

УДК 004.773.2

Голуб Зоряна ДмитрівнаАспірант кафедри соціальних комунікацій та інформаційної діяльності, orcid.org/0000-0001-7618-6153
Національний університет «Львівська політехніка», Львів

РОЗРОБЛЕННЯ ФОРМАЛЬНИХ МОДЕЛЕЙ ДЛЯ АВТОМАТИЗАЦІЇ ВИЯВЛЕННЯ ІНФОРМАЦІЙНО-ПСИХОЛОГІЧНОЇ МАНІПУЛЯЦІЇ

***Анотація.** Онлайн-спільноти користуються значною популярністю як платформи для здійснення інформаційної діяльності. Відповідно управлінська ланка онлайн-спільнот стикається з проблемою виявлення негативного для функціонування спільноти явища – інформаційно-психологічної маніпуляції (ІПМ). Виявлення ІПМ вручну є часо- та трудозатратним процесом. Запропоновано формальні моделі онлайн-спільноти, тактики ІПМ та вказівника ІПМ, які необхідні для розроблення автоматизованих методів і засобів виявлення ІПМ. Запропоновано формальну модель спільноти, виділено ключові ознаки онлайн-спільноти необхідні для виявлення ІПМ та поділено їх за етапами і методами виявлення ІПМ, у яких вони використовуються. Формальна модель тактики ІПМ складається з елементів, які дають змогу відобразити їх як послідовну схему використання шаблонів. Крім того, запропоновано схему визначення вказівника прийому ІПМ, який дає можливість швидко ідентифікувати фрагменти дискусії, які потребують перевірки на наявність ІПМ та визначити набір тактик ІПМ, які найімовірніше використовувались у прецеденті.*

Ключові слова: інформаційно-психологічна маніпуляція; онлайн-спільнота; вказівник прийому ІПМ; тактика ІПМ; прийом ІПМ

Вступ

Інтернет-засоби комунікації є основним каналом інформування в сьогоденні. Швидкість передачі інформації, широка аудиторія та різноманіття джерел, територіальна безмежність роблять їх привабливими як для подачі, так і для отримання інформації.

Перенесення негативних процесів та явищ пов'язаних зі спілкуванням в онлайн-середовище ставить перед адміністраторами та модераторами онлайн-спільнот завдання виявлення та нейтралізації деструктивної інформаційної діяльності. Відповідно зростає актуальність теми дослідження мовних [1], поведінкових [2; 3], інтенціональних [4] та психологічних механізмів та засобів реалізації комунікативних актів в інтернет-медіа. Дослідження різних аспектів спрямовані на підвищення ефективності інтернет-комунікації [5], а саме на збільшення та використання переваг та запобігання негативним явищам інтернет-комунікації.

Значна частина інтернет-комунікації відбувається на платформах онлайн-спільнот. Це пояснюється невеликою кількістю вимог до публікування інформації (наприклад, реєстрація) та значною кількістю реципієнтів та інформаторів.

Комунікація в онлайн-спільнотах найбільш наближена до усного діалогічного спілкування, яке завжди переважало серед інших традиційних видів інформування за кількістю переданої інформації та

довірою до неї. Інтернет-відповідник усного діалогічного спілкування відіграє важливу роль у інформаційному обміні.

Деструктивним явищем для спільноти, учасників та об'єктів дискусії (обговорюваних брендів, організацій, осіб) є інформаційно-психологічна маніпуляція. Інформаційно-психологічна маніпуляція (ІПМ) – це цілеспрямований вплив однієї людини на підсвідомість інших людей за допомогою компонент інформаційного простору та механізмів психологічного впливу з односторонньо вигідною метою.

Оскільки онлайн-спільнотам властиве швидке поширення інформації, то протидія ІПМ вимагає вчасного виявлення, що можливе лише за умови масштабного підрозділу кваліфікованих кадрів або автоматизованих методів та засобів.

ІПМ здійснюють на основі тактик розроблених як для традиційного спілкування, так і спеціально для онлайн-спільнот. Домінантним видом комунікації в онлайн-спільнотах є комунікація за допомогою текстових повідомлень. Відповідно методи виявлення ІПМ мають базуватися на формальних моделях спільноти, тактик ІПМ та вказівника прийому ІПМ.

