CHAPTER

Open Source Intelligence and Advanced Social Media Search

2

INFORMATION IN THIS CHAPTER

- Open source intelligence (OSINT)
- Web 2.0
- Social media intelligence (SOCMINT)
- Advanced social media search
- Web 3.0

INTRODUCTION

As we already covered the basic yet essential terms with little details in the previous chapter, it's time to move on to understanding the core topic of this book, that is open source intelligence also known by its acronym OSINT, but before that we need to recognize how we see the information available in public and up to what extent we see it.

For most of us internet is limited to the results of the search engine of our choice. If we talk about a normal user who wants some information from the internet he/she directly goes to a search engine; let's assume it's one of the most popular search engine Google and puts a simple search query. A normal user unaware of advanced search mechanisms provided by Google or its counterparts puts simple queries he/she feels comfortable with and gets a result out of it. Sometime it becomes difficult to get the information from search engine due to poor formation of the input queries. For example, if a user wants to search for a *windows blue screen error troubleshoot*, he/she generally enters in the search engine query bar "my laptop screen is gone blue how to fix this," now this query might or might not be able to get the desired result in the first page of the search engine, which can be a bit annoying at times. It's quite easy to get the desired information from the internet, but we need to know from where and how to collect that information, efficiently. A common misconception among users is that the search engine that he/she prefers has whole internet inside it, but in real scenario the search engines like Google have only a minor portion of the internet indexed. Another common practice is that people don't go to the results on page two of a search engine. We all have heard the joke made on this that "if you want to hide a dead body then Google results page two is the safest place." So we want all our readers to clear their mind if they also think the same way, before proceeding to the topic.

Hacking Web Intelligence. http://dx.doi.org/10.1016/B978-0-12-801867-5.00002-1

OPEN SOURCE INTELLIGENCE

Simply stated, open source intelligence (OSINT) is the intelligence collected from the sources which are present openly in the public. As opposed to most other intelligence collection methods, this form does not utilize information which is covert and hence does not require the same level of stealth in the process (though some stealth is required sometimes).

OSINT comprises of various public sources, such as:

- Academic publications: research papers, conference publications, etc.
- Media sources: newspaper, radio channels, television, etc.
- Web content: websites, social media, etc.
- Public data: open government documents, public companies announcements, etc.

Some people don't give much heed to this, yet it has proven its importance time and again. Most of the time it is very helpful in providing a context to the intelligence provided from other modes but that's not all, in many scenarios it has been able to provide intelligence which can directly be used to make a strategic decision. It is thought to be one of the simplest and easiest modes by many if not most, yet it does has its difficulties; one of the biggest and unique out of all is the abundance of data. Where other forms of intelligence starve for data, OSINT has so much data that filtering it out and converting it into an actionable form is the most challenging part.

OSINT has been used for long time by government, military as well as the corporate world to keep an eye on the competition and to have a competitive advantage over them.

As we discussed, for OSINT there are various different public sources from which we can collect intelligence, but during the course of this book we will be focusing on the part which only uses internet as its medium. This specific type of OSINT is called as WEBINT by many, though it seems a bit ambiguous as there is a difference between the internet and web (discussed in Chapter 1). It might look like that by focusing on a specific type we are missing a huge part of OSINT, which would have been correct few decades earlier but today where most of the data are digitized this line of difference is slowly thinning. So for the sake of understanding we will be using the terms WEBINT and OSINT interchangeably during this book.

HOW WE COMMONLY ACCESS OSINT

SEARCH ENGINES

Search engines are one of the most common and easy method of utilizing OSINT. Every day we make hundreds of search queries in one or more search engines, depending upon our preference and use the search results for some purpose. Though

the results we get seem simple but there is a lot of backend indexing goes on based on complex algorithms. The way we create our queries make a huge difference in the accuracy of the result that we actually seek from a search engine. In a later chapter we will discuss how to craft our queries so that we can precisely get the result that we desire. Google, Yahoo, and Bing are well-known examples of the search engines.

Though it seems like search engines have lots of information, yet they only index the data which they are able to crawl through programs known as *spiders* or *robots*. The part of the web these spiders are able to crawl is called as the surface web, the rest of the part is called as the dark web or darknet. This darknet is not indexed as it is not directly accessible via a link. Example of darknet is a page generated dynamically using the search option on a web page. We will discuss about darknet and associate terms in a later chapter.

NEWS SITES

Earlier the popular mediums of news were newspaper, radio, and television; but the advancement in the internet technology has drastically changed the scenario and today every major news vendor has a website where we can get all the news in a digital format. Today there even exist news agencies which only run online. This advancement has certainly brought news at the touch of our fingertips at anytime, anywhere where there is an internet connection available. For example, bbc.com is the news website for the well-known British Broadcasting Corporation.

Apart from news vendors, there are sites run by individuals or a group as well and some of them focus on topics which belong to specific categories. These sites are mainly present in form of blogs, online groups, forums, or IRCs (Internet Relay Chat), etc. and are very helpful when we need the opinion of the mass on a specific topic.

