Open-Source Intelligence (OSINT) Investigation in Facebook

Pranesh Kumar Narasimhan[1], *Chinmay Bhosale*[1] *, Muhammad Hasban Pervez*[4] ⓞ, *Najiba Zainab Naqvi*[4] ⓞ *, Mert Ilhan Ecevit*[4] ⓞ, *Klaus Schwarz*[1,3] ⓞ, *Reiner Creutzburg*[1,2] ⓞ

[1] *SRH Berlin University of Applied Sciences, Berlin School of Technology, Ernst-Reuter-Platz 10, D-10587 Berlin, Germany Email: npranesh26@gmail.com, chinmaybhosale@gmail.com, klaus.schwarz@srh.de, reiner.creutzburg@srh.de*

[2] *Technische Hochschule Brandenburg, Department of Informatics and Media, IT- and Media Forensics Lab, Magdeburger Str. 50, D-14770 Brandenburg, Germany, Email: creutzburg@th-brandenburg.de*

[3] *University of Granada, Faculty of Economics and Business, P.° de Cartuja, 7, ES-18011 Granada, Spain*

[4] *Kadir Has University, Cibali Mah. Kadir Has Cad., 34083 Fatih, Istanbul, Turkey Email: mert.ecevit@stu.khas.edu.tr, hasban.pervez15@gmail.com, najeebanaqvi94@gmail.com*

Keywords: Open Source Intelligence, OSINT, SOCMINT, Cybersecurity, Facebook, OSINT Investigation, Cybersecurity Training

Abstract

OSINT has come a long way now. It is still developing ideas, and many investigations are yet to happen shortly. The essential requirement for all OSINT investigations is information that is valuable data from an excellent source.

This paper discusses various tools and methodologies related to Facebook data collection and analyzing part of the collected data. At the end of the paper, the reader will get a deep and clear insight into the available techniques, tools, and descriptions of those tools present to scrape the data out of the Facebook platform and the types of investigations and analyses that the gathered data can do.

It should be noted that the paper presents the status of the OSINT Facebook investigation possibilities of November 2022.

Introduction

OSINT is an abbreviation for Open-Source Intelligence which is data collection or data gathering legitimately from publicly available sources about people, companies, or institutions used for connecting, evaluating, analyzing, and processing in the context of intelligence [1].

As years pass by and currently in the modern era, social media plays a massive role within all people, organizations, and institutions. This allows OSINT – Open-Source Intelligence to step in and grow on a big scale. This Introducing Social Media Intelligence which was shortly called SOCMINT. It becomes easy to get all the necessary information about each person and organization in a single place [2]. Not all the information can be scraped out of social media platforms and be used, there is a small line between legal and illegal scraping, and only the publicly available information can be scraped and used. Even social media platforms are having and allowing users to have types of profiles like complete public profiles and semi-public profiles, giving the users a lot of privacy options regarding their data and ensuring that not all the data will be publicly available for everyone.
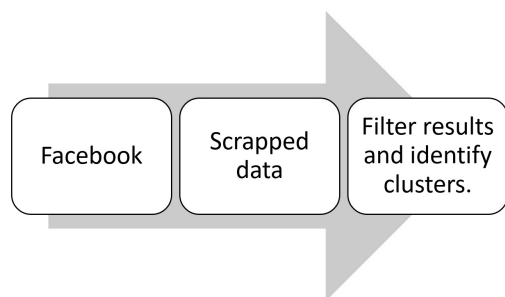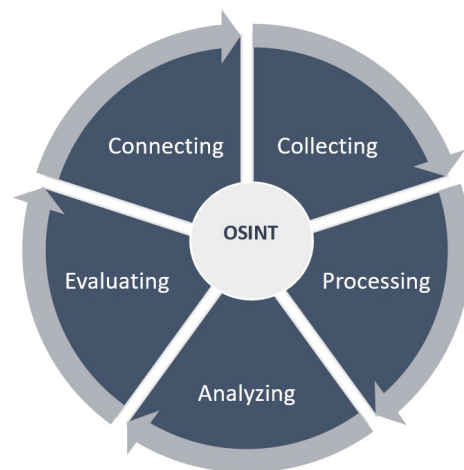

Figure 1. *OSINT scheme*

Figure 2. *Overview of Investigation*

IS&T International Symposium on Electronic Imaging 2023
Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications 2023

357-1

Facebook is a social media platform that also contributed on a large scale to the opportunity of SOCMINT. Facebook is the place where millions of people visit to open out their ideas and interests about all topics; this is the place where people digitally speak, argue, and show their emotions. Facebook is also a platform where general and informative topics like news, career options, and many more are widespread. Facebook has also widened its variety of options like Facebook Pages, Facebook Groups, Facebook Call, and Facebook Messenger [3]. All these options provided the SOCMINT to improve stronger and analyze better with related Facebook data.

Facebook Data

Facebook is the largest social media platform, with over 2.2 billion monthly active users. Since 2004, the organization's and its users' growth has been immense — the more users, the more data revolves around the platform.

The Facebook platform has a lot of data that are related to different kinds of topics. Some types of data that will be available or found by a user after logging in are listed in the table below [4].

List of Facebook Data

These are some of the Facebook data available on the platform and easily accessible by a normal user who logged in.

Data Scraping

Data scraping is the process of extracting data from the internet or document. Web scraping is a common way to extract valuable information from a website. There is a considerable difference between Data Scrapers and Data Crawlers. Data Crawling is like a search engine like Microsoft or Google. But Data Scrapers are different from them in the way it typically used and developed to extract or scrape data from particular websites. As we already discussed, a lot of valuable information is available on the websites. Numerous tools are available now in the market to scrape data from internet website platforms. Usually, there are a lot of scraper bots out there with different options in them like they only serve their purpose solely for that like

- Content Scraping
- Price Scraping
- Contact Scraping

Content Scraping Only the contents can be scraped out of the website and analyzed, and make a lot of statistics and decisions with those content.

