Swapnal Shahil

Mail: swapnalsahil@gmail.com
Portfolio: https://swapnalshahil.github.io/
Github: https://github.com/swapnalshahil
Linkedin: https://www.linkedin.com/in/swapnalshahil/

SCoRe Lab - Open MF

Google Summer of Code 2021

About Me

I am currently in the 2nd year of my Bachelor’s Degree from the Indian Institute of Technology Guwahati (IIT G). I have been tinkering with the codes since the first year of my college. I love working on open-source projects and learn every time new things from projects and the community. I have good knowledge of React, Js, ES7, HTML, CSS with my interest in the web and I am contributing to this organization since December and will continue further in all my possible way.

Analytics API and UI Development

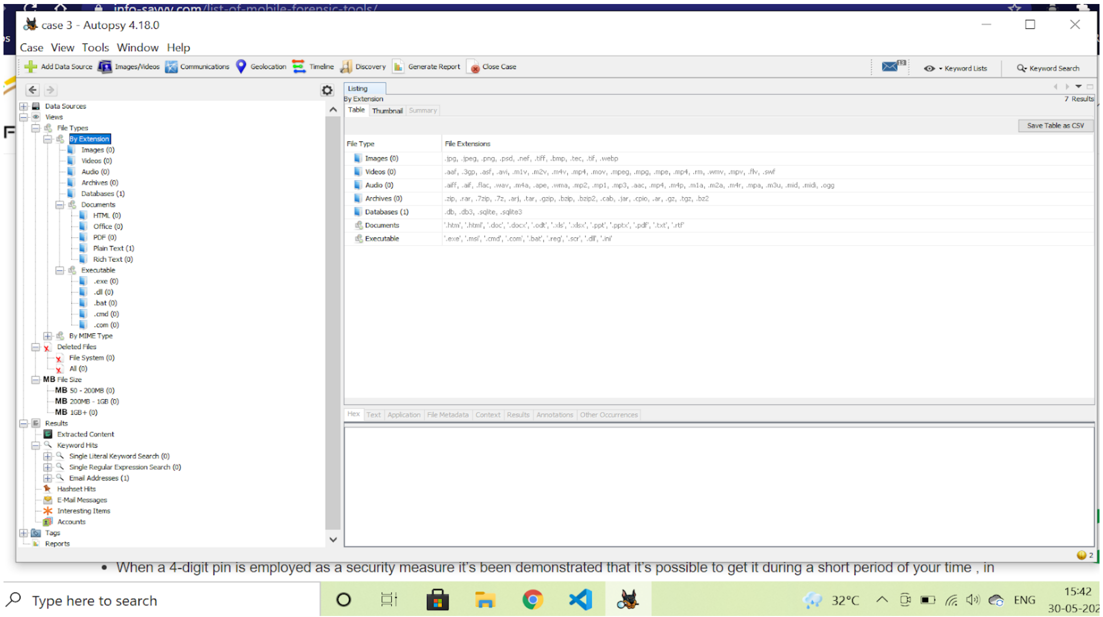
Research:

Available Analytics tool:

1. Sleuth Kit(+Autopsy):
    - It’s an open-source digital forensics toolkit that can be used for in-depth analysis of various file systems.

Features:

- List allocated and deleted ASCII and Unicode file names.
- Display file names and metadata structure
- Lookup file hashes in a hash database, such as the NIST NSRL, Hash Keeper, and custom databases that have been created with the 'md5sum' tool.
- Organize files according to their type.
- Multimedia - Extract EXIF from pictures and watch videos
- Give the final report in various modules like HTML, text, excel report.
- Also, it opens one case at a time and works on it.



