

# SearchOL: A Tool for Reconnaissance

Farhan Ahmed<sup>1</sup>, Pallavi Khatri<sup>1</sup>, Geetanjali Surange<sup>1</sup>, Animesh Agrawal<sup>2</sup>

<sup>1</sup> School of Engineering and Technology, ITM University,  
Gwalior, Madhya Pradesh, India  
farhanahmed2794@gmail.com, pallavi.khatri.cse@itmuni.ac.in,  
geetanjali.surange@itmuni.ac.in,

<sup>2</sup> Independent Researcher  
akag9906@gmail.com

**Abstract:** Organization and common person personal data is available online. System administrators, however neglect the quantity of system and user information that can be extracted anonymously from the content that is publicly available on the internet. This publicly available information is critical and of great use to Penetration testers who wish to exploit the system. This work proposes an OSINT based tool developed in Python for gathering user related data from social sites using multiple search engines, get IP and information related to IP address. Tool collects data passively and from the results proves to be a comprehensive data aggregator from multiple social platforms. The tool can be used for Information gathering and scanning which are the first and second phase respectively of ethical hacking. The novelty of this tool is, this tool gives most important, most relevant and concise results from various search engines. "SearchOL" is the simplest and fastest tool to gather information about anything. SearchOL reduces the efforts of the pentesters that they do by searching again and again on different search engines.

**Keywords:** Ethical Hacking, Reconnaissance, Penetration Testing, Vulnerabilities, Scanning, IP address.

## I. Introduction

We are in the age of internet. Now-a-days almost all things are available online, whether it is a small utility we use in our homes, car we ride on, book we read, to book our movie show, to book our traveling tickets, and even we can find good friends or our life partner using online websites for marriages. We are all covered or surrounded by internet. Today in every place we can easily find the wi-fi facilities. Now internet become as important as food, cloths and House. There is an old saying in Hindi that human must need three basic things to live that are "Roti (Food), Kapda (Cloths) and Makaan (House or Shelter)". But in 21st century this old saying is appended with Internet. Without internet we can't spend our day. Internet becomes our basic need.

We know that Internet is very important and without it we can't do many things that we can easily do with the help of internet. But we know that everything has two sides "Good and Bad". Same go with internet, it has some advantages and some disadvantages also. Our dependency on internet and extensive use of internet has made internet an area of focus for wrong people.

While surfing through Internet we visit various websites and

create accounts on various sites. We use social media and fill our personal information their such as date of birth mobile number, address email id etc. This information is publicly accessible means a wrong person can also use it. Therefore, it becomes so important to save guard our Personal Identifiable Information (PII).

The activity or method use by hacker to gather our precious information is called Foot printing and Reconnaissance. This is also use by Ethical Hacker to safeguard or information. This is the first step in Ethical hacking [1]. This can be done manually or by using tools. The foot printing and reconnaissance is of two types: Passive and Active.

In passive foot printing the hacker gather information without the knowledge of a target or without direct connection with target. In Active foot printing Hacker made a connection with targets network and gather more sensitive information.

The second step of ethical hacking [1] is Scanning, there are three types of scanning: Port scanning, Vulnerability scanning, Network mapping. In our proposed work we implement port scanning in our tool.

All the web application we interact with are managed and run on servers, Servers are the computer program or device that provides services to another computer, also known as clients. servers need to be of newer hardware or software or updated frequently. Hackers can hack servers if it uses older or vulnerable software or hardware.

A photograph we click and video we record, stores a lot of information about itself such as date/time, GPS location, Camera model etc. This data is called EXIF (Exchangeable Image File Format) data. As this provide the important information about an image or video many social media sites use to remove this EXIF data from images or videos.

We proposed an OSINT [2] based tool "SearchOL" which is an upgrade version of our previous tool "SearchOL", it has SearchOL features plus seven more features. It provides the easy way to gather information from internet which is publicly available.

Our tool "SearchOL" decrease the efforts done by ethical hackers for doing passive reconnaissance about a target by simply entering the keyword (person name or organization name) and our tool will search the most relative informative websites links from Google, Ask, Yahoo, Bing search Engines and save the links in a text file for further analysis.

Our tool also provides the functionality of getting IP address of target website, find location of IP address, Server information, Port scanning, identify the right file type, Image Exif data and video EXIF data This decreases the overhead that ethical hacker does by going one by one on different platforms for gathering the information about a target and note the details manually. These all things done by “SearchOL” automatically.

## II. Literature Survey

Paper [2] Describes the concept of Open Source Intelligence (OSINT) for solving human problems in today’s world. OSINT is a result of continuously growing open information from open internet which results in gathering of so sensitive information easily.

The process of ethical hacking starts with the gathering of more and more information about the target, we can gather simple and sensitive both information from social media and internet sites, the gathering of information about the target is known as Foot printing. There are many tools available to do foot printing. With the help of Foot printing, we can gather information about network such as Network ID, domain name, IP address, protocols, news articles, web server links etc. If hacker get some very sensitive information, he or she can use this information for its malicious activities [3].

Authors in [4] proposes a cyber-reconnaissance tool named SearchSimplified built using Java. This tool gather data related to the organization entered. This tool gather data using Google’s cache system, advanced query operators such as intitle: site: filetype: This tool gather data with the help of Google.

Work proposed in [5] provides survey and taxonomy of adversarial Reconnaissance Technique, this paper tells us about cyber kill chain, Open-Source Intelligence, Sniffing, Cyber Deception and case studies of cybercrimes, categories of Target information for reconnaissance, external an Internal Reconnaissance, Taxonomy of reconnaissance techniques, Defensive Measures against Reconnaissance Techniques, etc.

