УДК 004.5

Аналіз динаміки і взаємозв'язків об'єктів кібербезпеки

Ланде Д.В.

Постановка проблеми

В роботі представлено пропонується метод автоматичного екстрагування і виявлення об'єктів предметної області (зокрема, кібербезпеки) в інформаційних потоках [1], аналізу їх взаємозв'язків і візуалізації. Засоби виявлення об'єктів як іменованих сутностей будуються на основі концепцій машинного навчання. У подальшому вивчається динаміка згадувань цих об'єктів, після чого досліджуються взаємозв'язки об'єктів, визначаються їх окремі кластери [2]. Запропоновано форму візуального відображення інформаційного потоку в розрізі об'єктів і дат, що є прямокутною таблицею (діаграма Wordlet), комірки якої заповнені чисельними значеннями, що відповідають частотам появи найменувань об'єктів в інформаційних потоках у розрізі дат [3]. Розглянутий підхід може застосовуватись для вирішення питань аналізу та візуалізації розподілу об'єктів для будь-яких відібраних інформаційних масивів у розрізі питань, що цікавлять дослідника та мають значні часові рамки.

Мета

Метою доповіді є представлення методу аналізу динаміки і взаємозв'язків об'єктів предметної області. Первинною інформацією для заповнення такої системи виступають інформаційні потоки мережі Інтернет. Поряд із методом розглядається його застосування для визначеної предметної області – кібербезпеки.

Виклад основного матеріалу

Пропонується до розгляду метод, сутність якого полягає у виконанні таких технологічних операцій, як експертне створення запитів до наявних інформаційно-пошукових систем [1], що відповідає предметній області. В результаті опрацювання цих запитів створюються великі за обсягом масиви релевантних документів, в яких за допомогою спеціальних алгоритмів визначаються необхідні фрагменти. На базі відібраних масивів екстрагуються іменні сутності (об'єкти), що відносяться до різних періодів часу. У подальшому за допомогою сучасних методів аналізу мереж досліджуються взаємозв'язки об'єктів, визначаються їх окремі кластери.

Запропонована форма візуального відображення інформаційного потоку в розрізі об'єктів, що є прямокутною таблицею (діаграма Wordlet), в комірки якої занесені числа, що відповідають кількості згадувань обраного об'єкту в розрізі дат. Тобто. стовпцям цієї таблиці відповідають дати, а рядкам – об'єкти предметної області. При візуалізації світлі відтінки відповідають більшим значенням, темні – меншим. Також будується кореляційна мережа цих об'єктів, де зв'язками є значення кореляцій, інформаційних потоків, що відповідають цим суб'єктів (Рис. 1). Кластеризація (розфарбовування) мережі здійснено за алгоритмом модулярності. Кластери сформовані за кореляцією часу подій, в яких фігурували визначені об'єкти.

Таким чином, для реалізації запропонованого метода: 1) створюється набір стартових запитів до наявних інформаційно-пошукових систем; 2) розроблено програмне забезпечення (ПЗ) витягу необхідних фрагментів із вибраних документів; 3) розроблено ПЗ екстрагування об'єктів на базі моделей машинного навчання; 4) адаптовано ПЗ формування кореляційних мереж взаємозв'язку об'єктів, їх візуалізації, кластерного аналізу; 5) розроблено ПЗ візуалізації діаграм Wordlet.

