УДК 004.492

РОЛЬ МЕТОДІВ OSINT В КІБЕРБЕЗПЕЦІ ТА ЇХ ЗАСТОСУВАННЯ ПІД ЧАС ВОЄННИХ КОНФЛІКТІВ

І. О. Мірошніченко[1,a], Д. В. Ланде[1]

[1] Навчально-науковий Фізико-технічний інститут

Анотація

Ця стаття зосереджена на дослідженні ролі збору розвідувальної інформації із загальнодоступних джерел (OSINT) та її аналізу у контексті кібербезпеки і важливості даної процедури у військових конфліктах. Вирішення проблеми ефективного виявлення, аналізу та протидії кіберзагрозам у військових операціях. У тезах розглядається процес вилучення ключових подій кібербезпеки під час війни та те, як ідентифікувати конкретні суб'єкти, які відіграють важливу роль у кіберпросторі під час війни. Визначено ключові аспекти та методи, які можуть бути використані для ефективного забезпечення протидії пропаганді в ситуаціях військового протистояння.

Ключові слова: OSINT, NLP, Elasticsearch

Вступ

З появою соціальних медіа і впливом інтернету попри всі сфери життя, світ став свідком безлічі військових конфліктів, де фронт інформаційної війни грає найважливішу роль у формуванні думки мирних жителів.

Метою мого дослідження є аналіз використання методів Open Source Intelligence у контексті кібербезпеки під час збройної агресії росії проти України. Це завдання включає визначення ключових засобів OSINT і їх застосування для вилучення найважливіших подій у кібербезпеці, ідентифікації об'єктів для аналізу, побудови взаємопов'язаних мереж і аналізу динаміки змін у кіберпросторі за фрагмент періоду ведення війни.

В результаті, буде продемонстровано потенціал використання інформації з відкритих джерел, що допоможе визначити кібераспекти конфліктів сьогодення, яка може стати основою визначення та протистояння дезінформації та розробити стратегії ведення війни у майбутньому.

1. Сучасні методи ведення війни

Військові конфлікти в епоху цифрових технологій набирають нових обертів. Росія вступила в гібридну війну з Україною, використовуючи конвенційну зброю, тероризм, порушення прав і свобод людини, пропаганду та дезінформацію суспільства, вплив на громадську думку громадян, підрив стабільності в країні, цензуру та заборону культури та мови. Інформаційна війна стала ключовим моментом розвитку збройної агресії. Поширення фейкових новин, створення ботоферм у соціальних мережах, спроби дескридитації влади – все це становить неймовірну загрозу для мирних людей та держави. Open Source Intelligence в реаліях сьогодення дозволяє виявляти приховані місця розташування військової техніки і живої сили противника, виявляти маршрути пересування ворога, персонифікувати осіб, підтверджувати гіпотези та будувати стратегій бойових дій.

Наприклад, завдяки OSINT-агенції «Molfar» відбуваються розслідування воєнних злочинів на окупованих територіях, ідентифікуються російські військовослужбовці, які порушують права українців щодня.

Російський уряд також приділяє особливу увагу поширенню через державні ЗМІ негативної інформації про політичні та соціальні аспекти інших країн. Кампанії дезінформації мають великий вплив на демократичні процеси в різних країнах.

- У Німеччині, наприклад, німецькі політичні партії та урядові установи стали жертвами кібератак з російських джерел, включаючи публікацію електронних листів передвиборчої кампанії Клінтон хакерською групою Fancy Bear.
- Різні країни також стикаються з підривними рухами та дезінформацією, які поширюються різними каналами та призводять до порушень виборчих процесів, урядів і соціальних норм.[1]

2. Основні джерела і методи OSINT

Використання відкритих джерел інформації зазвичай передбачає три основні методи: пасивний, напівпасивний та активний.[2] Вибір найбільш ефективного методу залежить від конкретних умов збору інформації та типу даних, що отримуються.

