УДК 004.056:004.738.5(045)
DOI: 10.18372/2310-5461.38.12855

О. В. Барабаш, д-р техн. наук, проф.
Державний університет телекомунікацій
orcid.org/0000-0003-1715-0761
e-mail: bar64@ukr.net;

Р. В. Грищук, д-р техн. наук, старш. наук. співроб.
Житомирський військовий інститут імені С. П. Корольова
orcid.org/0000-0001-9985-8477
e-mail: Dr.Hry@i.ua;

К. В. Молодецька-Гринчук, канд. техн. наук, доц.
Житомирський національний агроекологічний університет
orcid.org/0000-0001-9864-2463
e-mail: kmolodetska@gmail.com

ВИЯВЛЕННЯ ЗАГРОЗ ІНФОРМАЦІЙНІЙ БЕЗПЕЦІ ДЕРЖАВИ У ЗМІСТІ ТЕКСТОВОГО КОНТЕНТУ СОЦІАЛЬНИХ ІНТЕРНЕТ-СЕРВІСІВ

Вступ

На сучасному етапі внаслідок інтенсивного розвитку інформаційних технологій відбулися процеси диверсифікації каналів поширення контенту і зростання ролі комунікативної складової у системі стратегічних комунікацій суспільства. Результатом постійного удосконалення інформаційного середовища є перетворення соціальних Інтернет-сервісів (СІС) на один з найбільш популярних засобів масової комунікації [1–3]. Сьогодні СІС забезпечують користувачів, яких прийнято називати *акторами*, інструментами створення і поширення мультимедійного контенту, організації акторів у віртуальні спільноти, координації їх взаємодії в реальному житті тощо. Тому СІС являють собою дієвий засіб впливу на процеси управління в державі, що безпосередньо пов'язані з інформаційною безпекою людини, суспільства, держави [3]. У світлі останніх подій в Україні та світі СІС виступають інструментом проведення інформаційних операцій унаслідок цілеспрямованого поширення недостовірного чи викривленого контенту та ведення негативної пропаганди. Наслідками таких дій є розпалювання суспільної ворожнечі, загострення соціальних конфліктів, зростання протестних настроїв, міжнаціональної та міжетнічної ворожнечі тощо [2]. Тому одним із важливих завдань системи забезпечення інформаційної безпеки держави (ІБД) є своєчасне виявлення контенту в СІС, який містить деструктивний інформаційний вплив на акторів та становить загрозу ІБД. Проблема виявлення загроз ІБД у текстовому контенті СІС пов'язана з існуванням об'єктивного протиріччя між рівнем розвитку сучасних технологій інформаційного впливу на акторів і науковим базисом їх детектування на основі змістовного аналізу природомовного тексту. При цьому окреслена проблема не обмежується проведенням лінгвістичного аналізу текстового контенту і належить, зокрема, до задач інформаційного пошуку. Таким чином, складність процедур завчасного виявлення загроз ІБД у змісті текстового контенту віртуальних спільнот зумовлена відсутністю комплексного підходу до їх пошуку та детектування. Отже, розроблення дієвих підходів до виявлення інформаційного впливу на акторів у СІС за результатами аналізу змісту текстового контенту є актуальним науково-прикладним завданням.

Аналіз останніх досліджень і публікацій

Критичний аналіз публікацій за напрямком дослідження продемонстрував, що для оброблення природомовних текстів використовуються методи статистичного або лінгвістичного аналізу [4–8]. Суть статистичних методів полягає в аналізі змісту текстового контенту на основі частоти використання окремих слів або *Bag of Words*. Такий підхід не дозволяє врахувати зв'язність текстового контенту, що є однією із ключових вимог ефективності виявлення загроз у змісті текстового контенту, які мають прихований характер. Тому для розв'язку поставленої задачі необхідно використати методи лінгвістичного аналізу, метою яких є виявлення структурно-смислової єдності текстового контенту, його комунікативної спрямованості та інтерпретації для встановлення смислу. Лінгвістичний аналіз складається з декількох основних етапів, серед яких найбільший інтерес для досліджень становить семантичний аналіз. Процедура семантичного аналізу спирається на використання баз

знань і тезаурусів та призначена для відображення зв'язку між окремими словами й словосполученнями у текстовому контенті.

