УДК 004.738.5:004.056.5(045)

Молодецька-Гринчук К.В.
Житомирський національний агроекологічний університет

МЕТОДИКА ВИЯВЛЕННЯ МАНІПУЛЯЦІЙ СУСПІЛЬНОЮ ДУМКОЮ У СОЦІАЛЬНИХ ІНТЕРНЕТ-СЕРВІСАХ

У статті теоретично обгрунтовано і узагальнено ознаки маніпуляцій суспільною думкою акторів у соціальних інтернет-сервісах. Запропонована методика виявлення маніпуляцій у текстовому контенті грунтується на сучасних методах інтелектуального аналізу і раннього виявлення загроз інформаційній безпеці держави у соціальних інтернет-сервісах. Розроблена методика відрізняється від відомих врахуванням інформаційної невизначеності, що дозволило підвищити ефективність функціонування системи забезпечення інформаційної безпеки держави.

Ключові слова: маніпуляції, соціальні інтернет-сервіси, загрози, машинне навчання, моніторинг.

Постановка проблеми у загальному вигляді та її зв'язок із важливими практичними завданнями. Соціальні інтернет-сервіси (СІС) є ефективною платформою взаємодії учасників віртуальних спільнот, яких називають акторами [1, 2]. Завдання сучасних СІС полягає у забезпеченні акторів дієвим інструментарієм для спілкування, обміну контентом різного типу, самовираження тощо. В публікаціях [1–8] значна популярність СІС серед громадян пояснюється високою довірою до неофіційних інтернет-ресурсів, безпосереднім залученням аудиторії до обговорення шляхів вирішення тих чи інших проблем, об'єднання акторів у групи за принципом співчуття або емпатії. В свою чергу, СІС активно використовуються громадянським суспільством для самоорганізації з метою впливу на політичні й суспільні процеси в державі, формування суспільної думки тощо. Проте, позитивні комунікаційні характеристики СІС перетворили віртуальні спільноти на ефективний інструмент проведення інформаційних операцій проти людини, суспільства держави.

Під час взаємодії акторів у СІС виникає низка психологічних явищ [4, 7], використання яких зловмисниками для проведення інформаційних операцій створює передумови для маніпулювання суспільною думкою, впливу на свободу вибору акторів, емоційний і психічний стан, дискредитацію існуючої системи управління в державі тощо. У публікації [9] встановлено, що маніпуляції представляють собою спосіб інформаційно-психологічного впливу у прихованому вигляді для спонукання суб'єктів впливу до реалізації заданих дій і досягнення об'єктом впливу однобічних переваг. У свою чергу, високі темпи розвитку сучасних засобів і технологій комунікації у мережі Інтернет призвели до створення якісно нових технологій маніпулятивного впливу на учасників віртуальних спільнот. Це призвело до появи протиріччя між новими загрозами інформаційній безпеці держави у СІС та існуючими науковими методами автоматизованого їх виявлення. Тому аналіз маніпулятивних технологій, які застосовуються для впливу на акторів, розробка методів своєчасного виявлення маніпулятивних ознак інформаційних операцій у СІС є актуальним теоретико-прикладним завданням на шляху забезпечення інформаційної безпеки держави.

Аналіз останніх досліджень і публікацій. У публікаціях [3, 8–11] професора Л. Компанцевої розглянуто методи маніпулятивного впливу з використанням сугестивних технологій. Встановлено, що сугестія є процесом впливу на психічну сферу акторів для зниження критичності сприйняття та реалізації змісту, який навіюється. Особливу увагу приділено технологіям нейролінгвістичного програмування і чорному піару, викладено особливості сугестивного впливу в мережі Інтернет. Автором встановлено, що ознаками сугестивного впливу на акторів віртуальних спільнот є [10]: лавиноподібне поширення контенту в СІС завдяки появі «позаколективної» поведінки; підвищення ролі інтелектуальної рецепції, яка визначає здатність впливати на поведінку акторів подібно їх самостійному мисленню; формування нових символічних систем як ознаки встановлення влади; психічне зараження, яке проявляється у несвідомій здатності актора мимовільно схилятися до психічних впливів; переконання акторів у деяких ідеях за відсутності критичного сприйняття віртуальною спільнотою контенту; вплив на свідомість акторів з використанням відомих брендів.

Також проблемі маніпулювання суспільною свідомістю у СІС присвячено публікації професора Г. Почепцова [12–14]. Він відзначає, що сучасні інформаційні операції проводяться з активним залученням інтелектуального ресурсу, а безпосередньо СІС є ефективним інструментом для їх реалізації. Автор досліджує технологію мікротаргетингу, яка зводиться до спрямування повідомлень окремим акторам завдяки їх персоналізації. Результатом таких дій є збільшення уваги актора до поширюваного контенту і його прихильність до заданих подій. Дослідження професора С. Расторгуєва [15, 16] пов'язані з інформаційною війною і, зокрема, виявленням прихованих загроз, моделюванням інформаційних впливів через засоби масової інформації. Автор вводить поняття «сугестивного шуму», який представляє собою цілеспрямовану приховану дію на об'єкт управління, що надходить на нього з потоком іншої фонової інформації – запитань, фактів, правил. Реакція об'єкта на сугестивний шум визначається індивідуально в межах власних можливостей осмислення і структури його знань. В свою чергу, запропоновані авторським колективом в публікаціях [17, 18] підходи до виявлення інформаційних впливів у СІС носять дескриптивний характер, автори не наводять методику їх реалізації для подальшого практичного використання і нейтралізації негативного впливу на віртуальні спільноти.

Проведений критичний аналіз показав, що на сучасному етапі питанню детектування маніпулятивних ознак інформаційних операцій в СІС в розрізі інформаційної безпеки людини, суспільства, держави приділено недостатньо уваги. Відсутність дієвого переліку ознак маніпулятивного впливу на акторів СІС, загальноприйнятих методів і технологій виявлення сугестивних технологій у віртуальних спільнотах, недостатній рівень врахування досвіду гібридної війни з Російською Федерацією додатково актуалізують обраний напрямок наукових досліджень.

