УДК 004.056

Котляров А.В., Кушніренко Н.І.

Національний університет «Одеська політехніка»

АЛГОРИТМ ПОШУКУ ПРОФІЛІВ КОРИСТУВАЧІВ ЗА НІКНЕЙМОМ У СОЦІАЛЬНИХ МЕДІА ЯК ЕЛЕМЕНТ ПРОТИДІЇ КІБЕРЗЛОЧИННОСТІ

Вступ. У сучасному цифровому світі соціальні медіа стали важливою частиною повсякденного життя мільярдів людей, забезпечуючи швидку комунікацію та обмін інформацією. Водночас їх широке використання поєднується з високим рівнем анонімності, що створює умови для здійснення кіберзлочинів. В боротьбі з такими загрозами ключову роль почав відігравати OSINT - збір та аналіз даних із відкритих джерел для використання як розвідданих або доказів. Це зумовлює потребу у розробці спеціалізованих алгоритмів і застосунків для пошуку користувачів за цифровими слідами.

Мета: Розробка алгоритму автоматизованого пошуку профілів користувачів за нікнеймом у соціальних медіа, що сприяє протидії кіберзлочинності.

Основна частина

Для ефективного пошуку користувачів у соціальних медіа необхідно мати вихідні дані, що слугують відправною точкою. Такі дані називають ідентифікаторами - унікальними або частково унікальними відомостями, які допомагають звузити пошук або однозначно ідентифікувати особу. У OSINT-розвідці по соцмедіа зазвичай виділяють п'ять основних ідентифікаторів: номер телефону, електронну адресу, прізвище та ім'я, нікнейм [1].

Кожен із цих ідентифікаторів має свої обмеження. Телефони та електронні адреси унікальні, але зазвичай приховані й рідко доступні для відкритого пошуку. Прізвище та ім'я часто поширені та мають багато варіантів написання, що ускладнює точну ідентифікацію. Натомість найбільш універсальним і зручним для пошуку є нікнейм, оскільки він поєднує відкритість та певний рівень унікальності

Нікнейм (від англ. nickname - прізвисько) - це вигадане ім'я, яке користувач обирає для використання у соціальних медіа. Воно може базуватися на власному імені чи прізвищі, надихатися вигаданими персонажами або відомими особистостями, а також мати інше значення, пов'язане з власником. У соцмедіа нікнейм виконує роль унікального ідентифікатора, замінюючи числовий ID.

Здійснити ручний пошук можна двома способами: перший - ввести запит у пошукове поле будь-якої соціальної медіа; другий (лише для веб-версій) - змінити в адресному рядку значення нікнейму на потрібне. Наприклад, у URL https://twitter.com/user1 частина «user1» є нікнеймом, і її можна замінити на будь-який інший для пошуку потрібного. Зрозуміло, що ручний пошук займає багато часу і є незручним, адже доводиться перевіряти кожну соцмедіа окремо, а їхня велика кількість значно ускладнює процес. Тому для спрощення було розроблено алгоритм, який автоматизує процес пошуку.

Алгоритм працює за такою послідовністю кроків.

Крок 1. Введення даних. Користувач задає нікнейм (наприклад, user2).

Крок 2. Формування посилання. З бази шаблонів URL вибирається перше

(наприклад, https://instagram.com/{}) і замість {} підставляється введений нікнейм - утворюється повна адреса, наприклад https://instagram.com/user2.

Крок 3. Перевірка доступності профілю. Виконується запит за сформованим URL, після чого аналізується отриманий HTTP-статус.

Крок 4. Обробка відповіді сервера. Якщо код належить до груп 3xx або 4xx - профілю із таким нікнеймом не існує; якщо код 2xx - додатково аналізується HTML-код сторінки.

Крок 5. Аналіз вмісту сторінки. Якщо у HTML немає фраз, що вказують на помилку (наприклад, "User not found"), URL вважається успішним; у протилежному випадку - невдалим.

Крок 6. Перехід до наступного шаблону. Якщо всі шаблони ще не перевірені, алгоритм повторюється для наступного; якщо перевірено останній - формуються та виводяться результати перевірки.

Для програмної реалізації цього алгоритму використовуються бібліотеки Requests для отримання HTTP-запитів та BeautifulSoup для аналізу HTML-коду сторінки. Замість Requests також можна використовувати Selenium, і цей вибір залежить від можливостей користувача. Головні проблеми Selenium - висока ресурсозатратність та повільність, проте він має суттєві переваги: імітує дії людини, тому CAPTCHA не розпізнають його як бота, і дозволяє отримувати доступ до соцмедіа з обов'язковою авторизацією [2].

Слід зазначити, що подібний, хоча й спрощений, алгоритм уже реалізовано в деяких застосунках. Найчастіше вони не виконують перевірку фраз-помилок. Це, звісно, прискорює роботу алгоритму, проте ігнорує так звані soft 404 (м'яка помилка 404) [3] - випадки, коли не повертаються коди 4xx навіть при відсутності користувача. Замість цього може повернутися код 200 з повідомленням про помилку або відбутися перенаправлення на головну сторінку. У результаті алгоритм може помилково визначити, що профіль існує, хоча насправді його немає. Через це такі застосунки зазвичай уникають перевірки тих соціальних медіа, де спостерігається подібна поведінка.

Висновок. Розроблений алгоритм забезпечує автоматизоване визначення наявності профілів користувачів за нікнеймом у соціальних медіа, що значно спрощує процес пошуку у порівнянні з ручними методами. У ході було продемонстровано його структуру, описано логіку функціонування та запропоновано способи програмної реалізації. Його застосування може стати базою для створення програмних засобів OSINT, підвищуючи ефективність аналізу цифрових слідів та підтримуючи заходи протидії кіберзлочинності.

Перелік використаних джерел.

1. Walkow M., Pohn D. Systematically Searching for Identity-Related Information in the Internet with OSINT Tools, 2024. 8 p. DOI: https://doi.org/10.48550/arXiv.2407.16251.

2. Python Selenium vs Python Requests. [Електронний ресурс]. - Режим доступу: https://scrapeops.io/python-web-scraping-playbook/python-selenium-vs-python-requests/.

3. Prieto, V.M., Alvarez, M., Cacheda, F. Soft-404 Pages, A Crawling Problem. 2014 J. Digit. Inf. Manag., 12, 73-92.