Також встановлено, що проблема виявлення загроз ІБД у змісті текстового контенту СІС нерозривно пов'язана з використанням методів інформаційного пошуку [9]. Тому перспективним є застосування методів семантичного пошуку і аналізу для врахування змісту текстового контенту віртуальних спільнот.

У публікації [4] представлено метод виявлення загроз ІБД у СІС у змісті текстового контенту з використанням технологій семантичного аналізу даних на базі онтологій й латентно-семантичного аналізу.

Перевагою запропонованого методу є підвищення ефективності виявлення прихованих ознак загроз у текстовому контенті й наповнення онтологічних баз знань системи забезпечення ІБД новими та невідомими до цього семантичними шаблонами загроз.

Однак, такий метод не забезпечує ранжування релевантності виявленого текстового контенту СІС на етапі інформаційного пошуку відповідно до сформованого семантичного ядра пошукового запиту до сервісу. Таким чином, перспективним напрямком досліджень є підвищення релевантності досліджуваного текстового контенту, що дозволить збільшити оперативність та ефективність процедур раннього виявлення ознак загроз системою забезпечення ІБД у СІС.

Мета статті (постановка завдання)

Необхідно розробити підхід до виявлення загроз ІБД у змісті текстового контенту СІС для своєчасного виявлення деструктивного інформаційного впливу на акторів віртуальних спільнот. Такий підхід повинен забезпечити підвищення релевантності досліджуваного текстового контенту СІС, виявлення відомих семантичних шаблонів загроз ІБ, наповнення онтологічних баз знань невідомими до цього шаблонами загроз і реалізованість в умовах обмеженості наявних ресурсів системи забезпечення ІБД.

Виклад основного матеріалу

Особливістю функціонування СІС є використання такої архітектури поширення контенту, яка забезпечує вплив на емоційний сферу акторів. Так, публікації акторів або віртуальних спільнот є загальнодоступними і викликають прояв у інших користувачів СІС емоційної оцінки такому контенту у формі коментарів, лайків, повторних публікацій.

Використання прихованого інформаційного впливу на акторів сумісно з технологіями маніпулятивного впливу призводить до маніпулювання суспільною думкою. На основі результатів звіту *Information disorder: toward an interdisciplinary framework for research and policy making* [10] встановлено, що здійснення інформаційного впливу в СІС реалізується на основі таких складових концептів:

1) три типи інформаційних впливів — дезінформація, неповний контент, шкідливий контент;

2) три етапи інформаційного впливу — створення, виробництво, споживання;

3) три компоненти інформаційного впливу — актор, контент, інтерпретація.

Розглянуті складові інформаційного впливу, відповідно за запропонованого у публікації [11] підходу, формують концептуальну модель, яка складається з множини відповідних концептів.

Кожен з визначених концептів описується відповідною змінною, яка описує стан відповідного концепту. При цьому окремі концепти представляються собою об'єкти інформаційного простору СІС, які впливають на інформаційну безпеку людини, суспільства, держави.

Взаємозв'язок розглянутих складових інформаційного впливу у СІС подано на рис. 1.

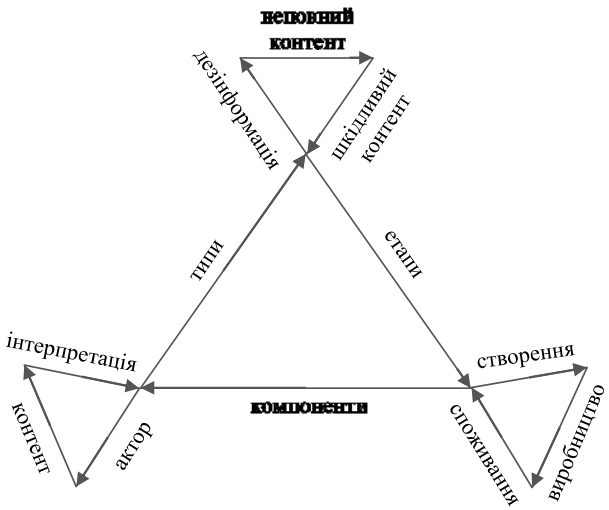

Рис. 1. Взаємозв'язок концептів інформаційного впливу у СІС

Узагальнена характеристика концептуальної моделі подана в таблиці. Таким чином, інформаційний вплив на акторів СІС описується запропонованою концептуальною моделлю, що дозволяє розробити системну модель виявлення загроз ІБД у змісті текстового контенту віртуальних спільнот. Використання системного підходу до процесів виявлення інформаційного впливу на акторів віртуальних спільнот у змісті текстового контенту дозволив реалізувати комплексний підхід до пошуку та виявлення загроз ІБД у СІС. Для цього запропоновано системну модель виявлення загроз ІБД у змісті текстового контенту СІС, подану на рис. 2.