File Type	File Extensions
Images (0)	.jpg, .jpeg, .png, .psd, .nef, .tiff, .bmp, .tec, .tif, .webp
Videos (0)	.aaf, .3gp, .asf, .avi, .m1v, .m2v, .m4v, .mp4, .mov, .mpeg, .mpg, .mpe, .mp4, .rm, .wmv, .mpv, .flv, .swf
Audio (0)	.aiff, .aif, .flac, .wav, .m4a, .ape, .wma, .mp2, .mp1, .mp3, .aac, .mp4, .m4p, .m1a, .m2a, .m4r, .mpa, .m3u, .mid, .midi, .ogg
Archives (0)	.zip, .rar, .7zip, .7z, .arj, .tar, .gzip, .bzip, .bzip2, .cab, .jar, .cpio, .ar, .gz, .tgz, .bz2
Databases (1)	.db, .db3, .sqlite, .sqlite3
Documents	'.htm', '.html', '.doc', '.docx', '.odt', '.xls', '.xlsx', '.ppt', '.pptx', '.pdf', '.txt', '.rtf'
Executable	'.exe', '.msi', '.cmd', '.com', '.bat', '.reg', '.scr', '.dll', '.ini'

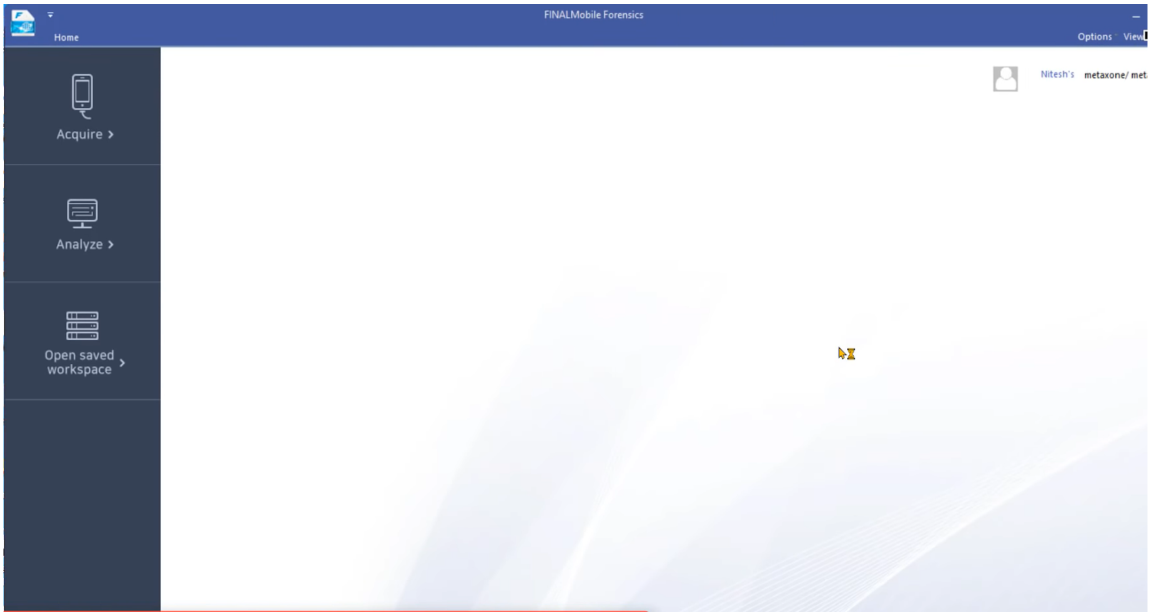
- When a 4-digit pin is employed as a security measure it's been demonstrated that it's possible to get it during a short period of your time , in

