**НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ**

ІНСТИТУТ ПРОБЛЕМ РЕЄСТРАЦІЇ ІНФОРМАЦІЇ НАН УКРАЇНИ

# ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА БЕЗПЕКА

## МАТЕРІАЛИ XXII МІЖНАРОДНОЇ НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ

**ВИПУСК 22**

Київ – 2022

*Рекомендовано до друку Вченою радою*
*Інституту проблем реєстрації інформації НАН України*
*(протокол № 14 від 20 грудня 2022 р.)*

**Інформаційні технології та безпека. Матеріали XXII Міжнародної науково-практичної конференції ІТБ-2022.** – Київ: Інжиніринг. – 132 с. ISBN: 978-966-2344-85-1

До збірника увійшли матеріали доповідей, представлених на XXII Міжнародній науково-практичній конференції «Інформаційні технології та безпека» (ІТБ-2022, 16 листопада 2022 року, м. Київ, Україна).

У збірнику представлені матеріали, присвячені питанням функціональної стійкості інформаційних систем, безпеки та живучості критичних інфраструктур, комп'ютерного моделювання складних систем, технологій аналітики даних великих обсягів (Big Data), створення аналітичних систем на основі відкритих джерел інформації (OSINT), моделювання, аналізу та прогнозування процесів мережевої взаємодії, методів і засобів підтримки прийняття рішень.

Для фахівців в області інформаційних технологій, інформаційної і кібернетичної безпеки, а також для аспірантів і студентів старших курсів вищої школи відповідних спеціальностей.



ISBN 978-966-2344-85-1

# МЕТОДИКА ВИЯВЛЕННЯ ОБ‘ЄКТІВ КІБЕРБЕЗПЕКИ НА БАЗІ ТЕХНОЛОГІЇ OSINT

Д.В. Ланде [1,2][0000-0003-3945-1178], О.О. Пучков [1][0000-0002-8585-1044], І.Ю. Субач [1][0000-0002-9344-713X]

[1] Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна
[2] Інститут проблем реєстрації інформації НАН України, Київ, Україна

dwlande@gmail.com, igor_subach@ukr.net

В інформаційних ресурсах мережі Інтернет міститься багато прихованих знань. Ці знання вносяться користувачами, які у сукупності утворюють своєрідне експертне середовище. У зв’язку з цим, основна задача технологій розвідки у відкритих джерелах (OSINT) полягає у виявленні та екстрагуванні прихованих експертних знань, їх узагальненні, а також подальшої аналітичної обробки. Для цього застосовуються лінгвістичні і статистичні методи, а також методи кластерного аналізу. В роботі запропонована методика екстрагування понять із текстів повідомлень мережевих джерел, що стосуються предметної області кібербезпеки, фільтрування цих понять за статистичними ознаками, рейтингування, формування мережі їх взаємозв’язків, кластеризації і візуалізації цієї мережі. Для створення програмної реалізації запропонованих підходів використовується мова програмування Perl у середовищі ОС Linux, а також засоби програмного забезпечення для моделювання, аналізу та візуалізації графів – Gephi.

**Ключові слова:** OSINT, об’єкти кібербезпеки, екстрагування понять, мережа термінів, веб-ресурси.

## Постановка проблеми

Фахівцям, що працюють у визначеній предметній області, зазвичай відомі її основні поняття та об’єкти. Проте, з плином часу виникають нові поняття та нові об’єкти. У сфері кібербезпеки такими об’єктами можуть бути нові види кібератак, нові хакерські угрупування, нове деструктивне програмне забезпечення, аналітичні групи тощо. Можуть з’являтися нові змістовні зв’язки між такими об’єктами, що також потребує додаткового аналізу. В окремих групах об’єктів,

наприклад, злочинних хакерських угрупуваннях, можуть зміщуватися центри та об'єкти особливої уваги фахівців із кібербезпеки. Таким чином, виникає завдання постійного моніторингу інформації у межах визначеної предметної області. Така інформація може бути представлена в мережі Інтернет, до контенту якої, зокрема, документів, розміщених на веб-сайтах, може бути застосована технологія розвідки у відкритих джерелах (OSINT) [1].

Мета цієї роботи – створення і апробування методики визначення основних об'єктів кібербезпеки і зв'язків між ними на базі аналізу змістовної складової веб-простору, а також формування, кластеризація та аналітична обробка сформованих мереж об'єктів кібербезпеки. При вирішенні цих завдань в рамках запропонованої методики має бути проаналізовано тематичну частину кириличного сегменту веб-простору і соціальних мереж щодо публікацій у сфері кібербезпеки.

## Основні кроки методики

1. На першому етапі методики формується інформаційний масив релевантних тематиці документів, для чого мають використовуватись наявні системи контент-моніторингу, наприклад система Cyber Aggregator [1].

Для отримання інформаційного масиву публікацій щодо кібербезпеки необхідно опрацювати тематичний запит до такої системи, наприклад, застосовувався запит:

***Кібератака | кибератака |***
***Кібербезпека | кибербезопасность***

У результаті отримується інформаційний масив релевантних документів великого обсягу (декілька тисяч документів на місяць).

2. На другому етапі на основі лінгвістичного і статистичного аналізу здійснюється екстрагування понять із предметної області, що містяться в документах отриманого на першому кроці інформаційного масиву. Основна ідея екстрагування об'єктів полягає у тому, що на цей час більшість нових понять в повідомленнях українською, російською, білоруською мовами позначаються латиницею або кириличними літерами, але в лапках. При цьому для екстрагування іменних сутностей також застосовується словник відомих іменних сутностей об'єктів кібербезпеки, які відшукуються в інформаційному масиві. Крім того,

виявляються не кириличні короткі словосполучення в інформаційному масиві.

3. На третьому етапі здійснюється сортування відібраних понять за частотою та фільтрація цих понять фахівцем-експертом.

4. На четвертому етапі здійснюється формування мережі відібраних понять [2]. Для цього визначаються неспрямовані зв'язки між поняттями. Два поняття вважаються зв'язаними, якщо вони входять в той самий сегмент документу із відібраного інформаційного масиву.

5. На п'ятому етапі здійснюється кластеризація відібраної мережі та їх кластеризація за алгоритмом модулярності, а також візуалізація із застосуванням системи Gephi [3].

## Висновки

Запропоновано методику виявлення іменних сутностей об'єктів кібербезпеки із документів, представлених у мережі Інтернет. Методика враховує приховані знання, внесені експертним мережевим середовищем.

Кластерний аналіз і візуалізація отриманої мережі об'єктів кібербезпеки дозволяють наглядно спостерігати за станом і динамікою розвитку понятійної бази предметної області.

## Перелік посилань


1. D. Lande, O. Puchkov, I. Subach, M. Boliukh, D. Nahornyi OSINT investigation to detect and prevent cyber attacks and cyber security incidents // Information Technology and Security. Том 9, N 2 (2021). - С. 209-218. DOI: doi.org/10.20535/2411-1031.2021.9.2.249921.
2. Lande, D., Dmytrenko, O. Creating Directed Weighted Network of Terms Based on Analysis of Text Corpora. 2020 IEEE 2nd International Conference on System Analysis and Intelligent Computing, SAIC 2020, 2020, 9239182
3. Cherven K. Mastering Gephi Network Visualization. – Packt Publishing, 2015. – 378 p. ISBN 78-1-78398-734-4.

# ЗМІСТ