Характеристика концептуальної моделі

Множина концептів	Змінні стану	Примітка
$X^{type} = \{X_i^{type}\}$	Y_i^{type}	Типи інформаційних впливів у СІС: *дезінформація* — навмисно створений штучний контент, який має на меті деструктивний вплив на акторів, групу акторів, суспільство, державу; *неповний контент* — недостатній для прийняття обґрунтованих рішень за повнотою контент, не має на меті деструктивний вплив на акторів; *шкідливий контент* — контент, в основу якого покладено реальні факти, спрямований на нанесення шкоди актору, групі акторів, суспільству, державі
$X^{phase} = \{X_i^{phase}\}$	Y_i^{phase}	Етапи здійснення інформаційного впливу: *створення* — зводиться до вироблення наративу для подальшого інформаційного впливу на визначену цільову аудиторію; *виробництво* — полягає у формуванні контенту, що містить інформаційний вплив на акторів у відкритому або прихованому вигляді; *споживання* — складається з поширення і публікації у СІС контенту, який містить інформаційний вплив
$X^{comp} = \{X_i^{comp}\}$	Y_i^{comp}	Компоненти інформаційного впливу: *актор* — це користувач, який задовольняє інформаційні та комунікаційні потреби з використанням СІС; *контент* — інформаційне наповнення віртуальних спільнот СІС для зацікавлення акторів, вираження точки зору на актуальні події, поширення оперативної та корисної інформації тощо; *інтерпретація* — тлумачення акторами спожитого контенту, їх реакція на нього, дії у віртуальному чи реальному просторі внаслідок здійсненого інформаційного впливу.

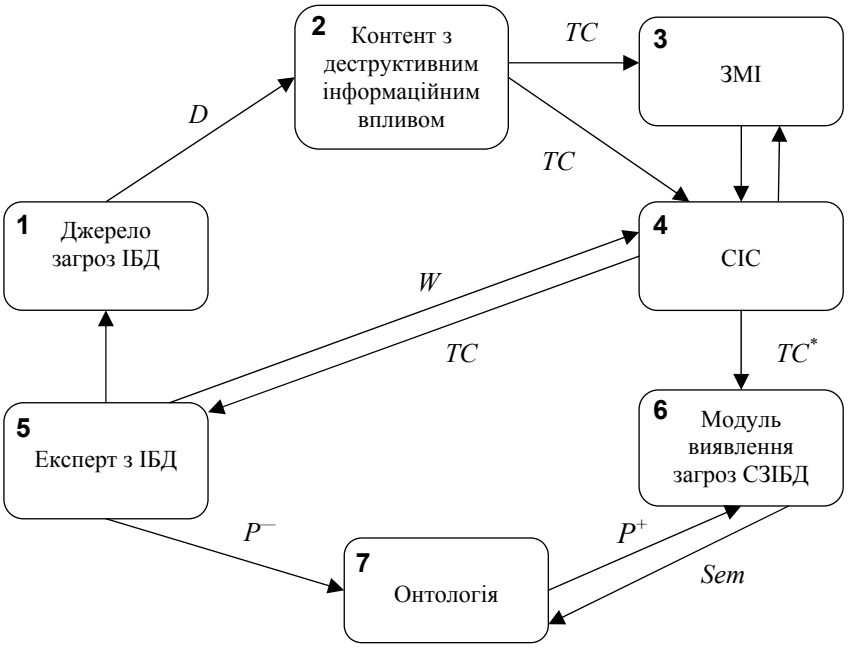

Рис. 2. Системна модель виявлення загроз ІБД у змісті текстового контенту СІС

Системний аналіз процесів виявлення інформаційного впливу на акторів у СІС дозволив виявити його сім складових компонентів: 1 — джерело загроз ІБД; 2 — контент з деструктивним інформаційним впливом; 3 — засоби масової інформації (ЗМІ); 4 — СІС; 5 — експерт з ІБД; 6 — модуль виявлення загроз системи забезпечення ІБД у СІС; 7 — онтологія.