Мета статті

Передумовою нейтралізації деструктивного впливу за допомогою ІПМ є своєчасне виявлення ІПМ в онлайн-спільнотах. Мета цієї статі

запропонувати формальні моделі спільноти, тактики маніпуляції та вказівника прийому ППМ, які б стали основою для розроблення автоматизованих засобів та методів виявлення ППМ.

Виклад основного матеріалу

Формальна модель онлайн-спільноти

Визначення формальної моделі базових елементів онлайн-спільноти залежить від завдання дослідження. У [6] описана формальна модель онлайн-спільнот, з метою розроблення методів та засобів побудови ефективних віртуальних спільнот на основі Веб-форумів. Формальна модель онлайн-спільноти для побудови методів і засобів виявлення інформаційних загроз віртуальних спільнот в інтернет-середовищі соціальних мереж описана у [7]. Оскільки завданням цього дослідження є розроблення методів і засобів виявлення інформаційно-психологічної маніпуляції, ключовими об'єктами дослідження є інформаційне наповнення, а також автори його елементів, оскільки за допомогою інформації про авторів можна ідентифікувати джерело небезпечних повідомлень.

За основу взято формальну модель віртуальної спільноти запропоновану у [6], адже вона відповідає загальній структурі, характерній для усіх типів віртуальних спільнот.

Формальна модель віртуальної спільноти подана як кортеж:

$$Community = \langle Content, Member \rangle,$$

де $Content = \{Discussion_j\}_{j=1}^N$ – множина дискусій, в яких беруть участь учасники спільноти;

$Member = \{Member_j\}_{j=1}^N$ – множина учасників.

Ця формальна модель виділяє два відмінні з точки зору виявлення ППМ формати інформації в онлайн-спільнотах: інформація зосереджена навколо певної теми та інформація щодо профілю учасника. Відповідно до цієї формальної моделі критерії виявлення ППМ поділяються на динамічні та статичні. Статичні критерії розраховуються відносно профілів учасників онлайн-спільноти, динамічні для конкретних повідомлень, які містить спільнота [8]. Статичні критерії використовуються для ідентифікації підозрілих фрагментів онлайн-дискусії, натомість динамічні для виявлення прецедентів ППМ та потребують глибинного аналізу текстового наповнення дискусії.

Формальна модель дискусії

Інформаційне наповнення онлайн-спільноти складається з множини дискусій. Кожній дискусії відповідають унікальні характеристики, які вирізняють цю дискусію з-поміж інших, а саме: назва, тема та

створене користувачами інформаційне наповнення.

Крім того, формальна модель дискусії містить елементи, необхідні для створення черги з дискусій для перевірки на наявність ППМ. Для сортування дискусій за привабливістю для здійснення ППМ та створення черги важливо знати автора дискусії, кількість учасників, дату створення дискусії та дату останнього повідомлення. Також ці дані необхідні для виявлення підозрілих фрагментів дискусії.

Формальна модель дискусії задана таким кортежем:

$$Discussion_i = \left\langle \begin{array}{l} DiscussionTitle_i, DiscussionTopic_i, \\ DiscussionAuthor_i, Messages_i, \\ FirstMessageDate_i, LastMessageDate_i, \\ ParticipantCount_i, \end{array} \right\rangle,$$

де $DiscussionTitle_i$ – назва i -ї дискусії; $DiscussionAuthor_i$ – автор i -ї дискусії; $FirstMessageDate_i$ – дата першого повідомлення в i -й дискусії; $LastMessageDate_i$ – дата останнього повідомлення в i -й дискусії; $ParticipantCount_i$ – кількість учасників дискусії; $Messages_i = \{Message_{ij}\}_{j=1}^{N^{Discussion_i}}$ – множина повідомлень, що належать до i -ї дискусії, де $n^{Discussion_i}$ кількість повідомлень в i -й дискусії.

Формальна модель дискусії однозначно ідентифікує дискусію в онлайн-спільноті та дозволяє розрахувати необхідні дані для сортування дискусій за наявністю передумов до ППМ та виявлення підозрілих фрагментів дискусії.

Формальна модель учасника спільноти

Здійснюючи маніпуляцію, маніпулятор залишає по собі слід, цей слід проявляється на різних організаційно-структурних рівнях, а саме: в спільноті, в дискусії чи повідомленні. Тому аналізувати процес комунікації в інтернет-дискусіях потрібно через призму кожного з рівнів. Рівні відрізняються підходом до виявлення ознак ППМ, передбачають різну складність обчислення та точність результатів.

