiny bit of information may lead to the box of Pandora. In the next chapter we will move forward and learn about the exciting world of metadata. We will deal with topics like what is metadata, how is it useful, how to extract it, etc. We will also deal with topics like how it can be used against us and how to prevent that from happening.