Джерело загроз ІБД характеризується безпосередньо загрозами D_j, $j = \overline{1,k}$, які формалізуються у вигляді кортежу [12]

$$D = \langle R, S, C, T, Sph, M, F, Sr, Pos, I \rangle,$$

де R — відношення загрози до акторів СІС; S — вид суб'єкта загроз; C — характер загрози відносно СІС; T — мета реалізації загрози;

Sph — сфера суспільної діяльності, на яку впливає загроза; M — спосіб дії загрози; F — частота повторюваності; Sr — прихованість прояву; Pos — можливість реалізації загрози у СІС; I — рівень впливу на акторів у СІС.

Унаслідок вироблення інформаційних впливів на акторів віртуальних спільнот з урахуванням актуальності, рівня обговорення та критичності для суспільства інформаційних приводів у ЗМІ формується відповідний текстовий контент

$$TC = \{TC_i\},\ i = \overline{1,n}.$$

Після цього такий контент поширюється у інформаційному просторі ЗМІ та СІС для впливу на суспільну думку.

Експерт з ІБД на основі аналізу інформаційного простору СІС формує семантичне ядро

$$W = \{w_m\},\ m = \overline{1,l},$$

для пошуку текстового контенту за заданим інформаційним приводом, пов'язаним із необхідністю забезпечення захисту національних інтересів. Відібраний текстовий контент

$$TC^* = \{TC_b^*\},\ b = \overline{1,d},$$

надходить до модуля виявлення загроз системи забезпечення ІБД у СІС.

Дослідження текстового контенту TC^* на предмет наявності інформаційного впливу на акторів віртуальних спільнот виконується з використанням онтології для проведення семантичного аналізу. Виконується виявлення семантичних шаблонів відомих загроз ІБД

$$P^+ = \{Pattern_z^+\},\ z = \overline{1,r}.$$

У випадку високого ступеня прояву релевантності відібраного текстового контенту TC^* та відсутності у ньому відомих семантичних шаблонів загроз P^+ він передається для додаткового дослідження експертом з ІБД. У разі виявлення нових невідомих шаблонів загроз

$$P^- = \{Pattern_q^-\},\ q = \overline{1,v},$$

вони доповнюють універсальну множину семантичних шаблонів загроз $P^{\pm}$.

В основу запропонованого підходу до виявлення загроз покладено метод детектування інформаційних впливів у СІС за змістовними ознаками, представлений у публікації [4].

У результаті проведених досліджень встановлено, що перспективним напрямком вдосконалення є використання зважених компонентів семантичного ядра [13] для підвищення ступеня релевантності відібраного контенту та його подальшого аналізу на предмет загроз ІБД. Таким чином, суть запропонованого підходу полягає у такому.

Етап 1. Пошук текстового контенту в СІС на основі зважених компонентів семантичного ядра запиту. На цьому етапі експерт з ІБД визначає семантичне ядро $W = \{w_m\}$, $m = \overline{1,l}$, для пошуку текстового контенту в СІС за критерієм актуальності, критичності та рівня обговорення у суспільстві його тематики. Зважаючи на особливості публікацій текстового контенту у СІС використано метод латентно-семантичного індексування (*LSI*) [9], який забезпечує пошук контенту в СІС на основі його змісту, а не щільності ключових слів, і знаходження прихованих семантичних зв'язків між семантичним ядром й безпосередньо контентом. Суть першого етапу полягає у виконанні таких кроків:

1.1) попередня підготовка досліджуваного контенту СІС TC^* шляхом видалення стоп-слів, стеммінгу або лематизації слів. Стоп-слова створюють «інформаційний шум» і представлені у природомовному тексті сполучниками, частками, прийменниками тощо. Стеммінг полягає у виділенні основи слова виключенням закінчень і суфіксів та є необов'язковим на великих наборах публікацій контенту СІС. Лематизація — це приведення слова до словникового виду;

1.2) виключення з досліджуваного текстового контенту СІС TC^* слів, які вживаються тільки один раз;

1.3) формування частотної матриці M ключових слів W, які індексуються. Рядками i цієї матриці є ключові слова W семантичного ядра, за яким виконується моніторинг текстового контенту СІС, а стовпцями j — публікації акторів чи віртуальних спільнот. Елементи матриці m_{ij} представляють собою частоту вживання деякого ключового слова w_i в j-й публікації;

