УДК 004.056.5:004.738.5(045)

ВИЯВЛЕННЯ ІНФОРМАЦІЙНИХ ВПЛИВІВ У СОЦІАЛЬНИХ ІНТЕРНЕТ-СЕРВІСАХ НА ОСНОВІ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ТЕКСТОВОГО КОНТЕНТУ

Молодецька-Гринчук К. В.
к.т.н., доцент,
доц. каф. комп'ютерних технологій і моделювання систем
Житомирський національний агроекологічний університет
kmolodetska@gmail.com

Анотація. *Для вирішення проблеми ефективного моніторингу інформаційного середовища соціальних інтернет-сервісів у публікації запропоновано метод виявлення деструктивних інформаційних впливів на основі їх змістовних ознак у текстовому контенті. Завдяки розробленому методу автоматизуються процедури ідентифікації у віртуальних спільнотах інформаційних операцій, направлених проти інформаційної безпеки людини, суспільства, держави. В результаті досягається підвищення ефективності й швидкодії системи забезпечення інформаційної безпеки держави в соціальних інтернет-сервісах.*

На сучасному етапі значну роль в процесах комунікації суспільства відіграють соціальні інтернет-сервіси (СІС), які забезпечують учасників віртуальних спільнот – акторів, новітніми засобами взаємодії. Завдяки комунікаційним перевагам [1] 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. Дослідження показали, що найбільшу частку інформаційного середовища СІС займає текстовий тип даних. Зміст текстового контенту може містити деструктивний інформаційний вплив у явному або прихованому вигляді [2]. Висока складність процедур аналізу змісту текстового контенту й особливо автоматичного аналізу, призводить до ускладнення процесу виявлення початку та самого факту інформаційного впливу в СІС. Тому розроблення ефективних алгоритмів функціонування системи забезпечення інформаційної безпеки держави для вирішення проблеми моніторингу текстового контенту СІС є актуальним теоретико-прикладним завданням.

Аналіз останніх досліджень і публікацій показав, що для обробки і аналізу текстового контенту використовуються методи статистичного й лінгвістичного аналізу [3]. Перша група методів ґрунтується на аналізі змісту контенту за частотою слів, які в ньому використовуються. Спільним недоліком групи статистичних методів є неможливість врахування зв'язності текстового контенту. Для усунення цього недоліку використовують методи лінгвістичного аналізу, які включають, зокрема, семантичний аналіз для змістовного розуміння текстового контенту. Семантичний аналіз є складною процедурою, яка ґрунтується на використанні баз знань і тезаурусів для відображення зв'язку між окремими словами й словосполученнями [3]. Проблема виявлення ознак інформаційних впливів за змістовними ознаками не обмежується лінгвістичним аналізом текстового контенту СІС. Детектування загроз системою забезпечення інформаційної безпеки держави в СІС належить, зокрема, до задач інформаційного пошуку [3]. Дослідження [3-5] показали, що перспективним напрямком є використання моделей семантичного пошуку та аналізу, які враховують зміст текстового контенту.

Мета досліджень полягає в розробленні методу виявлення інформаційних впливів у СІС за змістовними ознаками, який дозволить підвищити ефективність функціонування системи забезпечення інформаційної безпеки держави.

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, який зводиться до такого.

Етап 1. Пошук текстового контенту в СІС за заданим інформаційним приводом. На цьому етапі визначаються ключові слова для пошуку текстового контенту в СІС за критерієм актуальності, критичності та рівня обговорення у суспільстві його тематики. У даному випадку поставлена задача зводиться до вибору методу інформаційного пошуку текстового контенту, який задовольняє вимогам релевантного пошуку текстового контенту, ефективній обробці, зокрема, коротких публікацій. Тому пропонується скористатися методом латентно-семантичного індексування (LSI) [3]. Особливостями LSI є пошук контенту в СІС на основі його змісту, а не щільності ключових слів, і пошук прихованих семантичних зв'язків між ключовими словами й безпосередньо контентом.

Етап 2. Виявлення ознак інформаційних впливів у СІС на основі сигнатурного методу і методу виявлення аномалій. Суть даного етапу полягає у виявленні загроз інформаційній безпеці державі в СІС, які містяться в текстовому контенті. Контент віртуальних спільнот, проіндексований і відібраний на попередньому етапі, підлягає семантичному аналізу на базі онтологій [4, 5]. Для цього складається онтологія функціонування віртуальної спільноти в СІС, будується семантичний опис текстового контенту, виявленого на першому етапі методу. Далі виявляють ознаки загрози у попередньо проіндексованому на першому етапі текстовому контенті СІС, що зводиться до такого [4]:

1) сигнатурний метод – виявлення зв'язку між об'єктом публікації і його характеристиками у контенті та негативними ознаками для цього об'єкта внаслідок реалізації загрози;

2) виявлення суперечностей на основі методу аномалій шляхом співставлення семантичного опису проіндексованого текстового контенту і семантичного шаблону загрози.

Розгорнутий аналіз експертами проіндексованого текстового контенту на предмет наявності загроз необхідний, якщо контент СІС є релевантним семантичному ядру пошукового запиту до контенту СІС з етапу 1, однак на етапі 2 не виявлено суперечливих фрагментів такого контенту і онтології. Тоді ідентифіковані експертами небезпечні семантичні конструкції додаються до шаблонів загроз і використовуються в онтології. У результаті таких дій метод LSI забезпечує ефективне виявлення залежностей між лексичними одиницями в контенті СІС для наповнення онтологічних баз знань.

Перевагою розробленого методу є застосування LSI для пошуку текстового контенту віртуальних спільнот, який містить деструктивний інформаційний вплив в 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, забезпечується підвищення ефективності й швидкодії системи забезпечення інформаційної безпеки держави в СІС.

Література:

1. Fuchs Chr. Social Media: Critical Introdaction / Christian Fuchs. – Sage, 2013. – 304 p.

2. Грищук Р. В. Основи кібернетичної безпеки : моногр. / Ю. Г. Даник, Р. В. Грищук; за заг. ред. проф. Даника Ю. Г. – Житомир : ЖНАЕУ, 2016. – 636 с.

3. Manning Chr. Introduction to Information Retrieval / Chr. Manning, P. Raghavan, H. Schütze. – Cambridge University Press, 2008. – 544 p.

4. Чернишук С. В. Методика виявлення кібернетичних загроз у природномовних текстах / С. В. Чернишук // Проблеми створення, випробування, застосування та експлуатації складних інформаційних систем. – 2013. – Вип. 8. – С. 112–121.

5. Марченко О. О. Порівняння методів онтологічного семантичного аналізу та алгоритмів латентного семантичного аналізу / О. О. Марченко // Вісн. Київського нац. ун-ту ім. Т. Шевченка. Сер. фіз.-мат. науки. – 2012. – 2. – С. 169–174.