Мета статті полягає у теоретичному обґрунтуванні й розробленні методики виявлення маніпулятивних ознак інформаційних впливів у СІС для підвищення ефективності процесів моніторингу інформаційного простору системою забезпечення інформаційної безпеки держави.

Для досягнення поставленої мети необхідно розв'язати частинні завдання:

визначити особливості використання маніпулятивних технологій в СІС і встановити ознаки використання маніпулятивних технологій впливу на суспільну думку акторів;

розробити методику виявлення маніпуляцій суспільною думкою у СІС;

навести приклади застосування технологій маніпуляції суспільною думкою у СІС;

провести експериментальне дослідження запропонованої методики.

Виклад основного матеріалу дослідження. У сучасних умовах СІС створюють умови для ефективного проведення інформаційних операцій із застосуванням маніпулятивних технологій, серед яких виділяють [8]: низьку вартість поширення контенту у віртуальних спільнотах і мобільність процесів взаємодії акторів; наявність засобів для приховування джерел інформаційних акцій, комплексне застосування різних СІС і віртуальних спільнот; використання гіперпосилань для організації доступу акторів до контенту маніпулятивного змісту безпосередньо із СІС; прихований вплив на акторів, анонімність і безкарність зловмисників; проведення у СІС експериментів з метою дослідження процесів соціальної комунікації, ефективності впливу на суспільну думку; здійснення процедур моніторингу інформаційного середовища СІС для встановлення даних актора, який поширює той чи інший контент.

У даній статті під *маніпулятивними ознаками інформаційних операцій* у СІС будемо розуміти наявність у контенті віртуальних спільнот прихованого впливу на акторів з метою зміни їх поведінки, цілей, намірів чи інших психологічних характеристик в інтересах суб'єкта впливу. Дослідження [4, 7, 8, 19] показують, що кількість сучасних технологій маніпулятивного впливу на акторів СІС постійно зростає завдяки новим досягненнями у галузі психології, лінгвістики, журналістики, комунікативних стратегій, теорії маніпулятивного впливу та інших наук. Серед найбільш дієвих технологій, які використовуються у віртуальних

спільнотах СІС під час проведення інформаційних операцій, узагальнивши, виділимо наступні (*рис. 1*).

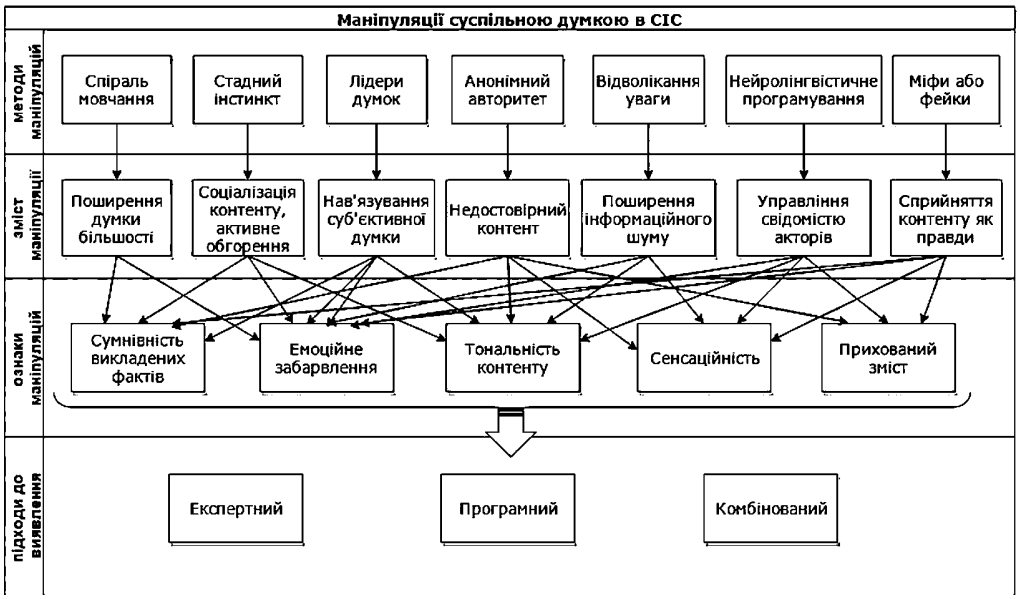

Рис. 1. Методи маніпуляцій суспільною думкою у СІС

«Спіраль мовчання» – це модель комунікації, запропонована Е. Ноель-Нойманн, яка показала свою ефективність в СІС і описує особливості процесів висловлювання та поширення громадської думки. Суть моделі полягає в приховуванні акторами своєї громадянської позиції, якщо вона не співпадає з точкою зору більшості. В окремих випадках актори схильні погоджуватися із твердженнями, раніше неприйнятними для них.

Стадний інстинкт акторів СІС пов'язаний з колективною поведінкою особистості й полягає в тому, що більша увага приділяється публікаціям контенту або віртуальним спільнотам з великою кількістю коментарів, «лайків», репостів, учасників тощо. Такими діями зловмисники виконують соціалізацію заданого контенту чи віртуальної спільноти, створюють ілюзію активного обговорення, їх значущості і критичності для учасників віртуальної спільноти. Для реалізації даної технології маніпуляції часто використовуються соціальні боти [20].

Лідери думок в СІС представляють собою акторів або віртуальні спільноти акторів, які обізнані в деякій галузі. Вони публікують контент з власною оцінкою, поясненнями і аргументацією подій, а менш активні актори сприймають його як пояснення явищ й фактів. Автор теорії «Лідерів думок» П. Ф. Лазарсфельд стверджує, що сприйняття контенту відбувається через два рівні – спочатку факти оцінюються лідерами думок через призму їх знань і вмінь, а потім, застосовуючи міжособистісну комунікацію, пропонують іншим групам акторів своє бачення ситуації, думки, висновки. Таким чином, лідери думок в СІС опосередковано впливають на сприйняття фактів більшістю учасників віртуальних спільнот. Використовуючи таку технологію маніпуляції суб'єкт інформаційної операції може поширювати конкретні ідеї, нав'язуючи бажану точку зору на події в державі і суспільстві.

