JUSTICE LEAGUE

SMART INDIA HACKATHON 2024

SMART INDIA HACKATHON 2024

NIRMAS: NIA's Investigation Reporting and Multidimensional Analysis System

Problem Statement ID : SIH1743

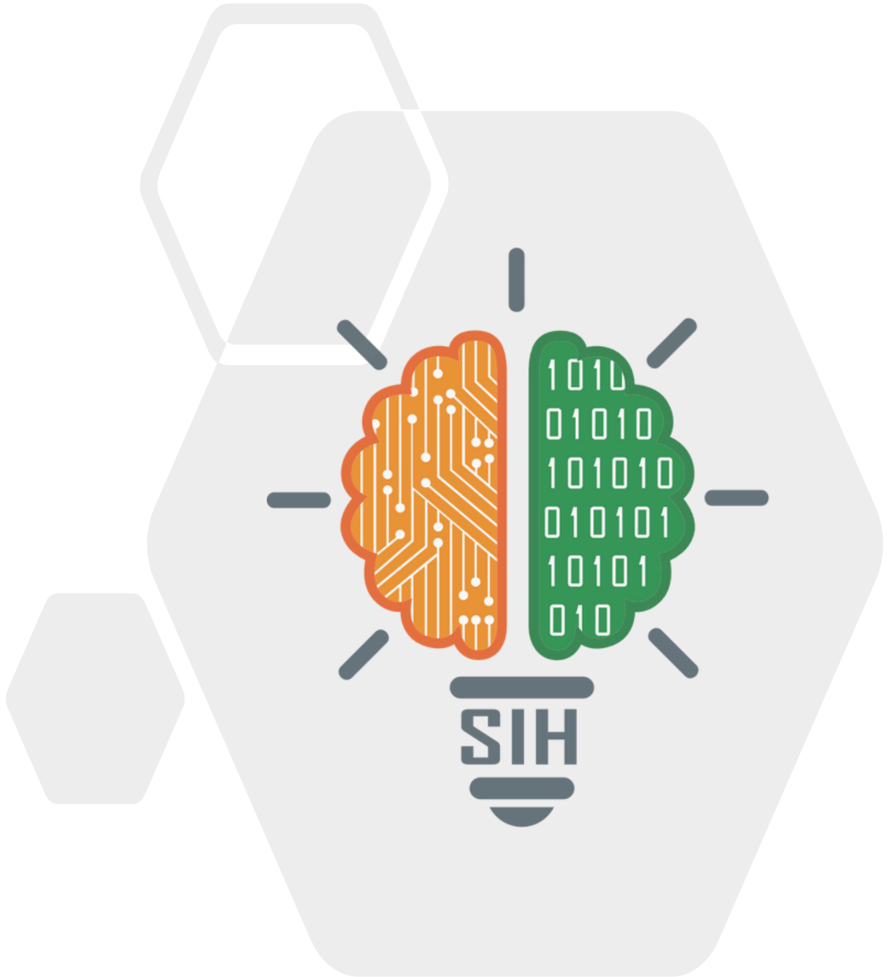
Problem Statement Title : Parsing of Social Media Feeds

Theme : Miscellaneous

PS Category : Software

Team ID : **25074**

Team Name : Justice League



NIRMAS: NIA's Investigation Reporting and Multidimensional Analysis System

Step-by-Step Approach

- **Automated Data Capturing, compatible across Android and Windows**:
  - **When device is physically accessible,** we use **ADB automation with Python** for collecting screenshots by sending commands to the connected device (via USB or Wi-Fi) through scripts.
  - **When Physical device is not accessible,** we use **Selenium WebDriver** for automated screenshot capture when access is remote. It works by mimicking user interactions such as logging in, navigating profiles, and capturing data by interacting with DOM elements on the HTML page.
  - **Supported Social Media Platforms**: The tool can investigate Facebook, Twitter (now X), Instagram, Telegram, WhatsApp, and Google accounts. **Separate automation scripts** will be written for each platform based on the unique structure and functionality of each platform.
  - **In Case Login Credentials are not available, Cookie data from a logged-in session** can be utilized to gain login access. When running Selenium to automate tasks on the website, instead of performing a login, the automation script directly injects the saved cookies into the Selenium browser instance.

- **Optical Character Recognition(OCR) Integration & Sentiment Analysis**:
  - **OCR** extracts text from screenshots, converting them into searchable, analyzable data.
  - **Machine Learning** performs sentiment analysis on messages and posts, identifying key content.

- **Secure Data Storage via Blockchain** by storing hashes of screenshots for **tamper-proof verification**, ensuring evidence integrity during court proceedings.

- **Optimized Storage** by converting non-essential data into **lower resolution formats** to conserve space without compromising critical details.

- Investigators can **filter** data based on **date, keywords, contacts, or location**, allowing focus on relevant evidence.

