Національний університет «Львівська політехніка»

Міністерство освіти і науки України

Кваліфікаційна наукова
праця на правах рукопису

ЛАЗУРАК ЗОРЯНА ДМИТРІВНА

УДК 004.773.2

ДИСЕРТАЦІЯ

М Е Т О Д И І З А С О Б И В И Я В Л Е Н Н Я І Н Ф О Р М А Ц І Й Н О - П С И Х О Л О Г І Ч Н О Ї М А Н І П У Л Я Ц І Ї В О Н Л А Й Н - С П І Л Ь Н О Т А Х

10.02.21 Структурна, прикладна і математична лінгвістика

Подається на здобуття наукового ступеня кандидата технічних наук

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____ /Лазурак З.Д./

Науковий керівник Пелещишин Андрій Миколайович, д.т.н., професор

Львів - 2018

АНОТАЦІЯ

Лазурак З. Д. Методи і засоби виявлення інформаційно-психологічної маніпуляції в онлайн-спільнотах. — Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук (доктора філософії) за спеціальністю 10.02.21 «Структурна, прикладна та математична лінгвістика». — Національний університет «Львівська політехніка», Львів, 2018.

У дисертаційній роботі розв’язано важливе наукове завдання — розроблення методів і засобів виявлення інформаційно-психологічної маніпуляції в онлайн-спільнотах, що дало змогу вчасно виявляти прецеденти ІПМ, запобігати розповсюдженню ІПМ, підвищило якість комунікації в онлайн-спільнотах та кількість учасників онлайн-спільноти.

У першому розділі «Аналіз онлайн-спільнот як платформ для здійснення ІПМ» проведено аналіз онлайн-спільноти як платформи для здійснення ІПМ, а саме проаналізовано: типову структуру онлайн-спільноти, текстові види спілкування та їхні характеристики. Розглянуто особливості використання вербальних, невербальних та паравербальних засобів в онлайн-спільнотах. Розглянуто підходи до класифікації онлайн-спільнот та на основі класифікацій окреслено область дослідження дисертаційної роботи: відкриті онлайн-спільноти з інтерактивною текстовою комунікацією.

Проведено порівняльний аналіз традиційної маніпуляції та ІПМ. На основі виявлених спільних ознак розглянуто можливості адаптування існуючих схем тактик маніпуляції до онлайн-комунікації. Детально розглянуто специфічні види невербальних та паравербальних засобів комунікації в онлайн-спільнотах, які відповідають невербальним та паравербальним маркерам маніпуляції в офлайн-середовищі.

Розглянуто інструментарій тональнісного аналізу як засіб для ідентифікації ознак, необхідних для встановлення психічних станів учасників дискусії. Розглянуто існуючі класифікації емоцій та психічних станів людини.

Проаналізовано онлайн-дискусію як діалог та розглянуто можливість використання діалогічних актів як одного з видів маркерів для виявлення ІПМ в онлайн-спільнотах.

У другому розділі «Формальні моделі онлайн-спільноти, інформаційно-психологічної маніпуляції та фільтрів для виявлення підозрілих дискусій» представлено розроблені формальні моделі: онлайн-спільноти, тактики ІПМ, фільтрів для виявлення підозрілих фрагментів дискусії та описано структуру семантичних змінних. Кожна з формальних моделей містить характеристики, необхідні для виявлення ІПМ в онлайн-спільнотах.

Формальна модель онлайн-спільноти складається з інформаційного наповнення та учасників онлайн-спільноти. Ці два елементи є важливими з погляду виявлення ІПМ і вимагають різних підходів до аналізу та використання з метою виявлення ІПМ. Інформаційне наповнення складається з дискусії, а ті, своєю чергою, з повідомлень.

Повідомлення, крім характеристик, які дають змогу його однозначно ідентифікувати, та інших обов’язкових характеристик (автор повідомлення, дата й час розміщення повідомлення, тип повідомлення за спрямуванням та показник релевантності повідомлення), мають також базові обов’язкові характеристики та додаткові характеристики, необхідні для виявлення ІПМ.

Для формального представлення тактики ІПМ, як послідовної зміни станів реципієнта внаслідок зовнішніх впливів, використано кусково-лінійний агрегат. Кусково-лінійний агрегат у кожний часовий момент характеризується одним із внутрішніх станів, які належать до множини внутрішніх станів. Агрегат сприймає вхідні сигнали та змінює стан залежно від сигналу і відповідно до функції переходів. Особливістю кусково-лінійного агрегату є те,

що множини станів і вхідних сигналів конкретизуються за допомогою векторів параметрів.

Формальна модель фільтрів для виявлення підозрілих фрагментів дискусії розроблена на основі формалізованих статичних та динамічних критеріїв наявності ІПМ. Фільтр дає змогу окреслити область потенційної наявності ІПМ і, таким чином, підвищує ефективність виявлення ІПМ.

Критерії, які свідчать про потенційну наявність ІПМ, поділяються на два темпоральні види: динамічні і статичні. В основу такої класифікації покладено часовий період, необхідний для збору інформації задля розрахунку критерію. Статичні критерії – це критерії, які розраховуються на основі діяльності учасника протягом установленого періоду часу. Динамічні критерії – це критерії, на основі яких можна робити висновки зразу ж після їхньої ідентифікації, які не потребують спостережень протягом певного періоду часу.

За допомогою статичних критеріїв розглядають інформаційну діяльність учасника в проекції на структуру онлайн-спільноти, тобто відносно трьох рівнів організації інформаційного наповнення спільноти. Відповідно до формальної моделі онлайн-спільноти цими трьома рівнями є рівень спільноти, дискусії та повідомлення. Критерії цих трьох організаційно-структурних рівнів відрізняються механізмом розрахунку та значимістю.

Поставивши за мету підвищити ефективність алгоритму виявлення ІПМ в онлайн-спільнотах, вдалось уникнути перебору всього інформаційного наповнення дискусії. Цього досягнуто за рахунок окреслення областей дискусії, які мають найбільше ознак наявності ІПМ, тобто виділення підозрілих фрагментів дискусії.

Підозрілі фрагменти дискусії – це набори логічно пов’язаних повідомлень. Кількісні і якісні характеристики цих повідомлень та профілів їх авторів властиві повідомленням з ІПМ.

Виявляючи маніпуляцію, потрібно враховувати той факт, що кожна інтернет-дискусія має свої правила, стиль спілкування та аудиторію, а також

передбачені платформою структуру та засоби спілкування. Від цього залежатиме формування набору фільтрів для виявлення ІПМ у конкретній дискусії, вага критеріїв, на яких ґрунтуються обрані фільтри та порогові значення, які не може перевищувати результат аналізу дискусії за допомогою фільтра.

Формальна модель семантичних змінних дає змогу виявляти ІПМ на основі ознак лексичного рівня. Семантична змінна (СЗ) – це частина вербального наповнення повідомлення, яка представляє певний концепт та варіює залежно від дискурсу, стилю, теми обговорення. СЗ може бути представлена за допомогою одного слова або словосполучень.

СЗ позначають не лексеми, а поняття. У повідомленнях однакові поняття можуть бути представлені за допомогою різних лексем, а різні поняття за допомогою лексико-семантичних варіантів одної лексеми (різних змістових планів багатозначного слова).

У третьому розділі «Методи виявлення інформаційно-психологічної маніпуляції в онлайн-спільнотах» розроблено алгоритм виявлення ІПМ, а також методи, необхідні для виконання етапів алгоритму.

Алгоритм виявлення ІПМ складається з підготовчого етапу, етапу виявлення, етапу нейтралізації та етапу формування результатів і рекомендацій.

Підготовчий етап полягає у пошуку релевантних онлайн-спільнот у www та у Facebook. Це досягнуто за допомогою агрегатора пошукових систем, засобів пошуку у Facebook та видобування інформації з бази даних «Онлайн-спільноти».

Етап виявлення передбачає виконання двох ключових завдань: виділення фрагментів дискусії з найбільшою концентрацією ознак ІПМ та виявлення прецедентів застосування тактик ІПМ у дискусії. Відповідно до цих завдань його поділено на підетапи: виявлення фрагментів дискусії з найбільшою концентрацією ознак ІПМ, виявлення прецеденту ІПМ. Для

виявлення фрагментів дискусії з найбільшою концентрацією ознак ІПМ застосовуємо фільтри для підозрілих фрагментів дискусії. Виявлення прецеденту ІПМ починається з встановлення за допомогою інструментарію тональнісного аналізу емоцій учасників, а на основі останніх визначаємо психічні стани учасників дискусії. Відповідно до інформації про стани та послідовність їхньої зміни, робимо припущення щодо тактики ІПМ у дискусії. Припущення перевіряють за рахунок пошуку у дискусії прийомів ІПМ, які можуть переводити учасника у відповідні стани. Якщо прийоми виявлені, то припущення щодо тактики ІПМ підтверджено.

Етап нейтралізації полягає у виявленні груп профілів, з яких відбувається маніпулятивна діяльність, на основі визначення характеристик мовного стилю, поведінки та профілю маніпулятора та порівняння їх із відповідними характеристиками учасника. Крім того, на цьому етапі встановлюються можливі шляхи поширення ІПМ на основі побудованого соціального графа онлайн-спільноти та визначення його метрик.

Етап формування результатів та рекомендацій полягає у поданні результатів моніторингу, а також рекомендацій щодо нейтралізації актуальних прецедентів ІПМ та запобігання майбутнім ІПМ.

У четвертому розділі «Розроблення програмно-алгоритмічного комплексу виявлення ІПМ» побудовано програмно-алгоритмічний комплекс виявлення ІПМ на основі розроблених формальних моделей онлайн-спільноти і тактики ІПМ, а також алгоритму виявлення ІПМ в онлайн-спільнотах. Програмно-алгоритмічний комплекс розроблено для виявлення ІПМ у певній онлайн-спільноті або щодо певної організації в онлайн-спільнотах. Він передбачає виконання завдань виявлення ІПМ із різним ступенем деталізації умов.

Програмно-алгоритмічний комплекс дає змогу:

- задати параметри завдання з виявлення ІПМ;

- відстежити результати проміжних етапів та налаштувати дії і параметри, необхідні для виконання дій наступних етапів відповідно до отриманих проміжних результатів;
- уточнити параметри, необхідні для виявлення ІПМ відповідно до особливостей комунікацій та структури певної спільноти;
- подати кілька варіантів результатів роботи алгоритму, які відрізняються рівнем деталізації певних ознак чи проміжних етапів.

Споживачами комплексу є відділи інформаційної діяльність організації та адміністративна ланка онлайн-спільнот (адміністратори, модератор, контент-менеджери і т.д.).

Ефективність та результативність програмно-алгоритмічного комплексу виявлення ІПМ оцінюємо за швидкістю та точність виявлення прецедентів ІПМ. Крім того, використовуємо опосердковані показники, а саме: показник позитивного розвитку онлайн-спільноти та показник захищеності від ІПМ. Показники позитивної динаміки взято до уваги за умови відсутності інших заходів для розвитку спільноти. Показники рівня захисту від ІПМ — за відсутності необґрунтованих діяльністю об’єкта деструктивних змін інформаційного образу.

Важливим фактором у розробленні програмно-алгоритмічного комплексу з виявлення ІПМ в онлайн-спільнотах є його апробація для онлайн-спільнот з інтерактивною текстовою комунікацією та з метою виявлення ІПМ з встановленою тематикою. Зокрема впровадження результатів відбулось при виявленні ІПМ у онлайн-спільноті «Молодіжного націоналістичного конгресу (МНК) – Львів» та при виявленні ІПМ щодо діяльності Львівської молодіжної крайової скаутської організації «Білі Горвати».

Ключові слова: онлайн-спільнота, інформаційно-психологічна маніпуляція, тактика інформаційно-психологічної маніпуляції, маркер інформаційно-психологічної маніпуляції.

Список опублікованих праць за темою дисертації

1. Huminskyi R.V., Peleshchyshyn A.M., Holub (Lazurak) Z.D. Suggestions for informational influence on a virtual community // International Journal of Computer Science and Business Informatics. 2015. Vol. 15, No. 1. P.47-65. Available at: http://ijcsbi.org/index.php/ijcsbi/article/view/512/147

2. Holub (Lazurak) Z. The algorithm for detecting online discussion fragments containing information and psychological manipulation // Regional interuniversity compendium of scientific works «System technologies» [Системные технологии]. Dnipro, 2017. № 6 (113). P. 85-91.

3. Голуб (Лазурак) З.Д. Формалізація прийомів інформаційно-психологічної маніпуляції // Вісник Національного технічного університету «ХПІ». Збірник наукових праць. Серія: Нові рішення в сучасних технологіях. Харків: НТУ «ХПІ», 2017. № 32 (1254). С. 55-61.

4. Голуб (Лазурак) З.Д. Система критеріїв для виявлення фрагментів онлайн-дискусій з підозрою на наявність інформаційно-психологічної маніпуляції // Вісник Національного технічного університету «ХПІ». Збірник наукових праць. Серія: Нові рішення в сучасних технологіях. Харків: НТУ «ХПІ», 2018. 9 (1285). С. 106-111.

5. Голуб (Лазурак) З.Д. Структура словника маркерів лексичних змінних для виявлення інформаційно-психологічних маніпуляцій // Вісник Хмельницького національного університету. Серія: Технічні науки. 2017. № 2 (259). С. 264-268.

6. Pełeszczyszyn A., Holub (Lazurak) Z. Development of the system for detecting manipulation in online discussions // Advances in Intelligent Systems and Computing (AISC). 2017. Vol. 543. P. 111-117. Available at: https://link.springer.com/chapter/10.1007/978-3-319-48923-0_15

7. Голуб (Лазурак) З.Д. Розроблення алгоритму виявлення шкідливих інформаційно-психологічних маніпуляцій в онлайн-спільнотах ВНЗ //

Інформатизація вищого навчального закладу: Вісник Національного університету «Львівська політехніка». Львів, 2017. № 879. С. 33-41.

8. Голуб (Лазурак) З.Д. Розроблення формальних моделей для автоматизації виявлення інформаційно-психологічної маніпуляції // Управління розвитком складних систем: зб. наук. пр. / Київський нац. університет будівництва і архітектури. Вип. 34. Київ, 2018. С. 85-91.

9. Peleschyshyn A., Holub I., Holub (Lazurak) Z. Methods of real-time detecting manipulation in online communities // Proceedings of the XIth International Scientific and Technical Conference on Computer Science and Information Technologies (CSIT 2016). Lviv Polytechnic Publishing House, 2016. P. 15-17.

10. Peleschyshyn A., Holub I., Holub (Lazurak) Z. Formal model and key features of an online community fundamental for detecting informational and psychological manipulation // Proceedings of the XIIth International Scientific and Technical Conference on Computer Science and Information Technologies (CSIT 2017). Lviv Polytechnic Publishing House, 2017. P. 101-104.

11. Peleschyshyn A., Holub I., Holub (Lazurak) Z. The preliminary stage of the algorithm for detecting information and psychological manipulation in online communities // Proceedings of the XIIIth International Scientific and Technical Conference on Computer Science and Information Technologies (CSIT 2018). Lviv Polytechnic Publishing House, 2018. P. 30-33.

12. Korzh R., Peleschyshyn A., Holub (Lazurak) Z., Analysis of integrity and coverage completeness of the informational image of a higher education institution // Proceedings of the XIIIth International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET`2016), Lviv-Slavske. Lviv, 2016. P. 825-827.

13. Пелещицин А., Голуб (Лазурак) З. Виявлення маніпуляції щодо потенцій-них покупців в онлайн спільноті // Матеріали Міжнародної науково-практичної конференції «Інформаційне суспільство: тенденції регіонального розвитку» (ISRDT-2016). Львів: Редакція ”УП” , 2016. С. 58-59.

14. Голуб (Лазурак) З.Д. Огляд класифікацій онлайн спільнот // Інформаційна діяльність, документознавство, бібліотекознавство: історія, сучасність, перспективи : матеріали III Всеукр. наук.-практ. конф., Київ, 25–26 квіт. 2017 р. Київ: НАКККіМ, 2017. С. 17-19.

15. Голуб (Лазурак) З., Пелещишин А. Моделювання маніпулятивної тактики за допомогою кусково-лінійного агрегата // Матеріали 5-ї міжнародної науково-практичної конференції «Інформація, комунікація, суспільство − 2016». Львів, 2016. С. 80-81.

16. Голуб (Лазурак) З., Пелещишин А. Виявлення прийомів ІПМ реалізованих за допомогою посилань // Матеріали 6-ї міжнародної наукової конференції «Інформація, комунікація, суспільство − 2017». – Львів, 2017. – С. 65-66.

17. Голуб (Лазурак) З., Пелещишин А. Види спілкування в онлайн-спільнотах та їхні характеристики // Матеріали 7-ї міжнародної наукової конференції «Інформація, комунікація, суспільство − 2018». – Чинадієво, 2018. – С. 45-46.

ABSTRACT

Lazurak Z.D. Methods and means for detecting information and psychological manipulation in online communities. — Qualification scientific work on the rights of manuscript.

Thesis for a Ph.D. degree in specialty 10.02.21 — structural, applied and mathematical linguistics. — Lviv Polytechnic National University Ministry of Education and Science of Ukraine, L'viv, 2019.

In the thesis, the important scientific task of the development of methods and means for detecting information and psychological manipulation (IPM) in online communities is solved. The methods and means are aiming at identifying IPM precedents in a timely manner in order to prevent dissemination of IPM, to increase the quality of communication in online communities and the number of participants in an online community.

The first chapter, "Analysis of online communities as platforms for IPM" considers an online community as a platform for IPM. The typical structure of an online community and its characteristic features, as well as text communication, are scrutinized. Peculiarities of using verbal, nonverbal and paraverbal means in online communities are considered. Approaches to the classification of online communities are considered. On the basis of the classification the research area of the thesis is defined, namely, public online communities with interactive text communication.

The comparative analysis of traditional manipulation and IPM is conducted. On the basis of the analysis results, the possibilities of adapting existing tactics of manipulation to online communication are considered. The specific types of nonverbal and paraverbal communication means that correspond to verbal, non-verbal and paraverbal manipulation markers in an offline communication are considered in detail.

Sentiment analysis tools are considered as means for identifying features necessary for provoking psychological states of the discussion participants. Existing classifications of emotions and psychological states are considered.

The analysis of an online discussion as a dialogue is conducted and the possibility of using dialog acts as one of the types of markers for identifying IPM in online communities is considered.

The second chapter, "Formal models of online communities, information and psychological manipulation and filters for detecting suspicious discussions" presents the formal models of an online community, an IPM tactic, filters for detecting suspicious discussion fragments, and semantic variables (SV). Each of the formal models contains the characteristics required to analyze the online community in order to detect IPM.

The formal model of an online community consists of two elements: content and members of an online community. These two elements are important from the point of view of identifying IPM and require different approaches to utilizing them for the purpose of identifying the IPM. The content consists of discussions, messages in their turn constitute a discussion.

A message is described with mandatory characteristics (author of the message, posting date and time, direction type of a message, and relevance indicator of the message). Besides, a message is characterized by the basic characteristics and additional characteristics required for IPM detection.

The piecewise linear aggregate is utilized for the formal representation of IPM tactics. The piecewise linear aggregate depicts consequent changes in recipient states due to external influences. The piecewise linear aggregate is characterized by one of the internal states belonging to the set of internal states at each time moment. The piecewise linear aggregate receives input signals and changes the state depending on the signal and according to the transitions function. The peculiarity of the piecewise linear aggregate is that the sets of states and input signals are specified using the parameter vectors.

The filter for detecting suspicious fragments of the discussion is developed on the basis of static and dynamic criteria for the presence of IPM. The filter is used to

outline the area of the potential presence of IPM and, in this way increases the effectiveness of detecting IPM.

The criteria that indicate the potential presence of IPM are divided into two temporal types, dynamic and static. The basis of this classification is the time period required for collecting information needed to calculate the criterion. Static criteria are calculated on the basis of the participant's activity over a fixed period of time. Dynamic criteria are the criteria which require observation over a period of time.

Static criteria consider participant's information activity in relation to the three levels of organization of the content in a community. According to the formal model of an online community, these three levels are the level of community, discussion, and message. The criteria of the three organizational and structural levels differ in significance and the mechanism of calculation.