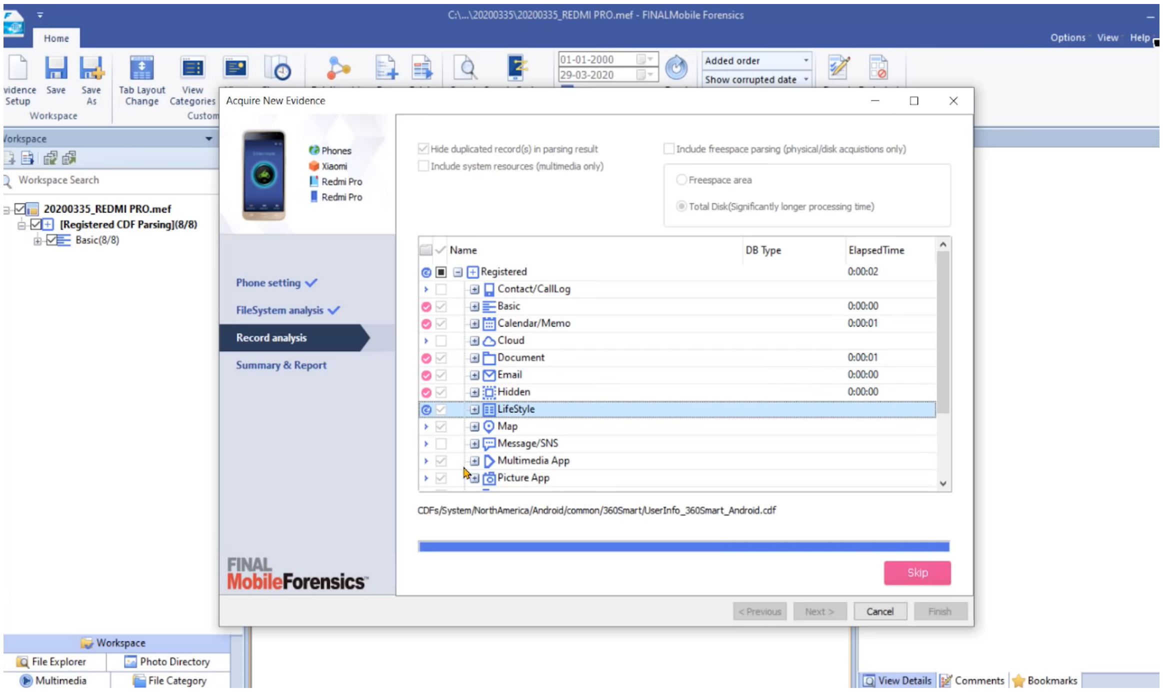
2.FINALMobile Forensics

- It simply captures data and analyzes it from mobile devices through logical and physical acquisitions.
- Not an open-source tool
- Subscription needed.

Features:

- Program is able to read its own generated phone images(MEFs)
- After data is loaded it auto-generates filename on basis of the model number of mobile devices.
- Results are generated in the HTML,pdf,excel.
- Display the file and all formats of files.





01-01-2000
29-03-2020

☑ Hide duplicated record(s) in parsing result
□ Include system resources (multimedia only)
□ Include freespace parsing (physical/disk acquistions only)
○ Freespace area
● Total Disk(Significantly longer processing time)

	Name	DB Type	ElapsedTime
■	Registered		0:00:02
□	Contact/CallLog		
☑	Basic		0:00:00
☑	Calendar/Memo		0:00:01
□	Cloud		
☑	Document		0:00:01
☑	Email		0:00:00
☑	Hidden		0:00:00
☑	LifeStyle		
☑	Map		
□	Message/SNS		
☑	Multimedia App		
☑	Picture App		