1.4) полягає в сингулярному розкладанні початкової матриці M на три компоненти

$$M = U \times S \times V^T,$$

де U і V^T — ортогональні матриці розмірністю $i \times k$ та $k \times j$ відповідно; S — діагональна матриця розмірністю $k \times k$, причому k — кількість сингулярних значень матриці (прихованих тематик контенту), а її елементи впорядковані за спаданням;

1.5) виділяють ті рядки матриці U і стовпці V^T, які відповідають найбільшим сингулярним числам k, а їх величина — ступеню прояву ключових слів у колекції публікацій в СІС. З метою підвищення ефективності виявлення контенту з деструктивним інформаційним впливом введено критерій релевантності відібраних публікацій [13].

$$RP_j = \sum_{i=1}^{m} \omega_{ij} \alpha_i ,$$

де ω_{ij} — коефіцієнт важливості виявленої прихованої тематики на базі частоти вживання ключових слів; α_i — експертна оцінка важливості.

Коефіцієнт ω_{ij} розраховується відповідно до закону Зіпфа на основі виразу [13]

$$\omega_{ij} = \frac{m_{ij}}{\lg \frac{D}{f_i}},$$

де D — загальна кількість досліджуваних публікацій; f_i — кількість публікацій, у яких використовується ключове слово w_i.

Експертна оцінка α_i визначається у результаті обробки результатів проведеного опитування

$$\alpha_i = \frac{1}{L} \sum_{l=1}^{L} \alpha_{lj} , \quad \alpha_{lj} = \frac{\sum_{l=1}^{L} \varphi_{lj}}{\sum_{l=1}^{L} \sum_{j=1}^{D} \varphi_{lj}},$$

де φ_{ij} — ранг, який присвоюється l-тим експертом j-му ключовому слову з семантичного ядра.

Після цього відбувається впорядкування досліджуваних публікацій у СІС на основі введеного критерію релевантності документів, які аналізуються.

Етап 2. Виявлення ознак загроз ІБД у СІС на основі сигнатурного методу і методу виявлення аномалій, який полягає в такому:

2.1) складається онтологія функціонування віртуальної спільноти в СІС

$$Ont = \langle P_n, R_n \rangle ,$$

де Ont — онтологія; P_n — скінченна множина концептів; R_n — скінченна множина відношень між концептами.

При цьому в онтології Ont виділяються такі підмножини [8]: $P_s(p_n) \in P_n$ — підмножина множини концептів P_n, суміжних до деякого концепту p_n; $P_{in}(r_n) \in P_n$ — підмножина множини концептів P_n, інцидентних до відношення r_n; $R_{in}(p_n) \in R_n$ — підмножина множини відношень R_n, інцидентних до деякого концепту p_n; $R_z(p_n) \in R_n$ — підмножина множини відношень R_n, вживання яких вказує на небезпеку для концепту p_n;

2.2) побудова семантичного опису текстового контенту, виявленого на першому етапі

$$Sem_t = \langle P_t, R_t \rangle ;$$

2.3) виявлення ознак загрози у попередньо проіндексованому на першому етапі текстовому контенті СІС. У формальному вигляді правила виявлення загроз ІБД зводяться до такого [8]:

– детектування на основі сигнатурного методу і семантичних ознак

$$\exists r_t(p_t) : p_t \in P_n \wedge r_t \in R_z(p_n),$$

де $r_t(p_t)$ — деяке відношення з текстового контенту СІС, що аналізується; $p_t \in P_n$ — концепт онтології Ont досліджуваної віртуальної спільноти; $r_t \in R_z(p_n)$ — множина відношень, які вказують на небезпеку для деякого концепту p_n;

– встановлення суперечностей на основі виявлення аномалій шляхом співставлення семантичного опису проіндексованого текстового контенту і семантичного шаблону загрози. При цьому встановлюється невідповідність фактів у контенті СІС, що проявляється як:

– протиріччя понять в змісті контенту віртуальних спільнот — вживання концепту в конкретному відношенні не передбачене онтологією

$$\exists r_t(p_t) : p_t \in P_n \wedge r_t \in R_n \wedge p_t \notin P_{in}(r_n);$$

– протиріччя відношень в контенті — вжите відношення між концептами не визначене онтологією

$$\forall p_n \in P_n \neg \exists r_n \in R_n : r_t = r_n ;$$

2.4) розгорнутий аналіз експертами з ІБД проіндексованого текстового контенту TC^* на предмет наявності загроз.