Посилання на анонімний авторитет зводиться до згадування в якості джерела контенту авторитетних осіб, наприклад, політиків, науковців, духовенства тощо. З метою збільшення переконливості контенту наводяться оцінки експертів, свідчення учасників подій, документи. Однак, в таких випадках джерело фактів не ідентифіковане і відповідальність за поширення такого контенту ніхто не несе.

Емоційний резонанс в СІС використовується для створення у акторів віртуальних спільнот заданого емоційного стану і одночасної передачі контенту. Такий підхід забезпечує сприйняття контенту на рівні емоцій і вимкнення механізмів логіки і критичного мислення. В

його основі лежить феномен соціальної індукції, який полягає в поширенні емоційного стану окремих акторів на інших учасників віртуальних спільнот завдяки емпатії.

Відволікання уваги акторів спрямоване на їх перефокусування від першочергового контенту до другорядного, який поданий як сенсація. Таким чином створюється інформаційний шум в СІС, який приховує важливі події.

Міфи або *фейки* – це прийом поширення в СІС контенту, який містить викривлені, спотворені, вигадані факти про дійсність. Метою даної технології маніпуляції є забезпечення сприйняття акторами контенту як правди без критичного осмислення і перевірки фактів. Поширення фейків часто поєднується з іншими технологіями для досягнення бажаного ефекту суб'єктами маніпуляцій свідомістю.

Нейролінгвістичне програмування застосовується у СІС для управління свідомістю акторів з використанням спеціальних лінгвістичних конструкцій контенту, образів, зображень, відео тощо. Засновники теорії нейролінгвістичного програмування Р. Бендлер і Дж. Гріндер досліджували методи впливу на особистість для легкої і швидкої стійкої зміни програм її поведінки.

Результати аналізу академічної літератури [8, 10, 18] показали, що узагальнюючи частинні ознаки використання технологій маніпулятивного впливу на акторів СІС для їх виявлення доцільно виділити наступні:

– *сумнівність викладених фактів*, що визначається приховуванням джерел і авторів інформації, недостатньою аргументацією, посиланнями на думку широкого загалу, наявністю риторичних запитань;

– *емоційне забарвлення* контенту, що використовується для відображення емоційного стану його автора і проявляється у перенасиченні контенту образними засобами, прикметниками, порівняннями тощо;

– *тональність* контенту по відношенню до деякого об'єкту чи події, яка відображає оцінювальні судження актора і може проявлятися у використанні зображень, смайлів тощо;

– *сенсаційність* контенту, яка має на меті привернути увагу акторів завдяки посиланням на заяви скандальних осіб, вживанню слів, які підвищують тривожність та ін.;

– *прихований (імпліцитний) зміст* контенту пов'язаний з його глибинним змістом, отриманим в результаті розумової діяльності на основі співвідношення системи знань і цінностей актора з мовними одиницями й конструкціями.

Слід зауважити, що виявлення розглянутих ознак застосування маніпулятивних технологій для проведення інформаційних операцій у СІС є складноформалізованою задачею. Тому для їх детектування використовують експертний, програмний і комбінований підходи (див. рис. 1). Суть експертного виявлення полягає у залученні до процесу моніторингу експертів або співробітників спеціальних підрозділів для прийняття рішень щодо наявності у контенті СІС маніпулятивних технологій. Недоліками даного підходу є суб'єктивність оцінок експертів, складність процедур виявлення прихованих маніпуляцій у контенті, які спираються на досвід роботи експерта. Для підвищення швидкості функціонування системи забезпечення інформаційної безпеки держави доцільно використовувати програмні методи детектування загроз у СІС, спрямованих на маніпулювання суспільною думкою. Встановлено, що такі методи ґрунтуються на сучасних технологіях контент-аналізу, інтелектуального аналізу контенту та методів машинного навчання [17]. Недоліками контент-аналізу, який проводиться засобами спеціалізованого програмного забезпечення, є складність процесів уточнення мети опублікованого контенту, що призводить до появи невизначеності результуючої оцінки; відображення суджень конкретного розробника у інформаційному забезпеченні програмних комплексів, а саме базах даних словників і семантичних ядер пошуку; прихований характер лінгвістичних конструкцій маніпуляцій у текстовому контенті. Застосування інтелектуального аналізу контенту обмежується складністю вилучення даних з великих масивів даних і високою вартістю за умов обмеженості ресурсів. Комбіновані методи представляють собою поєднання експертних і програмних таким чином, щоб компенсувати їх недоліки. Перевага використання цієї групи методів полягає в зниженні суб'єктивності оцінки контенту СІС, зростанні

швидкодії прийняття рішень щодо наявності маніпулятивних технологій і підвищені загальної ефективності функціонування підсистеми моніторингу інформаційного середовища віртуальних спільнот.

З метою розв'язку поставленої задачі розроблено методику виявлення маніпуляцій суспільною думкою акторів у СІС в результаті аналізу текстового контенту, яка грунтується на сучасних методах обробки даних – контент-аналізу і машинного навчання, не суперечить дослідженням [18] В. М. Панченко та полягає в такому.