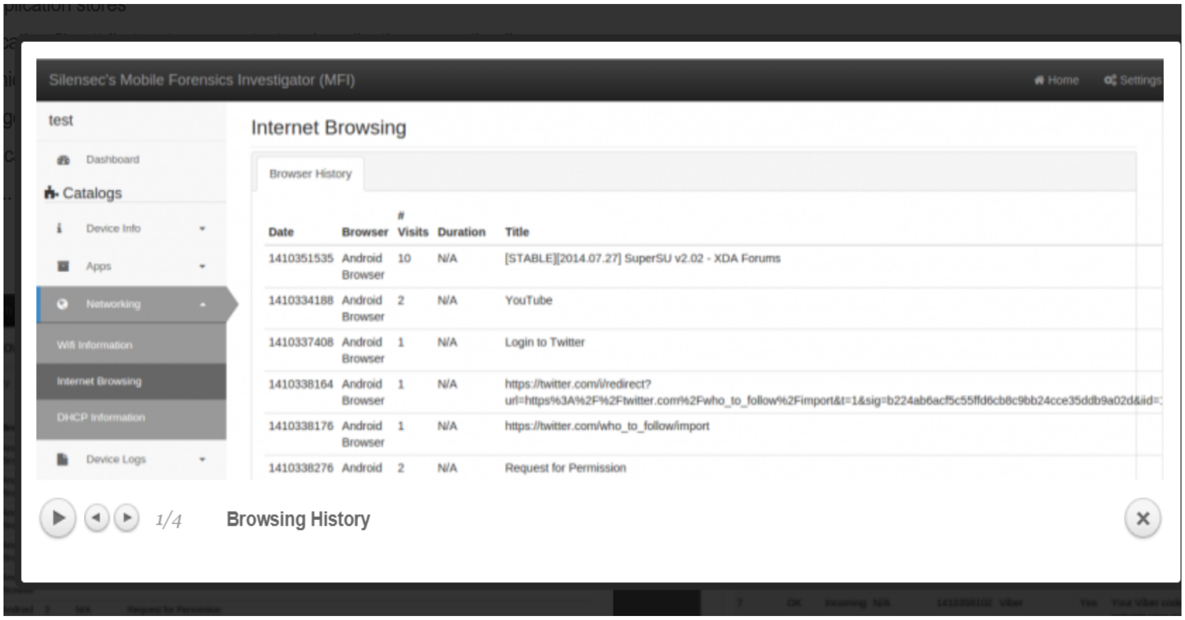
CDFs/System/NorthAmerica/Android/common/360Smart/UserInfo_360Smart_Android.cdf

3.Nyuki Forensic Investigator (NFI)

- It’s an open-source application that provides analysis for mobile device artifacts.

Features:

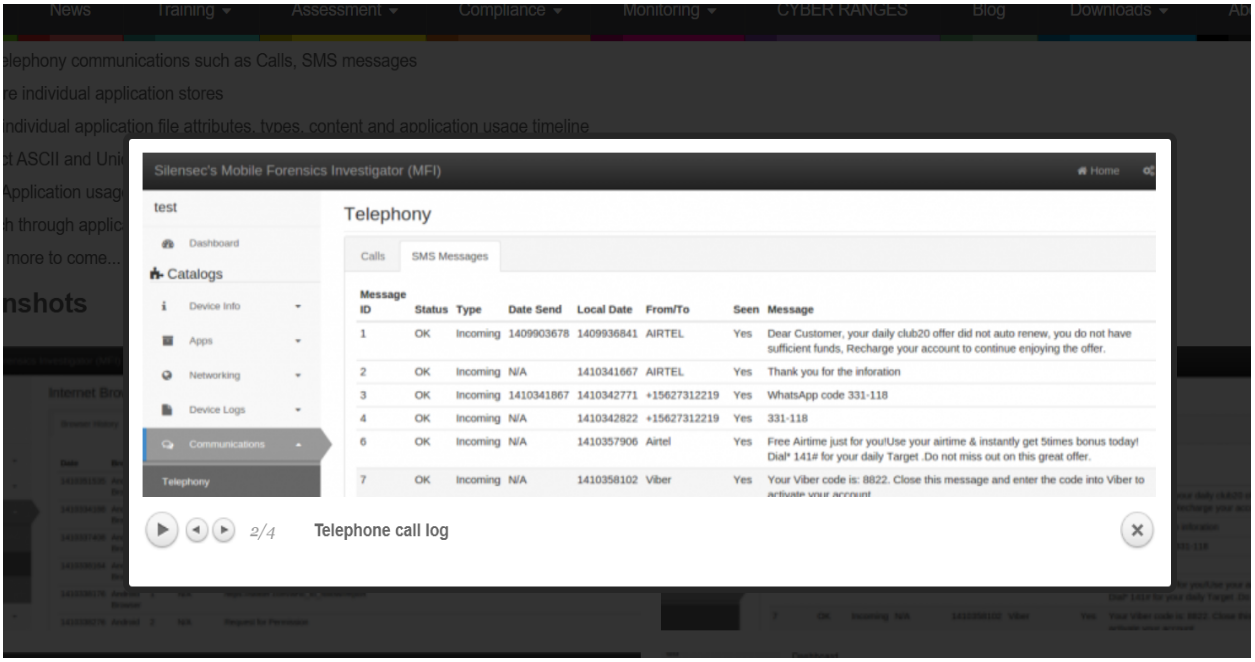
- Explore the device contacts, Facebook friends, LinkedIn connections, Whatsapp contacts
- Read through Facebook, LinkedIn, Skype chats
- List Telephony communications such as Calls, SMS messages
- Information related to Bluetooth devices, wifi networks, IP address leases.
- Extract ASCII and Unicode strings from Application files
- Files in a systematic way.