Статичні критерії ППМ поділяються відповідно до цих трьох рівнів і пов'язані з конкретним профілем учасника, тому формальна модель учасника спільноти має містити дані для розрахунку цих критеріїв [8].

Формальна модель учасника онлайн-спільноти [9] містить ім'я, емейл, посилання на профіль у Facebook (за наявності), роль учасника щодо ППМ, характеристики поведінки учасника, характеристики профілю та характеристики активності учасника. Елементи формальної моделі необхідні для однозначної ідентифікації учасника спільноти, виявлення підозрілого профілю та підозрілих фрагментів дискусії.

Формальна модель учасника виглядає так:

$$Member_i = \left\langle \begin{array}{l} MemberName_i, Email_i, \\ FacebookAccount_i, IPMRole_i, \\ BehaviouralCharacteristics_i, \\ ProfileFeatures_i, ActivityHistory_i \end{array} \right\rangle,$$

де $MemberName_i$ – псевдо учасника; $Email_i$ – адреса електронної пошти учасника; $FacebookAccount_i$ – це посилання на профіль учасника у Facebook; $BehaviouralCharacteristics_i$ – характеристики поведінки учасника; $ProfileFeatures_i$ – характеристики профілю учасника; $ActivityHistory_i$ – історія діяльності учасника.

$IPMRole_i$ – роль учасника у дискусії з точки зору виявлення ППМ, цей елемент може набувати одне зі значень вектора:

$$IPMRoleValue = \left(\begin{array}{l} Neutral; Manipulator; \\ Victim; Zombie \end{array} \right),$$

де елементи відповідають таким значенням: нейтральний, маніпулятор, жертва, зомбі.

Аналізуючи поведінкові характеристики профілю, можна виявити аномалії, наприклад, розміщення повідомлень виключно з 9 до 5, відсутність реакції на коментарі до власного допису, участь у дискусіях лише певної тематики. Якщо в поведінкових ознаках профілю виявлено кілька таких аномалій, то його діяльність у спільноті варто проаналізувати на наявність ППМ.

Аналізуючи характеристики профілю учасника, можна виявити маніпуляторів спільників, за однаковим чи подібним заповненням полів профілю. Також підозрілого учасника за низьким ступенем наповненості профілю.

Характеристики діяльності несуть інформацію про важливі події в інформаційній діяльності учасника у спільноті, наприклад, дата реєстрації чи дата порушення правил спільноти.

Наведені характеристики профілю необхідні для однозначної ідентифікації профілю, аналізу з метою виявлення підозрілих профілів та встановлення можливих шляхів поширення ППМ. Спільнота не має містити дані щодо всіх наведених характеристик, оскільки критерії сортування дискусій та набір фільтрів для виявлення підозрілих уривків дискусії формуються на основі даних, передбачених структурою дискусії.

Формальна модель повідомлення

Формальна модель повідомлення складається з автора повідомлення, дати й часу розміщення повідомлення, типу повідомлення, та показника релевантності повідомлення, а також базових обов'язкових характеристик та додаткових характеристик необхідних для виявлення маніпуляції:

$$Message_i = \left\langle \begin{array}{l} MessageID_i, AuthorNickname_i, \\ TimeStamp_i, MessageContent_i, \\ DirectionType_i, Relevance_i, \\ FeaturesBasic_i, FeaturesAdditional_i \end{array} \right\rangle,$$

де $MessageID_i$ – ідентифікатор повідомлення; $AuthorNickname_i$ – псевдо автора повідомлення; $TimeStamp_i$ – дата і час розміщення повідомлення; $MessageContent_i$ – текст повідомлення; $DirectionType_i$ – множина типів повідомлень за спрямуванням; $Relevance_i$ – це показник релевантності повідомлення до теми дискусії; $FeaturesBasic_i$ – кортеж обов'язкових характеристик повідомлення; $FeaturesAdditional_i$ – кортеж додаткових характеристик повідомлення.

$DirectionType_i \in \{Stimulus, Reaction\}$ – значення, які може набувати елемент множина типів повідомлень за спрямуванням.

Множина типів повідомлень за спрямуванням містить два типи (4): *Stimulus* – це повідомлення, в яких надається певна інформація і *Reaction* – це повідомлення, які є відповіддю на інше повідомлення. У повідомленні типу *Reaction* – є посилання на повідомлення, яке спровокувало цю відповідь. Визначення типів повідомлення необхідне для відтворення логічної структури інформаційного наповнення обговорення.