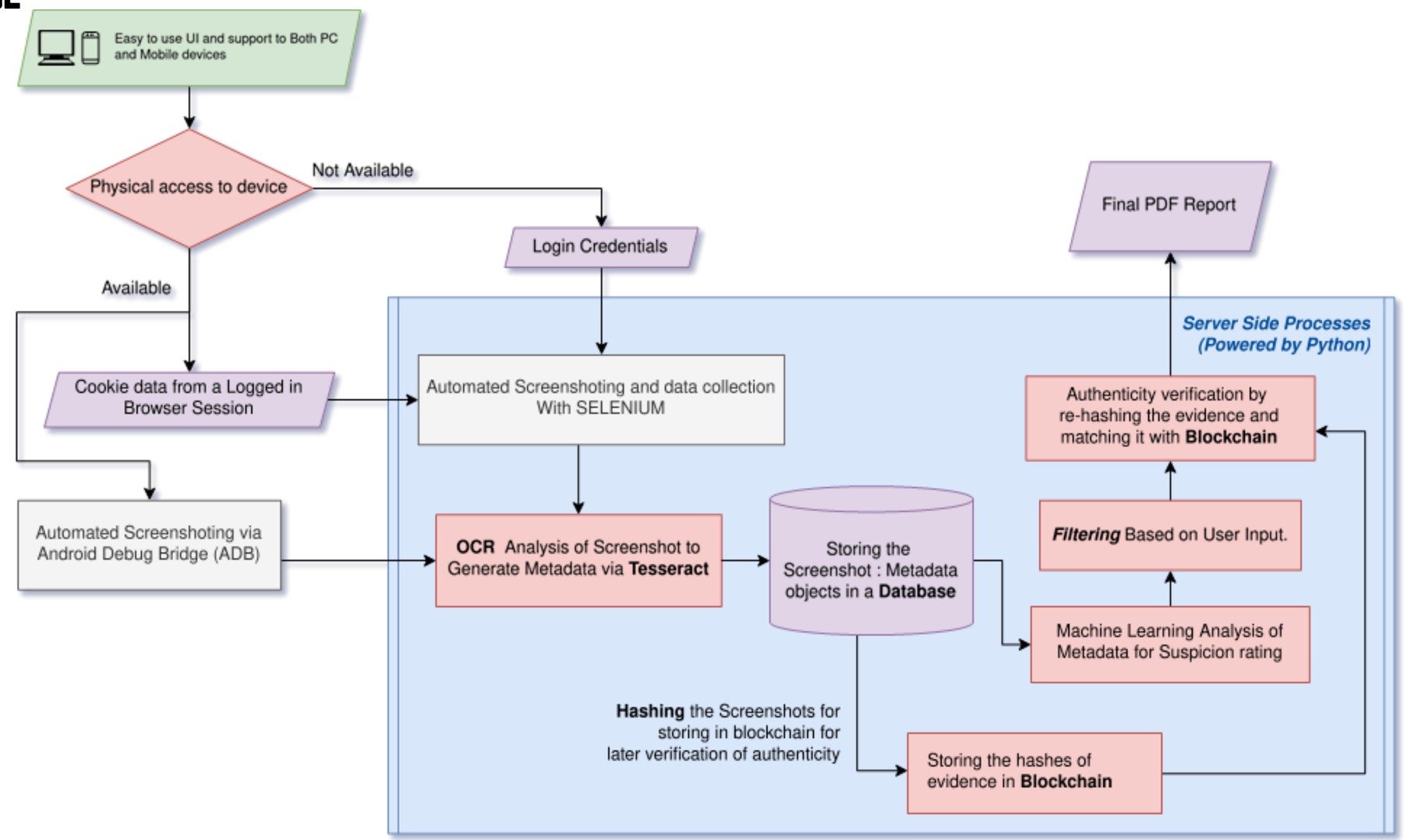
- **Structured Report Generation**:
  - Uses **ReportLab** to automatically compile captured data into **PDF reports**, ready for court.

Key Features

- **Easy** to use, modify and extend.
- **Minimal Training needed** for Investigator
- Powered by Fully **Free and Open-Source** Tech stack
- **Cross-Platform** Compatibility (Android + Windows Support)
- **Seamless Login** with **Stored Browser Cookies**.
- Comprehensive **Automated Data Capture**
- Performance **Optimization**
- Integrated **OCR** Technology
- **Advanced Filtering** Options
- Automated **Suspicion Rating System**
- **Blockchain-Backed Security** and **Verification**.
- Professional **Report Generation**

Demo: https://youtu.be/XqN7UhhuRL8
FAQs: https://sih-faq.vercel.app/

Technical Approach



Tech Stack

- **Languages:** Python, Javascript, Kotlin.
- Automated Screenshot Capturing via **Selenium and Android Debug Bridge (ADB).**
- **Tesseract OCR** for image processing and analysis.
- **Spacy for Sentiment Analysis.**
- Blockchain: **HyperLedger Fabric**
- Database: **MySQL**
- Cloud Storage: **Google Cloud**
- Frontend: **Next.js**
- Backend: **Python (Django)**
- Reporting & Documentation: **ReportLab Python Library**

Demo: https://youtu.be/XqN7UhhuRL8
WHY USE BLOCKCHAIN?: https://sih-faq.vercel.app/

JUSTICE LEAGUE

Feasibility and Viability

SMART INDIA HACKATHON 2024

Feasibility

- **Minimal Training Needed:** User-friendly interface requires little to no training for investigators.
- **Security:** Robust Security by **Role Based Access Control (RBAC)** ensure the system is well protected from cyberattacks.
- **Evidence Verification:** Blockchain ensures **tamper-proof evidence** at report generation by **storing cryptographic hashes** of each screenshot in a blockchain.
- **Human Error:** Mitigation strategies reduce risks from human errors.
- **Scalability:** By leveraging **cloud-based storage** and processing along with **data compression and optimization**, it can dynamically scale to handle vast amounts of data, including high-resolution images and lengthy chat histories.
- **Cross-Platform Compatibility**: The tool works across Android and Windows, ensuring investigators can perform their tasks on both mobile devices and desktops.