In paper [6] author shows various web-based platforms for collecting and tracking IP information. Author performed an experiment in specialized university computer lab, Connect all hosts machine in the lab in Local Area Network (LAN). The results provide host name, Autonomous system, Internet Service Provider, country, continent etc.

A Comparative Study on Web Scraping is done in [7]. In these various practices of web scraping is shown. The author shows the multiple web scraping techniques by we can easily scrap websites, and compare various web scraping software. This paper gives the knowledge of various web scraping tools and techniques.

We can easily and efficiently gather data from publicly available sources using OSINT [2] (Open-Source Intelligence) tools of OSINT [2] used in investigation phase for collecting information about target. The use of OSINT [2] to gather information is shown in [8]. The proposed work uses the API keys of social media platforms and python libraries to check usernames exists or not, if exists gather data, store the results in database and display results in UI.

In research article [1], author describes the Hacking as author describes the Hacking as "Hacking is a common

practice that causes one's private and personal information to be violated." . Research article [1] also describes the various phases of Ethical Hacking.

Port scanning is described in paper “The Design of Large-Scale IP Address and Port Scanning Tool” [9]. This paper proposed a tool named HIRFL for port scanning. This paper describes the ICMP protocol, TCP Full Connection Port Scanning Technology and Classification

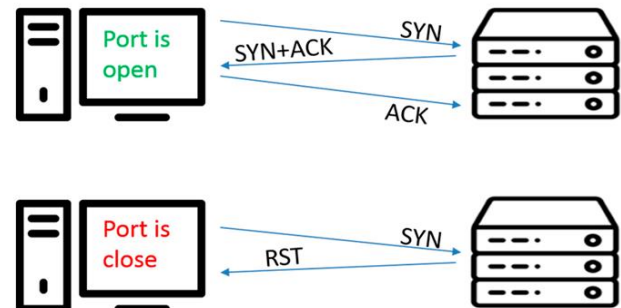


Figure 1. TCP three-way handshake port scanning [9]

EXIF (Exchangeable Image File Format) data is briefly described by author in paper “Metadata Practices for Consumer Photos” [10]. Paper describe Exif as "The EXIF standard for digital still cameras (DSC) image files uses regulated vocabulary to convey administrative metadata. It is widely used as a global standard for DSCs and numerous other disciplines. By including new tags to capture information necessary for new DSC capabilities, printer output processing, and extra GPS information, this standard accommodates recent technological improvements. A wide range of DSCs are intended to give as much capture data as possible in order to enhance information retrieval and picture management.

In paper “Sudomy: Information Gathering Tools for Subdomain Enumeration and Analysis” [11] proposed to develop applications to support the Information Gathering stage which makes it easier for Cyber Security researchers/analysis. This paper also briefly describes Domain Name System (DNS), National Institute of Standards and Technology SP 800-42 and Information Systems Security Assessment Framework (ISSAF).



Figure 2. NIST: Penetration Testing Methodology [11]

In Review Article "Current Status and Security Trend of OSINT", [12] Describes the today's situation that how OSINT can affect Security easily, The paths of information gathering is drastically increases by the evolvement of IoT and Big data. This article also through light on some important terms related to OSINT, such as:

### A. ISR:

Every nation in the world collects information about other nations. The amount of information that countries collect is

known as Intelligent Monitoring and Recognition (ISR). Methods of collecting information such as the ISR are three-fold:

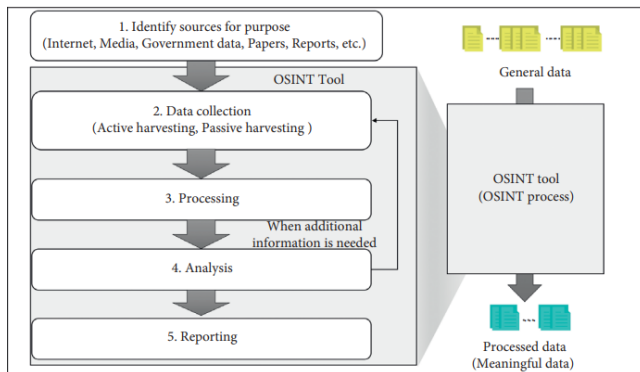
- 1) Open source intelligence (OSINT)
- 2) Human intelligence (HUMINT)
- 3) Technical intelligence (TECHINT)

It also explain it well by using examples:

*1) Open source intelligence ( OSINT ):*

Gather information using open internet, data, and software.

(a) *Open information:* In day-to-day life we share so much information on social media, we get so much information by newspapers and many other ways.



**Figure 3.** Structure of OSINT [12]

*2) Human intelligence ( HUMINT ):*

Intelligence that human collect by performing undercover operation and by spying.

(a) *White agent:* For white agent espionage is not allowed but can gather information which is available openly.

(b) *Black agent:* secret military activities, in particular illegal activities, which are ordered by a government or an organisation, but which they will not admit to having ordered: an agent/team of covert operations.

*3) Technical intelligence ( TECHINT ):*

In Technical intelligence we can use Technology to collect information about our enemies or target.

Legal aspects of criminal OSINT has been discussed by the the authors in [13] based on national and international regulations. Key application areas of criminal OSINT tools along with the tools used for the development of the methods are focused in their work. Their work also focuses on establishing a OSINT units within the imvestigaying organization. Criminal OSINT tool is gaining popularity along with traditional tools used for investigations and has become an integral part of any cybercrime investigation process.