$Relevance_i \in \{Relevant, Irrelevant\}$ – показник релевантності повідомлення, може набувати значення *Relevant, Irrelevant* відповідно пов'язані та не пов'язані за змістом повідомлення з тематикою онлайн-дискусії. Визначається на основі кількості наявних ключових слів та відповідності діалогічного акту до діалогічного акту в попередньому повідомленні.

Для кожного обговорення експерти визначають порогові значення параметрів повідомлення. Повідомлення, що їх перевершує, визначають як підозріле та відправляють на перевірку на наявність вказівників прийомів ППМ.

Формальна модель тактики ППМ

Тактики ППМ базуються на психологічних механізмах сприйняття інформації та здійснюються за допомогою наявних засобів передачі інформації.

Оскільки у цій статті розглянуто ППМ в онлайн-спільнотах, яка реалізовується у формі текстових повідомлень, то наявні засоби інформації це вербальні, частково невербальні (а саме емотикони та посилання, тобто лише ті, які реалізовано за допомогою текстових символів або, які мають однозначні коди, за допомогою яких їх можна зіставити з певним текстовим значенням).

Психічний та емоційний стан реципієнтів впливає на сприйняття повідомлення. Тому ключовою характеристикою тактик ПМ є послідовне переведення жертви ПМ зі стану в стан, з поданням інформації, яка засвоюється відповідним до стану чином і в кінцевому результаті приводить до того, що світогляд чи дії маніпулятора відповідають меті ПМ. Цей процес змодельовано за допомогою кусково-лінійних агрегатів [10], адже вони відображають зміну внутрішнього стану системи під дією зовнішніх сигналів.

Тактики ПМ вибудовуються з прийомів, які несуть приховану інформацію, яка змушує реципієнта перейти в інший психічний стан. Прийоми реалізуються за допомогою звичних для текстових повідомлень одиниць комунікації: текстових символів, параграфічних об'єктів, емотиконів та посилань. Комбінації вищенаведених одиниць є маркерами маніпулятивних прийомів.

У той же час маркери можуть траплятися і у звичайних, незловмисницьких повідомленнях. Тому, констатувати факт здійснення ПМ можна на основі повідомлень з маркерами прийомів ПМ, послідовність появи яких відповідає послідовності переходів між станами одної з тактик ПМ.

Ключовими елементами формальної моделі тактики ПМ є стан реципієнта та крок ПМ.

Стан, в якому перебуває людина має значний вплив на адаптування ідеї чи сприйняття. Психічний стан – один із можливих перцептивних режимів життєдіяльності людини, що на фізіологічному рівні відрізняється певними енергетичними характеристиками, а на психологічному рівні – системою соціально-психологічних фільтрів, що забезпечують специфічне сприйняття себе та зовнішнього світу. Психічні стани є реакцією на зовнішні подразники [11].

Перехід реципієнта у інший стан переважно може спричинити лише кількаразове виконання кроку маніпуляції. Отже, послідовна зміна психічних станів реципієнта є ознакою маніпуляції, а ознакою переходу в інший психічний стан є кількаразове виконання кроку маніпуляції. Кількість кроків маніпуляції можна визначити за кількістю виявлених маркерів. Якщо кроків було достатньо, реципієнт перейшов в інший психічний стан.

Модель тактики маніпуляції, представлена за допомогою кусково-лінійного агрегату, виглядає таким чином:

$$TacticModel = \left\langle \begin{array}{l} TacticState, TacticStep, \\ ChangeStateFunction \end{array} \right\rangle,$$

де $TacticState = \{TacticState_i\}_{i=1}^{N^{TacticState}}$ – множина станів даної тактики, де $n^{TacticState}$ – кількість можливих станів. До множини станів кожної тактики

належить початковий стан $InitialState_i$, всі параметри якого дорівнюють нулю, а множина кроків, які переводять у початковий стан – пуста;

$$TacticState_i = \left\langle \begin{array}{l} StateTitle_i, StateParameters_i, \\ TacticStep_i \end{array} \right\rangle,$$

$TacticState_i$ – стан характеризується назвою, вектором параметрів і множиною кроків, які спричиняють перехід у цей стан.

$StateParameters_i = (pState_1^{(i)}, \dots, pState_k^{(i)})$ – це вектор параметрів, який характеризує внутрішній стан. Кожен із параметрів вектора стану може набувати значення із заданого інтервалу $TacticState_i \hat{I} [min^i, max^i]$.