Даний крок необхідний, якщо контент СІС характеризується високим ступенем релевантності семантичному ядру пошукового запиту з етапу 1, однак на етапі 2 не виявлено відповідності фрагментів такого контенту і онтології.

Тоді ідентифіковані експертами нові семантичні конструкції загроз $\{Pattern^-\}$ доповнюють універсальну множину загроз $Pattern^{\pm}$, яка містить відомі шаблони загроз $\{Pattern^+\}$, тобто

$$\{Pattern^-\} = \{pattern_i^+ \mid pattern_i^+ \notin Pattern_i^+, pattern_i^+ \in Pattern^{\pm}\}.$$

Перевагами запропонованого підходу є комплексування семантичного і латентно-семантичного аналізу, що забезпечує взаємну компенсацію їх недоліків: полісемії, омонімії та інших видів лінгвістичних неоднозначностей для *LSI*; виявлення латентних залежностей між концептами для семантичного аналізу.

Висновки

Сьогодні СІС перетворилися на ефективний інструмент комунікації між акторами. Внаслідок глобалізації інформаційного простору завдяки своїм перевагам СІС використовуються як дієвий інструмент проведення інформаційних операцій, спрямованих проти ІБД. При цьому СІС здатні впливати на суспільні й політичні процеси й в реальному житті. Тому забезпечення заданого стану ІБД у СІС пов'язане із своєчасним виявленням загроз у контенті віртуальних спільнот. Запропонований концептуальний базис дослідження загроз ІБД у текстовому контенті СІС визначає, що здійснення інформаційного впливу на акторів в СІС реалізується на основі таких складових — типу впливу, етапів і компонентів. Концептуальний базис покладено в основу системної моделі виявлення загроз ІБД, яка використана при розробленні підходу до детектування загроз у змісті текстового контенту. Розроблений підхід ґрунтується на сучасних методах дослідження контенту — латентно-семантичному індексуванню і семантичному аналізі на базі онтологій, комбінація яких забезпечує взаємну компенсацію недоліків та ідентифікацію прихованих залежностей між мовними одиницями. Запропонований підхід відрізняється від відомих використанням зважених компонентів семантичного ядра пошукового запиту для підвищення ступеня релевантності відібраного текстового контенту. Внаслідок ранжування колекції відібраних текстових публікацій досягається підвищення ефективності виявлення загроз ІБД. Доповнення універсальної множини шаблонів загроз проводиться із залученням експертів з ІБД. У такому випадку дослідження текстового контенту із високим показником релевантності за умови відсутності в ньому відомих загроз забезпечує додавання до онтологічної бази знань нових невідомих семантичних шаблонів. Таким чином, запропонований підхід до виявлення загроз у текстовому контенті віртуальних спільнот дозволяє підвищити ефективність функціонування системи забезпечення ІБД у СІС.

ЛІТЕРАТУРА

1. **Грищук Р. В.** Основи кібернетичної безпеки: монографія / Р. В. Грищук , Ю. Г. Даник: під ред. проф. Ю.Г. Даника. — Житомир: ЖНАЕУ, 2016. — 636 с.

2. **Грищук Р. В.** Технологічні аспекти інформаційного протиборства на сучасному етапі / Р. В Грищук, І. О. Канкін, В. В. Охрімчук. Захист інформації. — 2015. — Т. 17. — № 1. — С. 80—86, DOI: 10.18372/2410-7840.17.8347.

3. **Онищенко О. С.** Соціальні мережі як інструмент взаємовпливу влади та громадянського сус-пільства: [монографія] / О. С. Онищенко, В. М. Горовий, В. І. Попик // НАН України, Нац. б-ка України ім. В. І. Вернадського. — К., 2014. — 260 с.

4. **Молодецька-Гринчук К. В.** Метод виявлення ознак інформаційних впливів у соціальних Інтернет-сервісах за змістовними ознаками / Радіоелектроніка, інформатика, управління. — 2017. — № 2(41). — С. 117—126, DOI: 10.15588/1607-3274-2017-2-13.

5. **Barabash O.**, Shevchenko G., Dakhno N., Neshcheret O., Musienko A. Information Technology of Targeting: Optimization of Decision Making Process in a Competitive Environment / International Journal of Intelligent Systems and Applications. — 2017. — Vol. 9, № 12. — PP. 1—9, DOI: 10.5815/ijisa.2017.12.01.