An open source tools for password cracking has been proposed in [14] that supports the cracking the passwords using OSINT investigation, Authors also describe the following:

*A. Password Tendencies:*

Users have a habit of creating the passwords that relates to familiar patterns. This usually includes names, important dates, feeling expressions, geographical locations etc. According to their survey conducted numbers play a significant role while framing a password and nearly 4.5% of all passwords found in Rockyou data set were dates. Authors

conducted a survey and users were asked to enter regular or mnemonic passwords formed using characters and phrases. Results proved that most mnemonic passwords contained external information, whereas only 13% of control group participants did the same.

*B. Password cracking techniques:*

Passwords can be retrieved easily knowing the fact that the majority of users use very evident information while creating the passwords. Success rate of retrieving a password from all the users instead of targeted one is always higher. Automated password cracking can be performed using wide range of tools available for investigations by law enforcement agencies. These tools can also be used for penetration testing and account recovery process by penetration testers and ethical hackers. The easiest approach often adopted to recover a password is to try all possible combinations by guessing, known as comprehensive search or brute force attack. It becomes tough to retrieve or guess the password using this technique when the length of password is long and it uses a combination of alphabets, special characters and numbers while forming the passwords.

*C. OSINT can be useful for law enforcement:*

Before the digital era, law enforcement agencies collected the information that was available and used it to produce information for investigation. They use and act on information they learn from conventional sources, like victim and witness narratives and physical evidence, throughout a regular crime investigation in order to solve a crime. Through the use of current OSINT techniques, such a collection of evidence can currently be supplemented by online sources. Additionally, these tools have low manpower and financial costs when used in an investigation.

*D. Digital forensic intelligence:*

Information collected through OSINT is used in addition to the artifacts collected through traditional surveys to reach to conclusions with more accuracy.

*E. Password Strength:*

Attackers are aware of that the passwords chosen by ordinary man can be guessed easily. One of the easiest ways to safeguard oneself the user must ensure that the strength of the password chosen is string enough to be broken by an attacker and is tough to guess.

"A venerable source in a new era: Sailing the sea of OSINT in the information age" written by Stephen C. Mercado is a part of the book "Secret Intelligence: A Reader". This chapter focuses on how the OSINT can provide us the most valuable information just by open information. the author said "Collecting intelligence these days is at times less a matter of stealing through dark alleys in a foreign land to meet some secret agent than one of surfing the Internet under the fluorescent lights of an office cubicle to find some open source. As trade and technology move forward, so does the world. Today, mouse clicks and online dictionaries are often more helpful than elegant clothing and majestic weapons in harvesting the data required to support analysts and cyber investigators in discovering the world. In combination with stolen secrets, diplomatic relations and technical collections,

open sources are what a former deputy director of intelligence called the "complex mosaic" of intelligence. [15]

In paper[16] author stated that OSINT open source intelligent is becoming increasingly popular now a days and is a powerful tool for information gathering. In this paper the author describe the working and setup of Maltego tool, The concept of MALTEGO consists of a combination of entities, transformations, and machines. Entities are real objects, such as an individual, DNS name, telephone number, e-mail address. A feature is visually depicted as a node on the graph. The MALTEGO client (Classic/XL) has about twenty entities, especially for online surveys. However, it is also possible to create your own entities. Paterva created the open-source, proprietary MALTEGO intelligence and forensic science software. The primary objective of MALTEGO is to offer a library of transformations for locating and connecting data from open sources and visualising this data in a graphical style appropriate for link analysis and data mining [17].

MALTEGO is a visual link analysis tool that comes with open-source Intelligence plugins called transformations. The tool offers real-time data mining. Collected information is displayed on a node-based graph that makes patterns and connections of multiple orders between the information easily identifiable. MALTEGO focuses on analysing real relationships between publicly accessible information about Internet infrastructures, individuals, and organizations [18].

The outcome of this study offers a review on web scraping techniques, software which can be used to extract data from web sites, phases of ethical hacking, port scanning, Exif data, current trends of OSINT, OSINT in Criminal Investigation, OSINT for password cracking, and tools in same domain.

### III. Proposed Work

*A. From the extensive literature survey, following conclusions are derived:*

- Existing system cannot Search using multiple search engines.
- They only search the usernames in different social sites.
- They require multiple dependencies to be installed before data searching.
- They search data from Organizations only and not from social accounts.
- Existing systems are complex.

*B. Our proposed tool "SearchOL" comprises of eight functionalities or tools, its functionalities are given below:*

- SearchOL – An Information Gathering Tool (For gathering important web links from internet)
- Get IP address
- Locate the Ip address
- Server information
- Port Scanning
- Identify File Types
- Extract Image Exif data
- Extract Audio/video metadata

To extract precise and more information from the web this work proposes a Python based web scraping tool called

SearchOL that will work on Google, Bing, Yahoo, Ask and will retrieve most relatable URLs from various websites.

The tool will also store the retrieved information in a text file that can be further used by an attacker to exploit the system or a user. The proposed tool used Advanced Search technique of Google search engine called Google Dorking [19] to discovers the data.

"SearchOL" get the IP address of a target website using socket programming, with the help of socket module in python language. It extracts IP address of given Domain name, Hostname of IP address such as amazonaws.

"SearchOL" uses geocoder library of python to find the location of IP address. It provides city, state country and Latitude-Longitude information.

Our tool also provides the facility of scanning for open ports of given IP address, this is done using Socket programming. We can scan for all ports or we can scan for some ports that we want to scan.