Price Scraping This is a unique purposed way of scraping only the price data from a particular website where these kinds of data will be a huge help for a particular organization to know a lot about their competition and can make vice decisions with this information

Contact Scraping There are many numbers of websites that have contact information like email addresses and phone numbers with them. This tool is used to get many contact addresses and is used by most organizations for advertisement calls. This kind of information is a tremendous gift for all the spammers and scammers out there for having contact with you [5].
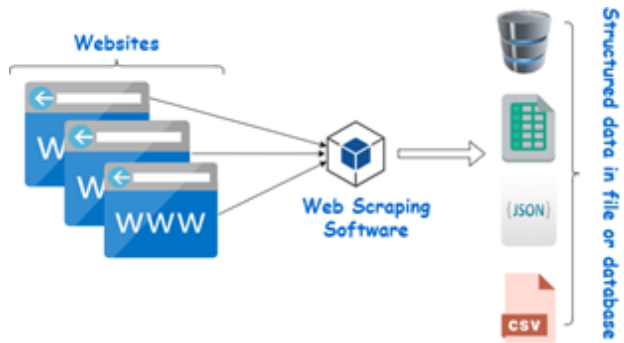

Figure 3. Web Scraping [3]

Facebook Scraping tools

Now, our main agenda is to concentrate mainly on the Facebook platform. Try to Scrape essential and valuable information from the platform and use those data for analyzing, evaluating, and predicting. Many tools are available online, particularly for Facebook data scraping. But not all the tools that are available are working. Currently, many tools are outdated and not working correctly, or in some cases, they don't give the exact output that we need. This paper will evaluate some of the currently working tools and their ease of use. Some of the tools that are working correctly are listed here.

- APIFY
- OSINT Combine
- Sowsearch
- LookupID
- Whopostedwhat
- Nairaland
- Facepager

Figure 4. Facebook Scraping (source: Facebook Scraper, [25]

APIFY:

APIFY is an online website/platform where some web scraping techniques are available. They also provide web scraping for our Facebook platform like • Facebook Page Scraper • Facebook Latest Posts Scraper • Facebook Latest Comments Scraper Let's see Facebook's latest comments on scraper methods from the above three options. This tool gives us a satisfactory output, like what we need and what is appropriate. This tool is very easy to use and does not take much time to compute and give us the output [6].

Step 1:

Login to the console using this website `https://console.apify.com/`

Step 2:

357-2

IS&T International Symposium on Electronic Imaging 2023
Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications 2023

Search and select the "Facebook Latest Comments Scraper". The select will redirect and pop up you to the scraper the page looks like. Figure – 5. This page has many options, like Input, and many options for the actors; actors are nothing but cloud-based software programs that can do many tasks that a human can do using a web browser. They can do anything from small tasks like filling in forms or unsubscribing from online services to more complex tasks like scraping and processing vast web pages. Many options for the actors, like Build – (latest, beta, or any versions), Timeout, and Memory (More memory means more CPU horse-power). They provide these many options so that the actors run smoothly and give us the necessary output.

Step 3:

There is an input field with the type URL. You need to type the URL of the post which you needed to be scraped and get the comments from it. And Click Start. Note: Type the full URL.
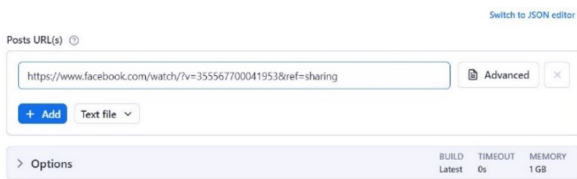

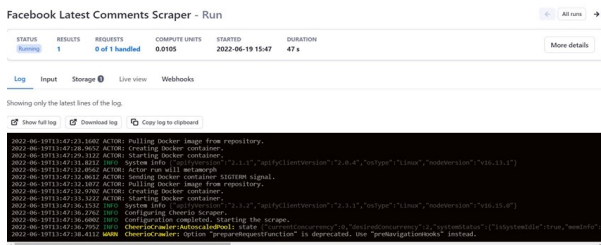**Figure 5.** *APIFY - Input field (Source: APIFY) [19]*

Figure 6. *APIFY – Running Page (Source: APIFY) [19]*

Step 4: After clicking Start, Figure - 6 shows up. The image indicates the ongoing process and shows you some code as it is running. You might notice they use the docker container and the cheerio scraper for scraping. This also shows us the Compute units, the time and date the process was initiated, and the process's duration. Step 5: The next page which loads up is the success page. Where they give the final status of the process, the number of results acquired, Requests handled by the process while executing, total compute units, and the total duration that had taken for the process to be executed. There is an Export tab on this page with many options for the dataset that scraped out. You have options like Excel, JSON, CSV, XML, HTML Tables, and RSS. The preview is only available for the JSON option. You can download the dataset in any option you want.

OSINT Combine

OSINT combine is a globally recognized company owned by an Australian veteran. This has a lot of free tools in it like

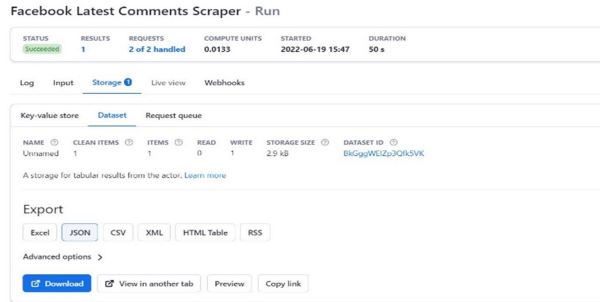- Sanctions Search
- TikTok Quick Search

Figure 7. *APIFY Success Page (Source: APIFY) [19]*

Figure 8. *APIFY – Output - JSON (Source: APIFY) [19]*

- Social Geo Lens
- Instagram Explorer
- World Social Media Platforms
- Snapchat Multi-viewer
- Google Analytics ID Explorer
- OSINT Collection Schema
- Reverse Image Analyzer
- Whatsmyname Username Tool
- Data Visualization Tool
- Reddit Post Analyzer

This tool helps search multiple keywords from a list and multiple profiles simultaneously. This tool provides different options for scraping data from the Facebook platform [7].

- Multi-search mutual friends
- Multi-keyword search on a profile
- Multi-keyword search on a page
- Multi-keyword search on a group
- Multi-keyword search for photos

Multi-Search mutual friends:

There are two input fields in this option. The first field should be filled by the Target Profile ID, which is the person's ID that one needs to find out. The next field is the Friends IDs to search. Now the tool will run a search of Target ID on all the profile IDs one provides in the Friends IDs field and will give the output in another tab in one's logged-in Facebook account.