Aiming to improve the effectiveness of the IPM detection algorithm, the discussion fragments that have the greatest number of identification features of IPM are detected.Suspicious fragments of the discussion are sets of logically related messages, the quantitative and qualitative characteristics of which and the quantitative and qualitative characteristics of the profiles of the authors of these messages are characteristic for messages containing IPM.

When detecting manipulation, it should be taken into account that every discussion has its own rules, style of communication and audience, as well as the structure and means of communication provided by the platform. This affects the formation of the set of filters for identifying IPM in a particular discussion.

The formal model of a semantic variable is used to detect IPM based on the lexical features. A semantic variable (SV) is a part of the verbal content of a message that represents a certain concept and varies depending on the discourse, style, and topic of discussion. SV can be represented by a single word or phrases.

SV denotes not lexemes, but concepts. In messages, identical concepts can be represented using different lexemes, and different concepts by means of lexical-semantic variants of the same lexeme.

In the third section, "Methods for detecting information and psychological manipulation in Online Communities", an algorithm for detecting IPM is developed, as well as methods used at the algorithm stages.

The IPM detection algorithm consists of a preparatory phase, a stage of detection, a stage of neutralization and a stage of formation of results and recommendations.

The preparatory stage is aimed at finding relevant online communities. This is accomplished by means of the search engines aggregator, Facebook search tools and extracting information from the "Online Community" database, which can be used as a kernel for the search.

The stage of detection is aimed at identifying fragments of online discussions that are characterized by the presence of the IPM features. Sentiment analysis tools are used to identify emotions of discussion participants. On the basis of the latter psychological states are defined The sequence of psychological states is used to assume that there is IPM in the discussion. The assumption is verified by finding an IPM tool that provokes the detected psychological stages.

The phase of neutralization lies in identifying groups of profiles, which are used to conduct IPM. This is done by determining the characteristics of the language style, behavior, and profile of the manipulator, and comparing them with the corresponding characteristics of other participants. In addition, at this stage, possible ways of IPM are determined. This is one by means of the social graph of an online community.

At the stage of the formation of the results and recommendations monitoring results are presented, as well as recommendations for the neutralization of detected IPM precedents and the prevention of IPMs are suggested.

In the fourth section "Development of the software and algorithmic complex for IPM detection", the construction of the software and algorithmic complex for IPM detection on the basis of the developed formal models of an online community and IPM tactics, as well as the algorithm for detecting IPM in online communities is

described. The software and algorithmic complex is designed to detect IPM in a particular online community or for a specific organization in discussions of various online communities. The software and algorithmic complex involves performing tasks for detecting IPM with varying degrees of conditions detailing.

The software and algorithmic complex provides the following opportunities:

- to set parameters for the IPM detection task and
- to track the results of the interim stages and to set necessary actions and parameters in accordance with the obtained interim results;
- to adjust parameters necessary for identifying the IPM in accordance with the peculiarities of communication and the structure of a particular community;
- to choose the way of presenting the results of the algorithm, which differ in the level of detail of certain features or intermediate stages.

Consumers of the complex are the departments of information activities of the organization and the administration of online communities (administrators, moderators, content managers, etc.).The architecture of the software and algorithmic complex of IPM detection is based on the methods and means described in the thesis. The complex includes the database.

The efficiency of the IPM detection software and algorithmic complex are evaluated by the speed and accuracy of detecting IPM precedents. In addition, the indicator of the positive development of the online community and the IPM protection score are used to assess results.

The efficiency of the complex is assessed by using it for detecting IPM in the online community of the “Youth Nationalist Congress–Lviv” and detecting IPM directed on the activities of the Lviv Regional Youth Scout Organization “White Horvats”.

Key words: online community, information and psychological manipulation, model of online community, IPM tactics, marker of information and psychological manipulation, software and algorithmic complex.

List of publications by the subject of dissertation

1. Huminskyi R.V., Peleshchyshyn A.M., Holub (Lazurak) Z.D. Suggestions for informational influence on a virtual community // International Journal of Computer Science and Business Informatics. 2015. Vol. 15, No. 1. P.47-65. Available at: http://ijcsbi.org/index.php/ijcsbi/article/view/512/147

2. Holub (Lazurak) Z. The algorithm for detecting online discussion fragments containing information and psychological manipulation // Regional interuniversity compendium of scientific works «System technologies» [Системные технологии]. Dnipro, 2017. № 6 (113). P. 85-91.

3. Holub (Lazurak) Z. D. Formalization of the tools of information and psychological manipulation // Bulletin of National Technical University «KhPI». Collection of scientific works. Series: New solutions in advanced technologies. Kharkiv: NTU «KhPI», 2017. № 32 (1254). P. 55-61.

4. Holub (Lazurak) Z. D. The system of criteria for detecting fragments of online discussions suspected of information and psychological manipulation presence // Bulletin of National Technical University «KhPI». Collection of scientific works. Series: New solutions in advanced technologies. Kharkiv: NTU «KhPI», 2018. 9 (1285). P. 106-111.

5. Holub (Lazurak) Z. D. The structure of the dictionary of lexical variable markers for detection of information and psychological manipulation // "Bulletin of Khmelnytskyi national university. Series: Technical science. 2017. № 2 (259). P. 264-268.

6. Pełeszczyszyn A., Holub (Lazurak) Z. Development of the system for detecting manipulation in online discussions // Advances in Intelligent Systems and Computing (AISC). 2017. Vol. 543. P. 111-117. Available at: https://link.springer.com/chapter/10.1007/978-3-319-48923-0_15

7. Holub (Lazurak) Z. D. Development of the algorithm for detecting malicious information and psychological manipulation in online comunities // Informatization

of a higher education institution». Bulletin of Lviv Polytechnic National Unversity. Lviv, 2017. № 879. P. 33-41.

8. Holub (Lazurak) Z. D. Development of formal models for automation of information and psychological manipulation detection // Management of the development of sofisticated systems: collection of scientific works. / Kyiv Ntional University of consctructions and architecture. Vol. 34. Kyiv, 2018. P. 85-91.

9. Peleschyshyn A., Holub I., Holub (Lazurak) Z. Methods of real-time detecting manipulation in online communities // Proceedings of the XIth International Scientific and Technical Conference on Computer Science and Information Technologies (CSIT 2016). Lviv Polytechnic Publishing House, 2016. P. 15-17.

10. Peleschyshyn A., Holub I., Holub (Lazurak) Z. Formal model and key features of an online community fundamental for detecting informational and psychological manipulation // Proceedings of the XIIth International Scientific and Technical Conference on Computer Science and Information Technologies (CSIT 2017). Lviv Polytechnic Publishing House, 2017. P. 101-104.

11. Peleschyshyn A., Holub I., Holub (Lazurak) Z. The preliminary stage of the algorithm for detecting information and psychological manipulation in online communities // Proceedings of the XIIIth International Scientific and Technical Conference on Computer Science and Information Technologies (CSIT 2018). Lviv Polytechnic Publishing House, 2018. P. 30-33.

12. Korzh R., Peleschyshyn A., Holub (Lazurak) Z., Analysis of integrity and coverage completeness of the informational image of a higher education institution // Proceedings of the XIIIth International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET`2016), Lviv-Slavske. Lviv, 2016. P. 825-827.

13. Peleschyshyn A., Holub (Lazurak) Z. Detection of manipulation directed at potential buyers in online communities // Proceedings of International scientific and practical conference “Information society: tendencies of regional development” (ISRDT-2016). Lviv: Publishing house ”UP”, 2016. P. 58-59.

14. . Holub (Lazurak) Z. D. Review of online communities classification // Informational activities, document science, library science: history, modernity, perspectives : Proceedings III All-Ukrainian, scientific and practical conference, Kyiv, 25–26 April 2017. Kyiv : [NAAAKiM], 2017. P. 17-19.

15. Holub (Lazurak) Z., Peleschyshyn A. Modeling of manipulation tactic by means of piece-wise linear aggregate // Proceedings of the 5th international scientific and practical conference «Information, communication, sociaty − 2016». Lviv-Slavske, 2016. – P. 80-81.

16. Holub (Lazurak) Z., Peleschyshyn A. Detecting IPM tools realized by means of links // Proceedings of the 5th international scientific and practical conference «Information, communication, sociaty − 2017». Lviv-Slavske, 2017. – P. 65-66.

17. Holub (Lazurak) Z., Peleschyshyn A. Types of communication in online communities and their pecularities // Proceedings of the 5th international scientific and practical conference «Information, communication, sociaty − 2018». Chynadievo, 2018. – P. 45-46

Зміст

Перелік умовних скорочень

Скорочення, термін, позначення	Пояснення
ІПМ	Інформаційно-психологічна маніпуляція
ДА	Діалогічний акт
СЗ	Семантична змінна
БД	База даних

Вступ

Актуальність теми. Із перенесенням спілкування з традиційного офлайн в онлайн-середовище негативні явища комунікації набули нових форм. Зокрема, маніпуляція – вид психічного впливу, вміле здійснення якого спричиняє приховане збудження в іншої людини намірів, які не збігаються з її актуальними бажаннями. Оскільки спілкування в онлайн-спільнотах відбувається за допомогою засобів та механізмів, відмінних від спілкування в офлайн-середовищі, для прихованого впливу на процеси прийняття рішення, формування світоглядної позиції та дії, до учасників онлайн-спільнот застосовують інформаційно-психологічну маніпуляцію. 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.

ІПМ має негативні наслідки для учасників спільноти, обговорюваних об’єктів дійсності (установи, бренду, іміджу людини) та на онлайн-спільноту загалом. Наслідком ІПМ є несвідомі дії або зміна світоглядної позиції учасників онлайн-спільноти, ушкодження позитивного іміджу певного об’єкта або відхід учасників і смерть онлайн-спільноти. Тому відповідальні за інформаційну діяльність організації та адміністративна ланка онлайн-спільнот (адміністратори, модератори) потребують методи і засоби виявлення ІПМ для попередження і протидії переліченим вище загрозам.

Згідно з науковими працями з галузі психології нейтралізація маніпуляції полягає у доведенні факту маніпуляції. Реципієнт, який не володіє експертними знаннями про ІПМ, не може зробити це самостійно, оскільки до нього застосовують прихований вплив на підсвідомість. З погляду онлайн-спільнот це означає, що учасники дискусії мають бути поінформовані про прецедент ІПМ, зокрема про це має бути повідомлено у дискусіях, в яких виявлено ІПМ та у дискусіях, в яких жертви ІПМ беруть активну участь.

Аналізу комунікації в онлайн-спільнотах, зокрема виявленню негативних явищ у комунікації, присвячено багато досліджень, серед них: поширення неправдивої інформації в соціальних медіа досліджував S. Kumar, інформаційний вплив у соціальних мережах вивчав E. R. Smith. Виявленням інформаційних загроз віртуальних спільнот в інтернет-середовищі соціальних мереж займався Р. В. Гумінський. В. П. Горбулін та Д. В. Ланде досліджували інформаційні операції з погляду безпеки суспільства, а саме аналізували інформаційні операції та моделі. Аналізу достовірності соціально-демографічних характеристик учасників віртуальних спільнот та ідентифікації фейкових та неправдивих акаунтів присвячено праці С. С. Федушко та K. Dutta. Дослідженням типової і атипової поведінки учасників онлайн-спільнот займався C. Chen. Виявленням сарказму у діалогах онлайн-спільнот займались L. Eisenberg та S. Oraby. Виявлення згоди і незгоди у діалогах соціальних мереж досліджували A. Misra та M. A. Walker, а виявлення обману у синхронній інтернет-комунікації – D. P. Twitchell.

Результати перелічених вище досліджень можуть бути використані для виконання проміжних завдань, необхідних для виявлення ІПМ в онлайн-спільнотах, наприклад, виявлення підозрілих та впливових профілів або певних мовних явищ у дискусіях. Ці дослідження проводили для онлайн-спільнот, розміщених на конкретних платформах, наприклад, Twitter, для певного виду форумів, наприклад, форумів онлайн-ігор чи для онлайн-спільнот, структура яких не передбачає рівномірної розподіленості комунікації між учасниками або в яких текстове подання інформації не є домінантним. Тому більшість методів, запропонованих у цих дослідженнях, вирішують вузько спеціалізовані завдання і є не ефективними з погляду виявлення ІПМ в онлайн-спільнотах.

Актуальною є розв’язання науково-прикладного завдання: розроблення науково обґрунтованих методів і засобів виявлення ІПМ, які враховують

специфіку текстової комунікації в онлайн-спільнотах із домінантною текстовою формою подачі інформації.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконана в межах зареєстрованої тематики кафедри соціальних комунікацій та інформаційної діяльності «Лінгвістичне забезпечення консолідації відкритих інформаційних ресурсів» (номер державної реєстрації 0113U005274).

Мета і завдання дослідження. Мета дисертаційної роботи – підвищити рівень захисту від деструктивних інформаційних впливів суб'єктів та об'єктів спілкування в онлайн-спільнотах за допомогою розроблення методів та засобів виявлення ІПМ.

Мета дисертаційної роботи передбачає виконання таких завдань:

- аналіз характерних особливостей та відмінностей онлайн-спільнот від традиційного офлайн-середовища спілкування та від інших інтернет-засобів комунікації та огляд існуючих засобів, необхідних для опрацювання інформаційного наповнення онлайн-спільноти;
- розроблення формальної моделі ІПМ з урахуванням ознак ІПМ в онлайн-спільнотах;
- розроблення системи маркерів прийомів ІПМ для інтерактивної текстової комунікації в онлайн-спільнотах;
- розроблення системи фільтрів для виділення підозрілих фрагментів ІПМ у дискусіях онлайн-спільнот;
- розроблення алгоритму виявлення ІПМ в онлайн-спільнотах та опис методів застосованих на кожному з етапів алгоритму виявлення ІПМ в онлайн-спільнотах;
- розроблення методів нейтралізації ІПМ;
- реалізація програмно-алгоритмічного комплексу виявлення ІПМ в онлайн-спільнотах на основі описаних у роботі методів і алгоритмів.

Об'єктом дослідження є комунікативні процеси вразливі до ІПМ в онлайн-спільнотах із домінантним текстовим поданням інформації.

Предметом дослідження є моделі ІПМ, методи і засоби виявлення ІПМ в інтерактивному текстовому інформаційному наповненні онлайн-спільнот.

Методи дослідження. Для розроблення формальних моделей онлайн-спільноти, ІПМ та фільтрів для виявлення підозрілих фрагментів дискусії використано теоретико-множинні підходи, а також для побудови формальної моделі ІПМ використано теорію кусково-лінійних агрегатів. У розробленні системи маркерів ІПМ використано результати досліджень у галузі контент-аналізу, зокрема метод анотування за допомогою діалогічних актів. Для формалізації маркерів ІПМ, а саме для представлення синтаксичних структур, характерних для певних прийомів, використано дерева залежності. Для побудови соціального-графа спільноти та визначення можливих шляхів поширення ІПМ використано інструментарій аналізу соціальних спільнот та теорію графів. Для моделювання баз даних, необхідних для алгоритму виявлення ІПМ використано апарат баз даних, зокрема інфологічну модель «сутність-співвідношення», її подано у нотації Баркера.

Наукова новизна одержаних результатів полягає в науковому обґрунтуванні та побудові методів та засобів виявлення ІПМ в онлайн-спільнотах. Отримано такі результати:

1. вперше побудовано формальну модель тактики ІПМ на основі кусково-лінійних агрегатів, що дало змогу виявляти тактику ІПМ в онлайн-спільнотах на основі зміни психологічних станів учасників;
2. вперше розроблено систему маркерів прийомів ІПМ на основі діалогічних актів, семантичних змінних та синтаксичних структур, що дало змогу виявляти прийоми ІПМ в онлайн-спільноті без експертного аналізу;
3. отримали подальший розвиток характеристики мовного стилю, поведінки та профілю учасників онлайн-спільноти, а також подання онлайн-спільноти у

вигляді соціального графа, що дає змогу виявляти профілі-спільники та визначати можливі шляхи поширення ІПМ.

Практичне значення одержаних результатів полягає у забезпеченні можливості виявлення прецедентів ІПМ в онлайн-спільнотах без експертного аналізу та прогнозуванні можливих шляхів поширення ІПМ. Зокрема практично цінними є такі результати:

1. розроблено систему фільтрів, де застосування наступних фільтрів залежить від результатів попередніх на основі запропонованих у роботі критеріїв, що дає змогу збільшити ефективність та швидкість пошуку підозрілого фрагмента дискусії;;
2. розроблено алгоритм виявлення прецедентів ІПМ, який заснований на формальній моделі тактики ІПМ, заданій за допомогою кусково-лінійного агрегату та передбачає верифікацію виявленої тактики, на основі лінгвістичних маркерів прийому ІПМ, що дає змогу збільшити точність результатів виконання алгоритму;
3. розроблено програмно-алгоритмічний комплекс виявлення ІПМ в онлайн-спільнотах, який дає змогу аналізувати великий обсяг інформаційного наповнення онлайн-спільнот та оперативно виявляти ІПМ.

Розроблений алгоритм виявлення ІПМ в онлайн-спільнотах та методи реалізації окремих його етапів використано для виявлення ІПМ у онлайн-спільноті «Молодіжного Націоналістичного Конгресу (МНК) – Львів» та для виявлення ІПМ щодо діяльності Львівської молодіжної крайової скаутської організації «Білі Горвати».

На основі виконаних досліджень здобувач розробила методичне забезпечення, яке використовують у навчальному процесі Національного університету «Львівська політехніка», а саме методичне забезпечення до виконання лабораторних робіт із курсу «Технології інформаційного пошуку» для студентів, що навчаються за спеціальністю 029 «Інформаційна, бібліотечна та архівна справа».

Особистий внесок здобувача. Усі наукові результати дисертаційної роботи отримані автором самостійно. У друкованих працях, опублікованих у співавторстві, авторові належать: [1] — опис видів спілкування в онлайн-спільнотах; [3] — постановка та аналіз проблеми інформаційних впливів у спільнотах, характеристики вагомих з погляду інформаційного впливу дискусій та учасників віртуальної спільноти; [79] — загальна схема системи виявлення маніпуляції в онлайн-спільнотах; [83] —методи виявлення інформаційно-психологічної маніпуляції в онлайн-спільнотах; [84] — формальна модель онлайн-спільноти з урахування характеристик необхідних для виявлення ІПМ; [92] — методи пошуку релевантних онлайн-спільнот за допомогою глобальних пошукових систем та засобами Facebook; [93] — методи виявлення ІПМ за допомогою посилань; [113]— опис маркерів ІПМ щодо потенційних покупців; [114]— розроблення формальної моделі тактики ІПМ; [115] — розроблення ознак зміни інформаційного образу внаслідок зовнішніх впливів.

Апробація результатів дисертації. Основні результати дисертаційного дослідження неодноразово доповідалися на міжнародних та всеукраїнських наукових конференціях, зокрема на: 5–7 Міжнародних наукових конференціях «Інформація, комунікація, суспільство» (Львів, 2016–2018); XIII Міжнародній конференції «Сучасні проблеми радіоелектроніки, телекомунікацій, комп'ютерної інженерії» TCSET'2016 (Львів, 2016); XI–XIII Міжнародних науково-технічних конференціях «Комп’ютерні науки та інформаційні технології» (Львів, 2016–2018); III Всеукраїнській науково-практичній конференції «Інформаційна діяльність, документознавство, бібліотекознавство: історія, сучасність, перспективи» (Київ, 2017); Міжнародній науково-практичній конференції «Інформаційне суспільство: тенденції регіонального розвитку (Львів, 2016). Результати дисертаційних досліджень регулярно доповідалися на наукових семінарах кафедри соціальних комунікацій та інформаційної діяльності Національного університету «Львівська політехніка» (2016–2018).

Публікації. За результатами виконаних досліджень опубліковано 17 наукових праць з них: 1 стаття у закордонному науковому періодичному виданні, 6 статей у наукових фахових виданнях України, 10 публікацій у формі матеріалів і тез доповідей наукових конференцій (з них 5 публікацій у виданнях, що входять до наукометричної бази даних Scopus).