6. **Mukhin V.**, Loutskii H., Barabash O., Kornaga Ya., Steshyn V. Models for Analysisand Prognosticationof the Indicatorsof the Distributed Computer Systems Characteristics / International Review on Computers and Software (IRECOS). — 2015. — Vol. 10, № 2. — pp. 1216—1224.

7. **Mashkov V. A.**, Barabash O.V. Self-Testing of Multimodule Systems Based on Optimal Check-Connection Structures / Engineering Simulation. — 1996. — Vol. 13. — pp. 479—492.

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

9. **Manning Chr.**, Raghavan P., Schütze H. Introduction to Information Retrieval / Cambridge University Press, 2008. — 544 p.

10. **Wardle C.**, Derakhshan H. Information disorder: toward an interdisciplinary framework for research and policy making / Access mode: www.coe.int.

11. **Нестругина Е. С.**, Ларина Е. Ю., Чичикало Н. И. Концепция эволюции синтеза мультиагентной информационно-измерительной системы процесса реабилитации человека после травматизма / Восточно-Европейский журнал передовых технологий. — 2013. — № 2/10(62). — С. 49–55.

12. **Молодецька К. В.** Узагальнена класифікація загроз інформаційній безпеці держави в соціальних інтернет-сервісах / Защита информации: сб. науч. трудов НАУ. — 2016. — Вып. 23. — С. 75–87.

13. **Стенин А. А.**, Тимошин Ю. А., Мелкумян Е. Ю., Курбанов В. В. Латентно-семантический метод извлечения информации из Интернет ресурсов / Восточно-Европейский журнал передовых технологий. — 2013. — № 4/9(64). — С. 19–22.

Барабаш О. В., Грищук Р. В., Молодецька-Гринчук К. В.

ВИЯВЛЕННЯ ЗАГРОЗ ІНФОРМАЦІЙНІЙ БЕЗПЕЦІ ДЕРЖАВИ У ЗМІСТІ ТЕКСТОВОГО КОНТЕНТУ СОЦІАЛЬНИХ ІНТЕРНЕТ-СЕРВІСІВ

Соціальні Інтернет-сервіси (СІС) представляють собою один з найпопулярніших засобів масової комунікації суспільства завдяки ефективному інструментарію для задоволення інформаційних і комунікаційних потреб користувачів, яких називають акторами. Однак, в світлі останніх подій, СІС використовуються провідними державами світу як дієвий інструмент проведення інформаційних операцій. Тому розроблення дієвих підходів до виявлення загроз у контенті СІС є актуальним науково-прикладним завданням на шляху забезпечення інформаційної безпеки людини, суспільства, держави. У роботі запропоновано концептуальний базис дослідження інформаційного впливу на акторів у змісті текстового контенту СІС. Встановлено, що основу концептуальної моделі складають типи, етапи та компоненти інформаційного впливу на акторів віртуальних спільнот. На основі запропонованої концептуальної моделі розроблено системну модель виявлення загроз інформаційній безпеці держави у змісті текстового контенту СІС. Визначено її основні компоненти та вказано особливості їх взаємодії між собою. Розроблена системна модель використана при розробленні підходу до виявлення загроз у змісті текстового контенту СІС на базі комплексування семантичного і латентно-семантичний аналізу. Запропонований підхід відрізняється від відомих використанням зважених компонентів семантичного ядра пошукового запиту для підвищення ступеня релевантності відібраного текстового контенту. Ранжування відібраних текстових публікацій за їх відповідністю пошуковому запиту дозволяє підвищити ефективність виявлення загроз інформаційній безпеці держави на базі онтологій. Завдяки аналізу відібраного текстового контенту на базі онтологій з використанням сигнатурного методу і методу аномалій досягається виявлення відомих семантичних шаблонів загроз. Доповнення універсальної множини шаблонів загроз проводиться із залученням експертів з інформаційної безпеки держави. У такому випадку дослідження текстового контенту із високим показником релевантності за умови відсутності в ньому відомих загроз забезпечує додавання до бази знань нових невідомих семантичних шаблонів. Таким чином, запропонований підхід до виявлення загроз у текстовому контенті віртуальних спільнот дозволяє підвищити ефективність функціонування системи забезпечення інформаційної безпеки держави у СІС.