Крок 1. Встановлення ознак сумнівності викладених у контенті СІС фактів. На першому кроці виявляються ознаки недостовірності контенту віртуальних спільнот СІС, що зводиться до такого:

посилання на суб'єктивну точку зору F_1 – відносний показник вживання у контенті СІС оцінок фактів експертами, ученими, авторитетними джерелами тощо

$$F_1 = \frac{R_d}{W}, \qquad (1)$$

де R_d – кількість виявлених посилань; W – загальна кількість слів; відсутність аргументації F_2 – відносний показник використання лінгвістичних конструкцій, які відкидають необхідність підтвердження і доведення істинності контенту (наприклад, *вочевидь*, *незаперечний факт* тощо)

$$F_2 = \frac{G_d}{W}, \qquad (2)$$

де G_d – кількість лінгвістичних конструкцій із запереченням необхідності верифікації контенту; частка запитальних речень F_3 – відношення кількості запитальних речень S_{okl} до загальної кількості речень S_z у текстовому контенті СІС

$$F_3 = \frac{S_{okl}}{S_z}; \qquad (3)$$

числові дані F_4 – відносний показник вживання у публікації СІС числових даних

$$F_4 = \frac{B_z}{W}, \qquad (4)$$

де B_z – загальна кількість наведених у публікації чисел; сумнівні висловлювання F_5 – відносний показник використання лінгвістичних конструкцій, які припускають різні підходи до тлумачення (наприклад, *можливо*, *ймовірно*, *завжди*)

$$F_5 = \frac{F_z}{W}, \qquad (5)$$

де F_z – кількість неоднозначних висловлювань.

Крок 2. Визначення емоційного забарвлення контенту. Даний крок має на меті встановлення наявності у текстовому контенті проявів індивідуального настрою чи почуттів актора щодо досліджуваних об'єктів чи подій. Суть кроку полягає в детектуванні таких ознак [18]:

окличні речення F_6 – відносний показник кількості окличних речень S_d в текстовому контенті

$$F_6 = \frac{S_d}{S_z}; \qquad (6)$$

вигуки F_7 – показник наявності у текстовому контенті вигуків (наприклад, *ну-ну*, *овва*, *ага* тощо)

$$F_7 = \frac{E_d}{W}, \qquad (7)$$

де E_d – кількість виявлених вигуків у публікації; прислівники F_8 – відносна кількість прислівників A_d у текстовому контенті, які використовуються для порівняння, перефокусування читача публікації на його емоції (наприклад, *наче*, *більше*, *назавжди* тощо)

$$F_8 = \frac{A_d}{A_z}, \qquad (8)$$

де A_z – загальні кількість прислівників у публікації; емоційний словник F_9 – показник вживання у публікації лексем емоційного характеру (наприклад, *безкарний, блокада, ганебний* тощо)

$$F_9 = \frac{V_d}{W}. \qquad (9)$$

де V_d – кількість емоційних лексем.

Таблиця 1

Групи методів виявлення тональності контенту СІС

Назва	Сутність методу	Переваги	Недоліки
Підхід на основі правил	тональність контенту визначається в результаті співставлення з попередньо визначеними правилами	висока точність методу за умови повноти бази правил; простота програмної реалізації	правила визначення тональності створюються для конкретної предметної області; невисока швидкість оцінки
Підхід на основі словників	ґрунтується на використанні тональних словників, які містять слова зі значеннями їх тональності. Загальна тональність контенту обчислюється обраним методом (наприклад, середнє арифметичне, класифікатор, що навчається)	простота використання в заданій предметній області; можливість автоматизації процедур оцінки тональності	використовується в межах визначеної предметної області;
Машинне навчання з учителем	полягає у навчанні машинного класифікатора на колекції попередньо відібраного контенту, який в подальшому використовується для аналізу тональності контенту	висока точність і швидкодія; можливість автоматизації процедур оцінки тональності контенту; наявність засобів оцінки точності; розбиття тональності контенту на задану кількість класів	ґрунтується на навчальній вибірці; необхідність розробки класифікаційної моделі; критичність принципу формування навчальної та валідаційної вибірок і тестових даних
Машинне навчання без учителя	зводиться до розв'язування задачі визначення тональності контенту без втручання дослідника, встановлює зв'язок між об'єктами	простота автоматизації процедур оцінки тональності контенту; не потребує підбору навчальної вибірки; відсутність потреби у апріорній інформації	низька точність; висока ресурсоємність і вартість; низька швидкодія; заздалегідь невизначена кількість класів

Крок 3. Оцінка тональності контенту. Метою є визначення позиції актора відносно досліджуваних об'єктів або подій. Задача оцінки тональності контенту віртуальних спільнот

розв'язується шляхом застосування методів машинного навчання та інформаційного пошуку. Даний крок [21, 22] зводиться до віднесення тональності публікації d_j, $j=\overline{1,n}$ до попередньо визначеної категорії c_i, $i=\overline{1,m}$ – негативна, позитивна, нейтральна тощо

$$(d_j, c_i) \in D \times C, \qquad (10)$$

де D – колекція публікацій СІС;

C – множина класів тональності публікацій.

Аналіз сучасних підходів до класифікації тональності контенту СІС показав, що для задач аналізу тональності контенту доцільно використовувати групи методів, які наведені в табл. 1 [21–26].

У результаті виконання кроку 3 отримаємо визначений клас тональності текстового контенту Q_3 публікації і його нормоване числове значення відповідно до шкали табл. 2 [27].

Таблиця 2

Приклад нормованої шкали оцінки тональності

Клас тональності, який використовується		Інтервал нормованої шкали оцінок
негативна	позитивна	
виражено негативна	виражено позитивна	1,00–0,70
помірно негативна	помірно позитивна	0,71–0,50
нейтральна	нейтральна	0,51–0,40
помірно позитивна	помірно негативна	0,41–0,20
виражено позитивна	виражено негативна	0,21–0,00

Зауваження 1. На попередніх етапах дослідження необхідно задати пріоритет для класу тональності контенту, яка аналізується, залежно від об'єкта семантичного ядра контенту.