Розділ 1. Аналіз онлайн-спільнот, інструментарію та підходів до виявлення ІПМ

У першому розділі наведено загальний аналіз онлайн-спільнот, інструментарію для аналізу текстового інформаційного наповнення онлайн-спільнот, підходів до виявлення ІПМ. Проведено огляд існуючих підходів до класифікації онлайн-спільнот, на основі розглянутих класифікації окреслено область дослідження дисертаційної роботи, тобто онлайн-спільноти з інтерактивною текстовою комунікацією. Проаналізовано онлайн-спільноту як платформу для здійснення ІПМ, а саме: типову структуру онлайн-спільноти, текстові види спілкування та їхні характеристики. Розглянуто особливості використання вербальних, невербальних та паравербальних засобів в онлайн-спільнотах.

Проведено порівняльний аналіз офлайн-маніпуляції та ІПМ. На основі виявлених спільних ознак розглянуто можливості адаптування існуючих схем тактик ІПМ до онлайн-комунікації. Детально розглянуто специфічні види невербальних та паравербальних засобів комунікації в онлайн-спільнотах, які відповідають невербальним та паравербальним маркерам маніпуляції в офлайн-середовищі.

Розглянуто інструментарій тональнісного аналізу як засіб для ідентифікації ознак необхідних для встановлення психічних станів учасників дискусії. Розглянуто існуючі класифікації емоцій та психічних станів людини.

Проаналізовано онлайн-дискусію як діалог та розглянуто можливість використання діалогічних актів як одного з видів маркерів для виявлення ІПМ в онлайн-спільнотах.

Основні результати розділу опубліковані автором у працях [1, 2, 3].

1.1. Аналіз онлайн-спільнот як платформ для здійснення ІПМ

Масштаби аудиторії, доступність, обсяг та швидкість оновлення інформації зробили онлайн-спільноти зручними платформами для ІПМ. Онлайн-спільноти є дуже ефективними з погляду поширення та отримання інформації, тому їх використовують для досягнення широкого спектру інформаційних завдань. За допомогою онлайн-спільноти можна побудувати надійну мережу підтримки, вона є джерелом нових ідей і потужним маркетинговим інструментом.

Термін онлайн-спільнота є міждисциплінарним тому немає універсального визначення онлайн-спільноти. Нижче наведено визначення онлайн-спільноти з тлумачного словника та з досліджень зі сфери ІТ.

Онлайн-спільнота — це група людей, які мають спільні інтереси, спільну сферу діяльності, пов’язані спільною метою і впродовж певного часу познайомились один з одним за допомогою веб-технологій [4].

Часто термін «онлайн-спільнота» вживають як синонімічний до «віртуальна спільнота», «мережева спільнота» та «електронна спільнота» [5, 6]. Віртуальна спільнота, онлайн-спільнота — це соціальна група людей, які спілкуються та взаємодіють через Інтернет за допомогою спеціалізованих сервісів та сайтів у WWW [6].

Існують дослідження, які трактують онлайн-спільноту як підвид віртуальної спільноти [7]. Віртуальна спільнота — це група людей, які переважно спілкуються за допомогою засобів інтернету, аніж безпосередньо віч-на-віч. Якщо механізм, який забезпечує спілкування, це комп’ютерна мережа, то ці спільноти називаються онлайн-спільнотами.

У цій дисертаційній роботі використовуємо таке визначення: онлайн-спільнота — це група людей, які регулярно взаємодіють один з одним онлайн, переважно з метою обміну інформацією і думками заради спільного інтересу [8].

1.1.1. Огляд існуючих підходів до класифікації онлайн-спільнот

Спільною ознакою онлайн-спільнот є колективна комунікація через Інтернет, але онлайн-спільноти мають багато відмінностей, спричинених технічними характеристиками, структурною організацією, інтерфейсом, тематикою тощо. Тому онлайн-спільноти відрізняються ступенями вразливості до різних видів ІПМ. Розробляючи методи і засоби виявлення ІПМ, необхідно зважати на ці відмінності. Для того, щоб окреслити область цього дослідження та визначити особливості підходу до виявлення ІПМ в онлайн-спільнотах різних видів, розглянуто існуючі класифікації онлайн-спільнот.

За ступенем інтеграції у веб виділено такі типи онлайн-спільнот [6]:

- соціальні мережі (неінтегровані у веб);
- дискусійні листи (частково інтегровані в веб);
- публічні соціальні мережі (здебільшого інтегровані в веб);
- веб-спільноти (повністю інтегровані в веб).

За медіа-можливостями та соціальною присутністю спільноти поділяють на [9]:

- низькі (блоги та мікроблоги);
- середні (сайти соціальних мереж та спільноти обміну інформаційним наповненням);
- високі (віртуальні соціальні світи, віртуальні світи ігор).

За доступністю онлайн-спільноти поділяють на [6]:

- відкриті (інформаційне наповнення спільноти доступне для усіх);
- закриті (інформаційне наповнення доступне лише для зареєстрованих);
- приховані (вступ у спільноти можливий лише за запрошенням).

За домінантним видом подання інформації онлайн-спільноти можна поділити на:

- текстові (наприклад, Stack Exchange, LiveJournal);
- мультимедійні (наприклад, Youtube, Vimeo, Coub, Imgur);
- графічні (наприклад, Instagram,).

Мета дослідження — виявити ІПМ, які:

- можуть мати велику аудиторію;

- можуть значно поширюватись;
- впливають на дії та світоглядну позицію реципієнтів у реальному світі;
- відбуваються у інтерактивних, динамічних середовищах, що передбачають рівномірну комунікацію всіх суб'єктів;
- відбуваються в середовищах, де переважають текстові засоби передачі інформації.

Тому у цій роботі представлено методи і засоби виявлення ІПМ у спільнотах, які належать до вказаних нижче класів:

За ступенем інтеграції у веб проаналізовано віртуальні спільноти та спільноти, утворені на основі публічних соціальних мереж, наприклад, на основі Facebook, оскільки лише вони можуть мати велику аудиторію та уможливити значне поширення інформації.

За медіа-можливостями і соціальною присутністю — середні онлайн-спільноти. На відміну від віртуальних соціальних світів та світів ігор, у цих спільнотах обговорюють реальні події та сутності, тому наявність у них ІПМ може спричинити несвідомі дії або зміну світогляду реципієнтів. Комунікативну активність у блогах чи мікроблогах зосереджено на одному з суб'єктів спілкування, інші учасники комунікації розміщують повідомлення, інформативні лише в контексті головного тексту блогу.

За доступністю онлайн-спільноти розглядають клас відкритих онлайн-спільнот. На відміну від закритих і прихованих, які не виходять за межі окресленого кола суб'єктів комунікації, користувачі можуть знайти відкриті спільноти в інтернеті та долучитись до них.

Класифікація онлайн-спільнот за спільною ознакою учасників [10]:

- географічні — обмежені географічним розташуванням (місто, район);
- демографічні — обмеження за віком, статтю, расою чи національністю;
- тематичні — учасників пов'язують спільні інтереси, наприклад, захоплення, група дозвілля чи професійна структура.

Цю класифікацію варто брати до уваги, якщо в завданні виявлення ІПМ вказано певні демографічні обмеження цільової аудиторії ІПМ. У такому випадку треба відштовхуватись від спільної узагальненої ознаки спілкування,

щоб знайти релевантні дискусії на підготовчому етапі. Крім того, ІПМ має різну тематику і цільову аудиторію, тому для визначення потенційно зараженої ІПМ спільноти необхідно враховувати тип спільнот за метою та характеристиками учасників. На ці характеристики також можна зважати під час визначення стандартних для спільноти поведінкових характеристик та мовного стилю. Наприклад, школярам не властиво писати науковою термінологією.

1.1.2. Типова структура онлайн-спільноти

Структурні особливості онлайн-спільнот великою мірою визначають перебіг комунікації між учасниками. Оскільки маніпулятори реалізовують ІПМ під час спілкування, то розуміння структури онлайн-спільноти з погляду інтеракції між учасниками є необхідним для визначення закономірностей спілкування в онлайн-спільнотах, вироблення способів формалізації ІПМ та онлайн-спільнот.

З погляду спілкування онлайн-спільнота складається з учасників та змістового наповнення (див. розд. 2.1 «Спеціальна модель онлайн-спільноти з погляду виявлення ІПМ»). Динамічно створюване учасниками інформаційне наповнення спільноти — це дискусія.

Дискусії складаються з повідомлень — короткої текстової інформації, створеної учасником спільноти. Повідомлення є атомарною одиницею інформаційного наповнення Веб-форуму та складається зі заголовка, тексту, дати створення та автора [11].

Маніпулятори здійснюють ІПМ в онлайн-дискусіях, тому встановлення змістового зв’язку між повідомленнями дискусії є необхідним для виявлення ІПМ.

Для встановлення зв’язків між повідомленнями необхідно врахувати спосіб реалізації дискусії. В онлайн-спільнотах дискусії реалізовані двома способами [6]:

- повідомлення учасників відображаються лінійно (послідовно) залежно від часу їхньої публікації;

- деревоподібна (розгалужена) структура відображає не лише часову послідовність появи повідомлень, а і зв’язок повідомлення-реакції та ініціювального повідомлення.

При цьому платформи, на яких розміщені онлайн-спільноти, передбачають певні відмінності у реалізації дискусій. Часто вони обмежують кількість рівнів розгалуженої дискусії, наприклад, для Facebook — це пост, коментар, відповідь, а для StackExchange — за питанням може йти коментар або відповідь і коментар на відповідь.

З огляду на способи реалізації дискусії, зв’язки між повідомленнями можуть бути виявлені явно та неявно. Явні зв’язки встановлюють на основі розгалуженої структури дискусії, наявності в повідомленнях-реакціях посилань на ініціювальні повідомлення чи вкладених цитувань.

Явні зв’язки можна встановити за допомогою:

- розгалуженої структури онлайн-спільноти (див. рис. 1.1);
- наявності в повідомленнях-реакціях посилань на ініціювальні повідомлення (див. рис. 1.3);
- наявності в повідомленнях-реакціях вкладених цитувань ініціювальні повідомлень (див. рис. 1.2).

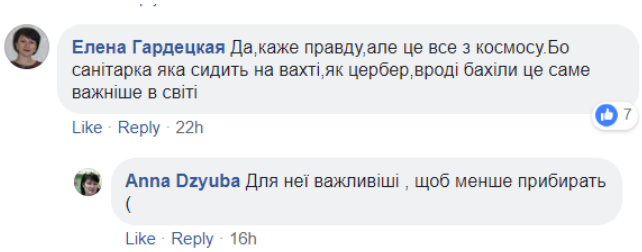


Рис. 1.1. Розгалужена структура дискусії

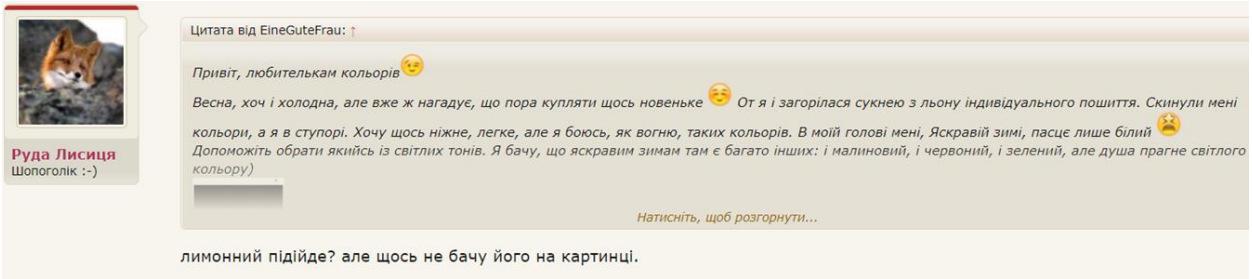


Рис. 1.2. Зв’язок між повідомленнями за допомогою вкладених цитувань

Рис. 1.3. Зв’язок між повідомленнями за допомогою посилань на ініціювальні повідомлення

Неявні зв’язки між повідомленнями виявляють аналізуванням інформаційного наповнення повідомлення:

- наявності імен учасників у заголовку, назві;
- наявності звернень на ім’я учасника в тексті повідомлення;
- повторень або синонімів ключових слів повідомлення [11];
- наявності особових та демонстративних займенників, слів зі значенням порівняння (наприклад, подібний, однаковий, інакший [12])
- наявності еліпсу [13].

Отже, дискусія є ключовим елементом онлайн-спільноти, застосувавши відповідні підходи до аналізу інформаційного наповнення дискусії, можна отримати інформацію необхідну для виконання проміжних завдань алгоритму виявлення ІПМ.

1.1.3. Текстові види спілкування в онлайн-спільнотах та їхні характеристики

Перенесення спілкування з офлайн в онлайн-середовище спричинило зміни у форматі та формулюванні висловлень. Причиною цих змін є відмінності засобів реалізації висловлень у обох середовищах. Ці відмінності необхідно враховувати під час адаптації методів та алгоритмів виявлення маніпуляції в офлайн-середовищі до онлайн-комунікації та під час розроблення нових методів та засобів виявлення ІПМ в онлайн-спільнотах.

Оскільки тема цього дисертаційного дослідження — методи та засоби виявлення ІПМ в онлайн-спільнотах, то необхідно враховувати відмінності засобів спілкування в онлайн-спільнотах від інших інтернет-медіа. Також відповідно до окресленої області дослідження розглянемо комунікацію в онлайн-спільнотах, в яких домінує текстова форма інформації.

В онлайн-спільнотах із домінантною текстовою формою інформації наявні такі засоби передачі інформації: повідомлення, пост, уподобання, поширення [12]. Ці засоби пропонуємо класифікувати за двома видами інформаційної активності, а саме — мовними і сигнальними (рис. 1.4.)

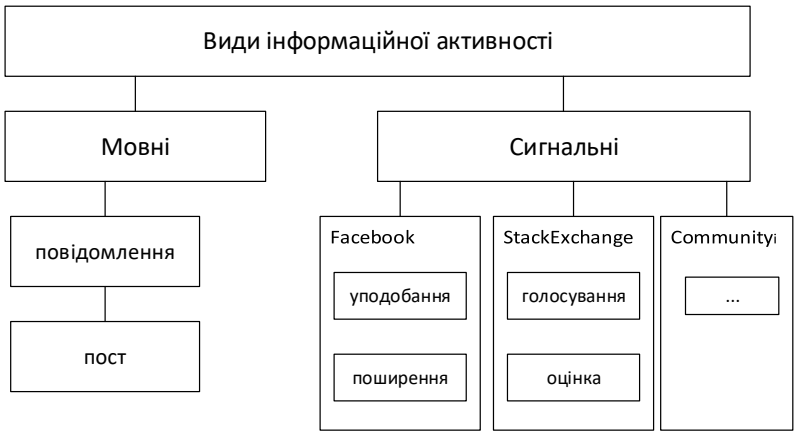


Рис. 1.4. Види інформаційної активності в онлайн-спільнотах

Мовними засобами інформаційної активності є, наприклад, пост та повідомлення. Універсальним мовним засобом інформаційної активності є повідомлення.

Види сигнальних засобів інформаційної активності варіюють залежно від структури онлайн-спільноти. До сигнальних належать, наприклад, уподобання та поширення у Facebook; голосування (яке потім враховують під час визначення рейтингу), оцінка повідомлення (можлива лише для користувачів, які мають високий рейтинг) на мережі форумів Stack Exchange.

Повідомлення є найбільш семантично необмеженою та контекстуально незалежною одиницею інформаторської активності. Повідомлення поділяють на два види: ініціювальне та повідомлення-реакцію.

Ініціювальне повідомлення — це повідомлення з певною інформацією, в якому ставиться питання, висловлюється позиція чи заклик до певних дій. Тоді як повідомлення-реакція є відповіддю на інше повідомлення, містить запитувану інформацію, доповнює аспект теми, який був заторкнутий у відповідному первинному повідомленні.

Повідомлення-реакція містить явно або неявно виражене посилання на повідомлення, яке спровокувало цю відповідь. Отже, визначити послідовність повідомлень на форумі можна на основі явних та неявних зв’язків між парами повідомлень (див. підрозділ 1.1.2 «Типова структура онлайн-спільноти»).

Розрізняти ці два види повідомлень потрібно для встановлення зв’язків між конкретними повідомленнями в онлайн-спільноті та відтворення логічної структури інформаційного наповнення дискусії.

Оскільки форма представлення та семантичні рамки мовних і сигнальних видів інформаторської активності відмінні між собою, то значення та можливість застосування у ролі індикатора ІПМ цих видів інформаційної активності різна. На основі засобів сигнального виду розроблено систему фільтрів, які вказують на уривок дискусії, який за певними шаблонами сигнальної активності свідчить про можливу присутність ІПМ. На основі засобів мовного виду розроблено систему маркерів ІПМ, які свідчать про наявність небажаних прихованих впливів (див. підрозділ 2.3 «Розроблення »).

1.1.4. Вербальні, невербальні та паравербальні засоби реалізації повідомлень

Як зазначено у меті дисертаційної роботи, розглядаємо комунікацію учасників онлайн-спільнот за допомогою інформації, поданої у тестовому форматі. Згідно з класичним визначенням текстової інформації — це інформація, передана за допомогою текстових символів (алфавітних, спеціальних символів та цифр) [14].

Текстову форму подання використовують переважно для передавання вербальної інформації. У випадку спілкування в онлайн-спільноті, перелічені вище знакові системи використовують для реалізації не лише текстового подання інформації, а й інтегрованих у повідомлення деяких видів невербальних та паравербальних засобів спілкування.

Вербальні засоби становлять базу повідомлення, оскільки за допомогою вербальної текстової інформації передається найбільша частина семантики повідомлення. Вербальні засоби комунікації — це лексичні одиниці (слова). Їх реалізовують за допомогою таких знакових систем, як алфавіти та спеціальні символи.

Невербальні засоби, які використовуються в онлайн-спільнотах, — це емотикони та гіперпосилання. Паравербальні засоби в онлайн-спільнотах реалізовані за допомогою метаграфеміки.

Маніпуляції у традиційних середовищах офлайн-спілкування з метою виявлення, протидії та захисту досліджують із погляду різних сфер науки, а саме: психології [15, 16], лінгвістики [17, 18], маркетингу [19, 21, 22] тощо. В цих дослідженнях основну роль відіграє аналіз інформації за допомогою вербальних, невербальних та паравербальних засобів. Для того, щоб адаптувати наявні алгоритми, методи та підходи до виявлення маніпуляції в офлайн-середовищі до реалій онлайн-спілкування, проведено порівняння використовуваних офлайн- і онлайн-засобів комунікації.

Як видно з рис. 1.5, вербальні засоби, які використовуються у повідомленнях онлайн-спільнот, збігаються з текстовими засобами офлайн-комунікації.

Рис. 1.5. Порівняння засобів передачі інформації в онлайн- та офлайн-комунікації

Невербальними засобами, які використовуються в онлайн-спілкуванні як своєрідні відповідники міміки, жестів та постави, є емотикони. Емотикони — це графічні символи, які використовують для передачі емоцій. Вони можуть бути реалізовані як кластери типографічних символів або зображень чи анімацій [23].

Емотикони відрізняються від класичних прикладів графічної інформації, тому що вони належать до обмеженої множини та їхні значення чітко прописані для кожної онлайн-спільноти. Ці характерні особливості емотиконів дають змогу легко конвертувати їх у текстову інформацію і, отже, брати до уваги їхню семантику під час виявлення ІПМ реалізованої за допомогою текстових засобів.

Емотикони — це зображення виразів обличчя, зокрема усмішки чи насупленості, сформовані з різних комбінацій символів розкладки клавіатури, які використовують для умисного і усвідомленого передавання почуттів

автора чи тону повідомлення [24]. Ключовим аспектом цього визначення є те, що за допомогою емотиконів автори повідомлень передають усвідомлені емоції. Тобто автор ідентифікує свої почуття та вибирає емотикон, який найближче відповідає його емоціям, свідомо хоче показати певні емоції.

У традиційному офлайн-спілкуванні прояв емоцій є свідомим та несвідомим. Інформацію про емоції передають за допомогою невербальних засобів, наприклад, міміки, жестів, постави. Саме несвідомо передана інформація допомагає легко виявити емоційні стани, а потім ідентифікувати маніпулювання. Такої можливості немає у спілкуванні в онлайн-спільнотах, адже вся інформативна активність комунікантів в онлайн-спільноті є усвідомленою. Навіть якщо повідомлення є наслідком прихованого впливу на підсвідомому рівні та відповідає спричиненому внаслідок ІПМ стану, виявлення емоцій є свідомим та умисним.