Internet Browsing

Date	Browser	# Visits	Duration	Title
1410351535	Android Browser	10	N/A	[STABLE][2014.07.27] SuperSU v2.02 - XDA Forums
1410334188	Android Browser	2	N/A	YouTube
1410337408	Android Browser	1	N/A	Login to Twitter
1410338164	Android Browser	1	N/A	https://twitter.com/i/redirect? url=https%3A%2F%2Ftwitter.com%2Fwho_to_follow%2Fimport&t=1&sig=b224ab6acf5c55ffd6cb8c9bb24cce35ddb9a02d&iid=
1410338176	Android Browser	1	N/A	https://twitter.com/who_to_follow/import
1410338276	Android	2	N/A	Request for Permission

Browsing History



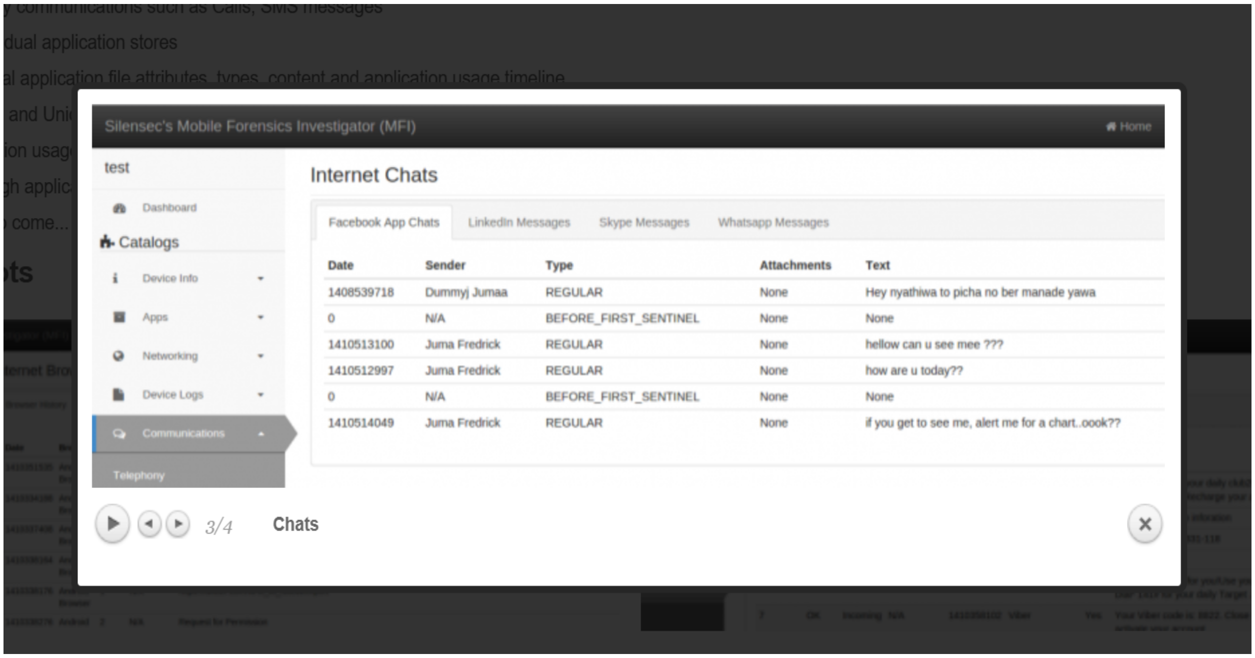
elephony communications such as Calls, SMS messages
re individual application stores
ndividual application file attributes, types, content and application usage timeline
t ASCII and Uni
Application usag
h through applic
more to come...
nshots