Крок 4. Сенсаційність контенту. На цьому кроці оцінюється здатність текстового контенту своїм змістом зацікавити, вразити і привернути увагу акторів СІС. Даний крок зводиться до виявлення таких ознак:

підвищення уваги F_{10} – відносний показник вживання слів, що привертають увагу актора, підвищують тривожність (наприклад, *вбивство*, *шок*, *сепаратизм*)

$$F_{10} = \frac{U_d}{W}, \qquad (11)$$

де U_d – кількість виявлених слів-інтенсифікаторів уваги;

оперативність F_{11} – показник використання слів для створення атмосфери швидкоплинності подій або явищ, їх терміновості (наприклад, *миттєво*, *швидко*, *несподівано*)

$$F_{11} = \frac{O_d}{W}, \qquad (12)$$

де O_d – кількість виявлених слів для позначення оперативності;

Крок 5. Виявлення прихованої теми контенту. Метою даного кроку є виявлення прихованої теми повідомлення в результаті тематичного моделювання. В даній статті під *темою контенту СІС* будемо розуміти його основний зміст, який автор доносить до читача. Встановлено, що для автоматизації процедур встановлення прихованої теми текстового контенту найбільш ефективними є методи ймовірнісного тематичного моделювання [22, 28]. Такі методи використовують для аналізу колекції документів і вилучають з них теми, зв'язки між темами та їх зміни у часі. При цьому досліджувані документи розглядаються як набір не пов'язаних між собою слів або *Bag of words*. Для кожної публікації у СІС d_j розраховуються ймовірності $P(t|d)$ її належності до набору тем $t \in T$.

У табл. 3 представлено сучасні методи, які доцільно використовувати для встановлення прихованої тематики контенту СІС.

Таблиця 3

Методи виявлення прихованої теми контенту СІС

Назва	Сутність методу	Переваги	Недоліки
Ймовірнісне латентно-семантичне індексування	функціонує на основі аспектної моделі, яка пов'язує приховані параметри теми у публікації з кожною змінною, яка спостерігається словом чи темою	кожна публікація відноситься до деякої теми із заданою ймовірністю; метод має статистичне обґрунтування; зручність практичної реалізації	кількість параметрів моделі лінійно залежить від кількості публікацій у колекції; можливе перенавчання моделі; додавання нової публікації у колекцію вимагає перебудови моделі; повільно сходиться на великих колекціях публікацій
Приховане розміщення Діріхле	належить до породжуючих моделей, які дозволяють побудувати речення відповідно до правил заданої мови. Публікація розглядається як набір різних тем, апріорно розподілених за Діріхле	ефективний для опису кластерних структур; спрощує вивід апостеріорних ймовірностей публікацій і їх тем	відсутність лінгвістичних обґрунтувань методу; можливе перенавчання моделі
Робастна тематична модель	ґрунтується на твердженні, що вживання терміну в документі пояснюється темою, є специфічним для даного документа (шум) або загальновживаним терміном (фон)	вилучення з публікації фону і шуму, що не впливають на тему публікації; кращі показники критерію прогнозування появи слів (перплексії)	необхідність зберігати значний об'єм додаткових параметрів моделі

Зауваження 2. Для подальшої оцінки використовується максимальне значення ймовірності належності документа до однієї із набору тематик контенту СІС, що становлять інтерес для дослідження.

Крок 6. Розрахунок інформаційної ентропії маніпуляції суспільною думкою в СІС. В узагальненому вигляді зв'язок між частинними ознаками маніпуляцій суспільною думкою в СІС, розглянутими на кроках 1-5, зобразимо у вигляді ієрархії (рис. 2).

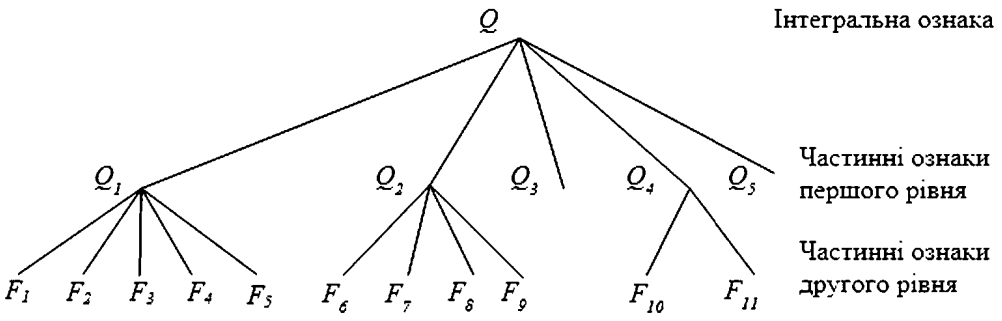

Рис. 2. Дерево рішень

Нехай маніпуляція суспільною думкою в текстовому контенті СІС проявляється k частинними ознаками. Позначимо ймовірності появи кожної з цих ознак як P_ν, $\nu = \overline{1,k}$. Припустимо, що проводиться N експериментів з виявлення ознак маніпуляцій у інформаційному потоці текстового контенту СІС. Для нього встановлюється кількість проявів ознак маніпуляцій суспільною думкою N_1, N_2, ..., N_k, а їх сума дорівнює N. Тоді загальну кількість виявленої інформації після проведення всіх експериментів можна оцінити за виразом [9, 29]

$$I = -\left(N_1 \log_2 P_1 + N_2 \log_2 P_2 + ... + N_k \log_2 P_k\right) = -\sum_{\nu=1}^{k} N_\nu \log_2 P_\nu . \qquad (13)$$

Якщо ліву і праву частину рівняння (13) поділити на N, отримаємо середню кількість інформації про присутність у досліджуваному текстовому контенті ознак маніпуляцій, отриманої за експеримент [30]

$$I_{avg} = \frac{I}{N} = -\sum_{\nu=1}^{k} \frac{N_\nu}{N} \log_2 P_\nu . \qquad (14)$$

Відношення $\frac{N_\nu}{N}$ представляє собою частоту появи f_ν відповідної ознаки маніпуляції суспільною думкою в інформаційному потоці текстового контенту віртуальних спільнот. При проведенні дослідження великої кількості публікацій акторів СІС, тобто при необмеженому зростанні N, частоти появи ознак f_ν будуть наближатися до відповідних ймовірностей P_ν. Розглянемо задачу пошуку мінімальної кількості інформації, граничне значення якої визначатиме критерій наявності у текстовому контенті СІС маніпулятивних технологій для управління суспільною думкою. Така задача відноситься до типових екстремальних задач з умовою, оскільки маємо функцію з k змінними. Розв'язок цієї задачі методом Лагранжа приймає вигляд [30]

$$f_1 \equiv P_1;\ f_2 \equiv P_2;\ \ldots;\ f_k \equiv P_k , \qquad (15)$$

тобто, найменша кількість середньої інформації про наявність в контенті СІС маніпуляцій суспільною думкою у випадку збігу ймовірності появи ν-ї ознаки P_ν з граничними значеннями відповідних частот f_ν

$$I_{avg\,\min} = -\sum_{\nu=1}^{k} f_\nu \log_2 f_\nu . \qquad (16)$$

Тоді критерій виявлення маніпуляцій у текстовому контенті СІС запишемо у вигляді нерівності

$$H = -\sum_{\nu=1}^{k} \sum_{l=1}^{g} Q_l^\nu \log_2 Q_l^\nu , \qquad (17)$$

де H – граничне значення інформаційної ентропії (невизначеності);