Ключові слова: соціальні Інтернет-сервіси; текстовий контент; інформаційний вплив; актор; інформаційна безпека держави; загрози.

Barabash O. V., Hryshchuk R. V., Molodetska-Hrynchuk K. V.

IDENTIFICATION THREATS TO THE STATE INFORMATION SECURITY IN THE TEXT CONTENT OF SOCIAL NETWORKING SERVICES

Social networking services are one of the most popular mass media society through effective tool to meet the information and communication needs of users, called actors. Social networking services are used by leading countries of the world as an effective tool for information operations. The development of effective approaches to detecting threats in the content of social networking services is an important scientific and applied task on the way of ensuring information security of man, society, and the state. The paper proposes a conceptual basis for the study of information influence on actors in the text content of social networking services. It is established that the basis of the conceptual model consists the types, stages and components of information influence on actors. On the basis of the proposed conceptual model, a system model for detecting threats to state information security in the text content of social networking services has been developed. Determined its main components and specified the features of their interaction with each other. The system model was used in developing an approach to detecting threats in the text content based on the integration of semantic and latent semantic analysis. The proposed approach differs from the known use of the weighted components of the semantic kernel in order to increase the relevance of the selected text content. Ranking of selected text publications according to their search query allows to increase the efficiency of threats detection to state information security on the ontologies. Addition of a multiplicity of threats templates is carried out with the involvement of experts in information security of the state. The proposed approach to detecting threats in the text content of virtual communities can improve the efficiency of the state information security system in social networking services.

Keywords: social networking services; text content; information influence; actor; information security of state; threats.

Барабаш О. В., Грищук Р. В., Молодецкая-Гринчук К. В.

ОБНАРУЖЕНИЕ УГРОЗ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ГОСУДАРСТВА В СОДЕРЖАНИИ ТЕКСТОВОГО КОНТЕНТА СОЦИАЛЬНЫХ ИНТЕРНЕТ-СЕРВИСОВ

Социальные Интернет-сервисы (СИС) представляют собой одно из самых популярных средств массовой коммуникации общества благодаря эффективному инструментарию для удовлетворения информационных и коммуникационных потребностей пользователей, которых называют акторами. Однако, в свете последних событий, СИС используются ведущими государствами мира как действенный инструмент проведения информационных операций. Поэтому разработка подходов к выявлению угроз в контенте СИС является актуальной

научно-прикладной задачей на пути обеспечения информационной безопасности человека, общества, государства. В работе предложен концептуальный базис исследования информационного воздействия на акторов в содержании текстового контента СИС. Установлено, что основу концептуальной модели составляют типы, этапы и компоненты информационного воздействия на акторов виртуальных сообществ. На основе предложенной концептуальной модели разработана системная модель выявления угроз информационной безопасности государства в содержании текстового контента СИС. Определены ее основные компоненты и указано особенности их взаимодействия между собой. Разработанная системная модель использована при разработке подхода к выявлению угроз в содержании текстового контента СИС на базе комплексирования семантического и латентно-семантического анализа. Предложенный подход отличается от известных использованием взвешенных компонентов семантического ядра поискового запроса для повышения степени релевантности отобранного текстового контента. Ранжирование отобранных текстовых публикаций за их соответствием поисковому запросу позволяет повысить эффективность обнаружения угроз информационной безопасности государства на базе онтологий. Благодаря анализу отобранного текстового контента на базе онтологий с использованием сигнатурного метода и метода аномалий достигается выявления известных семантических шаблонов угроз. Дополнение универсального множества шаблонов угроз проводится с привлечением экспертов по информационной безопасности государства. В таком случае исследования текстового контента с высоким показателем релевантности при отсутствии в нем известных угроз обеспечивает добавление в базу знаний новых неизвестных семантических шаблонов. Таким образом, предложенный подход к выявлению угроз в текстовом контенте виртуальных сообществ позволяет повысить эффективность функционирования системы обеспечения информационной безопасности государства в СИС.

Ключевые слова: социальные Интернет-сервисы; текстовый контент; информационное воздействие; актор; информационная безопасность государства; угрозы.

Стаття надійшла до редакції 19.03.2018 р.
Прийнято до друку 04.06.2018 р.
Рецензент — д-р техн. наук, с. н. с., Савченко В. А.