CORPORATE WEBSITES

Every major corporation today runs a website. It's not just a way to present your existence but also interact directly with customers, understand their behavior, and much more. For example, www.gm.com is the corporate website for General Motors. We can find out a plethora of information about a company from its website. Usually a corporate website contains information like key players in the organization, their e-mails, company address, company telephone, etc. which can be used to extract further information.

Today some of the corporate websites also provide information in the form of White Papers, Research Papers, corporate blogs, newsletters subscription, current clients, etc. This information is very helpful in understanding not only the current state of the company but also its future plans and growth.

CONTENT SHARING WEBSITES

Though there are various types of user-generated content out there which contains an amalgam of text as well as various different multimedia files, yet there are some

sites which allows us to share a specific type of content such as videos, photo, art, etc. These types of sites are very helpful when we need a specific type of media related to a topic as we know exactly where to find it. YouTube and Flickr are good examples of such sites.

ACADEMIC SITES

Academic sites usually contain information to some specific topics, research papers, future developments, news related to a specific domain, etc. In most cases this information can be very crucial in understanding the landscape for current as well as future development. Academic sites are also helpful in learning traits which are associated to our field of interest and also understand the correlation in between.

The information provided in the academic sites is very helpful in understanding the developments that are taking place in a specific domain and also to get a glimpse of our future. They are not only helpful in understanding the current state of development but also generating ideas based upon them.

BLOGS

Weblogs or blogs started as a digital form of personal diary, except they are public. Usually people used to write blogs in a simply way to express their views on some topics of interest, but this has changed in the past decade. Today there are corporate blogs, which talk about the views of the company and can reveal a lot about its pursuits; there are blogs on specific topics which can be used to learn about the topic; there are blogs related to events, etc.

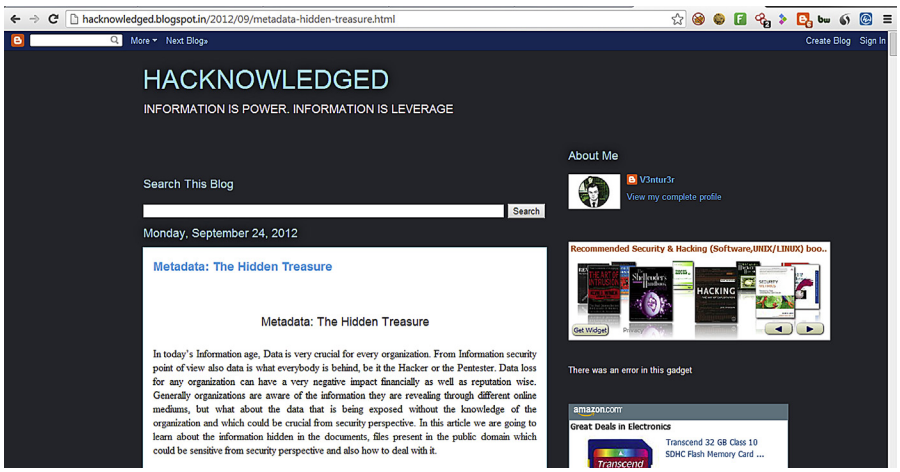

FIGURE 2.1

A blog on bylogger.in.

Blogs reveal a lot about not just the topic written about, but also about its author. In many job applications it is desired to have a blog for the applicant as it can be used to understand his/her basic psychological profile, communication skills, command over the language, etc.

GOVERNMENT SITES

Government sites contain a huge amount of public data. This includes not just information about the government but also about the people it is serving. Usually there are government sites which contain information about registered companies, their directors, and other corporate information; then there are sites which contain information about specific departments of the government; there are also sites where we can complain regarding public issues and check the status on it; etc.

From geopolitics perspective, government sites can be a good source of information for the development of a country, current advancements, its future plans, etc.

So now this is how we usually interact with the internet today, but it was not always like this. There were no blogs, no social media, no content sharing, etc. so how did we got here, let's see.

WEB 2.0

Earlier websites used to be mainly static, there was not much to interact with. Users simply used to open the web pages and go through the text and images and that was pretty much it. Around the late 1990, the web started to take a new form. The static pages were being replaced by user-generated content. The websites become interactive, people started to collaborate online. This was the advent of Web 2.0.

Web 2.0 drastically changed the way the web was interacted with. Earlier the content shared by the webmasters was the only information one could access, now people could post data on the web, opinions were being shared and challenged. This changed the way information was generated; now there were multiple sources to confirm or discredit a piece of data. People could share information about themselves, their connections, their environment, and everything they interacted with.

Now people were not just the viewers of the content of the web but the creators of it. The feature to interact and collaborate allowed people to create new platforms for information sharing and connecting around in the virtual reality. Platforms like content sharing portals, social networks, weblogs, wikis, etc. started to come into existence. Virtual world slowly started to become our second home and a source of plethora of information, which would have not existed earlier.