Multi-keyword search on a Profile, Page, or Group:

All three Multi-keyword searches on profile, page, and group are similar and take in similar input fields for their process. But gives the appropriate outputs as you expected for your entered inputs. This option actually has three input fields. The First field is a drop-down input field with Person, Page, and Group options.

IS&T International Symposium on Electronic Imaging 2023
Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications 2023

357-3

The second field is the Target ID that you want to be searched; let it be Person/Page/Group ID. The last field is the keyword list; one can enter as many keywords as one wants in this field. But one keyword in each line is the proper way to enter this field. Now this will process all the input fields that one has given and gives the appropriate output for the option one chooses in the first field in another tab in your own logged-in Facebook account.

Multi-keyword search for photos:

This option has only one input field, just the keywords list. One can enter as many keywords as possible but only one per line. This option searches all the keywords you enter line by line in the input field and gives the output like the other two options in a different tab in the logged-in Facebook account. To make all these multi-keyword searches work properly you must enable the pop-ups for this. You can also use the username or the page alphanumeric ID in the Target ID fields. The main advantage of this tool is that all the searches and processes are directly taking place on the Facebook platform, so nothing is stored anywhere, and there is no need to worry about storing and managing the process separately.
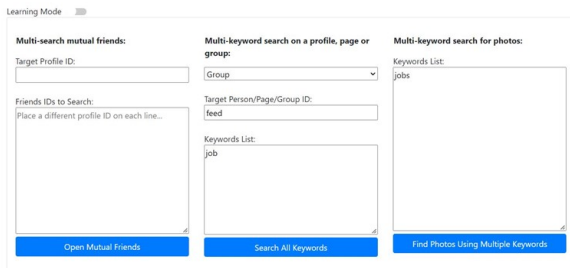

Figure 9. *OSINT Combine (Source: OSINT Combine) [20]*

Sowsearch:

Sowsearch is a tool that helps us find the correct data from the Facebook platform using all the appropriate filters. This tool is an experimental project for educational purposes and is still under development. This tool is nearly similar to the previously existing Graph search tool, which used the same process method. Sowsearch provides various options and filters for searching on the Facebook platform [8].

- Posts
- People
- Photos
- Pages
- Places
- Videos
- Events
- Top

These are the eight different types of searches that Sowsearch provides, each with unique filters for this search process.

Posts:

From Figure 10, one can see five filters have been mentioned and given for searching the posts.

- Most recent: This will search for all the recent posts
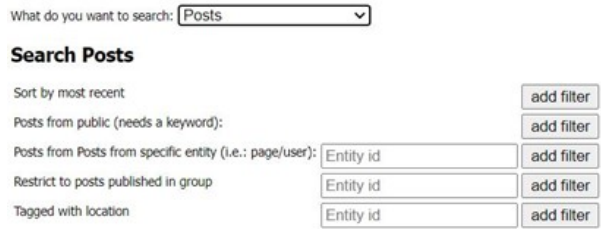- Posts from the public: All the posts from the public will be searched

Figure 10. *Sowsearch – Posts filters (Source: Sowsearch) [21]*

- Posts specific: Posts from a specific page or user will be searched by entering the specific ID.
- Restrict from group: Posts from the specific group will be searched and published by entering the specific group ID in the field.
- Tagged location: Search will be done based on the tagged location if the specific location ID is entered.

People:

From Figure 11, there are four different filters provided for finding people.

- City: Search of people based on the specific City ID.
- School: Search is based on the specific and unique school ID.
- Employer: Search is based on the correct employer ID.
- Friends with: This will search in the friend's list of this particular provided friends ID alone and provides the result.
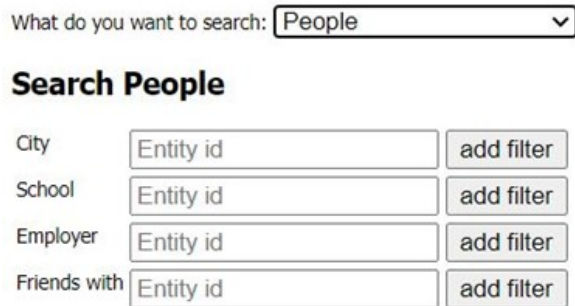

Figure 11. *Sowsearch – People filters (Source: Sowsearch) [21]*

Photos: To search Photos, Sowsearch provides a total of three filters with it. Figure – 12.

- Posted by: Search photos with the authors' profile ID.
- Tagged location: Photos searched with the unique location ID for the particular location tagged while posting.
- Photos seen: This filter will enable and show us the photos we have previously seen on the platform.

Pages: The search process of pages in Sowsearch has only two filters in it, namely,

- Verified: Searching only the verified pages

- Local business or place: Searching based on whether the page depends on company, organization or Institution, Brand or Product, Entertainment, Artist, Band, or Public Figure.

Videos: This is searching for videos from the platform using four different filters from Sowsearch. Figure – 13.

- Live videos: Searches and gives only the live videos.
- Episodes: Search is based only on the episodes available.
- Friends and group: This searches the videos only from your friends and groups of yours.
- Tagged location: This search is based on the unique location ID that is entered, which belongs to the location that has been used and tagged in the video while posting on the platform.

Events: This Events search has only one filter in the Sowsearch. Location- only the unique Location ID is added as the filter, and the search will be based on that. The output will be the events associated with that location will be listed.

Top:

Five different filters are present for this option in Sowsearch,

- Most recent: This will search for all the recent ones
- Posts specific: Posts from specific pages or users will be searched by entering the specific ID.
- Restrict from group: Posts from the specific groups will be searched and published by entering the specific group ID in the field.
- Tagged location: Search will be done based on the tagged location for the specific location ID that is entered

Places:

There are no specific or unique filters for Searching places that Sowsearch provides.

Figure 12. *Sowsearch – Photos filters (Source: Sowsearch) [21]*

For all the eight different varieties of options, Sowsearch had its own unique filters for itself. But there are also three common filters that are available for all the eight options, which are consistent and appropriate, • Start Date: Search is based on the start date • End Date: Search is based on the end date • Keywords: Keyword is used while searching The output of the Sowsearch is in two different ways. One is showing the URL on the same page in an alert pop-up box with an ok button on it, and the next way is Open the URL in a new window; this opens the output URL in a new tab and the search is shown in the logged in the Facebook platform in your browser.