- Пасивний збір інформації – це використання загальнодоступних ресурсів для збору даних.
- Напівпасивний метод надсилає обмежену кіль-

[a]ilomir-ipt24@lll.kpi.ua

Таблиця 1. Порівняльна таблиця

Інструмент	SHODAN	MALTEGO	RECON-NG	SPIDERFOOT
Порівняння функцій та можливостей	– Пошук підключених до Інтернету пристроїв за різними параметрами (наприклад, IPa адреса, порти, оервіои). – Виявлення вразливостей у підключених пристроях.	– Збір та аналіз великих обсягів даних з різних джерел. – Виявлення зв'язків між різними об'єктами. – Візуалізація аналізованих даних.	– Збір інформації з різних джерел, таких як соціальні мережі, пошукові системи тощо. – Аналіз зібраної Інформації для виявлення загроз та вразливостей.	– Збір інформації з різних джерел у мережі інтернет. – Аналіз зв'язків між об'єктами. – Виявлення загроз безпеці та вразливостей .

кість трафіку на цільовий сервер для отримання загальної інформації, для того, щоб збір інформації не привертав небажаної уваги від зовнішніх користувачів.

- Активне збирання розвідувальних даних передбачає безпосередню взаємодію із системою чи службою, де ведеться розвідка. Даний спосіб є прозорим для власника, він може переглядати та аналізувати його.

Для розвідки з відкритих джерел інформації існує безліч потужних інструментів, які надають різні функції, залежно від потреб. У Таблиці 1 зображені популярні інструменти збору та аналізу інформації з описаними функціями та можливостями.

3. Проблема використання відкритих джерел інформації для розвідки в соціальних мережах

Для успішного аналізу пропагандистських каналів було ухвалено рішення вивчення повідомлень і постів у соціальних мережах і месенджерах. Однак, серйозним випробуванням в автоматизації Open source intelligence стала ідентифікація суб'єктів і подальше прогнозування їхньої активності.

Завдання формулювання іменованих сутностей зводиться до двох головних кроків: визначення самого імені (особи, організації, місця розташування), яке можна позначити власним ім'ям, і категорії відповідно до онтології.[3]

Псевдоніми користувачів, які використовуються в онлайн-мережах, ускладнюють процес ідентифікації потрібних людей та організацій через використання кількох імен та анонімність користувачів, що призводить до неточності в аналізі джерел.[4]

Ще однією проблемою стала класифікація об'єктів за категоріями. Онтологія встановлює систему категорій та їхніх зв'язків, однак через різноманітність і динамічний характер контенту в сучасних інформаційних ресурсах потребує додаткового дослідження.

Розпізнавання імен означає асоціювання потрібної нам особи з кожною названою сутністю.

Множина суб'єктів розвідки може мати однакові ознаки: назву, ініціали, але вони відрізняються за зовнішнім виглядом, стилем написання, інтересами, мовою. Основна задача полягає в тому, щоб ідентифікувати особу точно з множини ймовірних співпадінь без додаткового втручання людини.

4. Практична реалізація методів Open source intelligence

4.1. Збір інформації з відкритих джерел

Критичним етапом для подальшої роботи з інформацією з пропагандистських джерел є дослідження актуальних матеріалів за відповідною темою. Під час мого розслідування було розглянуто варіанти API соціальних мереж, публічні сайти та набори даних. Вирішено використати датасет із тисячами постів із соціальної мережі «Twitter», де також зазначаються:

- Дата публікації повідомлення
- Час
- Ідентифікатор користувача
- Ім'я користувача
- Безпосередньо повідомлення з проросійським посилом.

4.2. Попередня обробка природної мови(NLP)

Нормалізація тексту являє собою перетворення тексту в придатну для роботи форму за допомогою регулярних виразів.[5] У моїй роботі було видалено розділові знаки, емодзі та інші символи, великі літери, а також цифри. Для прикладу, в табл. 2, зображено тестову інформацію з проросійських каналів до і після цієї обробки.

Після першого підготовчого етапу була проведена **токенізація тексту**, розбиття речення за словами для отримання уявлення про тенденції в текстовому наборі даних. На цьому етапі потрібна була обережність і врахування особливостей мови, адже під час токенізації деяких слів вони можуть просто втратити сенс для подальшого аналізу.

При видаленні **стоп-слів** був визначений список загальновживаних слів, які не несуть значущої інформації для аналізу – прийменники, займенники, сполучники тощо.