Potential Challenges

- **Anti-Bot Measures** in some Social media platforms might detect and block automated tools.
- **Risk of misuse** for unauthorized surveillance, potentially breaching public privacy.
- **Dynamic Content:** Posts, messages, and other data might be edited or deleted by the account owner or the platform itself during an investigation.
- **Over-reliance on Automation:** Excessive dependence on the tool may **reduce critical human oversight**.

Strategies to overcome them

- **Human-Like Behavior Simulation** and **Automated Captcha Solvers**.
- NIA should ensure **safeguards to prevent potential misuse**.
- **Cross-verification using Blockchain** is done to ensure the content is genuine.
- **Routine Inspections** can put a check on problems caused by Over-Reliance on automation technologies.

•Android App UI for ease of use:
https://www.figma.com/design/2f5Mvq1ZKXEy3GdWJGyPRv/SIH1743-Android-prototyp?node-id=0-1&t=lwqw3iVS5RKxvPNR-1

•Windows UI for Software:
https://www.figma.com/design/2f5Mvq1ZKXEy3GdWJGyPRv/SIH1743-Android-prototyp?node-id=2368-125&node-type=canvas

Impact and Benefits

Impact on Target Audience

- **Increased Efficiency:** Investigators can collect and analyze **large volumes** of social media data **quickly** and **accurately**.
- **Enhanced Accuracy and Integrity:** By automating the screenshot and data capture process, the tool **eliminates human error.**
- **Legal Admissibility: Blockchain-backed** verification and automated reporting ensure the evidence's integrity.
- **Broad Applicability:** Law enforcement across different regions can use the tool to investigate criminal activities like financial fraud to cyberbullying.
- **Cross-Platform Compatibility: Maximum usability is ensured** as the tool seamlessly works across both Android and Windows devices,.

Benefits of the Solution

Social

- **Justice and Accountability:** Thorough, accurate, and reliable investigations.
- **Faster Case Resolution: Reduces delays** in court cases, benefitting the judicial system by **clearing case backlogs.**
- **Public Safety:** Assists in solving cases involving cybercrimes, hate speech, and other online crimes.
- **Data Privacy Compliance:** We follow stringent privacy laws, ensuring that sensitive data is handled with care, meeting legal standards.

Economic

- **Cost Efficiency:** Automation reduces the resources required for evidence collection.
- **Scalability for Large Case:** For complex investigations requiring collection of vast amounts of data, it **minimizes human errors**.
- **Economic Deterrence:** It can contribute to deterring **white-collar crimes** like financial fraud, reducing their economic impact.
- **Resource Optimization:** Leading to faster resolutions.

Environmental

- **Paperless Investigations:** The **digital** nature minimizes the need for printed evidence.
- **Energy Efficiency: Cloud-based storage** are more energy-efficient, especially if paired with data centers **powered by renewable energy.**
- **Reduced Carbon Footprint**

Research and References

Resources-

01. ADB: https://www.xda-developers.com/how-to-take-logs-android/

02. Selenium: https://www.browserstack.com/guide/architecture-of-selenium-webdriver

03. Wikepedia: https://en.wikipedia.org/wiki/Optical_character_recognition

04. TesserAct vs KerasOcr: https://thangasami.medium.com/tesseract-vs-keras-ocr-vs-easyocr-ec8500b9455b

05. Android Debug bridge: https://developer.android.com/tools/adb

06. Keras OCR: https://github.com/faustomorales/keras-ocr

07. BlockChain: https://hyperledger-fabric.readthedocs.io/en/release-2.0/blockchain.html

08. How does OCR work: https://www.klippa.com/en/blog/information/tesseract-ocr/

09. Solidity Docs: https://docs.soliditylang.org/en/v0.8.27/

10. Django Docs: https://docs.djangoproject.com/en/5.1/

11. PostgreSQL: https://www.postgresql.org/docs/current/index.html

Important Links-

01. Demonstration for Whatsapp when physical access to phone is available: https://youtu.be/XqN7UhhuRL8

02. Why use blockchain for verification of collected evidence: https://sih-faq.vercel.app/

03. Easy to Use UI for Mobile App: https://www.figma.com/design/2f5Mvq1ZKXEy3GdWJGyPRv/SIH1743-Android-prototyp?node-id=0-1&t=lwqw3iVS5RKxvPNR-1

04. Some FAQs: https://sih-faq.vercel.app/

05. Contact: https://sih-faq.vercel.app/