Sometimes the malicious user alters the file extensions to hide important stuff behind wrong extension so that it will not be identified by forensic investigator. "SearchOL" provides the functionality of identify right file type of a file. Now this feature can identify PNG, PDF, JPG, RTF, RAR, DOCX, PPTX, XLSX file types. This is done by analysing the file in hex form and see the first few bytes to check file signatures.

A photo we capture by our camera contains a lot of hidden information about the photo, specification of camera and sometimes GPS location of the place where it captured. "SearchOL" can extract this data from image using Python Imaging Library (PIL) that adds image processing capabilities to our Python interpreter.

Like Image audio and Video also holds some hidden data inside them, that can be useful for investigation purposes. This is easily done by "SearchOL" using tinytag, this is a library for reading audio/video meta data of most common audio/video files in pure python.

Above all information that we get can be store in a text file for further analysis, this reduces the efforts of penetration tester, ethical hacker and forensic investigator.

The working methodology of "SearchOL" is as follows:

As we discussed above that "SearchOL" has 8 features so we start from first "SearchOL – An Information Gathering Tool" It uses Python Requests Module [20], which allows us to send HTTP requests using Python, we use the requests.get(url) [20] method to send a GET request to the specified URL and returns a Response Object that contains the server's response to the HTTP request.

It uses Beautiful Soup; it's a Python library that analyzes structured data. It allows you to interact with HTML in a way similar to how you interact with a webpage with the help of development tools. It uses OS module [21], this python module provides functions for interacting with the operating system. We use this module to save the information that we gather from our search using SearchOL.

*1) The work flow of SearchOL is described below:*

- Take input
- Create URL for the input keyword
- Make request for the URL using Requests Module [20], one by one on Google, Bing, Yahoo, Ask Search engines [22].
- Find all links in the search result using Beautiful Soup [23] module.

- Filter the most related and useful links
- Append sites in a list named 'sitelist'
- Append the links in a list named 'links'
- Iterate through the 'sitelist' and 'links' to print the information
- Now, save the information gather in a text file using OS module [21].

#### 2) Get IP Address:

To get IP of any website it use socket library "socket.gethostbyname".

#### 3) Find Location of IP address:

To find the location of IP address, "geocoder.ip(IPADDRESS)".

#### 4) Extract Server Information:

Requests module help us to get server information, this is extract from header of requests object.

#### 5) Port Scanning:

Port scanning performed by "SearchOL" using socket library as "CurrentSocket = socket.socket(socket.AF\_INET,socket.SOCK\_STREAM)", the arguments that passed to socket() are constants used to specify the address family and socket type. Internet address family for IPv4 is specified by AF\_INET and TCP socket type is specified by SOCK\_STREAM.

The program ask user that he/she wants to scan all ports or scan for some specified ports. If user wants to scan for some ports than it ask to enter port numbers of ports to scan. This all scan results are automatically stored in a file named "Port Scan Results of [IP ADDRESS].txt".

#### 6) Identify File Type:

"SearchOL" Identify file types by reading the file in binary mode and converting first few bytes to hex then it matches the file signatures of predefined filetypes to file hex bytes to check for right file type. The list of file types identified by "SearchOL" is given below:

- PNG
- PDF
- JPG
- RTF
- RAR
- DOCX
- PPTX
- XLSX

#### 7) Extract Image EXIF data:

Hidden details of Image file can be extracted using "SearchOL". "SearchOL" do this by using PIL (Python Imaging Library) for extracting Exif Tags. Some images also contains GPS location but it is given in (26.0, 12.0, 7.11) format, "SearchOL" convert this to in float latitude and longitude format and supply it to a user defined function name "get\_location\_name(lat,lang)". This will give address of place where image captured.

#### 8) Extract audio/video metadata:

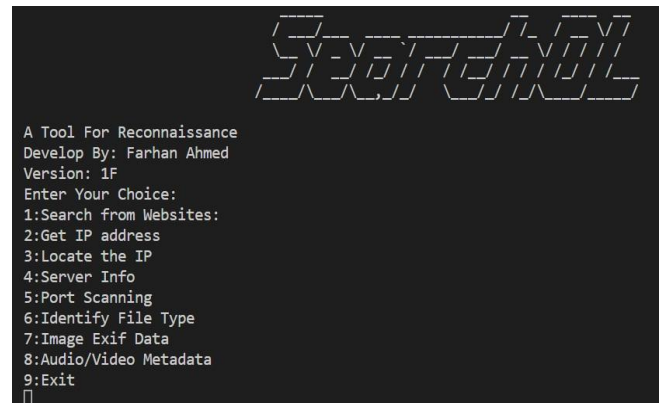
SearchOL can provide you the metadata related to audio or video. This is done by using TinyTag library of Python.

This data include artist, year, title, audio\_offset, comment etc.

"SearchOL" in the end will ask user to save the file in .txt format if user want to save the findings it will be saved successfully saved in a text file for future analysis.

## IV. Experimental Setup

The tool developed in Python version 3.9.7 [24]. is tested on system with configuration: Processor: Intel(R) Core (TM) i5-10210U CPU @ 1.60GHz 2.11 GHz, System type: 64-bit operating system, x64-based processor, RAM: 8.00 GB.



```

A Tool For Reconnaissance
Develop By: Farhan Ahmed
Version: 1F
Enter Your Choice:
1:Search from Websites:
2:Get IP address
3:Locate the IP
4:Server Info
5:Port Scanning
6:Identify File Type
7:Image Exif Data
8:Audio/Video Metadata
9:Exit
  
```

Figure 4. First View of "SearchOL"

When you Run the "SearchOL", first it shows you the options of features that you want to use as displayed in above screenshot.