This virtual world is now our reality. The ability to create content here allows us to share whatever information we want, our personal information, our professional information, our feelings, our likes/dislikes, and what not. Here we can tell others about ourselves and at the same time learn about others. We can share our views about anything and understand how other people perceive those issues. It allows us to interact with the whole world by sitting in a nook of it.

Today on these social platforms of ours it's not just individuals who exist, but there is much more. There are people in form of communities and/or groups; there are pages of political parties, corporates, products, etc. Everything we used to deal with in real life is being replicated in the virtual world. This certainly has brought the world closer in a sense and it does affect our lives.

The web at its current stage is not only a part of our life but it also influences it. By sharing our feelings, desires, likes/dislikes online we let others to know about us, understand our personality, and vice versa. Similarly the content posted online plays a huge role in our decision making. The advertisements we see online are personalized and depend upon our online behavior and these ads influence what we buy. Be it a political hashtag on Twitter or a viral video on YouTube we daily process a lot of online data and it does make a difference in our decisions.

Today the web has evolved to a level where there is abundance of data, which is a good thing as it increases the probability of us finding the answers to our questions. The issue with this is how to extract relevant information out of this mammoth and this is exactly what we will be dealing with in this book, starting from this chapter.

SOCIAL MEDIA INTELLIGENCE

Social media is an integral part of the web as we know it. It is mostly where all the user-generated content resides. Social media intelligence or SOCMINT is the name given to the intelligence that is collected from social media sites. Some of these may be open, accessible without any kind of authentication and some might require some kind of authentication before any information is fetched. Due to its partially closed nature some people don't count it as a part of OSINT, but for the sake of simplicity we will be considering it so.

Some social media types are:
- Blogs (e.g., Blogger)
- Social network websites (e.g., Facebook)
- Media sharing communities (e.g., Flickr)
- Collaborative projects (e.g., Wikipedia)

As now we have a clear idea about OSINT as well as social media from its perspective, let's move on to understand one of the integral part of social media and a common source of information sharing, i.e., social networks.

SOCIAL NETWORK

A social network website is a platform which allows its users to connect with each other depending upon their area of interest, location they reside in, real life relations, etc. Today they are so popular that almost every internet user has a presence on one or more of these. Using such websites we can create a social profile of our own, share updates, and also check profiles of other people in which we have some form of interest.

Some of the common features of social network websites are:
- Share personal information
- Create/join a group of interest
- Comment on shared updates
- Communicate via chat or personal message

Such websites have been very helpful in connecting people over boundaries, building new relations, sharing of ideas, and much more. They are also very helpful in understanding an individual, their personality, ideas, likes/dislikes, and what not.

INTRODUCTION TO VARIOUS SOCIAL NETWORKS

There are several popular social network sites where we are already registered, but why so many of these different social network sites, why not just couple of them? The reason is different social network focuses on different aspects of life. Some focus on generic real life relations and interests like Facebook, Google+, etc. Some focus on its business or professional aspect like LinkedIn and some on microblogging or quickly sharing of views like Twitter. There are lot more popular social networks with different aspects but in this chapter we will restrict to only these some of the popular ones, which are:

- Facebook
- LinkedIn
- Twitter
- Google+

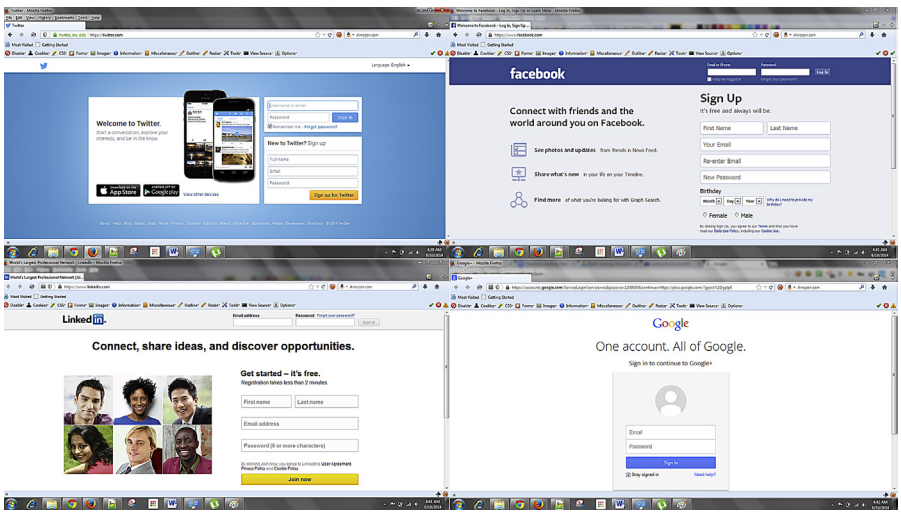

FIGURE 2.2

Social network sites.