Q_l^ν – числове значення прояву ознаки маніпуляції суспільною думкою;

$l = \overline{1,g}$ – індекси частинних ознак маніпуляцій другого рівня;

$\nu = \overline{1,k}$ – індекси частинних ознак маніпуляцій першого рівня.

Для зручності інтерпретації розрахованих значень введемо нормоване значення ентропії H_n

$$H_n = \frac{H_{\max} - H}{H_{\max}}, \qquad (18)$$

де $H_{\max}$ – максимальне значення ентропії.

Таким чином, зміст критерію виявлення маніпуляцій суспільною думкою СІС зводиться до оцінки інформаційної ентропії текстового контенту віртуальних спільнот, тобто встановлення рівня невизначеності щодо наявності у контенті прихованого впливу на акторів та порівняння його числового значення із допустимим граничним. Інформаційна ентропія (17) зменшується при зростанні частот появи ознак маніпуляцій суспільною думкою акторів СІС. У випадку малих частот прояву ознак маніпуляцій у текстовому контенті СІС інформаційна невизначеність зростає. Якісна шкала оцінки загроз маніпуляції суспільною думкою акторів віртуальних спільнот сформована в результаті обчислювального експерименту та узагальнення і адаптації підходів до оцінки загроз у галузі інформаційної безпеки (табл. 4) [9, 27].

Таблиця 4

Адаптована інтервальна шкала

Клас загрози	Інтервальні значення нормованої ентропії H_n
дуже високий	0,00–0,20
високий	0,21–0,49
значний	0,50–0,74
низький	0,75–0,90
дуже низький	0,91–1,00

Експерименти. Розглянемо застосування технологій маніпуляцій суспільною думкою у СІС на прикладі передвиборчої кампанії президента США у 2016 році. Для цього проаналізуємо зміну передвиборчих рейтингів кандидатів у президенти США Х. Клінтон і Д. Трампа, які отримано *HuffPost Pollster* з таргетингом на користувачів мережі Інтернет у період з 1 січня по 7 листопада 2016 року.

Моніторинг мікроблогу *Twitter* показав, що 10 червня 2016 року між кандидатами на посаду президента США відбулася словесна перепалка. Після підтримки чинним президентом Б. Обамою Х. Клінтон її суперник вчергове опублікував звинувачення у робочій переписці з домашнього комп'ютера, що є порушенням федерального законодавства. При цьому встановлено наявність у публікаціях ознак маніпуляцій – емоційності, негативної тональності, сенсаційності. В результаті інциденту підтримка кандидатів серед акторів СІС змінилася, як показано на рис. 3 (а). 16 вересня 2016 року Д. Трамп у результаті багаторічних суперечок і теорій змови про місце народження чинного президента США Б. Обами визнав, що він народився на Гавайях. Розвінчання міфу, який активно підтримувався командою Д. Трампа і обговорювався акторами у СІС, призвело до негативного впливу на рейтинг кандидата (рис. 3, б). Також 7 жовтня 2016 року в інформаційному просторі СІС було поширено записи зі зневажливими висловлюваннями Д. Трампа про жінок. Аналіз показав, що при цьому у соціальній мережі *Facebook* активно використовувалися лідери думок для оцінки події. Публікації мали емоційний характер з яскраво вираженою негативною тональністю і мали на меті дискредитацію кандидата на пост президента США. Після поширення такого контенту рейтинг Д. Трампа продовжив падіння, а його конкурента Х. Клінтон – зростати (рис. 3, в).