Figure 13. *Sowsearch – Videos filters (Source: Sowsearch) [21]*

Figure 14. *Sowsearch – Common filters (Source: Sowsearch) [21]*

Facepager:

Facepager is a tool that was developed to fetch and scrape valuable information from YouTube, Twitter, and other websites. Facepager is a tool that also works for Facebook, and it does the critical part where most of the tools do; it fetches the data with all the required filters and gives us and exports the output in an excel file. After installing Running the Facepager, the Software opens like Figure- 16. It has a lot of options in the window, but we need minimal options to properly execute and fetch the information from the Facebook platform [9].
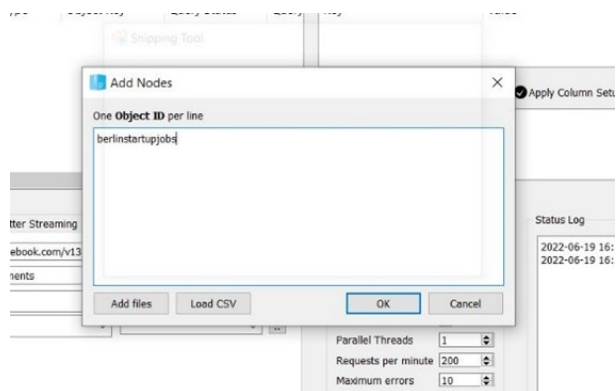
Figure 15. *Facepager – Add Nodes (Source: Facepager) [22]*

First, need to add the node, which is the basic requirement of the tool. The whole scraping process is done only with this unique node, which is nothing but the unique ID of a particular profile, page, group, or post that will be available at the end of the particular page's URL Figure - 15. After adding the node, choose the options Facebook in the list of mentioned websites. Then in

IS&T International Symposium on Electronic Imaging 2023
Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications 2023

357-5

the resource field, there will be lots of options, namely,

- Feed
- Likes
- Locations
- Milestones
- Photos
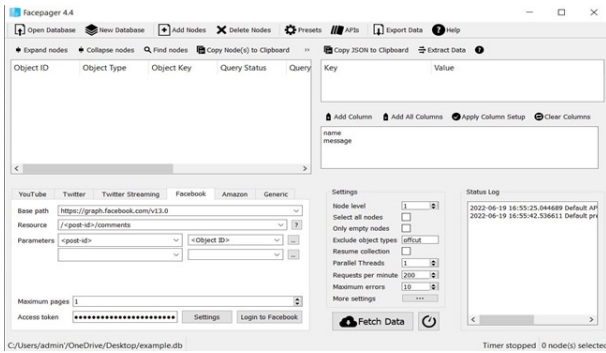- Posts
- Videos
- Comments
- Reactions
- Shared Posts.

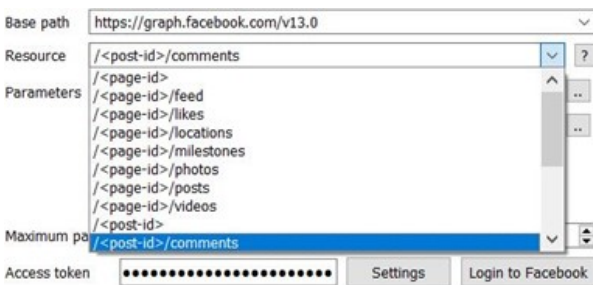**Figure 16.** *Facepager (Source: Facepager)* [22]

Figure 17. *Facepager – Resource (Source: Facepager)* [9]

After selecting the resources, the next important thing is to login to the Facebook account through the Facepager tool. So, clicking the login to Facebook button in the window opens a mini window of the Facebook login page like Figure – 18.

After logging in, the process is almost over with just one more click. Now click the Fetch Data button from the window. This will fetch and display the data from the Facebook platform in the above field. Now it's time to export the data; select Export data in the top tab and wait for the save screen to open. Now select the destination where you want the file and save it as CSV. This CSV file contains all the resources needed from the specific ID you added in the Add nodes section.

Practical Work

Selenium:

Apart from these tools, which are pre-existing out there on the internet, there is another way to get the data from the websites that one has wanted by using Selenium with Python and trying to access Facebook through webdriver to fetch all the essential information from the platform. You can access the Facebook
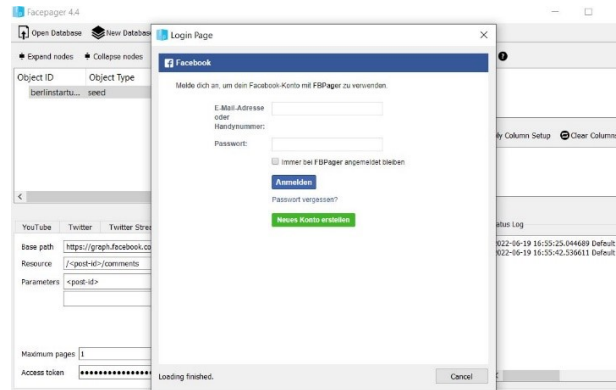

Figure 18. *Facepager – Login Page (Source: Facepager)* [22]

platform through the webdriver. For selecting and clicking the options, selenium in python has different find elements they are [10] • Locating by ID • Locating by Name • Locating by XPATH • Locating Hyperlinks by Link Text • Locating elements by Tag Name • Locating elements by Class Name • Locating elements by CSS Selectors With the help of their element finders, you can play around on the platform. One can log in, navigate through the platform, and search using all the provided Facebook filters. All the available elements in the platform can be selected. For example, I used selenium with Python to first login into the Facebook platform with my own credentials. Next is to navigate to a Facebook group that is more into career-providing opportunities —- then download all the media documents, like photos, to my local folder. My main idea is that usually, not all the job requirements and all the career opportunities will be written in sentences in a post; most will be like in a job opening advertisement in a photo. So, my main aim was to download all the photos of the advertisements into the folder. Then convert them into the written text format and append them to a file for further use. If you see in Figure - 19, this is the folder where all the images have been downloaded and saved in my local folder, where it can be used for the following conversion process.
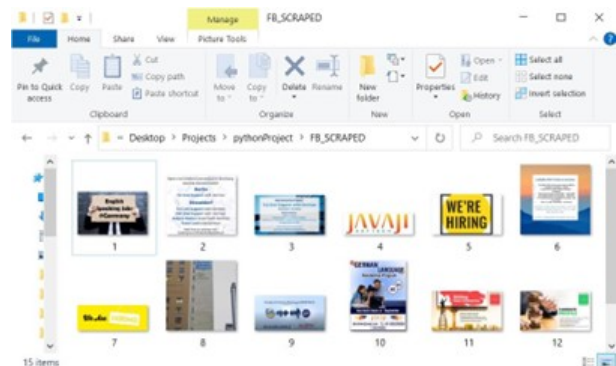