$StepParameters_i = (pStep_1^{(i)}, \dots, pStep_k^{(i)})$ – це вектор параметрів, який характеризує крок маніпуляції. Кожен з параметрів вектора стану може набувати значення із заданого інтервалу $TacticStep_i \hat{I} [min^i, max^i]$.

Отже, тактика ПМ подається як послідовність станів реципієнта та кроки маніпулятора, які переводять реципієнтів у відповідний стан. На базі цієї формальної моделі виявлення ПМ складатиметься з двох етапів. На першому етапі виявляються стани, їхня послідовність та встановлюються тактики, які були можливо застосовані. На другому етапі аналізуються вибрані тактики та прийоми, за допомогою яких можна реалізувати кроки тактик, з інформаційним наповненням повідомлення.

Визначення вказівника прийому ПМ

Прийоми ПМ будуються за допомогою звичайних мовних засобів. Прийоми ПМ реалізуються як висловлення, але внаслідок поєднання певної синтаксичної структури, лексики та параграфеміки мають вплив на підсвідомість реципієнта. Кожен прийом ПМ має мету, може приймати одну з форм реалізації. Прийом ПМ формалізовано таким чином:

$$Tool_k = \left\langle \begin{array}{l} ToolTitle_k, ToolGoal_k, \\ ToolParameters_k, ToolVariation_k \end{array} \right\rangle,$$

де $ToolTitle_k$ – назва прийому; $ToolGoal_k$ – маніпулятивна мета прийому, спрямована на зміни психічного стану реципієнта. $ToolParameters_k$ – це вектор параметрів, який характеризує дію прийому ПМ. Кожен параметр кроку має лише одне числове значення, $ToolVariation_k = \{ToolVariation_m\}_{m=1}^{m^{Tool_k}}$ – це множина форм реалізації прийому ПМ.

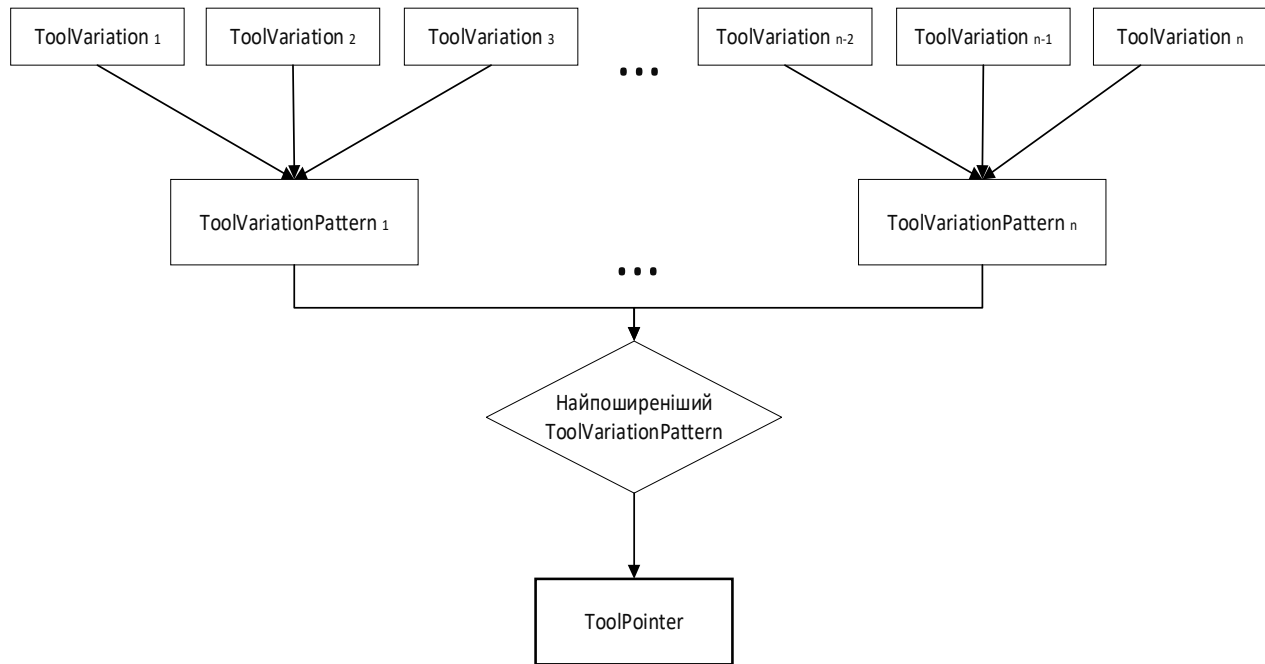


Рисунок – Схема визначення вказівника прийому ППМ

Внаслідок аналізу прецедентів ППМ визначають характерний набір лексичних, параграфічних елементів, емотиконів та посилань, який застосовується для реалізації прийому ППМ [12; 13].