Facebook

Facebook is one of the most popular and widely used social network sites. It was founded on February 4, 2004 by Mark Zuckerberg with his college roommates. Initially Facebook was restricted among Harvard University students, but now it's open for all to register and use Facebook whose age is above 13, though no proof is required. Among all other social network sites it contains the most wide-age group audience due to some of its popular features and generic aspects. Currently it has over a billion active users worldwide and adds over half a petabytes of data every 24 h.

It allows creating a personal profile where user can provide details like work and education, personal skills, relationship status, add family member details, add basic information like gender, date of birth, etc.; contact information like e-mail id, website details, etc.; and also life events. It also allows creating a page for personal or business use which can be used as a profile. We can also create groups, join a group of his/her interest, add other Facebook users based on relations or common interest, and categorize the friends. Like something we like, comment on something, share what we feel as status, check in where we were, share what we are doing right now, add pictures and videos. We can also share messages with someone or in a group publicly or privately and chat with someone. Adding notes, creating events, and playing games are some of its other features.

Now you might be thinking that why are we sharing all this information, because as a Facebook user most of us are aware of all these things. The reason to put a light on these features is it will help us in OSINT. As we discussed earlier, Facebook adds over half a petabyte data every 24 h, it has more than a billion active user and it allows users to share almost everything, combining all these three above statements we can say Facebook contains petabytes of structured data of over billion users like what a user likes, a user's basic information such as his/her name, age, gender, current city, hometown, work status, relationship status, current check-ins where he/she visited recently, everything which is a treasure hunt for any information gathering exercise. Now though mostly we don't use Facebook for hardcore intelligence purposes but still we use its search facility sometime to search for some person or page, etc. Like some day we remembered a school friend and we want to search for him/her in Facebook, so we search his/her name in Facebook or we search his/her name with the school name to get the result. Other option that can be used is if there is a group for the schoolmates, we can directly go there and search for the friends. Based on our preferences, location, studied school, college, colleagues, and friends, Facebook also recommends us some friends with people you may know option. This option also helps a lot to search someone in Facebook. We will cover advanced ways of searching in Facebook in an upcoming topic.

Facebook does allow setting privacy in most of the things mentioned above, like whom you want to share this information with, whether public, friends and friends of friends, only friends or only me. It also allows users to block inappropriate content or users, report spam and inappropriate content. But guess what most of us are unaware of these functionalities or simply ignore them.

LinkedIn

If you are a job seeker, jobholder, job provider, or business person, LinkedIn is the best place to stay active. It can be called as professional network where people are mostly interested in business-oriented stuffs. It has more than 259 million members in over 200 countries and territories.

LinkedIn allows us to register and create a profile. The profile basically consists of name, company name, position, current work location, current industry type, etc. Here we can also add details about our work like job position and responsibilities, add educational details, honor and award details, publications, certificates, skills, endorsement, projects undertaken, languages known, almost our complete professional life. Apart from that LinkedIn also allows us to add personal details such as date of birth, marital status, interests and contact information, which are a concern for certain employers.

Like Facebook it also allows us to connect with other users of similar interest or with whom we have some level of relationship. To maintain professional decorum LinkedIn restricts us to invite others if we have got too many of responses like "I don't know" or spam for our connection requests. Similar to Facebook there are also different groups present in LinkedIn where we can join to share common interest. It also provide feature to like, comment, and share whatever we want and also to communicate with other connections via private message. The one simple yet rich feature of LinkedIn over Facebook is that in Facebook we can only see the mutual friends between two users, but in LinkedIn it will show us how we are connected with a particular user just by visiting his or her profile. It also shows what are the things common between two of us so that we can easily understand on what and up to what extent the other user is similar to us. Other major thing is that in LinkedIn if we sneak into someone's profile, that user will get to know that someone has viewed his/her profile. Though this can be set to partial anonymous or full anonymous using privacy settings but still it's a very good feature in terms of professional network. Let's say we are a job seeker and some recruiter just sneaked into your profile, then we can expect a job offer. Like Facebook, LinkedIn also allows us to set privacy policy on almost everything.

LinkedIn is a great place for job seekers as well as job providers. The profile can be used as bio-data/resume or CV where recruiter can directly search for candidate based on the skill set requirement. Other than that it also has a job page where we can search or post jobs. We can also search for jobs based on his current industry type or company followed by him/her. Job seeker can search job based on location, keyword, job title, or company name.

Now from an OSINT perspective like Facebook, LinkedIn also has a lot of structural information or we can say structural professional information about a particular user and company such as full name, current company, past experience, skill sets, industry type, company other employee details, company details, etc. and using some advanced search techniques in LinkedIn we can collect all those information efficiently which we will discuss soon.

Twitter

Twitter is a microblogging service type social network. It allows us to read the short 140 character (or less) based messages as known as tweets without registration but after logging in we can both read as well as compose tweets. It is also known as SMS of the internet.