Figure 19. *Scraped Photos from Facebook*

Tesseract:

After downloading all the photos, my next process was to extract all the texts that were available in the photos and then append them to a text file. One can use Tesseract OCR, an open-source text recognition engine, for extracting. This OCR stands for

Optical Character Recognition, which means converting a two-dimensional image of text to a machine-printed text. OCR has many processes in it. They are,

- Preprocessing of Image
- Text Localizing
- Character Segmentation
- Character Recognition
- Post Processing

You can use Tesseract straight or the API to extract the printed text from the image - support for multiple languages. Tesseract does not have a built-in graphic interface. Many programming languages have support for Tesseract and also multiple frameworks too. One can mix it with an existing layout analysis to detect text in large documents or use an external text detector to detect text from an image of a single line of text [11][12].
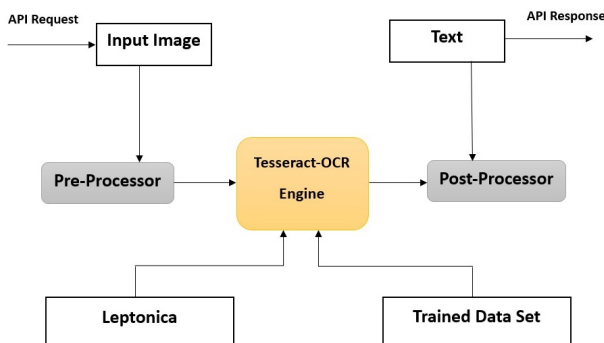
OCR Process Flow

Figure 20. *OCR Process Flow (Source: Nanonets [10]*
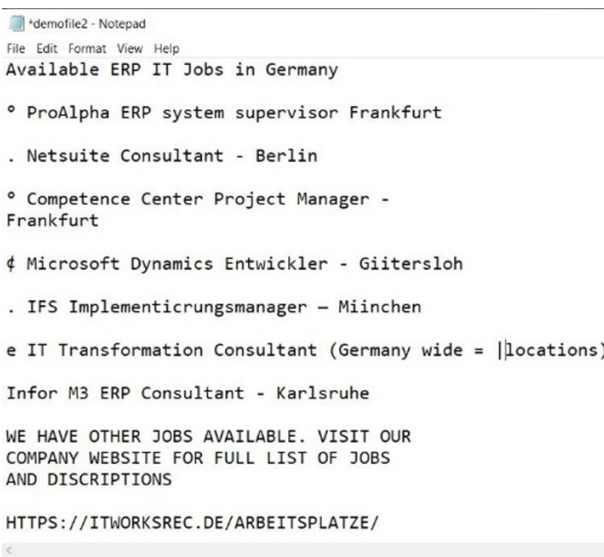
Figure 21. *Extracted text in the text file*

The above Figure - 20 is the process flow of the OCR process. Tesseract includes a NN (Neural Network system) configured as a text line recognizer. In Tesseract, CNN (Convolutional Neural Network) is used to detect images containing a single character. Figure – 21 shows the output texts printed in the text file, extracted from an image downloaded from the Facebook platform using selenium with Python as a tool data extractor. Basically, an Open Source Intelligence Investigation on an advertisement photo of career opportunities providing group.

The below table is the total overview of some of the OSINT tools that are present now and working correctly for extracting valuable and essential data from the Facebook platform. The table contains the name of the tool, the platform it works on, the type of data that can be extracted, the type of input needed for the process to perform, the prerequisites required for running the tool, and the output of the tool.

Figure 22. *Overview of the total concept*

IS&T International Symposium on Electronic Imaging 2023
Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications 2023

357-7

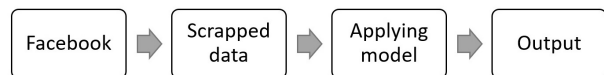| Type of Data	Features
1. Following	A list of people you follow
2. Likes on other's posts	Posts (Photos, Videos) that you have liked
3. Likes on your posts	Likes on your posts (photos, videos) from others
4. About me	All the personal information like Family and Relationships, Work and Education, and Life events in your profile under the About section.
5. About in others' profiles	All the personal information like Family and Relationships, Work and Education, and Life events in a friend or other users' profile under the About section.
6. Posts from you	Your posts (Photos, Videos)
7. Posts from others	Posts (Photos, Videos) from others.
8. Pinned posts	Posts pinned by you
9. Saved posts	Posts saved by you
10. Recent Activities	Your latest activity in the Facebook platform
11. Registration Date	Account opening date
12. Friends	Your Friends list
13. Mutual Friends	Mutual friend in between you and your other friend
14. Tagged posts	Posts (Photos, Videos) from others, and you are mentioned in it
15. Comments on other's posts	Comments on the posts (Photos, Videos) from others
16. Comments in your posts	Comments on the posts (Photos, Videos) of your own
17. Share other's posts	Share count and the people who shared the posts (Photos, Videos) from others
18. Share in your posts	Share and count the people who shared your posts (Photos, Videos)
19. Search	Searches made by you
20. Status	Status updated by you
21. Time zone	Selected time zone by you
22. Time Viewed	Individual amount of viewed video time
23. Time Spent	Time spent on watching videos
24. Shows	History of videos viewed
25. Show pages	History of pages visited
26. Profile visit	History of recent profile visits
27. Platforms	Platforms used to log in to Facebook
28. Phone numbers	Mobile numbers added to your profile
29. Mobile	Country code and service provider associated with the provided phone number used for logging in to the Facebook platform.
30. Religious views	Content added under the religious views in the profile
31. Political views	Content added under the political views in the profile
32. Friend request	Both sent and received pending friend requests.
33. Recommended	Pages, Groups, and events you recommend to others.
34. Pages	List of pages you own
35. Blocked contacts	Contacts you have blocked
36. Messenger	List of friends or people in the messenger and the list of blocked contacts only on messenger will also be available.
37. Messages	The sent and received messages on Facebook messenger. But the deleted messages won't be available as it is deleted from your account.
38. Marketplace services	History of services you have recently visited in the marketplace.
39. Marketplace categories	Recently searched or viewed categories in the marketplace.
40. Marketplace chat	Recent interactions on marketplace
41. Marketplace Items	All the items in the marketplace and the history of items you have searched and looked for in the marketplace will be available.
42. Notifications	All the notifications from Facebook friends, pages, groups, and marketplace will be available.
43. Login	The date, time, location, and IP address while login will also be available
44. Logout	The date, time, location, and IP address while logout will also be available
45. Games	History of the list of games that you play on the Facebook platform often.
46. Group	List of groups you own and list of groups you belong to.
47. Group visit	History of groups that you have visited recently.
48. Events	List of events you've joined and list of events you have been invited to attend.
49. Event History	List of event pages that were viewed recently.
50. Email	Your email address that is added to your account and the history of when that email got verified.