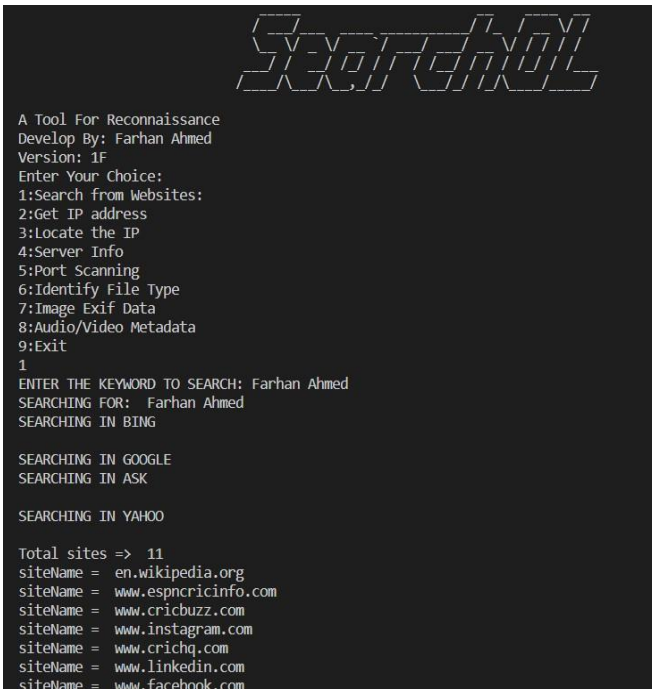
## V. Results

In this section we will one by one shows you the output of each 8 features of "SearchOL" tool –

### A. SearchOL:

A sample output from the tool is displayed as the user enters the name of an individual and proceeds to search on the social websites. The tool lists all the findings and allows to store complete data in a text file.





```

SearchOL
A Tool For Reconnaissance
Develop By: Farhan Ahmed
Version: 1F
Enter Your Choice:
1:Search from Websites:
2:Get IP address
3:Locate the IP
4:Server Info
5:Port Scanning
6:Identify File Type
7:Image Exif Data
8:Audio/Video Metadata
9:Exit
1
ENTER THE KEYWORD TO SEARCH: Farhan Ahmed
SEARCHING FOR: Farhan Ahmed
SEARCHING IN BING

SEARCHING IN GOOGLE
SEARCHING IN ASK

SEARCHING IN YAHOO


Total sites => 11
siteName = en.wikipedia.org
siteName = www.espnricinfo.com
siteName = www.cricbuzz.com
siteName = www.instagram.com
siteName = www.crichq.com
siteName = www.linkedin.com
siteName = www.facebook.com

```

**Figure 5.** Output results of ‘Search from Websites’

SearchOL provides you with the important and very relative links from web, it also provides the websites name from where it extract the links, this information will be saved in a text file for future analysis.

#### B. Get IP address:



```

SearchOL
A Tool For Reconnaissance
Develop By: Farhan Ahmed
Version: 1F
Enter Your Choice:
1:Search from Websites:
2:Get IP address
3:Locate the IP
4:Server Info
5:Port Scanning
6:Identify File Type
7:Image Exif Data
8:Audio/Video Metadata
9:Exit
2
Enter Domain Name: itmuniversity.ac.in
IP Address of itmuniversity.ac.in is 3.6.105.176
Hostname of 3.6.105.176 is ('ec2-3-6-105-176.ap-south-1.compute.amazonaws.com', [], ['3.6.105.176'])
Info of itmuniversity.ac.in is [(('AddressFamily.AF_INET: 2', 0, 0, ''), ('3.6.105.176', 80))]
ghbn of itmuniversity.ac.in is ('itmuniversity.ac.in', [], ['3.6.105.176'])
Do you want to save this in a file? (y/n)
y
Enter the file name : resultsfile
File already exists. Do you want to overwrite it? (y/n)
n
File is appending to the existing file.
File saved successfully! in folder "Info_Folder" with filename: RESULTSFILE.txt

```

**Figure 6.** Output of ‘Get IP address’

To get the IP address of the target website we can use this feature it will give us the IP address of the target site with the hostname, as we can see in the screenshot above Fig.6. the hostname of given IP is amazonaws.

#### C. Locate IP address:



```

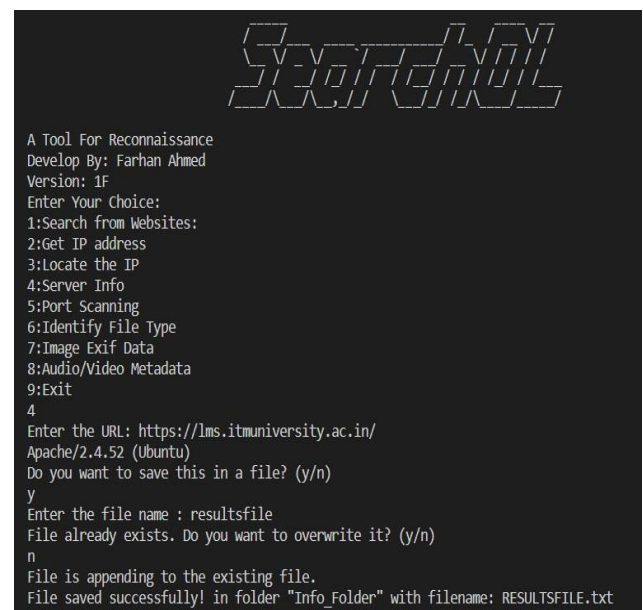
SearchOL
A Tool For Reconnaissance
Develop By: Farhan Ahmed
Version: 1F
Enter Your Choice:
1:Search from Websites:
2:Get IP address
3:Locate the IP
4:Server Info
5:Port Scanning
6:Identify File Type
7:Image Exif Data
8:Audio/Video Metadata
9:Exit
3
Enter IP Address: 3.6.105.176
CITY = Mumbai
STATE = Maharashtra
COUNTRY = IN
LatLng = [19.0728, 72.8826]
Do you want to save this in a file? (y/n)
y
Enter the file name : resultsfile
File already exists. Do you want to overwrite it? (y/n)
n
File is appending to the existing file.
File saved successfully! in folder "Info_Folder" with filename: RESULTSFILE.txt

