Automated Social Media Open Source Intelligence

UAA Computer Science & Engineering Department
UNIVERSITY of ALASKA ANCHORAGE

Tyler Lukacz
tylerlukaz@gmail.com

Logan Chamberlain
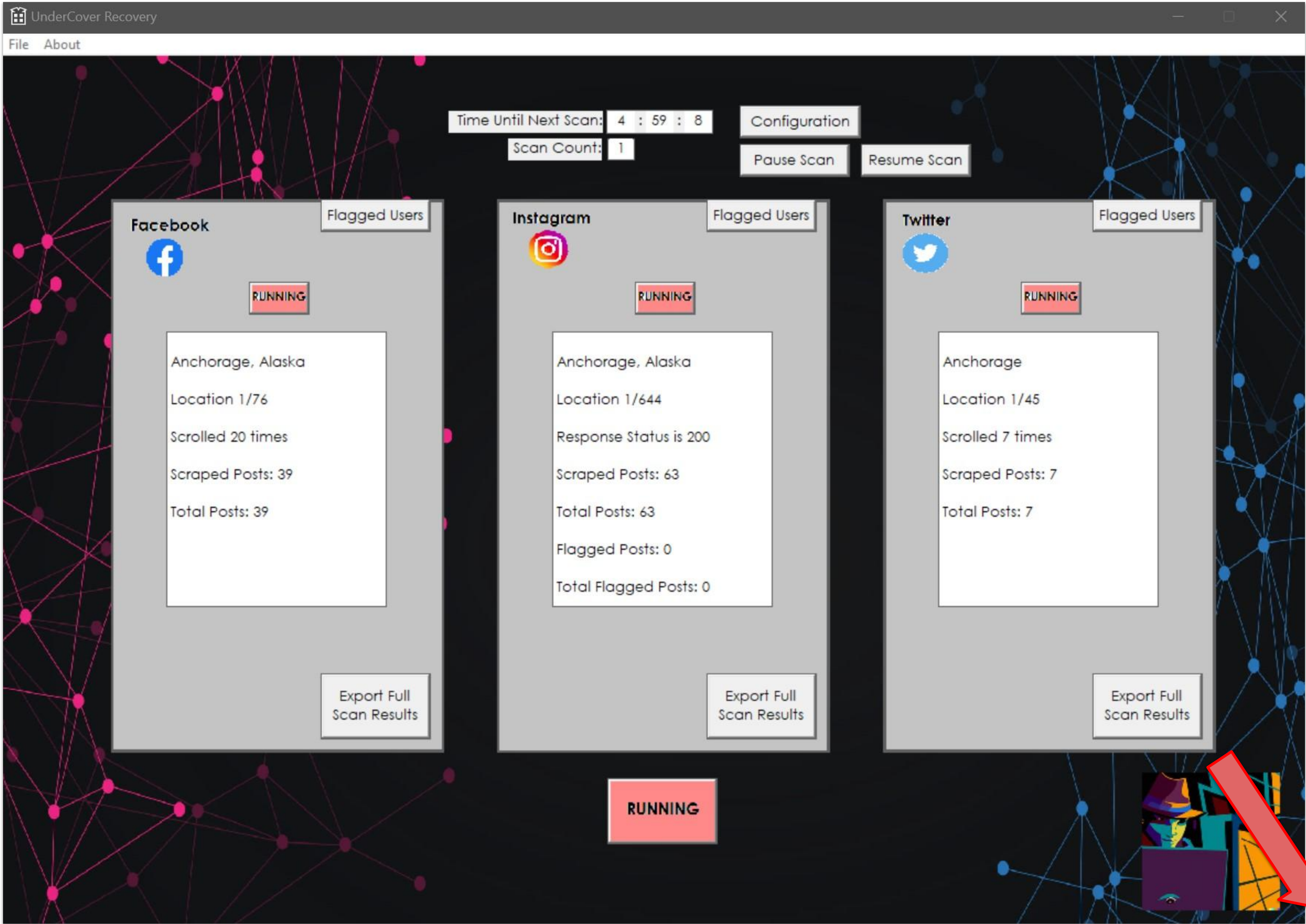lnchamberlain2@alaska.edu

PURPOSE

- Client was conducting manual reviews of social media sites.
  - Ineffective and Time Consuming.
- Operational Need:
  - Save time and close intel gaps through automated tool.