Table 1. List of Facebook Data

357-8

IS&T International Symposium on Electronic Imaging 2023
Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications 2023

Tool	Platform	Data Type	Input Type	Prerequisites	Output	Usage
APIFY	Web-Based	Unique Data like Posts, Comments, Pages	Facebook URL	Browser with Logged in Facebook, Facebook URL	Easy and simple output of data in the format of (CSV, JSON, XML, Excel, RSS)	Free/Paid
OSINT Combine	Web-Based	Filtered search for Mutual Friends, Photos, Persons, Pages, Groups	Unique Facebook ID	Browser with Logged in Facebook, a Unique Facebook ID	Facebook URL with all the searched filters on it.	Free
Sowsearch	Web-Based	Filtered search for Posts, People, Photos, Pages, Places, Videos, Events, and Top	Facebook unique ID	Browser with Logged in Facebook, Facebook ID	Facebook URL with all the searched filters on it	Free
Facepager	Host Computer Based	Unique Data like Posts, Comments, Pages, Likes, Shares, Shared posts, Locations, Milestones, Videos, and Reactions	Facebook Account unique Identifier	Properly Installed Facepager from the browser, Facebook account credentials, Unique Account Identifier	Easy and simple output data in the format of CSV in Excel	Free

Table 2. Overview of Tools

IS&T International Symposium on Electronic Imaging 2023
Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications 2023

357-9

Conclusions and Future work

This paper concludes and describes all the tools available for scraping all the valuable data and important information from the Facebook platform. Also, all the mentioned tool's technicalities are mentioned in this paper with clear explanations. In the end, a small practical example has been demonstrated, like scraping the advertisement photos from the career opportunities group through selenium with Python and downloading it to the local folder. Then apply, a small open-source algorithm Tesseract is used on the scraped and downloaded images to detect the sentences written on the image and append them to a text file. These are some of the Open-Source Intelligence Investigation like how many advertisements have all the valuable and exactly needed information for a Job Recruitment written on the posters.

The Future work from this can be done is getting to know more updated and paid versions of the tools that will help scrape the Facebook platform and extract the information in a proper format with multiple options. Then like the image-text conversion, many more AI models are out there like image classification, Image segmentation, and by selecting and downloading many more data other than images like posts, likes, comments, and shares and doing sentiment analysis on the comments to get a review for a post.

Acknowledgments

References

[1] Florian Schaurer and Jan Störger, "The Evolution of Open-Source Intelligence (OSINT)," From AFIO's The Intelligencer Journal of U.S. Intelligence Studies Volume 19, Number 3.

[2] Omand, David & Bartlett, Jamie & Miller, Carl. (2012). "Introducing social media intelligence (SOCMINT)." Intelligence and National Security. `https://doi.org/10.1080/02684527.2012.716965`, (Last access: Nov. 12, 2022).

[3] Carol A. Watson. "Social Searching: Using Social Media for Investigation, Litigation & Competitive Intelligence." University of Georgia School of Law, January 29, 2016.

[4] Facebook, Categories of available Facebook data, `https://www.facebook.com/help/930396167085762?cms_id=930396167085762`, (Last access: Nov. 12, 2022).

[5] Cloud Flare – Data Scraping, `https://www.cloudflare.com/`, (Last access: Nov. 12, 2022).

[6] APIFY – Web Scraping Solutions, Website: `https://apify.com/`, (Last access: Nov. 15, 2022).

[7] OSINT Combine – Facebook Search tool. Website: `https://www.osintcombine.com/facebook-search-tools`, ()Last access: Nov. 13, 2022).

[8] Sowsearch – Facebook Search Filter tool. Website: `https://www.sowsearch.info/`, (Last access: Nov. 12, 2022).

[9] Facepager – Facebook Data Extractor tool. Webpage: `https://github.com/strohne/Facepager`, (Last access: Nov. 12, 2022).

[10] Nanonets - Website, `https://nanonets.com/`, (Last access: Nov. 12, 2022).

[11] David Mathew Thomas, Sandeep Mathur, "Data Analysis by Web Scraping using Python." Amity Institute of Information Technology Amity University (AUUP), Sec-125, Noida.

[12] Tesseract – Open-Source text recognition Engine. Website: `https://github.com/tesseract-ocr/tesseract`, (Last access: Nov. 15, 2022).

[13] Ray Smith, "An Overview of the Tesseract OCR Engine". Google Inc. 2007.

[14] AlertMedia, "What Businesses Need to Know About Open Source Intelligence (OSINT)" `https://www.alertmedia.com/blog/open-source-intelligence/` (Last access: Nov. 12, 2022).

[15] Signal, "The Pivotal Role of OSINT for Effective Emergency Management Pivotal." `https://www.getsignal.info/blog/osint-emergency-management` (Last access: Nov. 12, 2022).