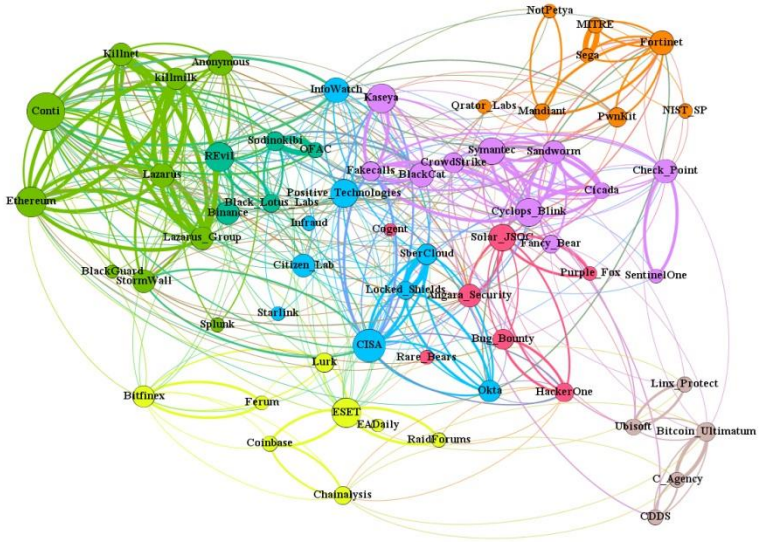

Рис. 1 – Мережа об’єктів кібербезпеки

На Рис. 2. наведено діаграму Wordlet для понять, що відповідають предметній області кібербезпеки, представленим на Рис. 1. Горизонтальні світлі риски на діаграмі відповідають періодам активності відповідного об’єкта в інформаційному полі Інтернету.

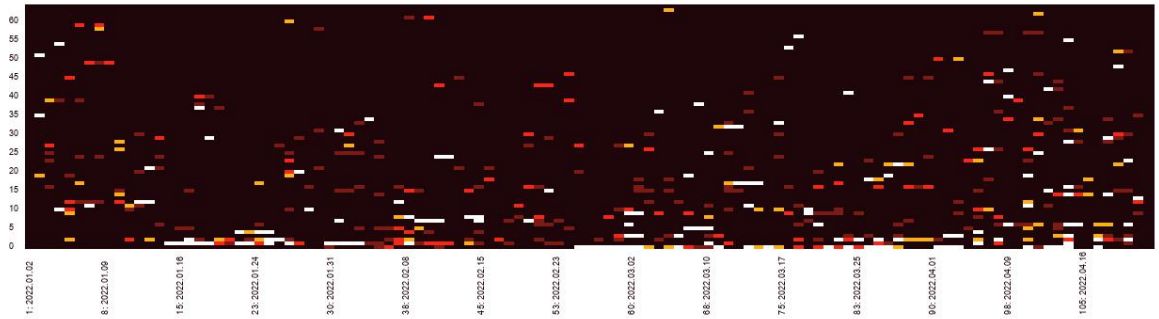

Рис. 2 – Діаграма, що відповідає активності об’єктів кібербезпеки

Висновки

Розроблено метод і побудовано діючий макет системи аналізу динаміки і взаємозв’язків об’єктів моніторингу інформаційних ресурсів мережі Інтернет. Первинною інформацією для заповнення такої системи виступають інформаційні потоки мережі Інтернет. Новизна підходу полягає в побудові загальної методології вирішення задачі аналізу інформаційних потоків, формуванні і кластерного аналізу кореляційних мереж об’єктів, а також візуалізації тенденцій предметної області за допомогою діаграм Wordlet.

Література

1. Ланде Д.В., Пучков О.О., Субач І.Ю. Агрегація інформації з різнорідних мереж як основа підготовки фахівців з кібербезпеки з питань оброблення надвеликих масивів даних // Information Technology and Security, 2021. – Том 9. – № 1. – С. 4-16. DOI: doi.org/10.20535/2411-1031.2021.9.1.247256.

2. Ланде Д., Страшной Л., Балагура І. Метод формування та кластеризації кореляційних мереж понять // Реєстрація, зберігання і обробка даних, 2021. – Том 23. – № 2. – С. 27-36. DOI: doi.org/10.35681/1560-9189.2021.23.2.239209.

3. Ландэ Д.В., Григорьев А.Н., Брайчевский С.М., Дармохвал А.Т., Снарский А.А. Объектная визуализация тематических информационных массивов // Труды 9-ой научной конференции "Электронные библиотеки: перспективные методы и технологии, электронные коллекции" – RCDL'2007, Переславль-Залесский, 2007. – С 148-150.