Ще одним видом невербальних засобів у онлайн-спільнотах є гіперпосилання. Гіперпосилання (посилання) — це адреса іншого мережевого інформаційного ресурсу [14]. У повідомленнях гіперпосилання реалізовані як виділений набір текстових символів. Хоч посилання і складається з символів, але їхня сукупність не має семантичного змісту, тому воно належить до невербальних елементів, інтегрованих у текстові повідомлення онлайн-спільнот.

Функція гіперпосилання — запропонувати швидкий доступ до релевантної до висвітлюваної теми інформації. Гіперпосилання переводить на сторінку, яка містить поширену інформацію, підтверджує інформацію у повідомленні, вказує на джерело тощо. У традиційному спілкуванні немає окремого класу засобів комунікації, функції та характеристики якого відповідали б посиланням в онлайн-спільнотах. Джерело інформації або ресурс із додатковою інформацією зазначають у традиційному спілкуванні засобами вербальної комунікації.

Текст повідомлення може бути по-різному оформлений за допомогою різних видів шрифтів, нахилу букв, великих та малих літер, відступів, порожніх рядків і т.д. Тобто повідомлення можуть містити різну

метаграфеміку. Метаграфеміка — це параметри форматування тексту (колонки, міжрядковий інтервал, шрифти), які використовуються для полегшення сприйняття тексту [25].

Метаграфеміка належить до паравербальних засобів спілкування. У традиційному спілкуванні паравербальві засоби утворюються за допомогою голосу та інтонації, ґрунтуються на тональних і тембрових особливостях мови. Паравербальні засоби комунікації виконують функції навмисної чи ненавмисної передачі інформації, впливають і на співрозмовника (свідомо чи несвідомо), і на мовця [26]. Замість мелодики, пауз, логічного наголосу, тембру голосу та тону мовлення у комунікації в онлайн-спільнотах таку функцію має метаграфеміка.

Кожен із видів засобів передачі інформації (вербальних, невербальних та паравербальних) може бути використаний для здійснення ІПМ в онлайн-спільнотах. У таблиці (табл. 1.1) зазначено приклади реалізації прийомів ІПМ за допомогою вербальних, невербальних та метаграфічних засобів та виділено сліди, за якими їх можна виявити.

Таблиця 1.1

Типи засобів комунікації у текстових повідомленнях

	Паравербальні	**Вербальні**	**Невербальні**
Прийоми	Прихована аналогія	Звернення до авторитетів	Псевдопосилання
Ознака	Необґрунтоване використання великої літери	Вживання назв визнаних компаній та структур.	Посилання, на джерела з нерелевантною інформацією
Контекст	*«Судячи з подібності стародавніх звичаїв у* Великоросів *і малоросів…»*	«За даними BBC Україна, відсоток громадян, які підтримують…»	«Син президента ухиляється від військової служби, в той час як усіх…»
Пояснення	Крім зневажливого ярлика, тут виражено меншовартість за допомогою великої і малої літер.	Люди неохоче приймають рішення та аналізують події, а схильні переймати погляди авторитетів та не перевіряти наявність інформації в джерелі.	Гіперпосилання імітують наявність достовірних джерел інформації. Посилання веде до сторінки, яка містить лише дотичну інформацію.

Отже, текстові повідомлення здебільшого реалізовані за допомогою вербальних засобів інформації, але, крім останніх, вони містять характерні для онлайн-середовища невербальні та паравербальні засоби, а саме: емотикони, гіперпосилання та метаграфеміку. Відповідно для виявлення прецедентів ІПМ, необхідно брати до уваги всі види засобів інформації в текстових посиланнях.

1.2. Порівняння офлайн-маніпуляції та ІПМ

Комунікація в онлайн-спільнотах поєднує риси міжособистісного (діалогічність, інтерактивність) і масового (масштабна неспеціалізована аудиторія) спілкування. Тому відмінності традиційної маніпуляції та ІПМ в онлайн-спільнотах потрібно розглядати в контексті комунікаційних особливостей реального і віртуального середовищ. Крім того, варто звернути увагу на відмінності останньої від маніпуляції за допомогою інших Інтернет-медіа.

Сприятливими для ІПМ характеристиками комунікації в онлайн-спільнотах є:

- динамічне розгортання, поєднане з відносно тривалим збереженням запису комунікації;
- незалежність від часових рамок, що дає змогу здійснювати ІПМ в будь-який сприятливий період доби;
- відкритість для вступу в дискусію для маніпуляторів та незловмисних учасників;
- широка аудиторія, тобто доступність інформаційного наповнення онлайн-спільноти не лише для зареєстрованих користувачів.

Спільними рисами офлайн-маніпуляції та ІПМ є:

- здійснення за певною схемою чи алгоритмом;
- наявність довільної кількості маніпуляторів та незловмисних учасників;
- застосування психологічних механізмів впливу.

Відмінні від офлайн-маніпуляції характеристики, які дають змогу виявити ІПМ в онлайн-спільнотах, такі:

- ІПМ є обмежена засобами комунікації, тобто наявні вербальні засоби і декілька видів не- та паравербальних.
- прийоми ІПМ залишають слід у інформаційному наповненні дискусії у текстовій формі.

Важливою характеристикою з погляду виявлення ІПМ є те, що ІПМ, як і офлайн-маніпуляція, ґрунтується на схемах та алгоритмах. Цю характеристику покладено в основу методів і засобів виявлення ІПМ, оскільки останні полягають у виявленні тактик ІПМ у створеному користувачами інформаційному наповненні онлайн-спільнот за допомогою системи маркерів. Отже, знання про алгоритми тактик маніпуляції варто використати для виявлення ІПМ в онлайн-спільнотах.

1.2.1. Проекція тактик маніпуляції на комунікацію в онлайн-спільнотах

Дослідженню алгоритмів та тактик маніпуляції в офлайн-середовищі присвячено праці у різних галузях: психології, політики, маркетингу, лінгвістики та ін. Результати цих досліджень є корисними з погляду виявлення ІПМ, оскільки після проекції тактик маніпуляції на комунікацію в онлайн-спільнотах, їх буде формалізовано та покладено в основу методів і засобів виявлення ІПМ в онлайн-спільнотах.

Маніпуляція є популярним інструментом впливу на реципієнтів інформації з метою досягнення широкого спектру корисливих цілей. Розробленням алгоритмів та вказівок для захисту від психологічної маніпуляції в офлайн-спілкуванні займалися [27, 28, 29]. Ці алгоритми успішно застосовувалися для виявлення маніпуляції в офлайн-спілкуванні, але є неефективними для онлайн-спільнот, оскільки вони виявляють тактики маніпуляції на основі маркерів, які передбачають аналіз комунікації людиною, а також маркерів, яких немає в онлайн-спільнотах (див. підрозділ 1.1.4 «Вербальні, невербальні та паравербальні засоби реалізації повідомлень»).

Оскільки у дисертаційній роботі розглядаємо онлайн-спільноти з домінантною текстовою інформацію, то доцільно виявляти ІПМ за допомогою

вербального сліду, який залишає маніпулятор. Цим слідом є мовні одиниці, за допомогою яких маніпулятор здійснює прийоми ІПМ, тобто звичайні мовні одиниці, які набувають маніпулятивних властивостей, лише коли вживаються в певних комбінаціях.

Тактики ІПМ, як і тактики офлайн-маніпуляції, складаються з маніпулятивних прийомів, які, здебільшого, перейшли в онлайн-комунікацію з офлайн-спілкування. Дослідженню цих прийомів та пошуку методів захисту присвячені наукові праці у сфері пропаганди [30], міжособистісних та масових маніпуляцій [27], а також нейролінгвістичного програмування (НЛП) [31].

Пропаганда є широко дослідженим явищем, яке у своїй сфері застосування перетинається з маніпуляцією: пропаганду часом проводять, застосовуючи психологічний вплив на підсвідомість реципієнта, водночас для маніпуляції часом використовують прийоми пропаганди, наприклад, блискуче узагальнення, навішування ярликів. Тому опис прийомів пропаганди використовуємо в цій роботі як основу для формалізованого запису прийому ІПМ.

Прийоми пропаганди змушують людей неусвідомлено здійснювати певні дії або змінювати світоглядну позицію. Існують класифікації прийомів пропаганди, переліки прийомів; виділено сім основних прийомів пропаганди. Для того, щоб розробити систему маркерів прийомів ІПМ, прийоми об’єднано в класи за психологічним механізмом здійснення. У табл. 1.2 наведено клас прийомів «Перехід на особисте». Таблиця містить інформацію про мету і психологічний механізм здійснення прийому та приклад застосування прийому в онлайн-комунікації.

Узагальнено психологічний механізм, використаний у прийомах цього класу, можна подати так: використання особистих рис реципієнта (учасника дискусії) або обговорюваної людини для спростування аргументу цієї особи.

Таблиця 1.2

Клас прийомів «Перехід на особисте»

Прийом	Мета	Психологічний механізм	Приклад
Ситуативний перехід на особистості	Змусити опонента зайняти необхідну позицію	Наголосити на ситуації, в якій перебуває опонент і подати необхідну позицію як єдино правильну у цій ситуації	«якщо ти українець, то ти маєш воювати» «якщо ти професіонал, то зациниш Lenovo ideapad 700!»
Імітація меркантильності	Заперечити чи спростувати твердження опонента	Подати, що позиція людини залежить від її специфічних одноосібних інтересів, а не випливає з раціональних фактів	«та він на тому хату побудував, то шо він буде проти?!» «можна подумати він би інакше сказав! Керовнік!»
Ілюзія суперечки	Відійти від теми обговорення	Змусити опонента виправдовуватись і пояснювати нерелевантні до теми обговорення питання	«Ти так кажеш, бо тобі так сказали!»
Необґрунтовані оціночні судження	Спровокувати емоційне виправдовування	Змусити опонента реагувати на необґрунтовані оцінки	«Це банально» «Дурня!» «Забобони!»

У класі прийомів «Перехід на особисте» можна виділити підклас прийомів, які використовуються лише для створення негативного ставлення до особи, предмету чи ідеї. На відміну від прийомів наведених у табл. 1.2, які використовуються для досягнення позитивного і негативного ставлення, тобто вони є універсальними, підклас прийомів «Перехід на особисте з негативною тональністю» використовують для провокування негативного ставлення реципієнта до певної сутності. Ці прийоми наведені нижче у табл. 1.3.

Таблиця 1.3

Клас прийомів «Перехід на особисте з негативною тональністю»

Прийом	Мета	Механізм	Приклад
Інвективний перехід на особисте	Викликати негативне ставлення до слів чи поглядів іншої людини	Упередження чи негатив непов'язаних із темою дискусії фактів перенести на твердження чи позицію людини	«не роби так, як він каже, бо він жид» «він бреше, бо він говорить російською» «та, слухай далі олігарха!»
Імітація необізнаності співрозмовника	Завершити, вийти або змусити опонента покинути небажане обговорення	Імітувати необізнаність чи відсутність досвіду опонента і, таким чином, завершити дискусію як недоцільну	«Ви не чули про…?! Про що тоді з Вами говорити?!» «ти не знаєш елементарних ...?» «тобі … нічого не каже?»
Отруєння джерела	Дискредитувати все, що сказав опонент	Негативно представити людину до, того як вона вступить в дискусію	«Запитаємо експерта, надіюсь цього разу його передбачення будуть успішнішими»

Отже, виявлення ІПМ в онлайн-спільнотах ґрунтуватиметься на тактиках офлайн-маніпуляції, для цього їх необхідно адаптувати до комунікації в онлайн-спільнотах, розробити маркери на основі засобів комунікації, наявних в онлайн-спільнотах, та погрупувати їх для створення класів прийомів, які можуть здійснюватись у певних тактиках ІПМ.

1.2.2. Види невербальних та паравербальних засобів, які використовують для реалізації ІПМ

Прийоми ІПМ в повідомленнях онлайн-спільнот реалізовуються за допомогою текстових, метаграфічних засобів, емотиконів як підвиду графічних засобів та посилань.

Емотикони впливають на сприйняття інформації реципієнтом, зокрема адекватна кількість емотиконів полегшує сприйняття тексту, створює неофіційний, часто дружній стиль спілкування. Оскільки учасники онлайн-спільнот не витрачають зусиль для опису свого емоційного стану, який у традиційному спілкуванні можна виявити за допомогою міміки, жестів, постави та голосу людини, то емотикони є засобами швидкої передачі базової

інформації про настрій та психологічний стан комунікантів. Також емотикони можуть заміщати слова або використовуватись для «розбавлення» слів повідомлення [32].

Емотикони належать до засобів передачі семантики повідомлення, які використовують здебільшого для передачі інформації двох видів: про об'єкти дійсності та пов'язані з ними процеси, а також про емоції. Відповідно емотикони поділяємо на два види: змістові та емоційні.

Змістові емотикони передають інформацію про сутності, дії, та процеси (поїзд , торт , танцівниця , стрижка), див. рис. 1.6. Емотикон може нести семантику дієслів та іменників, тому змістові емотикони більше вказують на тематику, а конкретне їхнє значення встановлюється з контексту (див. табл. 1.4). Наприклад, емотикон *Жінка, яка танцює* залежно від контексту може означати танцівниця, танцювати, танець, танцювальна вечірка і т.д.

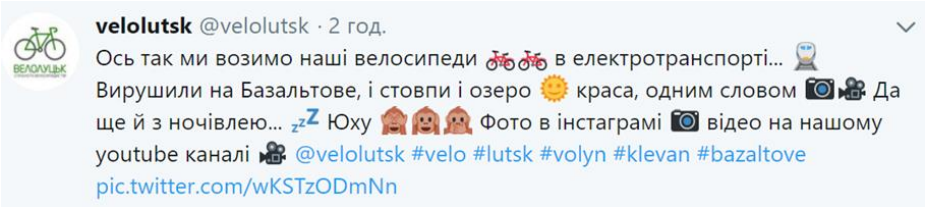


Рис. 1.6. Приклад вживання змістових емотиконів

Таблиця 1.4

Змістові емотикони

Графічний символ	Назва та приклади ключових слів	Приклади вживання з Twitter
	**Поїзд** дизельний поїзд пасажирський поїзд звичайний поїзд електричка	All aboard on the 955klothing train 7 days to get on the train. Be all in Go. Adventure awaits. getting the Eurostar to London
	**Торт до дня народження** день народження торт торт зі свічками	Заблочили @vykranyi вічна пам'ять Вітаємо, Тіна і Вітя!!! Happy Birthday Rawalpindi Express
	**Жінка, яка танцює** танцювання жінка в червоній сукенці танцівник/иця	Let's go Дааа, дєтка Upcoming is so exciting

Змістові емотикони можуть використовуватись в ІПМ для вираження тверджень, які у випадку реалізації вербальними засобами привертали б небезпечну для успішного здійснення ІПМ увагу.

Емоційні емотикони використовують для передавання настрою, почуттів чи душевного стану (див. табл. 1.5). Можна провести паралель між емоційними емотиконами в онлайн-середовищі та мімікою і жестами в традиційному. Міміку передають за допомогою «круглих личок», які ще називають смайликами, а жести — за допомогою зображень рук та долонь.

Таблиця 1.5

Емоційні емотикони

Графічний символ	Назва та приклади ключових слів	Приклади вживання з Twitter
	**Обличчя з перев'язаною головою** перев'язана голова незграба поранений(а)	I wanna get drunk tonight that's not healthy
	**Націлений вперед кулак** (середньосвітлий колір шкіри)	Sign the @VFPlus #FightBackSA petition today! Have a good bank holiday weekend
	**Чоловік, який жестикулює «ні»** (середньотемний колір шкіри)	I'm never exposing my pain .. Aint no debating it
	**Людина, яка закриває обличчя рукою** (світлий колір шкіри)	OMG what an old fashioned question.. So I've been using the wrong hashtag.

Емоційні емотикони можуть містити імпліцитну змістову інформацію. Набори емоційних емотиконів - це схематичні зображення лиць та кінцівок людей, відповідно кожен жест чи вираз обличчя має кілька варіантів, які передбачають кольори шкіри усіх рас. Учасники онлайн-спільнот послідовні у використанні емоційних емотиконів певного кольору. Коли ж учасники починають використовувати різного кольору емоційні емотикони – це означає, що таким чином вони хочуть передати інформацію щодо рас, яку не бажають висловлювати експліцитно.

Як було сказано вище, емотикони роблять стиль дружнім, неформальним та легким для прочитання, тому реципієнт не зосереджує на них увагу, а сприймає їх на підсвідомому рівні. Як наслідок за допомогою емотиконів маніпулятор може передати інформацію, однозначне висловлення, якої могло б спричинити провал ІПМ приклад (див. прикл. 1.1). Крім того, маніпулятор залишає за собою шанс у разі виявлення реципієнтом прихованого меседжу ІПМ, стверджувати, що емотикон неправильно потрактовано.

Приклад 1.1. Передача прихованої інформації за допомогою змістових емотиконів

«а вони вже давно 👬»

«🏃 sucks! Go 🏀!»

«Everyone has personal vocation 👨‍🔨🧑‍🌾»

Наведені вище особливості передавання інформації за допомогою емотиконів варто брати до уваги під час виявлення прийомів ІПМ (див. розд. 3.5 Методи виявлення прийомів ІПМ у дискусіях онлайн-спільнот). Крім того, емотикони використовуємо для встановлення емоційного стану учасника дискусії (див. розд. 3.4 «Методи виявлення послідовності зміни психічних станів учасників дискусії»). Для використання емотиконів під час виявлення інформації необхідно подати їхню семантику в текстовому вигляді, це можна зробити, знаючи, яку бібліотеку емотиконів передбачено на платформі для комунікації (рис. 1.7).

Зображення	Заголовок	Текст
	Girl Smile	:girl_smile:
	Girl Cray	:girl_cray:
	Girl Blum	:girl_blum:
	Girl Crazy	:girl_crazy:
	Girl Haha	:girl_haha:
	Girl Sad	:girl_sad:
	Girl Dance	:girl_dance:

Рис. 1.7. Таблиця інтерпретації емотиконів з форуму «Дівочі посиденьки»

Знайти емотикони за ключовими словами можна на сайті Emojipedia. Emojipedia [33] надає можливість пошуку емотиконів та показує їхні зображення в продукції компаній Google, Apple, Microsoft, Samsung, LG, HTC та платформах соціальних мереж WhatsApp, Facebook, Twitter та бібліотеці емотиконів Emojidex.

Крім того, для автоматизованого виявлення ІПМ можна скористатись таблицями анотацій емотиконів з Загального репозиторію мовних даних (Unicode Common Locale Data Repository [34]. Емотикони погруповані за змістом, наприклад «Посмішки і люди». Таблиці анотацій містять коротку анотацію емотиконів: їхню назву, ключові слова для пошуку, код у Unicode та зображення на платформах різних компаній (рис. 1.8).

Smileys

face-positive

№	Code	Browser	Appl	Goog	Twtr	One	FB	Sams.	Wind.	GMail	SB	DCM	KDDI	CLDR Short Name
1	U+1F600	😀	😀	😀	😀	😀	😀	😀	😀	😀	—	—	—	grinning face
2	U+1F601	😁	😁	😁	😁	😁	😁	😁	😁	😁	😁	😁	😁	beaming face with smiling eyes
3	U+1F602	😂	😂	😂	😂	😂	😂	😂	😂	😂	😂	—	😂	face with tears of joy

Рис. 1.8. Фрагмент таблиці повного списку емотиконів на базі репозиторію CLDR

Іншим видом невербальних засобів передачі інформації в повідомленнях онлайн-спільнот є посилання (див. розд. 1.2.2 «Види невербальних та паравербальних засобів, які використовують для реалізації ІПМ»). Посилання — це адреса Інтернет-ресурсу у форматі URL. Посилання використовують для зазначення джерела чи ресурсу з додатковою інформацією.

Внаслідок інформаційної насиченості Інтернету учасники онлайн-спільнот нечасто переходять за наведеними посиланням, особливо, якщо повідомлення містить багато посилань.