Unlike other social network sites Twitter has a user base which is very diverse in nature. Nowadays Twitter is considered as the voice or speech of a person. Tweets are considered as statements and are parts of news bulletin, etc.

The major reason why it is considered as voice of a person is because of its verified accounts. Verify account is a feature of Twitter which allows celebrities or public figures to show the world that it is the real account, though sometimes they also verify their account just to maintain control over the account that bears their name.

Like other social networking sites when we get registered in Twitter, it allows us to create a profile, though it contains very limited information like name, Twitter handle, a status message, website detail, etc.

A Twitter handle is like a username which uniquely identifies us in Twitter. When we want to communicate with each other we use this Twitter handle. Twitter handle generally start with a "@" sign and then some alphanumeric characters without space, for example, @myTwitterhandle. It allows us to send direct message to another user privately via messages or publicly via tweets. It also allows us to group a tweet or topic by using hashtag "#". Hashtag is used as a prefix of any word or phrase such as #LOL which is generally used to group a tweet or group a topic under funny category.

A word, phrase, or topic that is tagged mostly, within a time period is said to be a trending topic. This feature allows us to know what is happening in the world. Twitter allows us to follow other users. We can tweet, or simply share someone's tweet which is also known as retweets. It also allows us to favorite a tweet. Like other social network sites it also allows us to share images and videos (with certain restrictions). Tweets visibility is by default public but if a user wants then he/she can restricts their tweets to just their followers. Twitter is nowadays popularly used for announcement or giving verdict, statement, or replying to something online. The tweets of the verified account are taken as direct statement of that person. Corporates use this for advertising, self-promotion, and/or announcements.

Unlike the two social networks we discussed earlier Twitter does not contain much personal or professional data, yet the information it provides is helpful. We can collect information about social mentions, such as if you want to search details about infosec bug bounty, you can search in Twitter with a hashtag and you will get lots of tweets related to this where you can collect information such as which are the companies into bug bounty. What is the new bug bounty started? Who all are participating in bug bounty, etc. Unlike other social network sites Twitter has large amount of structured information based on phrases, words, or topics.

Google+

Google+ also known as Google+ is a social networking site by Google inc. It is also known as identity service which allows us to associate with the web-contents

created by us directly by using it. It is also the second largest social networking site after Facebook with billions of registered and active users. As Google provides various services such as Gmail, Play store, YouTube, Google Wallet, etc., the Google+ account can be used as a background account for these.

Like other social networking sites we just came across, Google+ also allows us to register ourselves, but the advantage that Google+ has over other social networking sites is that most Gmail (the popular e-mail solution by Google) users will automatically be a part of it just by a click. Like other social network sites we can create profile which contains basic information like name, educational details, etc.

Unlike other social networking sites a Google+ profile is by default a public profile. It allows video conferencing via Google hangout. It allows us to customize our profile page by adding other different social media links that we own like a blog.

We can consider it as one background solution for many Google services but yet it has its own demerit. Many users has one or more Gmail accounts that they actively use but in case of Google+ they might have the same number of accounts but they can use only one as active account. So in this way there is a chance that the total number of registered accounts and active user ratio might be very less as compared to other social networking sites.

Like its competitors Google+ also allows to create, join communities, follow or add friends, share photos, videos, or locations but one feature that makes Google+ a better social networking site is its +1 button. It's quite similar to like button in Facebook but the added advantage is that when the +1 count is higher for a topic or a link it increases its PageRank in Google also.

Now the OSINT aspect of Google+, like other social networking sites Google+ also have a huge amount of structured data of billion users. Other feature that makes Google+ a better source of information gathering is that the profiles are public. So no authentication required to get information. One another advantage of Google+ over other social sources is that it's a one stop solution; here we can get information about all the Google content a user is contributing or at least the other Google services details a user is using. This can be pandora of treasure.

ADVANCED SEARCH TECHNIQUES FOR SOME SPECIFIC SOCIAL MEDIA

Most of the social media sites provide some kind of search functionality to allow us to generally search for things or people we are interested in. These functionalities, if used in a bit smarter way, can be used to collect hidden or indirect but important information due to structural storage of user data in these social media.

FACEBOOK

We already discussed how Facebook can be a treasure box for information gathering. One functionality that helps us to get very precious information is Facebook graph search.

Facebook graph search is a unique feature that enhances us to search people or things that are somehow related to us. We can opt for graph search to explore location, places, photos, and search for different people. It has its unique way of suggesting what we want to search based on first letters or words. It starts searching an item from different category of Facebook itself such as people, pages, groups, and places, etc. and if sufficient results are not found it starts searching on Bing search engine and provides user with sufficient results. To provide most relevant results, Facebook also looks into our relation or at least area of interest and past experience, for example, we can get those things in higher ranking result those are either liked, commented, shared, tagged, check-in, or viewed either directly by us or by our friends. We can also filter the results based on social elements such as people, pages, places, groups, apps, events, and web results. The technology that Facebook is using in its graph search can be defined as the base of the semantic web, which we will discuss at the end of this chapter.