```

**Figure 7.** Output of ‘Locate IP address’

If we want to find the location of the IP address, we can easily find this using “Locate the IP” feature of SearchOL, it will give us the city, state, country and latitude-longitude value.

#### D. Server Information:



```

SearchOL
A Tool For Reconnaissance
Develop By: Farhan Ahmed
Version: 1F
Enter Your Choice:
1:Search from Websites:
2:Get IP address
3:Locate the IP
4:Server Info
5:Port Scanning
6:Identify File Type
7:Image Exif Data
8:Audio/Video Metadata
9:Exit
4
Enter the URL: https://lms.itmuniversity.ac.in/
Apache/2.4.52 (Ubuntu)
Do you want to save this in a file? (y/n)
y
Enter the file name : resultsfile
File already exists. Do you want to overwrite it? (y/n)
n
File is appending to the existing file.
File saved successfully! in folder "Info_Folder" with filename: RESULTSFILE.txt

```

**Figure 8.** Output of ‘Server info’

When performing penetration testing it is good to have the information of version of Services or software, as in above Fig.8. you can see that the server uses Apache/2.4.7 (Ubuntu).

### E. Port Scanning:

```

SearchOL
A Tool For Reconnaissance
Develop By: Farhan Ahmed
Version: 1F
Enter Your Choice:
1:Search from Websites:
2:Get IP address
3:Locate the IP
4:Server Info
5:Port Scanning
6:Identify File Type
7:Image Exif Data
8:Audio/Video Metadata
9:Exit
5
Enter the IP address to scan: 3.6.105.176
Do you want to scan for all ports? (Y/N): n
Enter the port numbers to scan:[1 2 3 4 5]
22 80 100 443
22 : Open 0
80 : Open 0
100 : Closed 10060
443 : Open 0

```

**Figure 9.** Output of ‘Port Scanning’

In second phase of ethical hacking that is scanning phase “Port scanning is very important”, SearchOL gives us the information that which port is open or which port is closed.

### F. Identify File Type:

```

SearchOL
A Tool For Reconnaissance
Develop By: Farhan Ahmed
Version: 1F
Enter Your Choice:
1:Search from Websites:
2:Get IP address
3:Locate the IP
4:Server Info
5:Port Scanning
6:Identify File Type
7:Image Exif Data
8:Audio/Video Metadata
9:Exit
6
Enter file name: F:\1 SearchOL - Journal\Code files Raw\b.exe
File type is: JPG
Do you want to save this in a file? (y/n)
y
Enter the file name : resultsfile
File already exists. Do you want to overwrite it? (y/n)
n
File is appending to the existing file.
File saved successfully! in folder "Info_Folder" with filename: RESULTSFILE.txt

```

**Figure 10.** Output of ‘Identify File Type’

Sometimes it happens that the malicious user hide the important stuff in a file and change the extension of file so that it will not capture by investigator. Our tool SearchOL can identify the true extension of a file. SearchOL can identify file types of following file types: PNG, PDF, JPG, RTF, RAR, DOCX, PPTX, XLSX

### G. Extract Image EXIF data:

```

SearchOL
A Tool For Reconnaissance
Develop By: Farhan Ahmed
Version: 1F
Enter Your Choice:
1:Search from Websites:
2:Get IP address
3:Locate the IP
4:Server Info
5:Port Scanning
6:Identify File Type
7:Image Exif Data
8:Audio/Video Metadata
9:Exit
7
Enter the image name with extension or image path: F:\1 SearchOL - Journal\Code files Raw\c.jpg
('Gwalior', 'Madhya Pradesh', 'India', 'in', '474001')
ImageWidth : 3120
ImageLength : 3120
ResolutionUnit : 2
ExifOffset : 210
Make : OPPO
Model : OPPO A33
Orientation : 0
DateTime : 2023:01:12 19:05:29
YCbCrPositioning : 1
XResolution : 72.0
YResolution : 72.0
ExifVersion : b'0210'
InteropIndex : R98
DateTimeOriginal : 2023:01:12 19:05:29
DateTimeDigitized : 2023:01:12 19:05:29
ComponentsConfiguration : b'\x01\x02\x03\x00'
ShutterSpeedValue : 4.058

```

**Figure 11.** Output of ‘Image Exif Data’

Images contains some hidden data about itself called EXIF data, this data can be extract using SearchOL, it can have GPS location, date, time information, etc.