Наявність посилань збільшує довіру до інформації, розміщеної в повідомленні, переконує в її достовірності. Отже, посилання зменшують сумніви учасників щодо якості наданої інформації та бажання витрачати час і

зусилля на перегляд ресурсів, на які ведуть посилання. Саме тому посилання часто використовують у реалізації прийомів груп звернення до популярної думки, звернення до авторитету.

Ми виділили три типи посилань, які використовуються у цілях ІПМ, а саме: недоступне посилання, нерелевантне посилання, посилання на неавторитетні джерела.

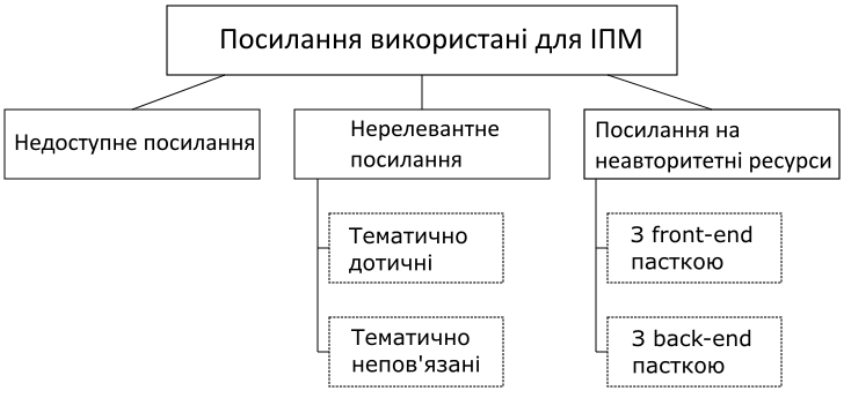


Рис. 1.9. Типи посилань, які використовуються для реалізації ІПМ

Недоступне посилання — це посилання на відсутній інформаційний ресурс. Виявити недоступне посилання можна, перейшовши за ним.

Нерелевантне посилання — це посилання на ресурси, які не містять зазначеної в повідомленні інформації, хоч і можуть бути тематично пов’язані. Нерелевантні посилання бувають двох видів: тематично дотичні, які ведуть на сайти з дотичною до дискусії тематикою, але не містять інформації, яку інформатор хоче підтвердити посиланням; тематично непов’язані ведуть, на ресурси, інформаційне наповнення яких тематично не відповідає дискусії.

Посилання на неавторитетні ресурси – ресурси, які розміщують інформацію, яка не відповідає вимогам (достовірність, об’єктивність, актуальність, надійність тощо). Часто ці сайти маскуються під надійні джерела інформації. Посилання на неавторитетні ресурси є двох видів: front-end пасткою - відмінності від посилання на оригінальний надійний ресурс можна помітити прочитавши посилання; back-end пасткою – потребує аналізу кодів (наприклад, Unicode) використаних символів.

На основі наведеної вище класифікації за реалізацією можна виявляти посилання на неавторитетні джерела (див. розд. 3.5 «Методи виявлення прийомів ІПМ у дискусіях онлайн-спільнот»).

Паравербальні засоби наявні у комунікації текстових онлайн-спільнот — це параграфеміка (метаграфеміка). Існують різні трактування цього поняття:

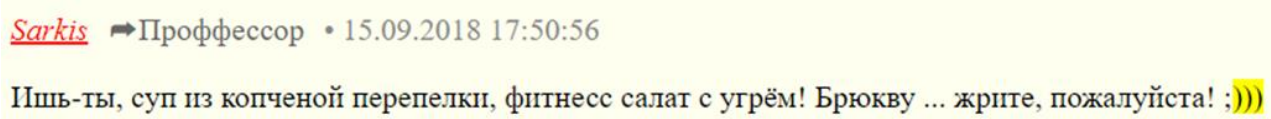
Параграфеміка — це система графічних елементів, які не входять у набір диференційно-графічних протиставлень, тобто до букв алфавіту [35].

Параграфеміка — це графічні засоби (наприклад, тире, перекреслений текст) застосовані не відповідно до правил орфографії [36].

У цій дисертаційній роботі використовуватимемо термін «параграфеміка» у вузькому значенні, тобто засоби пунктуації як засоби варіювання шрифтів та площинної організації тексту [37]. У широкому значенні, крім зазначених засобів, параграфеміка містить: засоби іконічної мови (рисунок, фотографія, таблиця, схема, графік) [38].

Функція параграфеміки полягає у полегшенні розуміння тексту, а з погляду виявлення ІПМ це означає, що за допомогою параграфеміки можна впливати на процеси сприйняття тексту та спричинити упереджене ставлення до інформації. Таким чином звичайні допоміжні текстові засоби передачі інформації, такі як: спеціальні символи, цифри, знаки пунктуації; розмір, колір, стиль шрифта; відступи, абзаци, міжрядковий інтервал можна використовувати для здійснення ІПМ.

До параграфеміки належить також неконвенційна розстановка знаків пунктуації. Діакритичні знаки, цифри, букви, символи можуть використовуватись для створення емотиконів. Функції і роль в ІПМ емотиконів, реалізованих як графічні об’єкти чи як набори букв та спеціальних символів - однакові, але за допомогою емотиконів-графічних об’єктів можна передати більше інформації (див. рис. 1.10).


Sarkis ➦Проффессор • 15.09.2018 17:50:56

Ишь-ты, суп из копченой перепелки, фитнесс салат с угрём! Брюкву ... жрите, пожалуйста! ;)))

Рис. 1.10. Приклад вживання параграфічних емотиконів

Параграфічні емотикони дуже рідко бувають змістовими – переважно вони передають емоційне навантаження повідомлення (див. табл. 1.6). Відповідно вони використовуватимуться на етапах виявлення станів учасників дискусії (див. підрозділ 3.4 «Методи виявлення послідовності зміни психічних станів учасників дискусії»).

Таблиця 1.6

Приклади параграфічних емотиконів

Параграфічний емотикон	**Вид**	**Значення**	**Приклади вживання**
:-D :D	емоційні	сміх	«ich hab die Durchsage nicht verstanden und bin auf die falsche Haltestelle ausgestiegen :-D»
:-0 :O		здивування	«Are you kidding?! :-0»
:-*	змістові	поцілунок	«Дякую! До вихідних :-*»
:-X		мовчати	«ок, я :-X не переживай»

Зміст метаграфеміки визначаємо на основі таблиць, які містять можливі трактування елементів метаграфеміки. Приклад трактування метаграфеміки наведений нижче (табл. 1.7).

Таблиця 1.7

Семантика елементів метаграфеміки

Назва	**Значення**	**Приклад**
Великі букви	Крик, підвищений тон	«давно пора плюс турнікети і за кожного перевезеного пасажира сплатити податок у міську казну. ДОСИТЬ ГОТІВОЧКОЮ НАБИВАТИ КИШЕНІ...» «Мапу магазинів фабрики.ЯКА торгує в ліпецку зробити СЛАБО? популісти хренові» «слухайте ВДОМА ту музику,яка вам подобається і нічого пропагувати усім узкій мір»
Перекреслений текст	«Випадково» висловлена думка. Це є часто ознака сарказму: в автор «зірвалось з язика»	«Выделение подразумеваемого смысла зачеркиванием стало в последние пару лет очень популярным среди ~~бездельников и экгибиционистов~~ авторов блогов и дневников» «Ну и муд~~ак~~рец»
Збільшений інтервал між буквами	Передає сповільнену дію (сприйняття)	«… бо тому шо то ж п о д у м а т и треба» «… поки з а д п і д ні м е – все вже готове буде»

Параграфічні емотикони виявляємо в тексті, як скупчення більше ніж двох спеціальних символів. Інші елементи метаграфеміки виявлятимемо за допомогою програм для аналізу оформлення тексту.

1.3. Аналіз підходів до виявлення емоцій та психічних станів в онлайн-комунікації

Реалізація тактик ІПМ передбачає послідовну зміну психічних станів реципієнта. Маніпулятори передають реципієнтові інформацію або мотивацію до дій, передбачену метою ІПМ, лише коли реципієнт перебуває в сприятливому для здійснення ІПМ психічному стані.

Психічний стан реципієнта визначають емоції та настрій; саме на основі цих ознак ми ідентифікуємо психічні стани учасників онлайн-спільнот. Оскільки в цій дисертаційній роботі розглянуто текстову комунікацію, то необхідно проаналізувати підходи до виявлення емоцій на основі текстового подання інформації.

Проаналізувати весь спектр людських емоцій і психічних станів та виявити їхні текстові ознаки неможливо, тому необхідно дотримуватись одної з існуючих класифікацій психічних станів і емоцій. З цією метою проведено аналіз класифікацій емоцій і психічних станів людини.

1.3.1. Аналіз наявних класифікацій емоцій і психічних станів людини

Сприйняття та реакція на зовнішній світ залежить від психічного стану людини. Цю закономірність маніпулятори використовують під час здійснення ІПМ. Тобто маніпулятори передають реципієнтам передбачену цілями ІПМ інформацію лише заздалегідь спричинивши перехід реципієнта в необхідний, згідно з обраною тактикою ІПМ, психічний стан.

Психічний стан - це змінний функціональний рівень психіки, спричинений чинниками внутрішнього та зовнішнього середовища людини, який визначає перебіг психічних процесів та вияв психічних властивостей людини у певний момент часу [39].

Згідно з класифікацією Ю. В. Щербатих, види психічних станів виділено за такими ознаками [40]:

0. за джерелом формування: ситуативно обумовлені, особистісно обумовлені;
1. за ступенем вираження: поверхневі, глибокі;
2. за емоційним значенням впливу: позитивні, нейтральні, негативні;
3. за тривалістю: короткочасні, середньої тривалості, тривалі;
4. за ступенем усвідомленості: неусвідомлені, усвідомлені;
5. за домінуванням раціонального чи емоційного компонента: емоційні, комбіновані, інтелектуальні;
6. за ступенем активації організму: астенічні, стенічні;
7. за основним рівнем систематичного прояву: фізіологічні, психофізіологічні, психологічні.

Ці ознаки пронумеровані від 0 до 7, відповідно кожен стан Ю. В. Щербатих задає як вектор з восьми елементів типу *x.y,* де *x* – це номер ознаки, а *y* – це номер конкретного значення ознаки.

Ця класифікація лише частково підходить для вирішення завдання виявлення ІПМ в онлайн-спільнотах, оскільки частина ознак не відображаються в спілкуванні в онлайн-спільнотах (наприклад, за основним рівнем систематичного прояву), а частина не інформативні з погляду виявлення ІПМ (наприклад, за джерелом формування). Ознаки 2, 5, 6 не є достатніми для ідентифікації психічного стану учасників онлайн-спільноти.

Корисну інформацію для процесу виявлення ІПМ несе класифікація психічних станів як виявів психічних процесів [39]:

- стани емоційні (наприклад, настрої, афекти, тривога);
- стани вольові (наприклад, рішучість, розгубленість);
- стани пізнавальні (наприклад, зосередженість, замисленість).

К. Шерер запропонував таку типологію емоційних станів: емоції, настрій, міжособистісні установки, ставлення, характер [41]. Кожен із цих класів характеризується певним рівнем інтенсивності, тривалості,

синхронізації, зосередженості на події, виявлення оцінки, мінливості і впливу на поведінку.

Для визначення психологічного стану учасника онлайн-дискусії важливо знати емоційні стани, які належать до таких видів:

- емоції (розлючений, сумний, радісний, переляканий, засоромлений, гордий, піднесений, у розпачі);
- настрій (життєрадісний, похмурий, дратівливий, млявий, бадьорий, депресивний);
- міжособистісні установки - характеризують ставлення людей під час інтеракції (холодні, теплі, близькі, віддалені відносини, підтримка, презирство);
- ставлення (ненависть, любов, бажання, цінування, подобається).

Перші два типи характеризуються меншою стабільністю, ніж наступні. Емоції та настрій легше змінювати, тоді як міжособистісні установки і ставлення є наслідком спілкування та аналізу інформації протягом тривалішого часу. Тому емоції та настрій змінюватимуться часто, а зміна ставлення та міжособистісної установки відбувається лише кілька разів, а той один раз протягом усієї ІПМ.

Відповідно, психічний стан учасника онлайн-спільноти можна задати кортежем, який міститиме елементи, Емоції і Настрій. Під час ІПМ значення елементу Емоції змінюватиметься частіше, ніж Настрій. Водночас інформацію про настрій не треба упускати, оскільки за нею можна визначити вразливість учасника спільноти до конкретної ІПМ (наприклад, коли настрій відповідає або близький до емоцій, які необхідно спровокувати).

Визначити точну кількість емоцій, які можуть переживати люди, неможливо, але згідно з деякими оцінками, людина може переживати до 1000 відтінків різних емоцій. Підібрати ознаки, семантичні одиниці та синтаксичні правила, щоб задати всі можливі реалізації всіх емоцій, які може відчувати людина, - неефективно і неможливо. Тому низка досліджень спрямована на визначення набору основних емоцій людини.

Р. Плутчік запропонував вісім базових емоцій. І ці базові емоції мають два рівні інтенсивності (низьку і високу інтенсивність) і відповідно утворюють ще 16 своїх відтінків [42]. Наприклад, низька інтенсивність довіри – це толерування, а висока - це прихильність.

Під час реалізації ІПМ інтенсивність емоцій реципієнтів змінюється переважно поступово; цю характерну ознаку тактик ІПМ зручно відстежити за колесом емоцій Р. Плутчіка.

Кожна базова емоція має протилежну, наприклад, протилежним до очікування є здивування. Р. Плутчік зобразив відношення між вісьмома базовими емоціями за допомогою колеса емоцій. У колесі емоцій протилежні розташовані діагонально на відстані трьох секцій один від одного.

Всі базові емоції, крім протилежних, можуть створювати діади і таким чином утворюють ще 24 емоції: наприклад, базові емоції очікування і довіра результують у вторинну емоцію надію.

Згідно з мовою анотування і репрезентації емоцій (EARL - emotion annotation and representation language) виділено 48 емоцій. Емоції поділено за десятьма класами, це зокрема: негативні неконтрольовані, негативні вольові, негативні пасивні, негативні помисли, знервованість, позитивні спокійні, позитивні жваві, позитивні помисли, турбота, реакція [43].

Під час тональнісного аналізу онлайн-комунікації дослідники виявили закономірну появу певних емоцій у повідомленнях-реакціях, спровокованих певними емоціями в повідомленнях-тригерах [44]. Одними із найпоширеніших у комунікації в онлайн-спільнотах є зазначені у табл. 1.8.

В ІПМ це має такий вигляд: наприклад, якщо маніпулятор хоче домогтися від читачів підримки його дій та їхнього залучення, то він буде писати про несприятливі обставини, з якими він наполегливо бореться; якщо потрібно дискредитувати когось, то автор висловлюватиме сум, причиною якого є дії особи чи організації, яку необхідно дискредитувати, та подаватиме певні дії об’єкта дискредитації як порушення.

Таблиця 1.8

Закономірності емоційного навантаження повідомлень

Повідомлення-тригер	**Повідомлення-реакція**
сум	солідарність, емпатія
несприятливі обставини	підтримка
наполегливість	заохочення
щастя	щастя
порушення	гнів

Вибір підходів до класифікації емоцій як складових психічних станів необхідний для розроблення формальної моделі стану учасника онлайн-спільноти під час ІПМ (див. розд. 2.2.1 «Формальна модель тактики ІПМ») та реалізації етапу виявлення ІПМ алгоритму виявлення ІПМ (див. розд. 3.1.2 «Етап виявлення»).

1.3.2. Класифікація інструментарію тональнісного аналізу

Аналіз тональності (sentiment analysis, opinion mining) полягає у виявленні емоційного навантаження висловлень та ідентифікації суджень авторів щодо певних об’єктів дійсності.

Тональнісний аналіз використовують у таких сферах:

- реклама і маркетинг (наприклад, системи аналізу відгуків та системи аналізу інтеракцій за допомогою технологій «другого екрана») [45];
- інформаційні війни (наприклад, для виявлення радикалізації поглядів певних груп населення [46])
- політика та менеджмент зв’язків із громадськістю [47];
- рекомендаційні системи [48];
- інтеграція у пошукові системи [49].

Загалом використання тональнісного аналізу можна класифікувати у двох напрямах за активністю позиції агента комунікації, це зокрема:

- тональнісний аналіз, спрямований на виявлення роздумів, емоцій, оцінок, ставлення людей до організації, продукту чи особи на основі створеного ними інформаційного наповнення;

- тональнісний аналіз, метою якого є підбір мовних структур та одиниць, які б забезпечили сприйняття інформації реципієнтами в необхідному інформатору емоційному стані.

Залежно від завдання застосовують різні види тональнісного аналізу. На рис. 1.11 наведено класифікацію видів тональнісного аналізу на основі відмінностей у застосованому методі технічної реалізації, рівні структурування тексту, системі оцінювання [50].

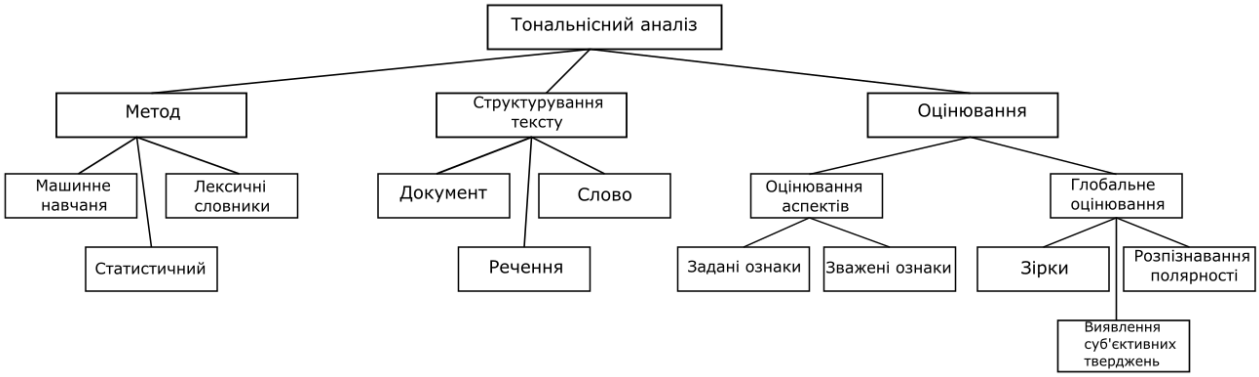


Рис. 1.11. Класифікація систем тональнісного аналізу

У різних видах тональнісного аналізу використано відмінні між собою способи вирішення таких проблем:

- оцінювання текстів, які містять сарказм, іронію, сленг, навмисні або випадкові орфографічні помилки, слова з кількома значеннями;
- оцінювання сутностей, які складаються з кількох аспектів;
- вибору емоційних шкал, за якими відбувається оцінювання (від класичної позитивної, нейтральної, негативної до шкалювання за багатьма емоціями).

За методом технічної реалізації є тональнісний аналіз, реалізований за допомогою:

- систем машинного навчання;
- тональнісних словників;
- статичних систем та систем правил.

Крім того, популярними є системи, які поєднують різні підходи до технічної реалізації для виконання конкретних типів завдань.

Системи тональнісного аналізу, які передбачають машинне навчання, поділяються на:

- з вчителем (на основі методу логічної регресії, методу опорних векторів, методу k-найближчих сусідів, наївного Баєсівського алгоритму, моделі максимальної ентропії) [51];
- без вчителя (автоматичний вибір початкового значення слів [52]).

Алгоритми без вчителя менш популярні для вирішення завдання аналізу тональності.

Системи, базовані на тональнісних словниках, оцінюють тональність тексту на основі емоційного навантаження слів, які містять спеціалізовані словники [53, 54].

Системи, базовані на правилах, полягають у правилах інтерпретації тональності тексту на основі синтаксичних структур, а не лише на основі лексики, на відміну від попередньої техніки. У таких системах враховано зміни семантики речення, які вносять заперечення, підсилювальні частки, ідіоми [50].

Статистичні моделі представляють тексти як суміш термів та оцінок. Терм – це одиниця мови, яка має хоча б одне семантичне значення та позначає певне поняття в контексті [55]. Припускається, що терм і його оцінка можуть бути представлені як поліноміальний розподіл. Основні терми кластеризують в аспекти, а тональнісні оцінки - в рейтинг.