Now though we learned about the feature that can allow us to search different things in Facebook but still the question is how? Now let's start with some simple queries.

Just put photos in search bar and you will be suggested by Facebook with some queries such as photos of my friends, photos liked by me, my photos, photos of X, etc. and similarly we can get lots of photo-related queries or we may create our own queries. So based on photos what we can get ultimately a query such as "Photos taken in Bangalore, India commented on by my friends who graduated before 2013 in Bhubaneswar, India" so it's basically about our own imagination what exactly we want to retrieve, then based on keywords we can create complex queries to get the desired results; though Facebook will suggest some of the unexpected queries based on keywords mentioned in the search bar. Similarly we can search for persons, locations, restaurant, employee of a particular company, music, etc.

Some basic queries related to different social elements are as follows:

1. Music I may like
2. Cities my friends visited
3. Restaurants in Macao, China
4. People who follow me
5. Single females who live in the United States
6. My friends who like X-Men movies
7. People who like football

Now let's combine some of these simpler queries to create a complex one

"Single women named 'Rachel' from Los Angeles, California who like football and Game of Thrones and live in the United States." Now isn't it amazing! And yes we can create query using following filters, like based on basic information such as name, age, gender, based on work, and education such as class, college passing year, degree name, based on likes and dislikes, based on tagged in, commented on, based on living, and also based on relationships. Now it's our wild imagination that can lead us to create different queries to get desired result.

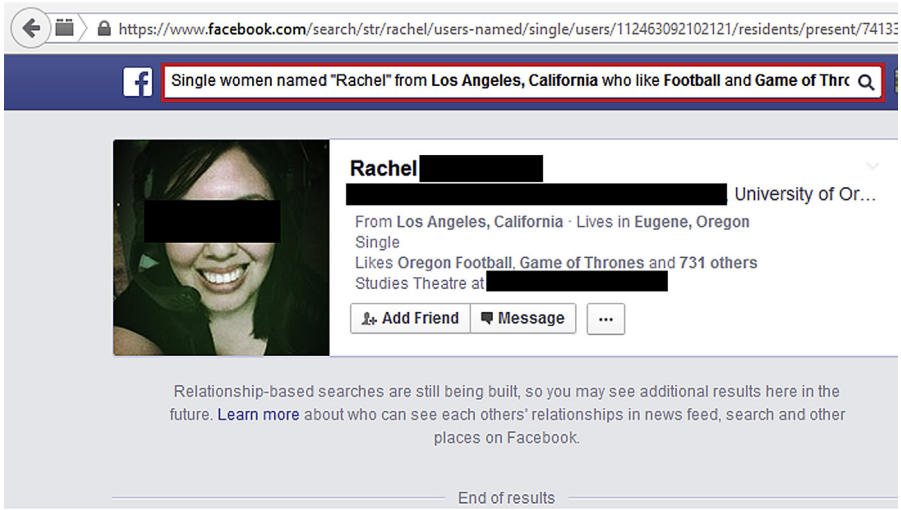

FIGURE 2.3

Facebook graph search result.

LINKEDIN

As we discussed how LinkedIn has its structural data of billions of users and what can we get if we search for something in particular, let's see how to search this particular platform. LinkedIn provides a search bar in top to search for people, jobs, companies, groups, universities, articles, and many more. Unlike Facebook, LinkedIn has its advance search page where we can add filters to get efficient result. Following is the page link for advanced search in LinkedIn:

https://www.LinkedIn.com/vsearch/p?trk=advsrch&adv=true

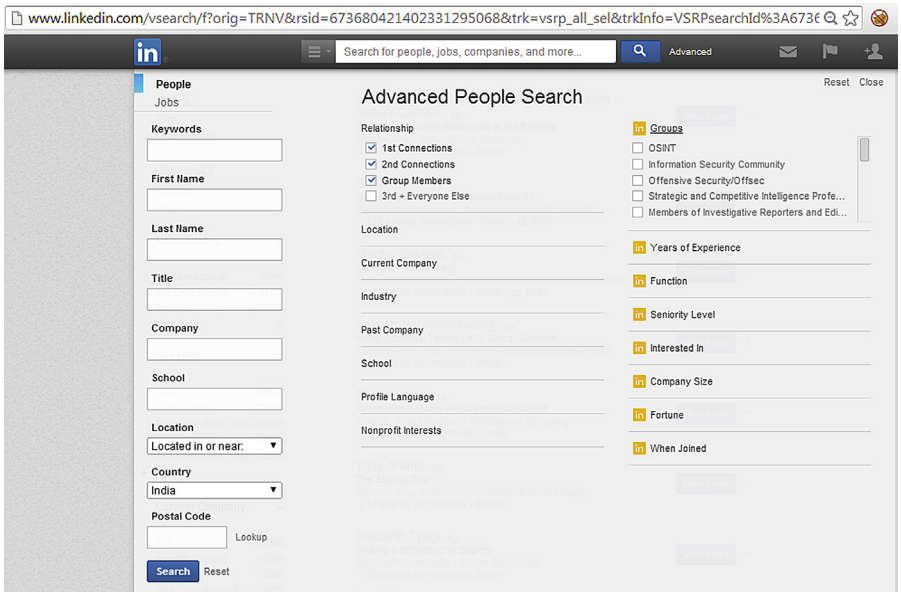

FIGURE 2.4

LinkedIn advanced search options.