Telephony

Calls | SMS Messages

Message ID	Status	Type	Date Send	Local Date	From/To	Seen	Message
1	OK	Incoming	1409903678	1409936841	AIRTEL	Yes	Dear Customer, your daily club20 offer did not auto renew, you do not have sufficient funds, Recharge your account to continue enjoying the offer.
2	OK	Incoming	N/A	1410341667	AIRTEL	Yes	Thank you for the inforation
3	OK	Incoming	1410341867	1410342771	+15627312219	Yes	WhatsApp code 331-118
4	OK	Incoming	N/A	1410342822	+15627312219	Yes	331-118
6	OK	Incoming	N/A	1410357906	Airtel	Yes	Free Airtime just for you!Use your airtime & instantly get 5times bonus today! Dial* 141# for your daily Target .Do not miss out on this great offer.
7	OK	Incoming	N/A	1410358102	Viber	Yes	Your Viber code is: 8822. Close this message and enter the code into Viber to activate your account

Telephone call log



y communications such as Calls, SMS messages
dual application stores
al application file attributes, types, content and application usage timeline
and Uni
ion usag
gh applic
o come...
ts

Internet Chats

Facebook App Chats | LinkedIn Messages | Skype Messages | Whatsapp Messages

Date	Sender	Type	Attachments	Text
1408539718	Dummyj Jumaa	REGULAR	None	Hey nyathiwa to picha no ber manade yawa
0	N/A	BEFORE_FIRST_SENTINEL	None	None
1410513100	Juma Fredrick	REGULAR	None	hellow can u see mee ???
1410512997	Juma Fredrick	REGULAR	None	how are u today??
0	N/A	BEFORE_FIRST_SENTINEL	None	None
1410514049	Juma Fredrick	REGULAR	None	if you get to see me, alert me for a chart..oook??

Chats

1. Correlation between different data present in the database:-
    - One-to-one(1:1): Ex- Account No. and Account holder name. Each account can have only one account holder.
    - One-to-many(1:N): Ex- Like a customer can order different things from the store.
    - Many-to-many(N:N): Ex - In a company N employees can work on M projects.

2.Analytics:-

A. Graphical: It basically shows data in a graphical way which saves lots of time.

Types:

- Centrality analysis: Estimates how important a node is for the connectivity of the network.
- Connectivity analysis: Determine how strongly or weakly connected two nodes are.
- Path analysis: Examines the relationships between nodes. Mostly used in shortest distance problems.

B. String-based:

Most of the digital investigations rely on textual evidence, this is obviously due to the fact that most of the stored digital data is linguistic. Ex- logged conversation, a lot of important text-based evidence can be gathered while dumping strings from images (smartphone memory dumps) and can include emails, instant messaging, address books, browsing history, and so on. Most of the currently available digital forensic tools rely on matching and indexing algorithms to search textual evidence at a physical level so that they search every byte to locate specific text strings.

C. Pattern Based:

Pattern Analytics can be defined as a discipline of Big Data that enables business leaders to understand how different variables of the business interact and are linked with each other. Variables can be of any kind and within any data source, structured as well as unstructured. Machine Learning(ML) is mainly used to handle such big data.