APPROACH

- Automated tool required 3 components:
  - Graphical User Interface.
    - Windows Desktop Application.
    - Collects login information, keyword and flagged user information, and provides real time scrape updates.
  - Scraping Scripts
    - Logon Authenticator that supports the three target sites.
    - Scripts to scrape and parse web content.
  - File Structure
    - Results in .html format
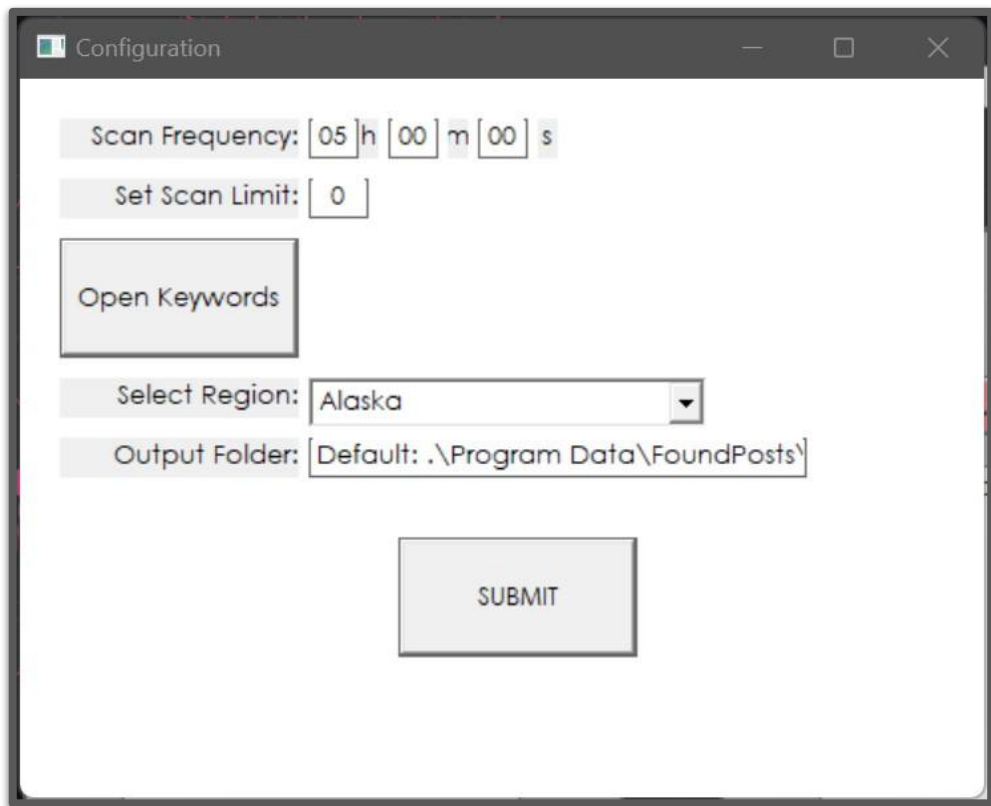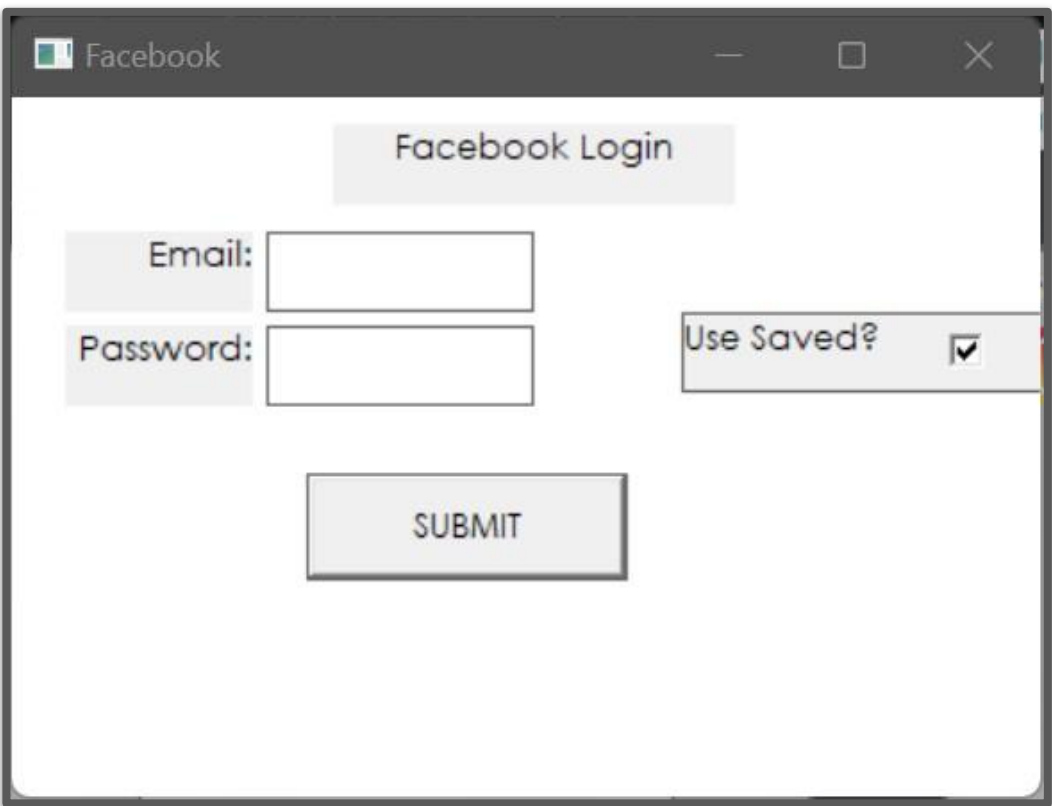   - Supports external storage

Project Report

CLIENT
Developed for Anchorage FBI - VC1

Ethics and Legal Considerations

All information collected is publicly available and open source material. Distribution and use controlled by the Developers, all rights reserved. Additionally, this tool does not use any social media APIs due to legal constraints.

Process

- The GUI allows the user to specify which target sites to scan, how many scans to run, and at what frequency.
  - The GUI displays the time between scans and how many have been completed.
  - Prints real time status updates from the scraper programs.
- The Authenticator submits logon requests to one of the social media sites using credentials collected by the GUI, verifies successful login.
  - On successful logins, the resultant cookies are stored for future use by scraper scripts.
  - Saves valid credentials to support streamlined future logins.
- Upon a successful login and once the program is configured with region selections and keywords, the GUI automatically launches scrapes of logged in sites at user defined intervals in perpetuity or until the Scan Limit is reached.
  - Automatically re authenticates using saved credentials when necessary
- Evidentiary copies of posts matching keywords or flagged users are downloaded locally and organized into .html documents