### H. Extract audio/video metadata:

```

SearchOL
A Tool For Reconnaissance
Develop By: Farhan Ahmed
Version: 1F
Enter Your Choice:
1:Search from Websites:
2:Get IP address
3:Locate the IP
4:Server Info
5:Port Scanning
6:Identify File Type
7:Image Exif Data
8:Audio/Video Metadata
9:Exit
8
Enter the path of the audio/video file: F:\1 SearchOL - Journal\Code files Raw\test.mp3
album: None
albumartist: None
artist: None
audio_offset: 0
bitrate: 320.0
channels: 2
comment: None
composer: None
disc: None
disc_total: None
duration: 4.226775
extra: {}
filesize: 169199
genre: None
samplerate: 48000
title: None
track: None
track_total: None

```

**Figure 12.** Output of ‘Audio/Video Metadata’

Like images audio/video also has metadata which is easily extract using SearchOL.

## I. Results text file:

```

RESULTSFILE.txt X
Info_Folder > RESULTSFILE.txt
1 SEARCH RESULT FOR: Farhan Ahmed
2 Total links => 20
3 Total sites => 11
4
5 SITE LINKS:
6 [+] :https://en.wikipedia.org/wiki/Farhan_Ahmed
7 [+] :https://www.espncricinfo.com/player/farhan-ahmed-1318270
8 [+] :https://www.cricbuzz.com/profiles/10879/farhan-ahmed
9 [+] :https://www.instagram.com/farhanahmdjovan/
10 [+] :https://www.cricq.com/players/2216078
11 [+] :https://www.linkedin.com/in/farhanahmed82
12 [+] :https://www.facebook.com/public/Farhan-Ahmed
13 [+] :https://www.espncricinfo.com/player/farhan-ahmed-950311
14 [+] :https://en.wikipedia.org/wiki/Farhan_Ahmed_Malhi
15 [+] :https://cricheroes.in/player-profile/256358/Farhan-Ahmed
16 [+] :https://www.instagram.com/farhanahmdjovan/?hl=en

```

Figure 13. Screenshot of text file of Results

```

RESULTSFILE.txt X
Info_Folder > RESULTSFILE.txt
41 RESULTS append:
42 *****
43
44 IP Address of itmunity.ac.in is 3.6.105.176
45 Hostname of 3.6.105.176 is ('ec2-3-6-105-176.ap-south-1.compute.amazonaws.com', [], ['3.6.105.176'])
46 Info of itmunity.ac.in is [(('AddressFamily.AF_INET: 2', 0, 0, ''), ('3.6.105.176', 80))]
47 ghn of itmunity.ac.in is ('itmunity.ac.in', [], ['3.6.105.176'])RESULTS append:
48 *****
49 Location of IP: CITY = Mumbai
50 STATE = Maharashtra
51 COUNTRY = IN
52 LatLng = [19.0728, 72.8826]
53 *****
54 Server Info of https://lms.itmunity.ac.in/ isApache/2.4.52 (Ubuntu)
55
56 RESULTS append:
57 *****file type is: JPG
58 *****
59 Image Exif Data:Location: ('Gwalior', 'Madhya Pradesh', 'India', 'in', '474001')
60 ImageWidth : 3120
61 ImageLength : 3120
62 ResolutionUnit : 2
63 ExifOffset : 210
64 Make : OPPO
65 Model : OPPO A33
66 Orientation : 0

```

Figure 14. Screenshot of text file of Results 2

## VI. CONCLUSIONS

As we see above that so much information is available on internet and this information if use for wrong purposes it will costs a lot. Many fraud calls, schemes, OTP frauds, Bank frauds done by just using your small-small information available online. Many hackers make fake accounts of victim user to deface him/her. The information we think useless, but it can have great impact on our life if goes in wrong hands. This paper provides the python-based OSINT tool "SearchOL" to gather important links related to the input keyword from Google, Ask, Bing and Yahoo search engines, get IP address, find the location of IP address, server information, port scanning of given IP, identify right file type, Extract image Exif data, Extract audio/video metadata and save them easily in a text file for further analysis. The tool we present can be used by penetration testers to look for the sensitive information released on internet. So that they can take appropriate measures to protect the sensitive information.

The future scope of this tool is that we can add more features related to footprinting and reconnaissance, can make a GUI

for this, also add some more search engines[22] to gather some more information, add secure web browsing[25] features and many more.

## Acknowledgment

First and foremost, I give thanks to Allah, the Almighty, on whom we ultimately rely for guidance and nourishment. Second, I want to thank my parents and sisters, who prayed for me to deal with every challenge and lead me in the right direction. I also want to express my sincere gratitude to my mentor, Prof. Pallavi Khatri, whose guidance, careful reading, and constructive criticism were helpful. I would want to express my sincere gratitude to her for her prompt and effective participation in helping me mould this research into its final form. Additionally, I like to thank Dr. Geetanjali Surange and Dr. Animesh Kumar Agrawal for their valuable contributions.

I also like to thank ITM University, Gwalior for giving me the intellectual foundation I needed to pursue this course of study.