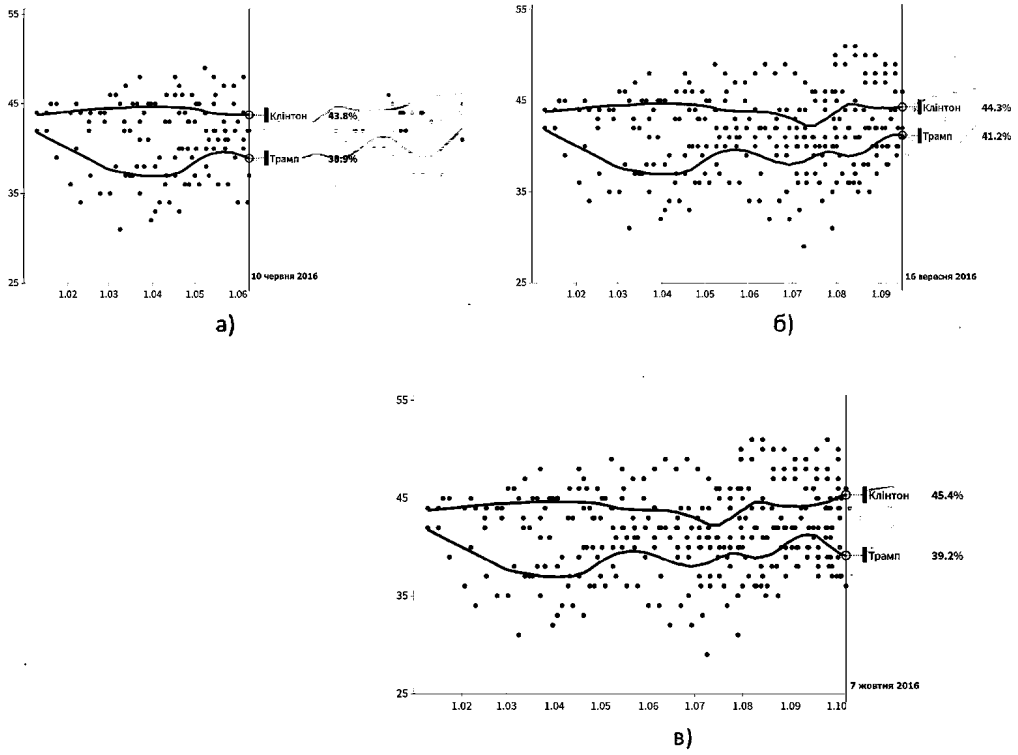

Рис. 3. Рейтинг кандидатів у президенти США

Комплексний аналіз передвиборної кампанії президента США 2016 року показав, що активно використовувалася технологія «спіралі мовчання». Рейтинги кандидатів від різних аналітичних агенцій істотно відрізнялися один від одного, а їх команди завчасно оголошували своїх керівників майбутніми переможцями. Це дозволило поширити серед виборців ідею невизначеності та невпевненості суспільства, відсутності прогнозованого переможця. Отже, проведений аналіз підтверджує дієвість використання механізмів маніпулятивних технологій у СІС для впливу на суспільну думку.

Також було проведено експериментальне дослідження запропонованої методики виявлення маніпуляцій суспільною думкою у СІС. Для аналізу було використано текстовий контент соціальної мережі *Вконтакте*, а також методи взаємодії із сервісом *API VK* та інтегроване середовище розробки *MS Visual Studio*. Визначення тональності текстового контенту реалізовано на основі мультиноміального наївного методу Байеса, а детектування прихованої тематики з використанням ймовірнісного латентно-семантичного індексування. У результаті розрахунку ентропії частинних ознак маніпуляцій суспільною думкою у СІС першого рівня (див. рис. 2) отримано такі числові значення, наведені у табл. 5.

Таблиця 5

Розрахункові значення ентропії

	H_{Q_1}	H_{Q_2}	H_{Q_3}	H_{Q_4}	H_{Q_5}
Значення	0,30	0,40	0,34	0,52	0,50

Візуалізація розрахункових даних подана у вигляді пелюсткової діаграми на рис. 4.

Відповідно до виразів (17)–(18) нормоване значення ентропії для даних табл. 5 приймає значення $H_n = 0{,}67$. Так, у текстовому контенті виявлено істотні прояви прихованої тематики і сенсаційності, присутні емоційна лексика й тональність контенту. Отже, досліджуваний контент СІС містить загрозу інформаційній безпеці значного рівня, тому вимагає вживання заходів захисту інформаційного середовища.

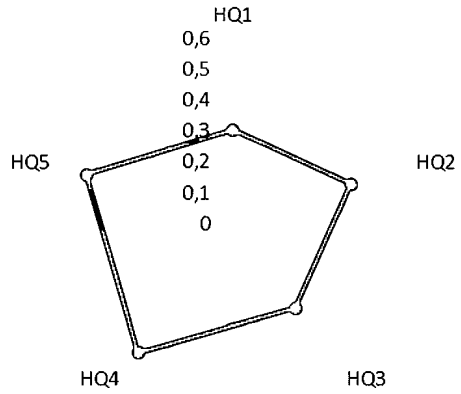

Рис. 4. Ентропія частинних ознак маніпуляцій

Висновки та перспективи подальших досліджень. Запропонована методика виявлення маніпулювання суспільною думкою у СІС ґрунтується на сучасних методах аналізу текстового контенту і забезпечує автоматизацію процедур детектування загроз інформаційній безпеці держави. Розроблена методика відрізняється від відомих врахуванням інформаційної невизначеності, яка виникає у разі використання маніпуляцій і сугестивних технологій зокрема, що дозволило підвищити ефективність моніторингу інформаційного простору СІС. Таким чином, досягається підвищення ефективності й швидкодії системи забезпечення інформаційної безпеки держави в СІС, що є сьогодні вкрай актуальним завданням для України.

Література:

1. Castells M. The Information Age: Economy, Society, and Culture : The Rise of the Network Society [vol. 1, 2nd edition] / Manuel Castells. – Wiley-Blackwell, 2011. – 656 p.

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

3. Компанцева Л. Ф. Сугестивний вплив в інтернеті: нові можливості лінгвістики / Л. Ф. Компанцева // Studia linguistica: зб. наук. праць. – 2011. – Вип. 5, Ч. II. – С. 213–220.

4. Ковалевич Б. В. Соціальні мережі як новий інструмент ведення інформаційних війн у сучасному світі / Б. В. Ковалевич // Грані. – 2014. – №4. – С. 118–121.

5. Milan S. From social movements to cloud protesting: The evolution of collective identity / S. Milan // Information, Communication & Society. – 2015. – 18. – PP. 887–900.

6. Using Twitter to mobilize protest action: Online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements / Y. Theocharis, W. Lowe, J. W. van Deth, G. García-Albacete // Information, Communication & Society. – 2015. – 18. – PP. 202–220.

7. Поліщук Ю. Я. Мас медіа як канал маніпулятивного впливу на суспільство / Ю. Я. Поліщук, С. О. Гнатюк, Н. А. Сейлова // Інформаційна безпека. – 2015. – Т. 21, Ч. 3. – С. 301–308.

8. Сугестивні технології маніпулятивного впливу : навч. посіб. / [В. М. Петрик, М. М. Присяжнюк, Л. Ф. Компанцева, Є. Д. Скулиш, О. Д. Бойко, В. В. Остроухов] ; за заг. ред. Є. Д. Скулиша. – [2-ге вид.] – К. : ЗАТ "ВІПОЛ", 2011. – 248 с.

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

10. Компанцева Л. Ф. Интернет-лингвистика: коммуникативно-прагматический и лингвокультурологический подходы : моногр. / Л. Ф. Компанцева. – Луганск : Знание, 2008. – 528 с.