3. Data can be presented in one of the three ways:

a) Text presentation: Text can be used to provide interpretation or emphasize certain data. If quantitative information to be conveyed consists of one or two numbers, it is more appropriate to use written language than tables or graphs.

b) Table presentation: Tables convey information that has been converted into words or numbers in rows and columns. Easy to understand. Tables are the most appropriate for presenting individual information and can present both quantitative and qualitative information.

c) Graph presentation: Graphs are effective for presenting large amounts of data, they can be used in place of tables to present small sets of data. A graph format that best presents information must be chosen so that readers and reviewers can easily understand the information.

Project Goals

1. Generate dummy dataset from a rooted device.
2. Extraction of the text messages.
3. Extraction of browser history.
4. Extraction of Bluetooth files.
5. Extracting data (EXIF) from images(like location, date, and time)
6. Extraction of location from the device.
7. Keyword search - Implementation of keyword search to get all the cases that have the searched keyword.
8. Implementing keyword search from cases.
9. Maximum occurred common word/data between two cases.
10. Returning all the common words between two cases.
11. Filtering Cases by Date or range of date.
12. Creating APIs for the above.
13. Tag association and tag-based search like searching cases that are tagged with theft to get all the theft-related cases.
14. Customized keyword search <kw1> + <kw2> + … + <kwN>
15. Analytics based on location.
16. Implementing analytics on the front end.

17. Presentation of observed analytics - graph, library, an API will be developed which will send the database data to the graph generating library to create required graphs and that will show in our OpenMF frontend.

18. Presentation of observed analytics between two cases using graphs and tables.

19. Update the project wiki.

20. Will write blogs and explain the project on Linkedin and Medium. Also, I have an idea to make video tutorials explaining the project and work, which further publish on Youtube and Linkedin to gain the crowd on the Website.

Timeline

Note: Each cell contains a time period of 1 week.

Week No.	Date	Work
0)	**17 May - 7 June**	Community Bonding (take feedback from mentors, Modify timeline and project details)
Coding officially begins!		
1)	**7 June - 13 June**	● Generate dummy dataset from a rooted device. ● Extraction of the text messages. ● Creating API which will show the extracted text messages.
2)	**14 June - 20 June**	● Extraction of browser history ● Extraction of Bluetooth files. ● Creating APIs which will show the extracted files for the above.
3)	**21 June - 27 June**	● Extracting data (EXIF) from images(like location, date, and time) ● Extraction of location from the device

		● Creating APIs which will show the extracted files for the above.
4)	**28 June - 4 July**	● Keyword search - Implementation of keyword search to get all the cases that have the searched keyword. ● Implementing keyword search from cases. ● Creating APIs for the above.
5)	**5 July - 11 July**	● Maximum occurred common word/data between two cases. ● Returning all the common words between two cases. ● Creating APIs for the above.

Evaluations (July 12- July 16)

6)	**12 July - 18 July**	● Taking feedback from mentors and working on bugs(if any). ● Leftover work(if any) ● Filtering Cases by Date or range of date. ● Creating API for that. ● Implementing frontend from week 4th,5th, and 6th in the analytics section.
7)	**19 July- 25 July**	● Tag association and tag-based search like searching cases that are tagged with theft to get all the theft-related cases. ● Customized keyword search <kw1> + <kw2> + … + <kwN> ● Analytics based on location ● Frontend implementation for this week.
8)	**26 July - 1 Aug**	● Presentation of observed analytics - graph, library, an API will be developed which will send the database data to the graph generating library to create required graphs and that will show in our OpenMF frontend.
9)	**2 Aug - 8 Aug**	● Presentation of observed analytics between two cases using graphs and tables.
10)	**9 Aug - 15 Aug**	● Leftover work (if any) ● Writing Wiki. ● Writing Medium Blog for documenting all progress and work done.

Final Evaluation week starts (Aug 16- Aug 23)

Final results of Google Summer of Code 2021 announced (August 31)

SCoRe Lab Contributions:

I have been contributing to this organization since last year and will continue as per need.