## References

- [1] Ushmani, Azhar. "Ethical hacking." *International Journal of Information Technology (IJIT)* 4.6 (2018).
- [2] Glassman, Michael, and Min Ju Kang. "Intelligence in the internet age: The emergence and evolution of Open Source Intelligence (OSINT)." *Computers in Human Behavior* 28.2 (2012): 673-682.
- [3] Shreya, Shruti & Kumar, N. Suresh & Rao, Konda & Rao, Bheesetty. (2020). Footprinting: Techniques, Tools and Countermeasures for Footprinting. *Journal of Critical Reviews*. 7. 2019-2025. 10.31838/jcr.07.11.311.
- [4] A. Roy, L. Mejia, P. Helling and A. Olmsted, "Automation of cyber-reconnaissance: A Java-based opensource tool for information gathering," 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST), 2017, pp. 424-426, doi: 10.23919/ICITST.2017.8356437.
- [5] Shanto Roy, Nazia Sharmin, Jaime C. Acosta, Christopher Kiekintveld, and Aron Laszka. 2022. Survey and Taxonomy of Adversarial Reconnaissance Techniques. *ACM Comput. Surv.* Just Accepted (April 2022).
- [6] Boyanov, Petar Kr. "Implementation of the web based platforms for collecting and foot printing IP information of hosts in the Computer networks and systems" *Space Research and Technology Institute-BAS, Bulgaria Konstantin Preslavsky University-Faculty of Technical Sciences Association Scientific and Applied Research* 16 (2019): 42.
- [7] Sirisuriya, Scm de S. "A Comparative Study on Web Scraping." (2015).
- [8] Sambhe, Nilesh, Piyush Varma, Arpan Adlakhiya, Aditya Mahakalkar, Nihal Nakade, and Renuka Lakhe. "Using OSINT to gather information about a user from multiple social networks, *Information Technology in Industry* 9, no. 2 (2021): 207-211.
- [9] Yuan, Chao, et al. "The design of large scale IP address and port scanning tool." *Sensors* 20.16 (2020): 4423.
- [10] Tesic, Jelena. "Metadata practices for consumer photos." *IEEE MultiMedia* 12.3 (2005): 86-92.



- [11] Ramadhan, Rizdqi Akbar, Redho Maland Aresta, and Dedy Hariyadi. "Sudomy: information gathering tools for subdomain enumeration and analysis." IOP Conference Series: Materials Science and Engineering. Vol. 771. No. 1. IOP Publishing, 2020.
- [12] Hwang, Yong-Woon, et al. "Current status and security trend of OSINT." Wireless Communications and Mobile Computing 2022 (2022).
- [13] Nyeste, Peter. "The use of the open source intelligence in the criminal investigations." CASOPIS NAUOA-SERIA PRAVO 21.1 (2020): 1-10.
- [14] A. Kanta, I. Coisel, and M. Scanlon, "A survey exploring open source Intelligence for smarter password cracking," Forensic Science International: Digital Investigation, vol. 35, Article ID 301075, 2020
- [15] Mercado, Stephen C. "Sailing the Sea of OSINT in the Information Age." Secret intelligence: A reader 78 (2009).
- [16] Schwarz, Klaus, and Reiner Creutzburg. "Design of Professional Laboratory Exercises for Effective State-of-the-Art OSINT Investigation Tools-Part 3: Maltego." Electronic Imaging 2021.3 (2021): 45-1.
- [17] Kalpakis, George, et al. "OSINT and the Dark Web." Open Source Intelligence Investigation. Springer, Cham, 2016. 111-132.
- [18] Steele, Robert D. "1997 OSINT What Is It – Why Is It Important to the Military (White Paper)." Academia.edu [www.academia.edu/9817888/1997 OSINT What Is It Why Is It Important to the Military White Paper](http://www.academia.edu/9817888/1997_OSINT_What_Is_It_Why_Is_It_Important_to_the_Military_White_Paper) .
- [19] Parmar, Mayur. (2019). Google Dorks -Advance Searching Technique. 10.13140/RG.2.2.24202.62404.
- [20] Chandra, R.V. and Varanasi, B.S., 2015. Python requests essentials. Packt Publishing Ltd.
- [21] Pilgrim, M., 2004. Exceptions and File Handling. In Dive Into Python (pp. 97-120). Apress, Berkeley, CA.
- [22] Croft, W.B., Metzler, D. and Strohman, T., 2010. Search engines: Information retrieval in practice (Vol. 520, pp. 131-141). Reading: Addison-Wesley.
- [23] Richardson, L., 2007. Beautiful soup documentation. Dosegljivo: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>. [Dostopano: 7. 7. 2018].
- [24] Van Rossum G, Drake FL. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace; 2009.
- [25] Tang, S., 2011. Towards secure web browsing. University of Illinois at Urbana-Champaign.



**Geetanjali Surange**, Associate Professor in Department of Computer Science and Applications, ITM University, Gwalior has done her MCA from Barkatullah University, Bhopal Ph.D. in the field of IoT forensic. She has 23 years of teaching experience in teaching UG and PG courses. Her area of interest is Cyber Security, Cyber Forensic, Computer Networks, Machine learning.



Dr Animesh Kumar Agrawal born on 21<sup>st</sup> Sept 1972 is a freelance researcher who was earlier working as an Associate Professor (Cyber Security) in National Forensic Sciences University, New Delhi. He has a PhD in Mobile Forensics. He is a Certified Ethical Hacker, Certified Hacking Forensic Investigator and a BSI Lead Auditor. His research interests are in the area of GPU programming, cyber security, drone security and mobile forensics. He has 40 research papers and counting to his credit which have appeared in IEEE Conferences, Springer Conferences, Springer Lecture Note Book and Scopus indexed Journals.

## Author Biographies



**Farhan Ahmed** was born in Gwalior, Madhya Pradesh, India on 16th of December 2002. He is currently doing his B. Tech. in Computer Science with a specialization in Cyber Forensics from ITM University, Gwalior, India (he will pass out in 2024).



**Dr. Pallavi Khatri**, born on 8<sup>th</sup> sept 1974 is currently Professor of ITM university, Gwalior. She earned her Ph.D. in 2014 in Mobile Adhoc networks. Her research interests include mobile ad-hoc network, Wireless Sensor Networks and Cyber security. More than 70 research articles of her has been published in renowned international conferences and renowned journals. Currently she is working in the area of Cyber security and Digital Forensics.