Вказівник прийому ППМ – це вербальні або невербальні засоби, які найчастіше використовують для реалізації прийому ППМ. Вказівник прийому ППМ використовується на перших кроках етапу виявлення ППМ та на підготовчому етапі з метою виділення підозрілих фрагментів ППМ.

Вказівник прийому ППМ визначають за схемою, поданою на рисунку. Спершу варіанти реалізації прийомів ППМ подають у вигляді шаблонів за допомогою нотації Бекуса-Наура – $ToolVariationPattern_i$. Потім визначають частоту використання шаблонів варіантів реалізації. Шаблон варіантів реалізації, який найчастіше використовується для здійснення цього прийому є $ToolPointer$.

$ToolPointer$ використовується з такою метою: на підготовчому етапі для виявлення підозрілого уривка дискусії, на етапі виявлення для швидкої ідентифікації застосованих прийомів ППМ [14].

Висновки

Ключовою передумовою ефективного функціонування онлайн-спільноти є виявлення та нейтралізація згубних інформаційних дій, які вигідні лише їхнім реалізаторам або замовникам. Наявність прецедентів ППМ у онлайн-спільноті загрожує зниженню якості інформаційного наповнення, рівня довіри учасників, а відповідно спричиняє відхід учасників та припинення активного функціонування спільноти.

Для вирішення цієї актуальної проблеми управління онлайн-спільнотами запропоновано формальну модель онлайн-спільноти та тактики ППМ. Ці моделі є основою для розроблення методів та засобів виявлення ППМ в онлайн-спільнотах. Формальні моделі враховують вербальні, невербальні елементи повідомлень, які є ознаками ППМ, а також поведінкові, діяльнісні та ознаки профілів учасників, які дають можливість окреслити область потенційно наявної ППМ.

Список літератури

1. *Unsupervised Modeling of Twitter Conversations* / A. Ritter, C. Cherry, B. Dolan // *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL*, pages 172–180, Los Angeles, California, June 2010.
2. *Angeletou S. Modelling and Analysis of User Behaviour in Online Communities* / S. Angeletou, M. Rowe, H. Alani [Electronic resource]. – <https://pdfs.semanticscholar.org/c8e4/1ccbfd8b953464b8eaa58c9af6c3cf851661.pdf>
3. *O'Donovan F. T. Characterizing user behavior and information propagation on a social multimedia network* / F. T. O'Donovan, C. Fournelle, S. Gaffigan, O. Brdiczka, J. Shen, J. Liu, K. E. Moore [Electronic resource]. – <https://arxiv.org/ftp/arxiv/papers/1305/1305.2091.pdf>