This advanced search page allows us to search for jobs and people based on current employee, past employee, job title, zip code radius, interested in, industry type, etc. It also allows us to search based on type of connection.

Different input boxes and their uses

- Keyword
  The keyword input box allows a user to insert any type of keyword such as pentester or author, etc.
- First Name
  We can search using first name.
- Last Name
  We can search using last name.
- Title
  Title generally signifies to the work title. Using which user will be provided with a drop down menu with four options to choose like current or past, current, past, past not current to enrich the search.
- Company
  We can search using company name. It also comes with a drop down menu with the options we just discussed.
- Location
  This drop down box comes with two options, i.e, located in or near and anywhere. User can use whatever he/she wants.
- Country
  Search based on country.
- Postal Code
  Search based on postal code. There is a lookup button present for user to check whether the entered postal code is for the desired location or not. By entering postal code automatically a within drop down box enables which contains following options to choose:

1. 10mi (15km)
2. 25mi (40km)
3. 35mi (55km)
4. 50mi (80km)
5. 75mi (120km)
6. 100mi (160km)

  This can be used to select the radius area you want to include in search along with the postal code.
- Relationship
  This checkbox contains options to enable direct connection search, connection of connection search, group search, and all search. User can enable the final option, i.e., 3rd+ everyone else to search everything.
- Location
  This option is for adding another location which is already mentioned in postal code.

- Current Company
  This option allows a user to add current company details manually.
- Industry
  It provides a user with different options to choose one or more at a time.
- Past Company
  This option allows to add past company details manually.
- School
  Similar to past company we can add details manually.
- Profile Language
  It provides a user to choose different languages one or more at a time.
- Nonprofit Interests
  It provides user to choose two options either bored services or skilled volunteer or both.
  The options which are present in the right side of the advanced search page are only for premium account members. There are other added functionality also present only for premium users.
  The premium member search filter options are
- Groups
- Years of Experience
- Function
- Seniority Level
- Interested In
- Company Size
- Fortune
- When Joined
  Apart from all these LinkedIn also allows us to use Boolean operators. Below are the operators with simple examples:
- AND: It can be used for the union of two keywords such as developer AND tester.
- OR: It can be used for options. Let's say a recruiter want to recruit a guy for security industry so he/she can search something like pentester OR "security analyst" OR "consultant" OR "security consultant" OR "information security engineer."
- NOT: This can be used to exclude something from other things let's say a recruiter wants fresher level person for some job but not from training domain so he/she can use developer NOT trainer.
- (Parentheses): This is a powerful operator where a user can group something from other such as (Pentester OR "Security Analyst" OR "Consultant" OR "Security Consultant" OR "Information Security Engineer") NOT Manager.
- "Quotation": It can be used to make more than one words as a single keyword such as "Information Security Engineer." Now if we use the same word without quotation. LinkedIn will treat it as three different keywords.

Unlike search engines which can hold a limited keyword in search box, LinkedIn allows unlimited keywords that is a major plus for the recruiters to search for skill

sets and other job requirement keywords in LinkedIn. So it provides the user freedom to use any number of keywords he/she wants with the use of operators wisely to create a complex query to get desired result.

Example of a complex query to look for information security professionals but who are not manager:

((Pentester OR "Security Analyst" OR "Consultant" OR "Security Consultant" OR "Information Security Engineer") AND (Analyst OR "Security Engineer" OR "Network Security Engineer")) NOT Manager.

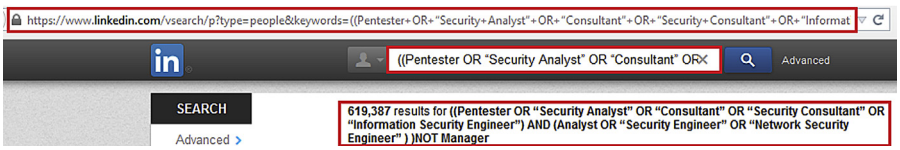

FIGURE 2.5

LinkedIn advanced search result.

TWITTER

So as we discussed earlier Twitter is basically about microblogging in the form of tweets and hence it allows us to search for tweets. Now simply inputting a keyword will get us the tweets related to that keyword but in case we need more specific results we need to use some advanced search operators. Let's get familiar with some of them.