Таблиця 2. Результат нормалізації тексту

tweet (**BEFORE**)
Украинские пограничники пересекли территорию России в Брянской области и Крыму..
Украинские системы ПВО были подавлены и дезактивирована военная инфраструктура военных авиабаз...
Бывший заместитель министра обороны Украины Алексей Селиванов призвал украинских военных вместе сзащитниками Донбасса свергнуть преступный режим в Киеве.
11 Колонны украинской военной техники в центре Киева в данный момент...
В Минобороны России заявили, что на территорию Украины не попали никакие политические объекты, только военная инфраструктура, мирному населению ничего не угрожает.

tweet (**AFTER**)
украинские пограничники пересекли территорию россии в брянской области и крыму
украинские системы пво были подавлены и дезактивирована военная инфраструктура военных авиабаз
бывший заместитель министра обороны украины алексей селиванов призвал украинских военных вместе с защитниками донбасса свергнуть преступный режим в киеве
колонны украинской военной техники в центре киева в данный момент
вминобороны россии заявили что на территорию украины не попали никакие политические объекты только военная инфраструктурамирному населению ничего не угрожает

Вибір оптимального алгоритму стемінгу. Стемінг є простим і практичним підходом до нормалізації тексту, який дозволяє об'єднувати різні варіанти слів в один формат для подальшого аналізу та порівняння. Цей процес дозволяє отримати кореневу форму слова, за допомогою відрізання кінців слів у іменниках та дієсловах, що позитивно впливає на якість обробки даних. В процесі роботи був використаний алгоритм «Snowball» через можливість його адаптації для різних мов і ситуацій.

Основними кроками алгоритму Сноубола є:

1. Визначення кореневого правила: Визначається правило, що визначає, які афікси можна видалити для певного типу слова.
2. Ітеративний процес стемінгу: Кожне слово розбивається на основу і закінчення згідно з визначеними правилами, і відкидається закінчення.
3. Перевірка на зупинку: Процес повторюється доти, доки не буде досягнуто стану, коли жодне з правил не може бути застосовано.
4. Вибір основи: Основа, яка залишається після видалення закінчення, стає стемом слова.

4.3. Аналізатор ElasticSearch

Для аналізу великих обсягів даних з датасету, ефективного пошуку та індексації було вирішено обрати високопродуктивний пошуковий движок «ElasticSearch». Основною перевагою даного рішення є підтримка REST API, що робить його легким у інтеграції з іншими системами та у використанні [6]. Аналізатор може бути розгорнутий на кластері серверів, це дозволяє розподілити навантаження системи на декілька вузлів, що поліпшить надійність та масштабованість. Важливим для аналізування та прогнозування великих обсягів даних у моєму випадку дослідження тисяч пропагандистських джерел є висока швидкість роботи.

5. Висновки

У даній статті розглянута проведено дослідження перспективного напрямку розвідки – OSINT. Було розглянуто вплив дезінформації та пропаганди на суспільство та ситуацію в країні. Наступним кроком у дослідженні став розгляд відкритих джерел інформації, а також методів збору даних. Також була підкреслена важливість подальших досліджень у цій галузі та потенціал для розробки нових стратегій і методів забезпечення кібербезпеки під час військових конфліктів.

Перелік використаних джерел

1. *Bennett W. L.*, *Livingston S.* The disinformation order: Disruptive communication and the decline of democratic institutions // European Journal of Communication. — 2018. — Vol. 33, no. 2. — P. 122–139. — DOI: 10.1177/0267323118760317 . — URL: https://journals.sagepub.com/doi/10.1177/0267323118760317.
2. Current Status and Security Trend of OSINT / Y.-W. Hwang, I.-Y. Lee, H. Kim, H. Lee, D. Kim. — 2022. — DOI: 10.1155/2022/1290129. — URL: https://www.hindawi.com/journals/wcmc/2022/1290129/.
3. *Layton R.*, *Watters P. A.* Automating Open Source Intelligence: Algorithms for OSINT. — Elsevier, 2016. — 205 p. — ISBN 9780128029169.
4. *Lande D.*, *Shnurko-Tabakova E.* OSINT as a part of cyber defense system // Theoretical and Applied Cybersecurity. — 2019. — Vol. 1, no. 1. — P. 103–108. — DOI: 10.20535/tacs.2664-29132019.1.169091. — URL: http://tacs.ipt.kpi.ua/article/view/169091.
5. *Jurafsky D.*, *Martin J. H.* Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. — Prentice Hall, 2023. — 577 p.
6. *Ланде Д. В.*, *Субач І. Ю.*, *Гладун А. Я.* Оброблення надвеликих масивів даних (Big Data) : навчальний посібник. — Київ: ТОВ "Інжиніринг", 2021. — 168 с. — ISBN 978-966-2344-83-7.