4. Liu D. User Interest and Interaction Structure in Online Forums / D. Liu, D. Percival, S. E. Fienberg [Electronic resource]. – <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1500/1876>
5. Krombholz K. Fake Identities in Social Media: A Case Study on the Sustainability of the Facebook Business Model / K. Krombholz, D. Merkl, E. Weippl [Electronic resource]. – <https://www.sba-research.org/wp-content/uploads/publications/krombholzetal2012.pdf>
6. Arrasvuori J. A Model for Analyzing Online Communities / J. Arrasvuori, Th. Olsson [Electronic resource]. – https://www.researchgate.net/publication/267704038_A_Model_for_Analyzing_Online_Communities
7. Пелецишин А.М. Аналіз існуючих типів віртуальних спільнот у мережі Інтернет та побудова моделі віртуальної спільноти на основі Веб-форуму / А.М. Пелецишин, Р.Б. Кравець, Ю.О. Серов // Інформаційні системи та мережі: Вісник Національного університету «Львівська політехніка». – 2011. – № 699. – С. 212-221.
8. Голуб З.Д. Система критеріїв для виявлення фрагментів онлайн-дискусій з підозрою на наявність інформаційно-психологічної маніпуляції / З.Д. Голуб // Вісник Національного технічного університету «ХПІ». Збірник наукових праць. Серія: Нові рішення в сучасних технологіях. – Х.: НТУ «ХПІ». – 2018. – № 9 (1285). – С. 106-111.
9. Tuna T. User Characterization for Online Social Networks / T. Tuna, E. Akbas, A. Aksoy, M. Abdullah Canbaz, U. Karabiyik, B. Gonen, R. Aygun [Electronic resource]. – <https://arxiv.org/abs/1611.03971>
10. Pranevicius H. Applications of finite linear temporal logic to piecewise linear aggregates / H. Pranevicius, S. Norgela [Electronic resource]. – <https://www.mii.lt/informatica/pdf/info869.pdf>
11. Gottschalk L. A. The measurement of psychological states through the content analysis of verbal behavior / L.A. Gottschalk, G.C. Gleser // Comprehensive psychiatry The Journal of Psychopatology – Issue 3 (Vol.12) – University of California Press, 1969. – P. 317.
12. Голуб З.Д. Формалізація прийомів інформаційно-психологічної маніпуляції / З.Д. Голуб // Вісник Національного технічного університету «ХПІ». Збірник наукових праць. Серія: Нові рішення в сучасних технологіях. – Х.: НТУ «ХПІ». – 2017. – № 32 (1254). – С. 55-61.
13. Peteszczyszyn A., Holub Z. Development of the system for detecting manipulation in online discussions // Advances in Intelligent Systems and Computing (AISC). – 2017. – V. 543. – С. 111-117.
14. Голуб З. Розроблення алгоритму виявлення шкідливих інформаційно-психологічних маніпуляцій в онлайн-спільнотах ВНЗ / З. Голуб // Вісник Національного університету «Львівська політехніка». Серія: Інформатизація вищого навчального закладу : збірник наукових праць. – Видавництво Львівської політехніки, 2017. – № 879. – С. 33-41.

Стаття надійшла до редколегії 20.04.2018

Рецензент: д-р техн. наук, проф. А.М. Пелецишин, Національний університет «Львівська політехніка».

Голуб Зоряна Дмитрієвна

Аспірант кафедри соціальних комунікацій та інформаційної діяльності
Національний університет «Львівська політехніка», Львів

РАЗРАБОТКА ФОРМАЛЬНЫХ МОДЕЛЕЙ ДЛЯ АВТОМАТИЗАЦИИ ВЫЯВЛЕНИЯ ИНФОРМАЦИОННО-ПСИХОЛОГИЧЕСКОЙ МАНИПУЛЯЦИИ

Аннотация. Онлайн-сообщества пользуются большой популярностью как платформы для осуществления информационной деятельности. Соответственно управленческое звено онлайн-сообществ сталкивается с проблемой выявления негативного для его функционирования явления – информационно-психологической манипуляции (ИПМ). Выявление ИПМ вручную является время- и трудозатратным процессом. Предложены формальные модели онлайн-сообщества, тактики ИПМ и указателя ИПМ, необходимые для разработки автоматизированных методов и средств обнаружения ИПМ. Предложена формальная модель сообщества, выделены ключевые признаки онлайн-сообщества, необходимые для выявления ИПМБ. Они разделены по этапам и методам выявления ИПМ, в которых они используются. Формальная модель тактики ИПМ состоит из элементов, позволяющих отразить их как последовательную схему использования шаблонов. Кроме того, предложена схема определения указателя приема ИПМ, дающего возможность быстро идентифицировать фрагменты дискуссий, нуждающиеся в проверке на наличие ИПМ, и определить набор тактик ИПМ, которые, вероятно, использовались в прецеденте.

Ключевые слова: информационно-психологическая манипуляция; онлайн-сообщество; указатель приема ИПМ; тактика ИПМ; прием ИПМ

Holub Zoriana

Ph.D student at the department of social communication and information activities
Lviv National Polytechnic University, Lviv

**DEVISING FORMAL MODELS FOR AUTOMATION OF IPM DETECTION WITH THE AIM
OF EFFECTIVE MANAGEMENT OF ONLINE COMMUNITIES**

Abstract. Popularity of online communities as platforms for informational activity surged. Administrative and managing officers of online communities face the problem of detecting and neutralizing destructive and malicious for functioning of an online community phenomenon, namely information and psychological manipulation (IPM). Manual detection of IPM is a time and efforts consuming process. The paper suggests formal models of online communities, IPM tactics and scheme for defining an IPM pointer, which are required for designing of automatic methods and means of IPM detection. A formal model of an online community is proposed, key features of online communities necessary for detecting IPM are defined, the features are grouped according to stages and techniques that are applied in the process of IPM detection. The formal model of IPM tactic consists of elements that make it possible to depict it as a sequence of used patterns of IPM tools. The scheme for defining an IPM tool pointer is detailed. An IPM tool pointer is used for prompt identification of discussion fragments that are to be checked for presence of IPM and for determining of a set of IPM tactics, that are the most probably used to implement an IPM precedent.