11. Інформаційно-психологічне протиборство (еволюція та сучасність) : моногр. / Я. М. Жарков, В. М. Петрик, М. М. Присяжнюк, Є. Д. Скулиш, Л. Ф. Компанцева ; Київ. нац. ун-т ім. Т. Шевченка. – К. : Віпол, 2013. – 247 с.

12. Почепцов Г. Г. Контроль над розумом / Г. Г. Почепцов. – К. : КМ акад., 2012. – 350 с.

13. Почепцов Г. Г. Теорія комунікації / Г. Г. Почепцов ; Київ. ун-т ім. Т. Шевченка. – [2-ге вид., доп.] – К., 1999. – 307 с.

14. Почепцов Г. Інформаційна війна як інтелектуальна війна [Електронний ресурс] / Г. Почепцов. – Режим доступу: http://osvita.mediasapiens.ua/material/13303. – Назва з екрана.

15. Расторгуев С. П. Математические модели в информационном противоборстве. Экзистенциальная математика / С. П. Расторгуев. – М. : АНО ЦСОиП, 2014. – 260 с.

16. Расторгуев С. П. Информационная война / С. П. Расторгуев. – М. : Радио и связь, 1999. – 221 с.

17. Панченко В. М. Методика виявлення ознак інформаційного впливу в засобах масової інформації / В. М. Панченко, В. І. Полевий // Інформаційна безпека людини, суспільства, держави. – 2011. – №3 (7). – С. 70–77.

18. Панченко В. М. Лінгвостатистичні ознаки маніпулювання суспільною свідомістю в засобах масової комунікації / В. М. Панченко // Сучасні інформаційні технології у сфері безпеки та оборони. – 2009. – № 1(4). – С. 81–85.

19. Рябий М. О. Модель виявлення PR-впливу через публікації в інтернет ЗМІ / М. О. Рябий, О. А. Хатян, С. П. Багацький // Інформаційна безпека. – 2015. – Т. 21, № 2. – С. 131–139.

20. Молодецька К. В. Підхід до виявлення організаційних ознак інформаційних операцій у соціальних інтернет-сервісах / К. В. Молодецька // Пріоритетні напрямки розвитку телекомунікаційних систем та мереж спеціального призначення. Застосування підрозділів, комплексів, засобів зв'язку та автоматизації в АТО : збірн. матер. IX наук.-практ. конф., 25 листоп. 2016 р. – Київ : ВІТІ, 2016. – С. 130–131.

21. Faraz A. A comparison of text Categorization methods / A. Faraz // International Journal on Natural Language Computing. – 2016. – 5(1). – PP. 31–44.

22. Ланде Д. В. Интернетика: навигация в сложных сетях: методы и алгоритмы / Д. В. Ланде, А. А. Снарский, И. В. Безсуднов. – М. : Книжный дом "ЛИБРОКОМ", 2009. – 264 с.

23. Волосюк Ю. В. Методи класифікації текстових документів в задачах Text Mining / Ю. В. Волосюк // Наукові записки Українського науково-дослідного інституту зв'язку. – 2014. – №6(34). – С. 76–81.

24. Fernández-Martínez F. Text categorization methods for automatic estimation of verbal intelligence / F. Fernández-Martínez, K. Zablotskaya, W. Minker // Expert Systems with Applications. – 2012. – 39(10). – PP. 9807–9820.

25. Sebastiani F. Machine learning in automated text categorization / ACM Computing Surveys (CSUR) // F. Sebastiani. – 2002. – 34(1). – PP. 1–47.

26. Воронина И. Е. Анализ эмоциональной окраски сообщений в социальных сетях (на примере сети «Вконтакте») / И. Е. Воронина, В. А. Гончаров // Вестник ВГУ, серия: системный анализ и информационные технологии. – 2015. – №4. – С. 151–158.

27. Грищук Р. В. Метод оптимізації розмірності потоку вхідних даних для систем захисту інформації / Р. В. Грищук, В. М. Мамарев // Інформаційна безпека. – 2012. – №2 (8). – С. 27–34.

28. Воронцов К. В. Модификации EM-алгоритма для вероятностного тематического моделирования / К. В. Воронцов, А. А. Потапенко // Машинное обучение и анализ данных. – 2013. – Т. 1. – №6. – С. 657–686.

29. Жураковський Ю. П. Теорія інформації та кодування : підр. / Ю. П. Жураковський, В. П. Полторак. – К. : Вища школа, 2001. – 255 с.

30. Тростников В. Н. Человек и информация : моногр. / Виктор Николаевич Тростников. – М. : Наука,1970. – 188 с.

Рецензент: д.т.н., проф. Поповський В.В. Надійшла 03.10.2016

Молодецкая-Гринчук К.В.

МЕТОДИКА ВЫЯВЛЕНИЯ МАНИПУЛЯЦИЙ ОБЩЕСТВЕННЫМ МНЕНИЕМ В СОЦИАЛЬНЫХ ИНТЕРНЕТ-СЕРВИСАХ

В статье теоретически обосновано и обобщено признаки манипуляций общественным мнением акторов в социальных интернет-сервисах. Предложенная методика выявления манипуляций в текстовом контенте основывается на современных методах интеллектуального анализа и раннего выявления угроз информационной безопасности государства в социальных интернет-сервисах. Разработанная методика отличается от известных учетом информационной неопределенности, что позволило повысить эффективность функционирования системы обеспечения информационной безопасности государства.

Ключевые слова: манипуляции, социальные интернет-сервисы, угрозы, машинное обучение, мониторинг.

Molodetska-Hrynchuk K.

DETECTION METHODS OF MANIPULATE PUBLIC OPINION IN THE SOCIAL NETWORKING SERVICES

The paper theoretically grounded and generalized signs of manipulation of public opinion actors in the social networking services. The technique of detecting manipulation of the text content based on modern mining methods and early detection of threats to information security of the state in social networking services. The method differs from the known information into account uncertainty, thus improving the efficiency of the system of information security.

Keywords: manipulation, social networking services, threats, machine learning, monitoring.