In case we want to search tweets for specific phrases we can use the "", for example, to search for the phrase *pretty cool* the query would be "pretty cool." To look hashtag we can simply type the hashtag itself (e.g., #hashtag). In case we want to search for a term but want to exclude another specific term, we can use the - operator. Say, for example, we want to search for hack but don't want the term security, then we can use the query *hack -security.* If we want the results to contain either one or both of the terms, then we can use the *OR* operator, such as *Hack OR Security*. To look for the results related to a specific Twitter account, we simply search by its Twitter handle (@Sudhanshu_C). The *filter* operator can be used to get specific type of tweet results, for example, to get tweets containing links we can use *filter:links. From* and *To* operators can be used to filter the results based upon the sender and receiver respectively, e.g., *From:sudhanshu_c, To:paterva*. Similarly *Since* and *Until* can be used to specify the timeline of the tweet, e.g., *hack since:2014-01-27, hack until:2014-01-27.* All these mentioned operators can be combined to get better and much precise results. To checkout other features we can use the Twitter advanced search page at https://Twitter.com/search-advanced, which has some other exiting features such as location-based filter.

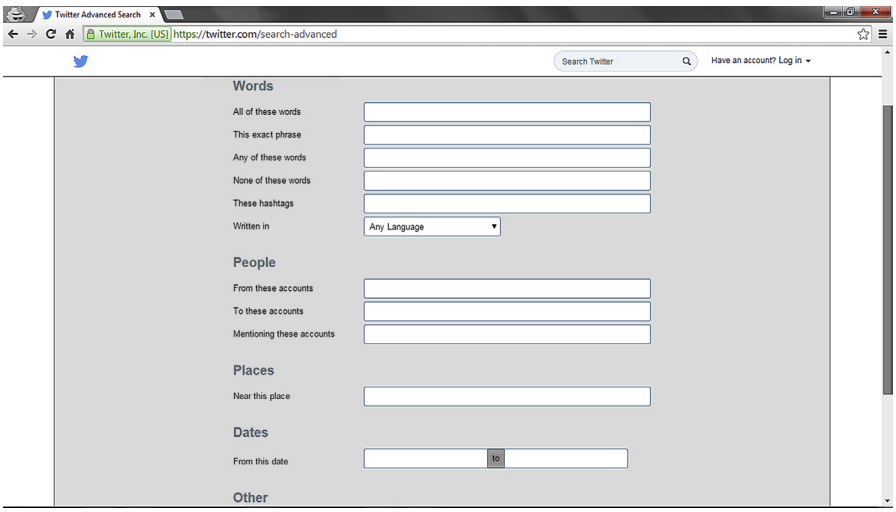

FIGURE 2.6

Twitter advanced search options.

SEARCHING ANY OPEN SOCIAL MEDIA WEBSITE

So we learned about social networks and how to search some of them, but what about the platforms that we need to search, but don't support any of the advanced search features we discussed about. Don't worry we have got you covered, there is a simple Google search trick which will help us out, it is the *site* operator. A Google search operator is simply a way to restrict the search results provided by Google within a specific constraint. So what the site operator does is that it restricts the search results to a specific website only, for example, if we want to search for the word "hack," but we only want results from the Japanese Wikipedia website, the query we will input in Google would be *site:ja.wikipedia.org hack*. This will give results for the word hack in the site we specified, i.e., ja.wikipedia.org. Now if we want to search in multiple platforms at once there is another Google operator which comes in handy, it is the *OR* operator. It allows us to get results for either of the keywords mentioned before and after it. When we combine it with the site operator it allows us search results from those specific platforms. For example, if we want to search the word "hack" in Facebook as well as LinkedIn the Google query would be *site:Facebook.com OR site:LinkedIn.com hack*. As we can see these operators are simple yet very effective, we will learn more about such operators for Google as well as some of the lesser known yet efficient search engines in the coming chapters.

WEB 3.0

So we discussed about Web 2.0 its relevance and how it affects us and also how to navigate through some of the popular social networks, now let's move forward and see what we are heading toward. Until now most of the data available on the web are

unstructured, though there are various search engines like Google, Yahoo, etc., which continuously index the surface web yet the data in itself has no standard structure. What this means is that there is no common data format which is followed by the entire web. The problem with this is that though search engines can guide us to find the information we are looking for yet they can't help us answer complex queries or a sequence of queries. This is where semantic web comes in. Semantic web is basically a concept where the web follows a common data format which allows giving meaning to the data. Unlike Web 2.0 where human direction is required to fetch specific data, in the semantic web machines would be able to process the data without any human intervention. It would allow data to be interlinked not just by hyperlinks but meaning and relations. This would not only allow data sharing but also processing over boundaries, machines would be able to make a relation between data from different domains and generate a meaning out of it. This semantic web is a crucial part of the web of the future, Web 3.0 and hence is also referred as semantic web by many.

Apart from semantic web there are many other aspects which would contribute toward Web 3.0, such as personalized search, context analysis, sentiment analysis, and much more. Some of these features are already becoming visible in some parts of the web, they might not be mature enough yet the evolution is rapid and quite vivid.