[16] Qadir, Junaid & Ali, Anwaar & Rasool, Raihan & Zwitter, Andrej & Sathiaseelan, Arjuna & Crowcroft, Jon. (2016): "Crisis Analytics: Big Data Driven Crisis Response." Journal of International Humanitarian Action. 1. `https://doi.org/10.1186/s41018-016-0013-9`, (Last access: Nov. 12, 2022).

[17] Echosec SYSTEMS LTD, "6 Reasons Why Open-Source Intelligence is Climbing the Priority Ladder" `https://www.echosec.net/blog/6-reasons-why-open-source-intelligence-is-climbing-the-priority-ladder` (Last access: Nov. 13, 2022).

[18] TechTarget, "Crisis Management" `https://www.techtarget.com/whatis/definition/crisis-management` (Last access: Nov. 15, 2022).

[19] InaSAFE, `http://inasafe.org/home/index.html` (Last access: Nov. 17, 2022).

[20] Australian Government, "New disaster management software released worldwide," `https://www.ga.gov.au/news-events/news/latest-news/new-disaster-management-software-released-worldwide` (Last access: Nov. 17, 2022).

[21] Pranantyo, Ignatius & Fadmastuti, Mahardika. (2014). "InaSAFE applications in disaster preparedness." `https://doi.org/10.1063/1.4915053`, (Last access: Nov. 12, 2022).

[22] JSTOR, "4 Case Study 3 – Ushahidi, An Open Platform for Situation Awareness" `https://www.jstor.org/stable/resrep12574.7?seq=1` (Last access: Nov. 17, 2022).

[23] Ushahidi, "Crisis Mapping Haiti: Some Final Reflections" `https://www.ushahidi.com/about/blog/crisis-mapping-haiti-some-final-reflections` (Last access: Nov. 17, 2022).

[24] Wald, D.J., Worden, C.B., Thompson, E.M., Hearne, M. (2019). "Earthquakes, ShakeMap." In: Gupta, H. (eds) Encyclopedia of Solid Earth Geophysics. Encyclopedia of Earth Sciences Series. Springer, Cham. `https://doi.org/10.1007/978-3-030-10475-7_182-1`, (Last access: Nov. 17, 2022).

[25] Facebook Scraper, Website: `https://pypi.org/project/facebook-scraper/#history`, (Last access:

357-10

IS&T International Symposium on Electronic Imaging 2023
Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications 2023

Sept. 10, 2022).

[26] reliefweb, "USGS ShakeMap: Haiti Region - Tue Jan 12, 2010 21:53:09 GMT" https://reliefweb.int/map/haiti/usgs-shakemap-haiti-region-tue-jan-12-2010-215309-gmt (Last access: Nov. 17, 2022).

[27] Verified Market Research, "Open Source Intelligence (OSINT) Market Size And Forecast" https://www.verifiedmarketresearch.com/product/open-source-intelligence-osint-market/ (Last access: Nov. 19, 2022).

[28] USGS, "ShakeMap" https://earthquake.usgs.gov/data/shakemap/ (Last access: Nov. 19, 2022)

[29] Github, https://github.com/usgs/shakemap (Last access: Nov. 19, 2022)

[30] Ida Norheim-Hagtun, Patrick Meier: "Crowdsourcing for Crisis Mapping in Haiti." Innovations: Technology, Governance, Globalization 2010; 5 (4): 81–89. https://doi.org/10.1162/INOV_a_00046, (Last access: Nov. 12, 2022).

[31] Dave Chaffey, "Global social media statistics research summary 2022" https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/ (Last access: Nov. 19, 2022)

[32] Sahana Foundation, "Eden" https://sahanafoundation.org/products/eden/ (Last access: Nov. 20, 2022)

[33] Talkwalker, "Try the best crisis management tools" https://www.talkwalker.com/blog/best-crisis-management-tools (Last access: Nov. 20, 2022)

[34] CSIRO, "ESA: Software for Emergency Situation Awareness" https://data61.csiro.au/en/Our-Research/Our-Work/Safety-and-Security/Disaster-Management/ESA (Last access: Nov. 18, 2022)

[35] CSIRO, "Emergency Situation Awareness" https://www.csiro.au/en/research/technology-space/ai/emergency-situation-awareness (Last access: Nov. 19, 2022)

[36] CSIRO ESA, "AUS tweets analysed by ESA over last 4 days" https://esa.csiro.au/aus/index.html (Last access: Nov. 17, 2022)

[37] Statista, "Number of deaths from natural disaster events globally from 2007 to 2021" https://www.statista.com/statistics/510952/number-of-deaths-from-natural-disasters-globally/ (Last access: Nov. 20, 2022)

[38] Our World in Data, "Global reported natural disasters by type, 1970 to 2019" https://ourworldindata.org/grapher/natural-disasters-by-type (Last access: Nov. 20, 2022).

[39] Our World in Data, "Natural Disasters by Hannah Ritchie and Max Roser" https://ourworldindata.org/natural-disasters (Last access: Nov. 20, 2022)

[40] Australian Government, "Australian Government Crisis Management Framework (AGCMF)" https://www.pmc.gov.au/resource-centre/national-security/australian-government-crisis-management-framework, (Last access: Nov. 21, 2022)

[41] Australian Government, "The use of social media in coun-trywide disaster risk reduction public awareness strategies" https://knowledge.aidr.org.au/resources/ajem-jan-2015-the-use-of-social-media-in-countrywide-disaster-risk-reduction-public-awareness-strategies/ (Last access: Nov. 20, 2022)

[42] Cabinet Office, "Disaster Management in Japan" https://www.bousai.go.jp/en/documentation/white_paper/pdf/2021/R3_hakusho_english.pdf (Last access: Nov. 21, 2022).

[43] National Library of Medicine, "Disaster Management in Japan" https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5059167/ (Last access: Nov. 21, 2022).

[44] PEARY, Brett & Shaw, Rajib & TAKEUCHI, Yukiko. (2012). "Utilization of Social Media in the East Japan Earthquake and Tsunami and its Effectiveness." Journal of Natural Disaster Science. 34. 3-18. https://doi.org/10.1186/s41018-016-0013-9, (Last access: Nov. 21, 2022).