За структуруванням аналіз тональності поділяють на аналіз тексту на рівні документу, речення, слова чи ознаки. Мета тональнісного аналізу на рівні документу — визначити полярність усього посту, повідомлення чи іншого виду документу, тоді як аналіз на рівні речення чи слова полягає у визначенні емоційної полярності цих одиниць.

За системою оцінювання вирізняють тональнісний аналіз різних аспектів сутності або оцінювання сутності на глобальному (загальному) рівні.

Одним із видів систем тональнісного аналізу за шкалами оцінювання є класичні системи, які розрізняють за ознаками позитивності, негативності, нейтральності. Такий тональнісний аналіз можна зробити на основі лексичного

ресурсу SentiWordNet [56]. SentiWordNet отримано внаслідок тегування множини синонімів лексичної бази даних WordNet 3.0 відповідно до їхнього ступеня позитивності, негативності та нейтральності.

Існують системи тональнісного аналізу, які детальніше класифікують емоції, зокрема виходять за рамки двох полярностей, а класифікують за емоційними станами, наприклад, злість, радість, сум. До них належать EffectCheck та Canvs.

EffectCheck — система для виявлення емоційного навантаження тексту відносно шести базових емоцій: тривоги, ворожості, депресивності, впевненості, співчуття і щастя. Система пропонує синоніми, які можна вжити замість слів, щоб змінити емоційний тон тексту [57].

Canvs — платформа для тональнісного аналізу, яка дозволяє відстежувати 56 різних емоцій аудиторії. Для визначення емоцій використовується словник з 500 млн термів [58].

Ще одним видом є системи, які класифікують тональність за двома полярностями, але передбачають визначення інтенсивності цих полярностей за градуйованою шкалою, яка означає силу емоцій. Часто використовують шкали з оцінками від 0 до 5, де 0 і 5 — екстремальні значення негативних та позитивних емоцій відповідно [59, 60].

Існують складні системи виявлення тональності в текстах, які поєднують у собі техніки та підходи для виявлення емоційної орієнтації на різних рівнях точності та щодо різних аспектів, а потім зводять проміжні результати у кінцевий.

У цій дисертаційній роботі інструментарій тональнісного аналізу використано для виявлення емоційного навантаження повідомлень реципієнтів. Пізніше на основі цієї інформації зробимо припущення щодо емоцій автора повідомлень та встановлюємо його психічний стан.

1.4. Аналіз дискусій онлайн-спільнот за допомогою діалогічних актів

Наявність змістових зв’язків між повідомленнями є однією з ознак дискусії як діалогу. Дискусія — це діалог, реалізований за допомогою текстових засобів комунікації, наданих платформами, на яких перебувають онлайн-спільноти.

Діалог передбачає вираження думок та реагування на думки інших. У традиційному середовищі зв’язок між репліками діалогів встановлюється за допомогою невербальних (поворот голови, зоровий контакт) та вербальних (звертання) засобів. Дискусії мають характерну для діалогів структуру, тобто почергові репліки учасників, які приймають ролі інформатора та реципієнта. Ці зв’язки можуть бути явно представлені в структурі дискусії або їх можна встановити за допомогою глибшого аналізу дискусії (див. підрозділ 1.1.2 «Типова структура онлайн-спільноти»).

Наявність учасників та теми є необхідними умовами діалогу. Вони дають змогу ідентифікувати діалог. Це правило справедливе і для дискусій: коли значно міняється набір учасників чи тема (наприклад, більше ніж на 50%), то це свідчить про завершення однієї дискусії та початок іншої.

Крім чергування висловлень та наявності кількох учасників, ключовою ознакою діалогу є альтернативний розвиток. Інакше кажучи, неможливо передбачити створення певним учасником конкретного повідомлення.

Незважаючи на альтернативність розвитку, діалог характеризується змістовою єдністю повідомлень. Тема відома учасникам дискусії, і вони сприймають усі повідомлення через призму конкретної теми.

Для реплік діалогу характерна цілеспрямованість мовленнєвої дії, тобто вони несуть явну чи приховану комунікативну мету адресанта чи реципієнта (наприклад, інформування, запит, наказ, пораду, обіцянку тощо). На основі цієї ознаки визначено такий компонент діалогу як діалогічний акт. Діалогічний акт — це фрагмент репліки учасника діалогу, який відповідає одній із комунікативних цілей [61].

Діалогічні акти засновані на мовленнєвих актах, виділених Дж. Серлем. Дослідник виділив такі типи МА та визначив функцію кожного з них [62]:

- асертиви — пропозиція, стверджування, присяга, вихваляння, підсумовування (висновків);
- директиви — наказування, прохання, питання, запрошування, порада, благання;
- комісиви — обіцянка, планування, давати клятву, битися об заклад;
- експресиви — дякувати, вибачатись, вітати, шкодувати, тобто комунікативні дії, які передбачають вираження психологічного стану;
- декларативи — заява, тобто негайне констатування зміни стану об’єкта та навколишнього середовища, наприклад, «Тебе звільнено».

Аналізуючи структуру діалогів за допомогою мовленнєвих актів, Дж. Остін ввів поняття діалогічного акту як діалогічної функції висловлення в широкому значенні, незалежно від її власного семантичного наповнення [63].

Оскільки комунікація в онлайн-спільнотах відбувається в діалогічному форматі, а однакові тактики ІПМ можна застосовувати для здійснення маніпуляцій у дискусіях різної тематики, для виявлення ІПМ потрібен засіб, за допомогою якого можна ідентифікувати тактики ІПМ незалежно від контексту та аналізувати комунікацію відносно типової структури онлайн-дискусії. Саме тому ми використовуємо інструментарій анотування на основі ДА як один із засобів для виявлення тактик ІПМ.

1.4.1. Підходи до аналізу спілкування на основі діалогічних актів

Діалогічні акти використовують для аналізу комунікації за допомогою відмінних між собою засобів, в різних середовищах та для різної кількості агентів комунікації, наприклад для аналізу телефонних розмов [64], аудіо-записів конференцій та зборів [63], аналізу дискусій в соціальних мережах [65]. ДА застосовують для виявлення сарказму [66], ставлення , обману [67] та інших явищ комунікації.

Дослідження у зазначених вище напрямках починаються з отримання характеристик елементів діалогу. Анотування діалогічних актів — це

зіставлення фрагментів діалогу з діалогічними актами, тобто зазначення комунікативних функцій для фрагментів діалогу [61].

ДА анотують вручну, дотримуючись чітких вказівок, також запропоновано алгоритми та засоби автоматизованого анотування. Автоматизовані засоби дають змогу анотувати ДА на високому рівні узагальнення [68].

Анотування ДА є важливим для функціонального аналізу людського спілкування. Навіть найбільш узагальнене анотування, яке дає змогу виявити в репліках діалогів такі ДА, як твердження, питання, відповідь, успішно використовується в багатьох програмних засобах, наприклад, для автоматизованих співрозмовників [69], автоматичного реферування діалогів [70], виявлення флірту [58].

Також підходи до розроблення систем автоматизованого анотування ДА можна поділити з вчителем та без. Системи, які ґрунтуються на докладно анотованих вручну наборах ДА – це система для віддаленого навчання студентів [71], функціонал якої передбачає класифікування відповідей студентів відповідно до ДА, що забезпечує ефективнішу обробку генерованої студентами інформації. До систем, які розроблені на основі створеної У системах, які ґрунтуються на машинному навчанні без вчителя, не використовують анотованих вручну стандартних міток [73], що створює певні труднощі в оцінюванні їхньої ефективності.

На основі діалогічних актів проводили виявлення обману в чатах комп’ютерної гри [74]. Це робили на основі корпусу телефонних розмов SwitchBoard. Це один із найбільших корпусів, протегованих за допомогою ДА, але аналіз онлайн-діалогів мав певні труднощі, оскільки останні відрізняються кількістю учасників, середовищем проведення та стилем мови. Оскільки метою було виявити обман у комп’ютерних іграх, до ДА корпусу SwitchBoard було додано такі спеціалізовані теги ДА: стратегія, розміщення активів, результат, нерелевантна репліка. З передбачених схемою тегування SwitchBoard, найчастіше в діалогах щодо комп’ютерних ігор траплялися такі ДА: твердження, звертання, питання, відповідь, розгубленість.

1.4.2. Підходи до анотування діалогічних актів

Поняття діалогічних актів є основним у дослідженні діалогів, а саме у інтерпретації комунікативної поведінки учасників діалогу [65], побудові анотованих корпусів діалогів [75, 76], розробленні систем автоматизованого ведення діалогів [69].

Як було зазначено вище, в основу ДА покладено мовленнєві акти; інакше кажучи, ДА є уточненими, деталізованими мовленнєвими актами. Часто ДА спеціалізовані відносно сфери використання чи тематики. Наприклад, Питання — це мовленнєвий акт, Питання_про_готель — це тематично спеціалізований діалогічний акт, який використовується в системах аналізу телефонної чи текстової діалогічної комунікації у сфері туризму.

Прикладами неспеціалізованих ДА, які відповідають мовленнєвому акту Питання, є:

- питання-привітання («Як життя?», «How do you do?», «Alles klar?»)
- метапитання («Що тут зробиш?», «Що тут скажеш?», «What can I say?», «Was soll ich tun?»)
- загальні питання («Чи їздить громадський транспорт після 24:00?», «Are you coming tomorrow?», «Reisen Sie gern?»)
- питання-прохання («Чи не могли б Ви допомогти?», «Could you tell me the time, please?», «Können Sie mir bitte sagen, wie…?»)
- спеціальні питання («Як пройти до центру?», «What would you like to start with?», «Welcher Tag ist heute?»)
- питання-пропозиція ідеї («Може поїдем у гори?», «Why don't you go ahead?», «Wollen wir spazieren gehen?»)
- питання-пропозиції («Хочеш яблуко?», «Would you like some coffee?», «Möchten Sie ein Stück Kuchen?»)
- питання про вибір («Пішки підеш чи на трамваї?», «Would you like white coffee or black?», «Einzel oder hin und zurück?»).

Для дослідження діалогічної комунікації в Twitter Е. Зарішева і Т. Шефлер розробили схему анотування ДА [61]. Схема ґрунтується на

універсальній таксономії діалогічних актів DIT++ (Dynamic Interpretation Theory) [68].

Е. Зарішева і Т. Шефлер адаптували таксономію діалогічних актів DIT++ для виявлення ДА у Twitter: з 88-ми ДА, які містить DIT++, залишили ДА, наведені на рис. 1.12 і рис. 1.13. ДА, які знаходились на найнижчих рівнях таксономії DIT++, було видалено, оскільки їх могли ідентифікувати лише експерти; метою Е. Зарішева і Т. Шефлер було забезпечити можливість анотування ДА некваліфікованим персоналом чи автоматизованими системами.

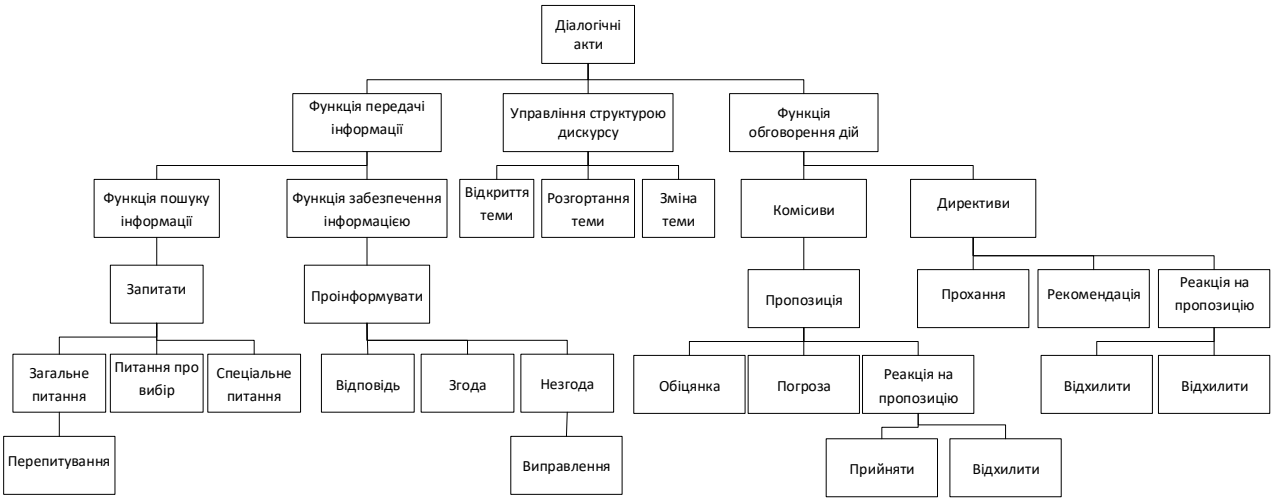


Рис. 1.12. Адаптована таксономія ДА DIT++ (Частина 1)

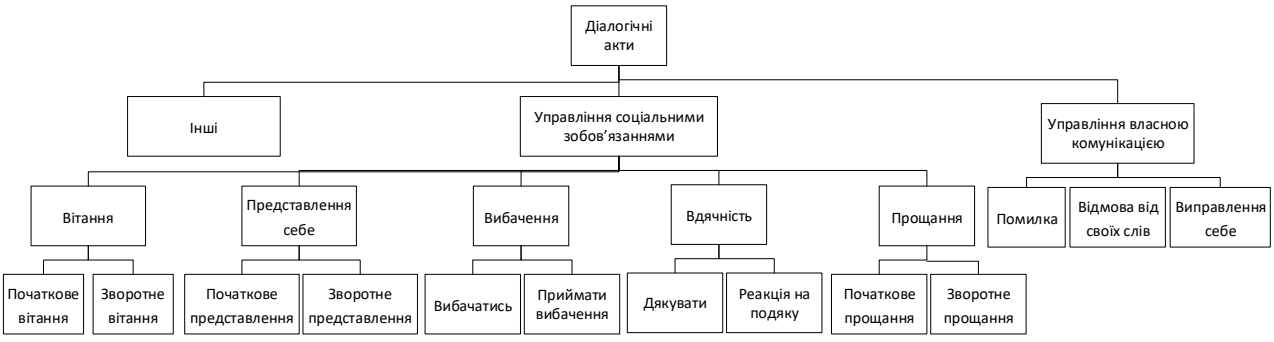


Рис. 1.13. Адаптована таксономія ДА DIT++ (Частина 2)

Оскільки наведена вище схема анотування ДА передбачена для ідентифікації ДА в тематично неспеціалізованій комунікації типи людина-людина та рівень деталізації ДА допускає автоматизоване анотування, то цю класифікацію візьмемо за базу для виявлення ДА в дискусіях онлайн-спільнот.

Існує засіб для автоматичного анотування за п’ятьма типами ДА [65]. Це анотування здійснюється на основі близько 530 маркерів, якими є n-грами (в цьому випадку 1-, 2-, 3-грами) зі слів та з символів. Засіб створений на основі методів машинного навчання із вчителем. Частково вручну, частково за допомогою автоматизованих засобів, дослідники проанотували 8613 повідомлень з Twitter, анотовані дані поділили на навчальну вибірку (90%) та вибірку для тестування (10%) та класифікували за допомогою лінійної опорно-векторної машини. Проблему сортування на кілька класів вирішили за допомогою методу «один проти решти».

Описаний вище метод доцільно застосувати для створення засобу для автоматичного анотування ДА для україномовного інформаційного наповнення. Оскільки поділ на п’ять класів надто узагальнений для розроблення маркерів для виявлення ІПМ, то пропонуємо поділити спершу на класи високого рівня ієрархії, а за таким самим способом поділити виділені класи на підкласи.

Висновки до розділу

У першому розділі проаналізовано існуючі види онлайн-спільноти, їхні структурні та функціональні відмінності. Розглянуто специфіку комунікації в онлайн-спільнотах, наявні засоби та способи комунікації. Проаналізовано існуючу класифікацію прийомів маніпуляції в традиційному офлайн-середовищі, виділено спільні та відмінні ознаки традиційної маніпуляції та ІПМ, запропоновано адаптацію схем тактик традиційної маніпуляції для виявлення ІПМ в інтерактивних текстових онлайн-спільнотах. Проведено огляд інструментарію тональнісного аналізу та визначено інструменти для застосування з метою виявлення психічного стану учасника онлайн-спільноти. Проаналізовано структуру онлайн-дискусії як діалогу, оглянуто існуючі класифікації діалогічних актів, запропоновано набір діалогічних актів, який підходить для вирішення проблеми виявлення ІПМ.

Розділ 2. Формальні моделі онлайн-спільноти, інформаційно-психологічної маніпуляції та фільтрів для виявлення підозрілих дискусій

У другому розділі на основі аналізу типових онлайн-спільнот запропоновано формальну модель онлайн-спільноти, яка відображає необхідну для виявлення ІПМ інформацію. Розроблено формальну модель тактики ІПМ, яка ґрунтується на кусково-лінійних агрегатах. Формальна модель тактики ІПМ передбачає маркери виявлення прийомів ІПМ на основі діалогічних актів, семантичних змінних та синтаксичних структур, а також формальний опис психічних станів учасника спільноти, які свідчать про вплив за допомогою ІПМ. Розроблено критерії перевірки на наявність ІПМ в онлайн-спільнотах, які поділено на статичні та динамічні за даними, які необхідні для їх розрахунку. На основі критеріїв побудовано фільтри для виявлення підозрілих фрагментів дискусії. Введено поняття семантичної змінної та описано структуру класів семантичних змінних.

Основні результати розділу опубліковані у працях [77, 78, 79, 80, 81, 82, 83, 84].

2.1. Спеціальна модель онлайн-спільноти з погляду виявлення ІПМ

Вибір 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. Формальну модель онлайн-спільноти для побудови методів і засобів виявлення інформаційних загроз віртуальних спільнот в інтернет-середовищі соціальних мереж описано в [85].

Оскільки завданням цієї дисертаційної роботи є розроблення методів і засобів виявлення інформаційно-психологічної маніпуляції, основними об’єктами дослідження є інформаційне наповнення, а також автори його елементів, адже за допомогою інформації про авторів можна ідентифікувати джерело небезпечних повідомлень.

За основу взято формальну модель віртуальної спільноти (2.1), запропоновану у [6], адже вона відповідає загальній структурі, характерній для усіх типів віртуальних спільнот. Елементи множин, з яких складається ця формальна модель, описано та деталізовано відповідно до завдання виявлення ІПМ.

Формальну модель віртуальної спільноти подано як кортеж:

$$Community = \langle Content, Member \rangle, \tag{2.1}$$

де $Content = \{Discussion_j\}_{j=1}^{N^{Discussion}}$ — множина дискусій, в яких беруть участь учасники спільноти; $Member = \{Member_j\}_{j=1}^{N^{Member}}$ — множина учасників.

Ця формальна модель виокремлює два відмінні з погляду виявлення ІПМ в онлайн-спільнотах види даних: дані щодо дискусії та дані щодо профілю учасника. Відповідно до цієї формальної моделі критерії виявлення ІПМ поділяються на динамічні та статичні. Статичні критерії розраховуються відносно профілів учасників онлайн-спільноти, динамічні для конкретних повідомлень, які містить спільнота (див. розд. 2.3 «Розроблення »). Статичні критерії використовують для ідентифікації підозрілих фрагментів онлайн-дискусії (див. розд. 3.3 «Методи виявлення підозрілих фрагментів дискусії»), натомість динамічні переважно для виявлення прецедентів ІПМ та потребують глибинного аналізу текстового наповнення дискусії (див. розд. 3.5 «Методи виявлення прийомів ІПМ у дискусіях онлайн-спільнот»).

2.1.1. Формальна модель дискусії

Інформаційне наповнення онлайн-спільноти складається з множини дискусій (2.1). З погляду змісту інформаційного наповнення описуємо дискусію за допомогою таких характеристик: назви, теми та створеного користувачами інформаційного наповнення, тобто множини повідомлень дискусії.

Крім того, кожну дискусію створює учасник онлайн-спільноти. Інформація щодо учасника-автора дискусії є важливою з погляду виявлення ІПМ, зокрема, на її основі можна віднести дискусію до списку дискусій, які необхідно перевірити на наявність ІПМ, наприклад, якщо автор дискусії є потенційним маніпулятором.