Keywords: information and psychological manipulation; online communities; IPM tool pointer; IPM tactic; IPM tool

References

- Ritter, A. (2010). *Unsupervised Modeling of Twitter Conversations* / A. Ritter, C. Cherry, B. Dolan // *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL.*, 172–180.
- Angeletou, S. *Modelling and Analysis of User Behaviour in Online Communities* / S. Angeletou, M. Rowe, H. Alani [electronic source]. – <https://pdfs.semanticscholar.org/c8e4/1ccbfd8b953464b8eaa58c9af6c3cf851661.pdf>
- O'Donovan, F. T. *Characterizing user behavior and information propagation on a social multimedia network* / F. T. O'Donovan, C. Fournelle, S. Gaffigan, O. Brdiczka, J. Shen, J. Liu, K. E. Moore [electronic source]. – <https://arxiv.org/ftp/arxiv/papers/1305/1305.2091.pdf>
- Liu, D. *User Interest and Interaction Structure in Online Forums* / D. Liu, D. Percival, S. E. Fienberg [electronic source]. – <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1500/1876>
- Krombholz, K. *Fake Identities in Social Media: A Case Study on the Sustainability of the Facebook Business Model* / K. Krombholz, D. Merkl, E. Weippl [electronic source]. – <https://www.sba-research.org/wp-content/uploads/publications/krombholzetal2012.pdf>
- Arrasvuori, J. *A Model for Analyzing Online Communities* / J. Arrasvuori, Th. Olsson [electronic source]. – https://www.researchgate.net/publication/267704038_A_Model_for_Analyzing_Online_Communities
- Peleshchyschyn, A.M. (2010). *Analysis of existing in the Internetnet work community types and devising a model of a virtual community on the basis of web-forum* / R.B. Kravets, Yu.O. Sierov // *Information systems and networks: Bulletin of Lviv Polytechnic National University*, 699., 212-221.
- Holub, Z.D. (2018). *System of criteria for detecting on line discussion fragments suspected of containing informational and psychologic manipulation*. *Bulletin of National technical University «KhPI». New solutions for modern technologies*, 9 (1285), 106-111.
- Tuna, T. *User Characterization for Online Social Networks* / T. Tuna, E. Akbas, A. Aksoy, M. Abdullah Canbaz, U. Karabiyik, B Gonen, R. Aygun [electronic source]. – <https://arxiv.org/abs/1611.03971>
- Pranevicius, H. *Applications of finite linear temporal logic to piecewise linear aggregates* / H. Pranevicius, S. Norgela [Electronic resource]. – <https://www.mii.lt/informatica/pdf/info869.pdf>
- Gottschalk, L. A. (1969). *The measurement of psychological states through the content analysis of verbal behavior* / L. A. Gottschalk, G. C. Gleser // *Comprehensive psychiatry The Journal of Psychopatology*, 3 (Vol.12), 317.
- Holub, Z.D. (2017). *Formalization of information and psychologic manipulation tools*. *Bulletin of National technical University «KhPI». New solutions for modern technologies*, 32 (1254), 55-61.
- Peleszczyszyn, A., Holub, Z. (2017). *Development of the system for detecting manipulation in online discussions*. *Advances in Intelligent Systems and Computing (AISC)*, 543, 111-117.
- Holub, Z.D. (2017). *Devising an algorithm for detecting malicious information and psychological manipulation sin online communities of HEI*. *Bulletin of Lviv Polytechnic National University. Informatization of an higher education institution*, 879, 33-41

Посилання на публікацію

- APA Holub, Zoriana. (2018). *Devising formal models for automation of ipm detection with the aim of effective management of online communities*. *Management of Development of Complex Systems*, 34, 85 – 91.
- ДСТУ Голуб, З.Д. *Розроблення формальних моделей для автоматизації виявлення інформаційно-психологічної маніпуляції [Текст]* / З.Д. Голуб // *Управління розвитком складних систем*. – 2018. – № 34. – С. 85 – 91.