- [Merged] https://github.com/scorelab/OpenMF/pull/47
  "Frontend documentation #47"
- [Merged]https://github.com/scorelab/OpenMF/pull/44
  "fixed typo in README.md #44"
- [Open]https://github.com/scorelab/OpenMF/pull/157
  "[Frontend] enabled lazy loading #157"
- [Open]https://github.com/scorelab/OpenMF/pull/117
  "[frontend] Know more button fixed #117"
- [Issue]https://github.com/scorelab/OpenMF/issues/46
  "Documentation needs update! #46"
- [Issue]https://github.com/scorelab/OpenMF/issues/116
  "Know More Button not working in login page #116"

Personal Information:

- Name: Swapnal Shahil
- Email: swapnalsahil@gmail.com
- GitHub: swapnalshahil
- LinkedIn: swapnalshahil
- Twitter: eulersgamma
- Duo: swapnalsahil@gmail.com
- Gitter nickname: swapnalshahil
- First language: Hindi, proficient in English
- Time zone: Indian Standard Time (GMT +5:30)
- Contact: +91 9205711402
- University: Indian Institute of Technology, Guwahati
- Year of study: 2nd year B. Tech.
- Major: Chemical Science and Technology (Batch of 2023)

Reference:

- /scorelab/OpenMF/

Questions

1. **Are you a SCoRe contributor/ Have you contributed to SCoRe before?**
   Yes, I have contributed to SCoRe Lab Organization since December 2020. Yes, My contributions are in **OpenMF** and **Fact-Bounty** projects
2. **How can we reach you (eg: email) if we have questions about your application?**
   Email: swapnalsahil@gmail.com

Gitter: swapnalshahil
Mob: 9205711402

3. **What is your GitHub username(s):**
   swapnalshahil

Project Specific Questions

4. **Which SCoRe GSoC project are you applying for (please submit separate applications for each project):**
   OpenMF - Analytics API
5. **What do you plan to accomplish over this summer for this project?**

   - Generate dummy dataset from a rooted device.
   - Extraction of the text messages.
   - Extraction of browser history.
   - Extraction of Bluetooth files.
   - Extracting data (EXIF) from images(like location, date, and time)
   - Extraction of location from the device.
   - Keyword search - Implementation of keyword search to get all the cases that have the searched keyword.
   - Implementing keyword search from cases.
   - Maximum occurred common word/data between two cases.
   - Returning all the common words between two cases.

- Filtering Cases by Date or range of date.
- Creating APIs for the above.
- Tag association and tag-based search like searching cases that are tagged with theft to get all the theft-related cases.
- Customized keyword search <kw1> + <kw2> + … + <kwN>
- Analytics based on location.
- Implementing analytics on the front end.
- Presentation of observed analytics - graph, library, an API will be developed which will send the database data to the graph generating library to create required graphs and that will show in our OpenMF frontend.
- Presentation of observed analytics between two cases using graphs and tables.
- Writing Wiki and blogs.

6. **If you have your own project to propose, please describe it here:**
   Not now but maybe in the future.
7. **List down any plans you have during this summer.**
   I don't have any other plan for this summer, so I can give my whole day for this project.
8. **Education:**
   - **University:** Indian Institute of Technology Guwahati
   - Department: Chemical Science and Technology
   - Current year of study: 2nd year of B.Tech
   - Programming courses:

   - Object-Oriented Programming
   - Data Structures and Algorithm
   - Interactivity with JavaScript
   - ReactJs
   - Python and C++ Programming
   - Git and Github

9. **Do you have work experience in programming?**
   Yes, I did some projects of mine till now but contributed to many organizations on regular basis. One of my projects (project-corona) shows the current number of COVID19 { active cases, death count, recovered count } using bar graph of all countries.
10. **Do you have previous open source experience? Briefly describe what you have done.**
   From my 1st year of my B. Tech I am involved with open-source. Also, I participated in Hactoberfest and able to create more than 10 valuable PRs. In my free time, I contribute to reactjs.org (Facebook React Community).
11. **Tell one interesting fact about yourself.**
   I am a very social person and I love photography.