Для сортування дискусій за сприятливістю для здійснення ІПМ та створення черги з дискусій для перевірки на наявність ІПМ (див. підрозділ 3.1.1 «Підготовчий етап») важливо знати дату створення першого повідомлення дискусії, дату останнього повідомлення та кількість учасників. Також ці дані необхідні для виявлення підозрілих фрагментів дискусії (див. підрозділ 2.3.2 «Формальна модель системи фільтрів для виявлення підозрілих фрагментів дискусій»).

Формальну модель дискусії задано таким кортежем:

$$Discussion_i = \langle DiscussionTitle_i, DiscussionTopic_i, DiscussionAuthor_i, Messages_i, FirstMessageDate_i, LastMessageDate_i, ParticipantCount_i \rangle, \quad (2.2)$$

де $DiscussionTitle_i$ — назва i-ї дискусії; $DiscussionTopic_i$ — тема 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.

Ця формальна модель дискусії дозволяє тематично ідентифікувати дискусію та дозволяє отримати дані про дискусію необхідні для виконання дій на всіх етапах алгоритму виявлення ІПМ.

2.1.2. Формальна модель учасника спільноти

Маніпулюючи, маніпулятор залишає слід, який проявляється на різних організаційно-структурних рівнях, а саме: в спільноті, дискусії, повідомленні. Тому аналізувати процес комунікації в онлайн-дискусіях потрібно через призму кожного з рівнів. Рівні відрізняються підходом до виявлення ознак ІПМ, передбачають різну складність обчислення, збору даних та точність результатів.

Статичні критерії ІПМ поділяємо відповідно до цих трьох рівнів, оскільки статичні критерії пов’язані з конкретним профілем учасника, то формальна модель учасника спільноти має містити дані для розрахунку цих критеріїв.

Модель учасника онлайн-спільноти (2.3) містить ім’я учасника, емейл, посилання на профіль у Facebook (за наявності), роль учасника дискусії з погляду ІПМ, набір характеристик поведінки учасника, характеристик профілю та історію дій. Ці характеристики необхідні для однозначної ідентифікації учасника спільноти, виявлення підозрілого профілю та підозрілих фрагментів дискусії.

Модель учасника виглядає так:

$$\begin{aligned} Member_i = \langle & MemberName_i, Email_i, FacebookAccount_i, IPMRole_i, \\ & BehaviouralCharacteristics_i, ProfileFeatures_i, RegistrationDate_i, RulesViolationCount_i \rangle, \end{aligned} \quad (2.3)$$

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; $RegistrationDate_i$ — дата реєстрації учасника в онлайн-спільноті; $RulesViolationCount_i$ — кількість зафіксованих модератором порушень правил онлайн-спільноти.

$IPMRole_i$ — роль учасника у дискусії з погляду виявлення ІПМ, цей елемент може набувати одне зі значень, які належать до множини:

$$IPMRoleValue = \{Neutral; Manipulator; Victim; Zombie\}, \quad (2.4)$$

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

Neutral — це роль звичайного учасника онлайн-спільноти, яка полягає в усвідомленому обміні інформацією.

Manipulator — це роль, яка полягає у застосуванні прихованого впливу до учасників дискусії, щоб нав’язати певні дії або світоглядну позицію.

Victim — це роль жертви маніпуляції, тобто це учасник спільноти, прояви емоційних станів якого відповідають послідовності змін емоційного стану, передбаченій тактикою ІПМ.

Zombie — це роль, яку може набути жертва маніпуляції (*Victim*), яка, засвоївши нав’язану маніпулятором ідею, сприяє поширенню цієї ідеї.

Роль учасника спільноти з погляду ІПМ важлива для виявлення можливих шляхів поширення ІПМ (див. розд. 3.1.3 «Етап нейтралізації»).

Аналізуючи поведінкові характеристики профілю (2.5), можна виявити аномалії, наприклад, розміщення повідомлень лише у робочий час (наприклад, з 9 до 17), відсутність реакції на коментарі до власного допису, участь у дискусії лише певної тематики. Якщо в поведінкових ознаках профілю виявлено кілька таких аномалій, то його діяльність у спільноті варто проаналізувати на наявність ІПМ (див. розд. 2.3.2 «Формальна модель системи фільтрів для виявлення підозрілих фрагментів дискусій»).

$$BehaviouralCharacteristics_i = \langle ReplyCount_i, PublishingFrequency_i, ActivityTime_i, SelfCentradActiity_i, EngagementDiscussionThemes_i, PublishedMessageCount_i, EngagementDiscussionThemes_i, InitiatedDiscussionCount_i \rangle, \quad (2.5)$$

де $ReplyCount_i$ — частота реакцій на дописи інших; $PublishingFrequency_i$ — кількість дописів учасника за певний відрізок часу; $ActivityTime_i$ — період доби, в який учасник активний у спільноті; $SelfCenteredActivity_i$ — частота відповідей на коментарі до власних дописів; $EngagementDiscussionThemes_i$ — кількість дискусій, в яких взяв участь учасник; $PublishedMessageCount_i$ — кількість опублікованих повідомлень; $InitiatedDiscussionCount_i$ — кількість розпочатих учасником дискусій.

$EngagementDiscussionThemes_i = \{EngagementDiscussionTheme_{ij}\}_{j=1}^{N^{EngagementDiscussionTheme_i}}$ — це множина тематик дискусій, в яких учасник узяв участь.

Деякі веб-форуми дозволяють учасникам увійти за допомогою профілю з соціальних мереж. Багато учасників користуються цією можливістю, як наслідок можна отримати адресу профілю учасника в соціальній мережі. Профілі учасників у соціальних мережах передбачають розміщення набагато ширшої інформації про учасника, ніж веб-форуми. Необхідно пов'язувати профіль учасника (2.6) у соціальних мережах з профілями учасника на веб-форумах, тому що ця інформація значно покращить ефективність виявлення ІПМ. Інформаційні структури соціальних мереж є відмінними між собою, але на основі базових елементів структури, які містять більшість соціальних мереж розроблено кортеж характеристик профілю учасника:

$$\mathrm{Pr}ofileFeatures_i = \langle FilledOutFieldsCount_i, FriendCount_i, PhotoCount_i, OwnPostInTimelineCount_i, OthersPostInTimelineCount_i \rangle, \quad (2.6)$$

де $FilledOutFieldsCount_i$ — кількість заповнених полів профілю учасника; $FriendCount_i$ — кількість друзів учасника; $PhotoCount_i$ — кількість фотографій

учасника; $OwnPostInTimelineCount_i$ — кількість власних постів у життєписі; $OthersPostInTimelineCount_i$ — кількість чужих постів у життєписі.

Наведені характеристики профілю необхідні для:

- однозначної ідентифікації профілю;
- розроблення критеріїв для виявлення підозрілих профілів та підозрілих фрагментів дискусії (див. розд. 2.3 «Розроблення »);
- для встановлення можливих шляхів поширення ІПМ (див. розд. 3.1.3 Етап нейтралізації).

Не у всіх спільнотах можна отримати інформацію щодо усіх характеристик учасника. Відповідно набір критеріїв для побудови фільтрів, які використовуватимемо під час аналізу комунікації з метою виявлення ІПМ, залежить від інформаційної структури платформи, на якій розміщена онлайн-спільнота.

2.1.3. Формальна модель повідомлення

Формальна модель повідомлення (2.7) складається з автора повідомлення, дати й часу розміщення повідомлення, типу повідомлення за спрямуванням та показника релевантності повідомлення, а також базових обов'язкових характеристик та додаткових характеристик, необхідних для виявлення ІПМ:

$$\begin{aligned} Message_i = \langle & AuthorNickname_i, TimeStamp_i, MessageContent_i, \\ & DirectionType_i, Relevance_i, BasicFeatures_i, AdditionalFeatures_i, IPMFeatures_i, Lexis_i \rangle, \end{aligned} \tag{2.7}$$

де $AuthorNickname_i$ — 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 тематики спільноти; $BasicFeatures_i$ — кортеж обов'язкових характеристик повідомлення; $AdditionalFeatures_i$ — кортеж

додаткових характеристик повідомлення; $IPMFeatures_i$ — маркери прийомів ІПМ на основі вербальних, невербальних та паравербальних ознак; $Lexis_i$ — це вектор, який характеризує наповнення специфічною лексикою даного повідомлення.

Тип спрямування повідомлення може бути таким:

$$DirectionType_i \in \{Stimulus, Reaction\}, \tag{2.8}$$

де *Stimulus* — це ініціюювальні повідомлення, в яких надається певна інформація, *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 дискусії (див. 1.1.3 «Текстові види спілкування в онлайн-спільнотах та їхні характеристики»).

Показник релевантності повідомлення може набувати значень, наведених у формулі (2.9). Релевантність розраховується на основі Tree Edit Distance — пошуку послідовності операцій редагування для перетворення одного дерева в інше з найменшою вагою [86].

$$Relevance_i \in \{\text{Re}levant, Irrelevant\}, \tag{2.9}$$

де *Relevant* — це повідомлення, пов’язані за змістом із тематикою онлайн-дискусії; *Irrelevant* — це повідомлення, не пов’язані за змістом із тематикою онлайн-дискусії.

Базові характеристики (2.10) повідомлення необхідні для виявлення підозрілих фрагментів ІПМ за статичними критеріями рівня повідомлення.

$$\begin{aligned} BasicFeatures_i = \langle & Length_i, TextSymbolCount_i, MetagraphemicsCount_i, \\ & EmoticonCount_i, LinkCount_i \rangle, \end{aligned} \tag{2.10}$$

де $Length_i$ — кількість текстових символів у повідомленні (літер, знаків пунктуації, цифр, пробілів); $TextSymbolCount_i$ — кількість літер, знаків пунктуації та цифр, використаних для передачі вербальної інформації; $MetagraphemicsCount_i$ — елементів метаграфеміки, наприклад: кластерів символів, які не містять жодної семантики з мовної точки зору і не входять у посилання, відступів, пробілів або регістру, які не відповідають певним нормам, наприклад, більше ніж два пробіли, більше ніж два пропущені рядки, слова написані великими буквами, перекреслені слова; $LinkCount_i$ — кількість посилань у повідомленні; $EmoticonCount_i$ — це кількість емотиконів у повідомленні.

Етап нейтралізації алгоритму виявлення ІПМ (див. розд. 3.1.3 «Етап нейтралізації») використовує для ідентифікації угрупувань маніпулятивних профілів дані про стиль учасника. Наприклад, порівнюючи стилі підозрілих учасників та профілів-маніпуляторів можна виявити профілі-спільники, тобто профілі, які контролює одна реальна особа-маніпулятор або профілі, які діють згідно з однаковими чітко прописаними вказівками.

Стиль визначають на основі аналізу повідомлень, які створив учасник. Характеристики повідомлення, за якими можна визначити стиль учасника, об’єднані в групу додаткові характеристики (6). Додаткові характеристики повідомлення подані як кортеж параметрів:

$$AdditionalFeatures_i = \langle WordCount_i, TermCount_i, SentenceLenght_i, ErrorPercentage_i, FrequentWords_i \rangle, \quad (2.11)$$

де $WordCount_i$ — кількість слів у повідомленні; $TermCount_i$ — кількість унікально вжитих слів у повідомленні; $SentenceLength_i$ — середня довжина речення в повідомленні; $ErrorPercentage_i$ — відсоток слів, написаних із помилками відносно всіх слів повідомлення; $FrequentWords_i$ — множина найчастіше вживаних слів.

Повідомлення, які входять до підозрілих фрагментів дискусії, перевіряють на наявність маркерів прийомів ІПМ. Для цього необхідно зібрати інформацію про синтаксичну структуру, ДА та семантичні змінні, наявні в повідомленні, та порівняти їх із вербальними маркерами прийомів ІПМ. Також виявити у повідомленні посилання, емотикони та метаграфеміку і порівняти їх із доповнювальними маркерами прийому ІПМ (2.12). Тобто в повідомленні треба виявити елементи, які відповідають елементам форми реалізації прийому ІПМ (див. розд. 2.2.2 «Формальна модель прийому ІПМ»).

$$IPMFeatures_i = \langle SyntacticStructureDetected_i, DialogActDetected_i, SemanticVariableDetected_i, ComplementaryMarkersDetected_i \rangle, \quad (2.12)$$

де $SyntacticStructureDetected_i$ — це множина синтаксичних структур, які містить повідомлення, $DialogActDetected_i$ — це множина діалогічних актів, які містить повідомлення, $SemanticVariableDetected_i$ — це множина семантичних змінних, які містить повідомлення; $ComplementaryMarkersDetected_i$ — це множина додаткових маркерів, які містить дане повідомлення.

Невербальні та паравербальні елементи повідомлення, за якими можна ідентифікувати наявність прийому ІПМ, подані таким кортежем:

$$ComplementaryMarkersDetected_i = \langle LinkDetected_i, EmoticonDetected_i, MetagraphemicsDetected_i \rangle, \quad (2.13)$$

де $LinkDetected_i$ — це множина кортежів, яка містить інформацію про тип і кількість посилань відповідного типу у цьому повідомленні; $EmoticonDetected_i$ — це множина кортежів, яка містить інформацію про назву і кількість відповідних емотиконів у цьому повідомленні; $MetagraphemicsDetected_i$ — це множина кортежів, яка містить інформацію про назву і кількість відповідних елементів метаграфеміки у цьому повідомленні.

Множина кортежів, яка містить інформацію про виявлені в повідомленні прийоми ІПМ, реалізовані за допомогою посилань:

$$LinkDetected_i = \left\{ \langle LinkType_{ij}, LinkNumber_{ij} \rangle \right\}_{j=1}^{N^{\left(LinkDetected_i \right)}}, \quad (2.14)$$

де $LinkType_{ij}$ — тип посилання, яке використовується для реалізації цього прийому; $LinkNumber_{ij}$ — кількість посилань певного типу використаних у повідомленні для реалізації прийому.

Множина кортежів, яка містить інформацію про виявлені в повідомленні прийоми ІПМ реалізовані за допомогою емотиконів:

$$EmoticonDetected_i = \left\{ \langle EmoticonName_{ij}, EmoticonNumber_{ij} \rangle \right\}_{j=1}^{N^{\left(EmoticonDetected_i \right)}}, \quad (2.15)$$

де $EmoticonName_{ij}$ — назва емотикону, який використовується для реалізації цього прийому; $EmoticonNumber_{ij}$ — кількість емотиконів використаних у повідомленні для реалізації прийому;

Множина кортежів, яка містить інформацію про виявлені в повідомленні прийоми ІПМ реалізовані за допомогою метаграфеміки:

$$MetagraphemicsDetected_i = \left\{ \langle MetagraphemicsName_{ij}, MetagraphemicsNumber_{ij} \rangle \right\}_{j=1}^{N^{\left(MetagraphDet_i \right)}}, \quad (2.16)$$

де $MetagraphemicsName_{ij}$ — назва елементу метаграфеміки, який використовують для реалізації цього прийому; $MetagraphemicsNumber_{ij}$ — кількість елементів метаграфеміки, використаних у повідомленні для реалізації прийому.

$Lexis_i$ — це кортеж, елементами якого є кількість ужитих у повідомленні одиниць специфічної лексики, а також пасивної лексики, зокрема неологізмів та застарілих слів. До видів специфічної лексики, які є в кортежі, належать: експресивна лексика, арго, жаргон, ненормативна лексика, просторіччя, діалектизми, терміни, неологізми. Ці види лексики виявляємо у дискусіях онлайн-спільнот, використовуючи спеціалізовані словники. Від наявності одиниць специфічної лексики залежить ступінь інтенсивності прийому $Intensity_i$ (див. розд. 2.2.2 «Формальна модель прийому ІПМ»).

Отже, лексичні особливості повідомлення характеризуються таким кортежем параметрів:

$$Lexis_i = \langle LoadedLanguage_i, Argot_i, Jargon_i, Profanities_i, SimpleLanguage_i, Dialect_i, Terminology_i, Neologism_i \rangle, \quad (2.17)$$

де $LoadedLanguage_i$ — кількість одиниць експресивної лексики у висловленні; $Argot_i$ — кількість арготизмів у висловленні; $Jargon_i$ — кількість жаргонізмів у висловленні; $Profanities_i$ — кількість одиниць ненормативної лексики у висловленні; $SimpleLanguage_i$ — кількість одиниць просторіччя у висловленні, $Dialect_i$ — кількість діалектизмів у висловленні, $Terminology_i$ — кількість термінів у висловленні, $Neologism_i$ — кількість неологізмів у висловленні.

Стиль спілкування у різних онлайн-спільнотах різниться між собою: для певних спільнот властиве використання жаргону, для інших — просторіччя. Тому необхідно розраховувати значущість специфічної лексики в повідомленні за ваговими показниками кожного виду специфічної та пасивної лексики, які визначили експерти для конкретної спільноти. Значущість специфічної лексики, як характеристики повідомлення з погляду ІПМ, розраховують за такою формулою:

$$WeightedLexis_i = \mathrm{Lexis}_i Weight_i, \quad (2.18)$$

де $WeightedLexis_i$ — вектор, який містить інформацію про значущість певних видів специфічної лексики у цьому повідомленні; Lexis_i — вектор, який характеризує наповненість цього повідомлення специфічною лексикою; $Weight_i$ — це вектор з ваговими показниками для кожного виду специфічної лексики.

Вагові показники для специфічних видів лексики встановлюють експерти, наприклад, на основі частоти вживання цих видів лексики у повідомленнях спільноти (для цього потрібен корпус повідомлень, в якому розмічено специфічні види лексики).

У прикладі 2.1 продемонстровано наявність просторіччя у прецеденті застосування прийому «ситуативний перехід на особисте».

Приклад 2.1. Прецедент прийому «ситуативний перехід на особисте»

«Якщо ти професіонал то зациниш Lenovo ideapad 700!»

Насиченість специфічною лексикою цього прецеденту є такою:

$$Lexis_i = \langle 0,0,0,0,1,0,0,0 \rangle . \quad (2.19)$$

2.2. Формальна модель інформаційно-психологічної маніпуляції

ІПМ ґрунтуються на психологічних механізмах сприйняття інформації та здійснюються за допомогою наявних засобів передачі інформації.

На підсвідомість реципієнтів впливають унаслідок застосування прийомів ІПМ, а досягнення мети ІПМ забезпечується завдяки дотриманню кроків тактики ІПМ.

Психічний та емоційний стан реципієнтів впливає на сприйняття повідомлення. Тому ключовою характеристикою тактик ІПМ є послідовне переведення жертви ІПМ зі стану в стан, з поданням інформації, яка засвоюється відповідним до стану чином і в кінцевому результаті приводить до того, що світогляд чи дії жертви відповідають меті ІПМ.

Зміни психічного стану реципієнта досягають унаслідок застосування прийомів ІПМ. Вони реалізовуються за допомогою звичних для текстових повідомлень одиниць комунікації: текстових символів, метаграфічних об’єктів, емотиконів та посилань.

Прийому ІПМ поставимо у відповідність множину наборів перелічених вище одиниць, які трапляються у текстових реалізаціях цих прийомів. З погляду виявлення ІПМ ці набори є маркерами прийомів ІПМ.

Водночас маркери прийомів ІПМ можуть траплятися і у звичайних, незловмисницьких повідомленнях.

Констатувати факт здійснення ІПМ можна на основі повідомлень з маркерами прийомів ІПМ, послідовність появи яких відповідає послідовності переходів між станами однієї з тактик ІПМ.

2.2.1. Формальна модель тактики ІПМ

Ключовими елементами формальної моделі тактики ІПМ є психічний стан реципієнта та крок ІПМ.

Стан, в якому перебуває людина, суттєво впливає на сприйняття ідеї. Психічний стан – один із можливих перцептивних режимів життєдіяльності людини, що на фізіологічному рівні відрізняється ступенем енергійності, а на психологічному рівні — системою соціально-психологічних фільтрів, що забезпечують специфічне сприйняття себе та зовнішнього світу [87]. Психологічні стани є реакцією на зовнішні подразники.

