

СЕМАНТИЧНИЙ АНАЛІЗ ТЕКСТОВОГО КОНТЕНТУ ДЛЯ ВИЯВЛЕННЯ ІНФОРМАЦІЙНИХ ВПЛИВІВ НА АКТОРІВ У СОЦІАЛЬНИХ ІНТЕРНЕТ-СЕРВІСАХ

На сучасному етапі завдяки комунікаційним перевагам соціальні інтернет-сервіси (СІС) перетворилися на потужний інструмент взаємодії громадянського суспільства і держави, формування суспільної думки з багатьох актуальних питань. Однак, СІС стали і дієвим засобом проведення інформаційних операцій проти людини, суспільства, держави [1, 2]. Виявлення ознак інформаційних операцій у першу чергу пов'язане з моніторингом контенту СІС. Під контентом у СІС будемо розуміти інформаційне наповнення віртуальних спільнот з метою зацікавлення акторів, вираження власної токи зору на події, поширення інформації тощо. Встановлено, що найбільшу частку інформаційного середовища СІС займає текстовий тип даних. Тому розроблення ефективних алгоритмів функціонування системи забезпечення інформаційної безпеки держави для вирішення проблеми моніторингу текстового контенту СІС є актуальним теоретико-прикладним завданням.

У результаті узагальнення відомих підходів до інформаційного пошуку і лінгвістичного аналізу контенту СІС розроблено метод виявлення інформаційних впливів на основі встановлення їх ознак у змісті текстового контенту. На першому етапі методу формується семантичне ядро для пошуку контенту відповідно до критерію актуальності, критичності та рівня обговорення у суспільстві. Індексация контенту виконується на основі латентно-семантичного індексування (LSI) [3]. Другий етап полягає у безпосередньому виявленні загроз інформаційній безпеці державі у СІС. Контент, проіндексований і відібраний на попередньому етапі, підлягає семантичному аналізу на базі онтологій. Спочатку складається онтологія нормального функціонування віртуальної спільноти в СІС. Після цього виконується побудова семантичного опису текстового контенту, виявленого на першому етапі методу. Далі виявляються ознаки загроз у попередньо проіндексованому на першому етапі методу текстовому контенті СІС на основі сигнатурного і методу аномалій [4].

Також проведено експериментальну перевірку достовірності розробленого методу і досліджено контент СІС Facebook на предмет наявності ознак інформаційного впливу. Було задано семантичне ядро $W = \langle \text{майдан}; \text{бандити}; \text{влада}; \text{поліція}; \text{корупція}; \text{диктатура}; \text{війна}; \text{злидні}; \text{Порошенко};$

олігархи> для його пошуку в колекції публікацій актора Миколи Гайдука. У результаті першого етапу отримано розподіл колекції публікацій у семантичному просторі. На другому етапі проведено виявлення загроз на базі сигнатурного методу і методу аномалій у відібраному текстовому контенті СІС. Це дозволило детектувати використання з концептом відношення, яке становить фізичну небезпеку для об'єкта публікації. Також виявлено протиріччя понять в змісті контенту. У досліджуваному документі використано атрибут концепту, який необхідно додати до семантичних шаблонів загроз онтологічної бази знань. Одержані результати збіжні, наприклад, з дослідженнями аналітичного центру Текст.орг, яким встановлено взаємопов'язані віртуальні спільноти у Вконтакте і Facebook, актори яких закликали до перевороту й протестів в Україні. У результаті розкрито адміністратора таких груп, зареєстрованого у СІС під іменем Микола Гайдук.

Перевагою розглянутого методу є застосування LSI для пошуку текстового контенту СІС, який містить деструктивний інформаційний вплив у явному або прихованому вигляді. Завдяки додатковому використанню сигнатурного методу і методу аномалій для детектування відомих та нових загроз інформаційній безпеці досягається взаємна компенсація недоліків таких підходів. Таким чином, досягається підвищення ефективності й швидкодії системи забезпечення інформаційної безпеки держави в СІС, що є сьогоднім вкрай актуальним завданням для України.

Список літератури

1. Гришук Р. В. *Основи кібернетичної безпеки : монографія* / Ю. Г. Даник, Р. В. Гришук; за заг. ред. проф. Даника Ю. Г. – Житомир : ЖНАЕУ, 2016. – 636 с.
2. Молодецька К. В. *Узагальнена класифікація загроз інформаційній безпеці держави в соціальних інтернет-сервісах* / К. В. Молодецька // *Защита информации*. – 2016. – Вып. 23. – С. 75–87.
3. Марченко О. О. *Порівняння методів онтологічного семантичного аналізу та алгоритмів латентного семантичного аналізу* / О. О. Марченко // *Вісн. Київського нац. ун-ту ім. Т. Шевченка*. – 2012. – 2. – С. 169–174.
4. Чернишук С. В. *Методика виявлення кібернетичних загроз у природномовних текстах* / С. В. Чернишук // *Проблеми створення, випробування, застосування та експлуатації складних інформаційних систем*. – 2013. – Вып. 8. – С. 112–121.