[45] B. Birregah et al., "Multi-layer Crisis Mapping: A Social Media-Based Approach," 2012 IEEE 21st International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, 2012, pp. 379-384, https://ieeexplore.ieee.org/document/6269762, (Last access: Nov. 20, 2022).

[46] ACM Digital Library, "A sensitive Twitter earthquake detector" https://dl.acm.org/doi/10.1145/2487788.2488101, (Last access: Nov. 20, 2022)

[47] Schwarz, Klaus; Franziska Schwarz, Reiner Creutzburg: "Conception and implementation of professional laboratory exercises in the field of open source intelligence (OSINT)". Proceed. Electronic Imaging Symposium 2020 (San Francisco, USA), Mobile Devices and Multimedia: Technologies, Algorithms & Applications Conference (MOBMU) 2020, https://doi.org/10.2352/ISSN.2470-1173.2020.3.MOBMU-278, (Last access: Nov. 22, 2022).

[48] Schwarz, Klaus; Reiner Creutzburg: "Design of Professional Laboratory Exercises for Effective State-of-the-Art OSINT Investigation Tools - Part 1: RiskIQ PassiveTotal". Proceed. Electronic Imaging Symposium 2021 (San Francisco, USA), Mobile Devices and Multimedia: Technologies, Algorithms & Applications Conference (MOBMU) 2021, https://doi.org/10.2352/ISSN.2470-1173.2021.3.MOBMU-043, Last access: Nov. 22, 2022).

[49] Schwarz, Klaus; Reiner Creutzburg: "Design of Professional Laboratory Exercises for Effective State-of-the-Art OSINT Investigation Tools - Part 2: Censys". Proceed. Electronic Imaging Symposium 2021 (San Francisco, USA), Mobile Devices and Multimedia: Technologies, Algorithms & Applications Conference (MOBMU) 2021, https://doi.org/10.2352/ISSN.2470-1173.2021.3.MOBMU-044, Last access: Nov. 22, 2022.

[50] Schwarz, Klaus; Reiner Creutzburg: "Design of Professional Laboratory Exercises for Effective State-of-the-Art OSINT Investigation Tools - Part 3: Maltego". Proceed. Electronic Imaging Symposium 2021 (San Francisco, USA), Mobile Devices and Multimedia: Technologies, Algorithms & Applications Con-

IS&T International Symposium on Electronic Imaging 2023
Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications 2023

357-11

ference (MOBMU) 2021, `https://doi.org/10.2352/ISSN.2470-1173.2021.3.MOBMU-045`, Last access: Nov. 22, 2022.

[51] Kant, Daniel; Reiner Creutzburg: 'Investigation of risks for Critical Infrastructures due to the exposure of SCADA systems and industrial controls on the Internet based on the search engine Shodan". Proceed. Electronic Imaging Symposium 2020 (San Francisco, USA), Mobile Devices and Multimedia: Technologies, Algorithms & Applications Conference (MOBMU) 2020 `https://doi.org/10.2352/ISSN.2470-1173.2020.3.MOBMU-253`, Last access: Nov. 22, 2022.

[52] M. S. Wong, N. Hideki and N. Yasuyuki, "The Incorporation of Social Media in an Emergency Supply and Demand Framework in Disaster Response," 2018 IEEE Intl. Conf. on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications `https://ieeexplore.ieee.org/document/8672243`, (Last access: Nov. 20, 2022).

[53] T. Sakaki et al., "The possibility of social media analysis for disaster management," 2013 IEEE Region 10 Humanitarian Technology Conference, 2013, pp. 238-243, `https://www.scopus.com/record/display.uri?eid=2-s2.0-84893406250&origin=inward&txGid=7adf7d88a2a5fe170927ab1110f2009f`, (Last access: Nov. 20, 2022).

Author Biography

Pranesh Kumar Narasimhan received his B. Tech in Computer Science from the Amrita School of Engineering (2020). He is pursuing his Master's in Computer Science focused on Big Data and Artificial Intelligence from SRH Berlin University of Applied Sciences (2021-2023). He has worked on more projects regarding Machine Learning since his undergraduate studies, and most of his work was focused on Big Data, Algorithms, and Machine learning.

Chinmay Bhosale received his M.Sc. degree in Computer Science from SRH Berlin University of Applied Sciences in 2022. His research work is focused on OSINT and Cyber Security.

Mohammad Hasban Pervez is pursuing a master's degree in management information systems while working as a researcher for KHAS CCIP (Kadir Has University Research Center on Cybersecurity and Critical Infrastructure Protection). He received his undergraduate degree in Telecommunications Engineering from N.E.D University of Engineering and Technology in Karachi, Pakistan, and has over two years of network experience.

Najiba Zainab Naqvi is pursuing a master's degree in Cybersecurity and submitted her Master's Project entitled Facebook investigation through OSINT for Entrepreneurs. She received her undergraduate degree in Electrical Computer Engineering from COMSATS University in Islamabad, Pakistan, and has over two years of IT experience.

Klaus Schwarz received his B.Sc. and M.Sc. in Computer Science from Brandenburg University of Applied Sciences (Germany) in 2017 and 2020, respectively. He is currently a Ph.D. student at the University of Granada, Spain. His research interests include IoT and smart home security, OSINT, mechatronics, additive manufacturing, embedded systems, artificial intelligence, and cloud security. As a faculty member, he is developing a graduate program in Applied Mechatronic Systems focusing on Embedded Systems at SRH Berlin University of Applied Sciences.

Reiner Creutzburg is a Retired Professor for Applied Computer Science at the Technische Hochschule Brandenburg in Brandenburg, Germany. Since 2019 he has been a Professor of IT Security at the SRH Berlin University of Applied Sciences, Berlin School of Technology. He is a member of the IEEE and SPIE and chairman of the Multimedia on Mobile Device (MOBMU) Conference at the Electronic Imaging conferences since 2005. In 2019, he was elected a member of the Leibniz Society of Sciences to Berlin e.V. His research interest is focused on Cybersecurity, Digital Forensics, Open Source Intelligence (OSINT), Multimedia Signal Processing, eLearning, Parallel Memory Architectures, and Modern Digital Media and Imaging Applications.

357-12

IS&T International Symposium on Electronic Imaging 2023
Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications 2023