УДК 65.012.8 + 004

МАНЖАЙ ОЛЕКСАНДР ВОЛОДИМИРОВИЧ

кандидат юридичних наук, доцент,

завідувач кафедри протидії кіберзлочинності факультету № 4 Харківського національного університету внутрішніх справ

http://orcid.org/0000-0001-5435-5921

МАНЖАЙ ІРИНА АНДРІЇВНА

завідувач навчального відділу Харківського університету

ОКРЕМІ АСПЕКТИ АВТОМАТИЗАЦІЇ АНАЛІЗУ САЙТІВ З ПРОТИПРАВНИМ КОНТЕНТОМ

Під час виконання оперативно-службових завдань різним підрозділам поліції доводиться здійснювати моніторинг мережних ресурсів на предмет наявності протиправного контенту. Нерідко за результатами такого моніторингу накопичується великий масив даних, який потрібно дослідити та обрати серед знайдених ресурсів найбільш перспективні з точки зору подальшого документування. Правильний підхід до вивчення переліку знайдених ресурсів з протиправним контентом передбачає застосування різноманітних статистичних та аналітичних методів. У результаті матимемо орієнтуючу інформацію не тільки по пріоритетних сайтах, але й первинні дані про можливих володільців і адміністраторів таких ресурсів.

Один із методів, який дозволяє «грубо» виокремити ресурси зі спільним корінням є аналіз так званих фавіконів (favicons /favorite icons/) – значків веб-сайтів чи веб-сторінок. У випадку збігу унікальних використовуваних зображень можна припустити пов’язаність розробників та/або адміністраторів відповідних ресурсів.

Візуально представити набір використовуваних фавіконів можна у вигляді таблиці із завантаженням відповідного адресі (окремій сторінці) сайту фавікона у комірку Google- чи MS Excel-таблиці. Описану операцію можна виконати за допомогою формули та використання в ній функції IMAGE.

=IMAGE(TEXTJOIN("",ИСТИНА,"http://www.google.com/s2/favicons?domain="
,A1),3)

або

=IMAGE(TEXTJOIN("",ИСТИНА,A1,"/favicon.ico"),3),

де A1 – комірка з адресою досліджуваного Інтернет-ресурсу.

Функція IMAGE вже тривалий час доступна у Google-таблицях, а також стала доступною в пакеті Microsoft 365 з другої половини 2022 року.

Сама автоматизація порівняння фавіконов може бути виконана за допомогою спеціалізованих застосунків, як от FavFreak (https://github.com/devanshbatham/FavFreak). Ця утиліта розраховує унікальний ґеш для фавікону кожного сайту з переліку та порівнює їх. Перелік доменних імен для вивчення формується у вигляді окремого текстового файлу. У підсумку матимемо звіт із відповідними збігами та текстовими файлами, що міститимуть інформацію про ресурси з однаковими фавіконами (рис. 1).

```
~ [40]  : [1198047028]
~ [1]   : [702403635]
~ [1]   : [-1965981700]
~ [1]   : [-1131104052]
~ [1]   : [214280171]
~ [1]   : [-1819991364]
~ [1]   : [561456715]
~ [12]  : [-1170068949]
```

Рис. 1. Кількість ресурсів з однаковим фавіконом та його ґеш

Під час аналізу фавіконів слід враховувати, що їх може бути декілька в межах одного сайту, як от окремі значки для окремих вебсторінок. Тому під час порівняння бажано використовувати декілька способів завантаження відповідних значків. Після виокремлення потенційно пов’язаних мережних ресурсів результати можна представити у вигляді графу за допомогою відповідного програмного забезпечення, наприклад Gephi.

Описаний спосіб автоматизації дозволяє значно скоротити час на первинну обробку інформації та створити умови для подальшого більш ретельного вивчення накопичених даних.