В основу багатьох схем маніпуляції покладено психічні стани, розумові стани, стани его, емоційні стани, настрій чи фрейми, тобто нестійкі психологічні стани. Ці всі поняття використовують у дослідженнях психологічних маніпуляцій у традиційному середовищі. Наприклад, згідно з транзакційним аналізом [28] жертва може перебувати в трьох его-станах: Батько, Дорослий. Дитина. Чалдіні у своїх дослідженнях не виділяє станів жертви, але описує емоції та хід думок реципієнта на кожному з етапів проведення маніпуляції [88]. Кожному з цих станів притаманний певний спосіб думок та сприйняття навколишньої дійсності. У цій дисертаційній роботі пропонуємо універсальну модель, за допомогою якої можна формалізувати тактики маніпуляції, описані за допомогою одного з описаних вище підходів, транзакційного аналізу чи нейролінгвістичного програмування, а також характерні для онлайн-спільнот тактики ІПМ, виявлені внаслідок експертного аналізу комунікації в онлайн-спільнотах.

Зміна психологічних станів під час комунікації не є частим та швидким явищем. Якщо кількість станів, у яких перебував реципієнт протягом певного проміжку часу, перевищує встановлену експертами кількість, то це свідчить про зміну станів унаслідок впливу на підсвідомість. Перехід реципієнта в інший стан може спричинити кількаразове застосування прийому ІПМ.

Отже, виділимо дві ознаки застосування тактики ІПМ:

- послідовна зміна психічних станів реципієнта;
- кількаразове застосування прийому ІПМ.

На основі зазначених вище ознак проводимо аналіз онлайн-дискусії на етапі виявлення ІПМ (див. розд. 3.1.2 «Етап виявлення»).

Для формального представлення тактики ІПМ як послідовної зміни станів реципієнта внаслідок зовнішніх впливів використано кусково-лінійний агрегат. Кусково-лінійний агрегат в кожний часовий момент характеризується одним із внутрішніх станів, які належать до множини внутрішніх станів. Агрегат сприймає вхідні сигнали та змінює стан залежно від сигналу і відповідно до функції переходів. Особливістю кусково-лінійного агрегату є те, що множини станів і вхідних сигналів, конкретизуються за допомогою векторів параметрів [89].

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

$$TacticModel = \langle TacticState, TacticStep, ChangeStateFunction \rangle, \quad (2.20)$$

де $TacticState = \{TacticState_i\}_{i=1}^{N^{TacticState_i}}$ — множина станів даної тактики ІПМ, де $N^{TacticState_i}$ — кількість психічних станів, які використовуються в даній тактиці; $TacticStep = \{TacticStep_i\}_{i=1}^{N^{TacticStep_i}}$ — це множина кроків тактики ІПМ, які переводять реципієнта в певний психічний стан передбачений даною тактикою, де $N^{TacticStep_i}$ — кількість кроків тактики ІПМ; $ChangeStateFunction$ — це функція зміни психічних станів реципієнта.

Психічний стан реципієнта, який передбачений тактикою ІПМ, представимо так:

$$TacticState_i = \langle StateTitle_i, StateParameters_i, TacticStep_i \rangle, \quad (2.21)$$

де $StateTitle_i$ — це назва психічного стану; $StateParameters_i$ — це вектор параметрів, який описує цей стан; $TacticStep_i$ — це крок тактики, який спричиняє перехід реципієнта у даний стан з попереднього.

Вектор параметрів, який описує психічний стан реципієнта, внаслідок застосування тактики ІПМ подамо так:

$$StateParameters_i = \left(pState_1^{(i)}, ... pState_k^{(i)}\right). \quad (2.22)$$

Кожен із параметрів вектору стану може набувати значення зі заданого інтервалу $pState_k^{(i)} \in [\min_k^i, \max_k^i]$.

До множини станів кожної тактики належить початковий стан $InitialState_i$, всі параметри цього стану дорівнюють нулю, а множина кроків, які переводять у початковий стан — пуста.

Крім множини психічних станів, яких може набувати реципієнт, тактика ІПМ складається з множини кроків, які спричиняють перехід реципієнта в один із цих станів. Крок тактики ІПМ характеризується кортежем таких елементів:

$$TacticStep_i = \langle StepTitle_i, StepParameters_i, Tool_i \rangle, \quad (2.23)$$

де $StepTitle_i$ — назва кроку; $StepParameters_i$ — це вектор параметрів, який описує даний крок; $Tool_i \subset \{Tool_k\}_{k=1}^{N^{Tool_i}}$ — це підмножина усіх прийомів ІПМ, за допомогою яких можна реалізувати даний крок, де N^{Tool_i} — це кількість прийомів ІПМ, які можна використати для реалізації даного кроку.

Вектор параметрів, який характеризує даний крок ІПМ подамо так:

$$StepParameters_i = \left(pStep_1^{(i)}, ... pStep_k^{(i)}\right). \quad (2.24)$$

Кожен із параметрів вектору стану може набувати значення зі заданого інтервалу $TacticStep_i \in [\min^i, \max^i]$.

У конкретній тактиці маніпуляції розмір вектору параметрів психічного стану дорівнює розмірові вектору параметрів кроку.

Функція переходів має такий вигляд:

$$TacticState_{i+1} = \langle StateTitle_{i+1}, pState_1^{(i)} + pStep_1^{(i)}, ..., pState_k^{(i)} + pStep_k^{(i)}, TacticStep_i \rangle. \quad (2.25)$$

Тобто параметри кроків тактики ІПМ додаються до відповідних параметрів вихідного психічного стану реципієнта, якщо хоч один із параметрів вихідного стану виходить за допустиму межу, реципієнт ІПМ переходить у наступний стан.

2.2.2. Формальна модель прийому ІПМ

Як було описано у попередньому підрозділі, для переведення реципієнта у певний стан ІПМ потрібно виконати крок ІПМ. Останній можна виконати, застосувавши прийом або кілька прийомів, сума відповідних параметрів яких дорівнює параметрам кроку тактики ІПМ.

Прийоми ІПМ реалізовуються як звичайне висловлення за допомогою наявних засобів текстових повідомлень. Прихований зміст повідомлень із прийомами ІПМ важко виявити людині без спеціальної підготовки, тому для автоматизованого виявлення прийомів ІПМ розробимо формальну модель прийому ІПМ, яка передбачає наявність маркерів.

Оскільки прийом ІПМ містить характерне поєднання синтаксичної структури, лексики та метаграфеміки, кожному прийому можна поставити у відповідність множину маркерів.

Кожен прийом ІПМ має назву, мету, описаний вектором параметрів, приймає одну з можливих форм реалізації та характеризується певним ступенем інтенсивності. Прийом ІПМ формалізовано так:

$$Tool_k = \langle ToolTitle_k, ToolGoal_k, ToolParameters_k, ToolVariation_k, Intensity_k, ToolPointer_k \rangle, \quad (2.26)$$

де $ToolTitle_k$ — назва прийому; $ToolGoal_k$ — мета прийому, спрямована на зміну психічного стану реципієнта; $ToolParameters_k$ — це вектор параметрів, який характеризує прийом ІПМ; $ToolVariation_k$ — це множина форм реалізації прийому ІПМ; $Intensity_k$ — це ступінь інтенсивності прийому ІПМ; $ToolPointer_k$ — це вказівник прийому.

Вектор параметрів прийому ІПМ подамо так (2.27). Кожному параметру прийому ІПМ відповідає одне числове значення.

$$ToolParameters_k = \left(pTool_1^{(k)},..., pTool_l^{(k)}\right). \quad (2.27)$$

Множину форм реалізації прийому подамо таким чином:

$$ToolVariation_k = \{Form_m\}_{m=1}^{N^{ToolVariation_k}}, \quad (2.28)$$

де $N^{ToolVariation_k}$ — це кількість форм реалізації даного прийому

Форми реалізації прийомів ІПМ отримуємо, проаналізувавши прецеденти ІПМ, тобто поділивши їх на класи за певними лексичними, синтаксичними та прагматичними, метаграфічними та невербальними ознаками. Виділені класи — це будуть форми реалізації прийомів ІПМ, а репрезентативні ознаки — це маркери форм реалізації ІПМ.

Вербальні засоби переважають за частотою використання та значущістю в текстовому поданні інформації, тому аналізуємо їх за допомогою трьох підходів, а саме: з погляду діалогічних актів, концептуального складу та синтаксичної структури. Відповідно форма реалізації прийому має таку структуру:

$$Form_i = \langle SyntacticStructure_i, DialogAct_i, SemanticVariable_i, ComplementaryMarkers_i \rangle, \quad (2.29)$$

де $SyntacticStructure_i$ — це множина синтаксичних структур, які можуть бути використані для реалізації даного прийому, $DialogAct_i$ — це множина діалогічних актів, за допомогою яких може бути реалізований прийом ІПМ, $SemanticVariable_i$ — це множина семантичних змінних, з яких може бути

вибудуваний прийом; $ComplementaryMarkers_i$ — це множина додаткових маркерів, які представляють інформацію передану за допомогою недомінантних засобів текстового подання.

Застосування трьох типів вербальних маркерів для виявлення прийомів ІПМ дає змогу вловити більше прецедентів ІПМ, ніж використання маркерів одного типу.

Кожній формі реалізації прийому ІПМ поставлена у відповідність множина синтаксичних маркерів.

$$SyntacticStructure_i = \left\{SyntacticStructure_{ij}\right\}_{j=1}^{N^{(SyntacticStructure_i)}}, \quad (2.30)$$

де $SyntacticStructure_i$ — маркер синтаксичної реалізації прийому ІПМ поданий у нотації Бекуса-Наура, $N^{SyntacticStructure_i}$ — кількість синтаксичних маркерів форми реалізації даного прийому.

Оскільки дискусія в онлайн-спільноті має діалогічний характер (див. 1.4 «Аналіз дискусій онлайн-спільнот за допомогою діалогічних актів»), то діалогічні акти використовуються як маркери прийомів ІПМ. Діалогічний акт — мінімальна одиниця діалогічного спілкування, — це процес створення конкретного висловлення, наділеного змістом та забезпеченого комунікативною метою, який передається від адресанта до адресата [90].

Множину діалогічних актів, які є маркерами певної форми реалізації прийому ІПМ, подаємо таким чином:

$$DialogAct_i = \left\{DialogAct_{ij}\right\}_{j=1}^{N^{(Dialog Act_i)}}, \quad (2.31)$$

де $DialogAct_i$ — діалогічний акт, який є маркером певної форми реалізації прийому ІПМ, $N^{DialogAct_i}$ — кількість діалогічних актів, які є маркерами даного прийому.

Крім синтаксичних структур та діалогічних актів як маркери форм реалізації прийомів ІПМ використовують семантичні змінні. Семантична змінна — це частина лексичного наповнення висловлення, яка варіює залежно

від дискурсу, стилю, теми обговорення, але належить до одного класу понять. Певній формі реалізації прийому ІПМ поставлена у відповідність множина маркерів на основі семантичних змінних (2.32). Семантичні змінні подані у нотації Бекуса-Наура (див. підрозділ 4.3 «Приклади подання синтаксичної структури »).

$$SemanticVariable_i = \left\{SemanticVariable_{ij}\right\}_{j=1}^{N^{(SemanticVariable_i)}}, \quad (2.32)$$

де $SemanticVariable_i$ — семантична змінна, яка є маркером форми реалізації прийому ІПМ, $N^{SemanticVariable_i}$ — кількість семантичних змінних, які є маркерами реалізації певного прийому ІПМ.

Застосуванню певних прийомів властиве використання специфічних одиниць лексики. Тому кожен прийом характеризується вектором, елементами якого є нулі і одиниці залежно від того чи є значущим цей вид специфічної лексики для даного прийому. На основі цього вектору встановлюється ступінь інтенсивності $Intensity_i$ конкретного прецеденту застосування прийому ІПМ, який розраховуємо як добуток векторів:

$$Intensity_k = ImpactVector_k \times WeightedLexis_i, \quad (2.33)$$

де $ImpactVector_k$ — вектор, який характеризує значущість певних видів специфічної лексики для даного прийому ІПМ; $WeightedLexis_i$ — це вектор, який характеризує наповнення і-го повідомлення специфічною лексикою.

Якщо під час аналізу онлайн-дискусії виявлено послідовну зміну станів, яка відповідає станам тактики ІПМ, то для підтвердження прецеденту ІПМ необхідно виявити прийоми, за допомогою яких здійснено кроки тактики ІПМ. Крок може бути здійснено за допомогою більше ніж одного прийому, а прийом має кілька форм реалізації, кожну з яких можна виявити за допомогою маркерів трьох типів. Застосувати всі маркери прийому ІПМ для ідентифікації використаного прийому ІПМ є ресурсо- і часозатратним процесом.

Для того щоб встановити, який прийом чи прийоми використав маніпулятор для реалізації кроку тактики ІПМ, доцільно використовувати вказівник прийому ІПМ.

Вказівник прийому ІПМ — це вербальні, невербальні чи паравербальні засоби, які найчастіше використовують для реалізації прийому ІПМ. Вказівник прийому ІПМ використовуємо на перших кроках етапу виявлення ІПМ та на підготовчому етапі з метою виділення підозрілих фрагментів ІПМ (див. підрозділ 3.5 «Методи виявлення прийомів ІПМ у дискусіях онлайн-спільнот»).

Вказівник прийому ІПМ визначають таким чином (рис. 2.1):

Спершу потрібно вибирати форми реалізації, які найчастіше використовуються для реалізації конкретного прийому, потім визначити по одному маркеру кожного типу, які трапляються у більшості форм реалізації прийому ІПМ.

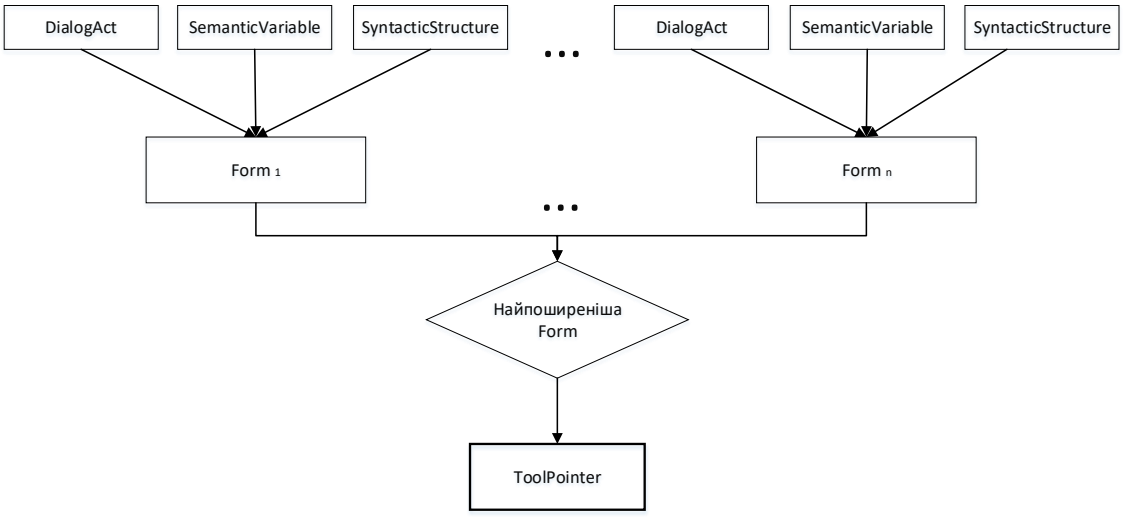


Рис. 2.1. Схема визначення вказівника прийому ІПМ

Крім описаних трьох типів маркерів ІПМ, які є головними, оскільки представляють домінантні засоби інформації у текстових повідомленнях онлайн-спільнот, для виявлення прийомів ІПМ залучають маркери, розроблені на основі невербальних та паравербальних засобів, тобто на основі емотиконів, посилань та елементів метаграфеміки. Останні об’єднано у такий кортеж:

$$ComplementaryMarkers_k = \langle Link_k, Emoticon_k, Metagraphemics_k \rangle, \quad (2.34)$$

де $Link_k$ — це множина типів посилань, які використовуються у формі реалізації прийому; $Emoticon_k$ — це множина емотиконів, які використовуються у формі реалізації прийому; $Metagraphemics_k$ — це множина елементів метаграфеміки, які використовуються у формі реалізації прийому.

Зміст невербальних і паравербальних елементів повідомлення визначають на основі таблиць для інтерпретації значення емотиконів, метаграфеміки та посилань (див. розд. 1.2.2 «Види невербальних та паравербальних засобів, які використовують для реалізації ІПМ»).

2.3. Розроблення системи фільтрів для виявлення підозрілих фрагментів дискусії

Фільтри для виявлення підозрілих фрагментів дискусії базуються на критеріях, які свідчать про потенційну наявність ІПМ. Ці критерії поділяються на два темпоральні види динамічні і статичні. В основу такої класифікації покладено часовий період, необхідний для збору інформації для розрахунку критерію. Введемо такі групи критеріїв.

- Статичні критерії — це критерії, які розраховуються на основі діяльності учасника протягом встановленого періоду часу.
- Динамічні критерії — це критерії, на основі яких можна робити висновки зразу ж після їх ідентифікації, які не потребують спостережень протягом певного періоду часу.

2.3.1. Статичні та динамічні критерії наявності ІПМ

Статичні критерії зосереджені навколо діяльності учасника онлайн-спільноти. Статичні критерії поділяються на два види залежно від спроможності учасника впливати на створення та створювати дані, необхідні для розрахунку критеріїв. Ці два класи ознак відповідають формальній моделі

учасника онлайн-спільноти (див. розд. 2.1.2 «Формальна модель учасника спільноти»).

Для розрахунку статичних критеріїв необхідні дані таких типів:

- дані профілю;
- поведінкові характеристики учасника;

Дані профілю — це дані, які учасник самостійно додає до профілю і змінює. Відсутність або мала кількість даних цього типу (аватар, місце навчання, робота і т.д.) є підставою для виникнення підозр щодо автентичності цього профілю.

Дані профілю відображають імідж учасника онлайн-спільноти. До них належать і дані, самостійно створені користувачем під час формування іміджу, і дані, які є своєрідними наслідками іміджу, тобто характеризують ставлення інших учасників спільноти до даного учасника, наприклад, кількість друзів та кількість чужих постів у життєписі учасника (дані отримано внаслідок аналізу профілю учасника у Facebook).

Дані профілю, зокрема, використовуються на підготовчому етапі алгоритму виявлення ІПМ в онлайн-спільнотах, з метою створення списку релевантних спільнот, адже на основі даних профілю визначають соціально-демографічні характеристики (див. розд. 3.1.1 «Підготовчий етап»).

Аналіз таких даних профілю, як імена учасників, використовується на підготовчому етапі під час сортування дискусій за сприятливістю до здійснення ІПМ (див. розд. 3.1.1 «Підготовчий етап») та на етапі виявлення, з метою ідентифікації підозрілих фрагментів дискусії (див. розд. 3.1.2 «Етап виявлення»). Наприклад, виявити потенційних маніпуляторів на основі імені можна так: якщо ім’я відповідає хоча б одному пункту з наведеного списку, то від цього профілю можна очікувати використання ІПМ.

- містить провокативний меседж;
- містить набори символів, які не мають жодного змісту;
- структура або елементи імені подібні до імен маніпуляторів.

Поведінкові характеристики учасника — це змінні в часі характеристики. З метою ефективної ідентифікації маніпуляторів, їх доцільно визначати через встановлені часові періоди.

Наприклад, до цієї групи характеристик, які виявляються на основі аналізу активності учасника в спільноті, належать:

- частота розміщення дописів у дискусії;
- частота реакцій на дописи;
- період доби, протягом якого користувач бере участь в інформаційних процесах.

Ефективність системи виявлення ІПМ залежить від актуальності даних для розрахунку значень статичних критеріїв. З усіх типів статичних критеріїв дані про поведінкові характеристики учасника вимагають регулярного оновлення, оскільки це найбільш змінні в часі статичні критерії.

За допомогою статичних критеріїв розглядають інформаційну діяльність учасника в проекції на структуру онлайн-спільноти, тобто відносно трьох рівнів організації інформаційного наповнення спільноти. Відповідно до формальної моделі онлайн-спільноти цими трьома рівнями є рівень спільноти, дискусії та повідомлення (див. розд. 2.1 «Спеціальна модель онлайн-спільноти з погляду виявлення ІПМ»). Критерії цих трьох організаційно-структурних рівнів відрізняються механізмом розрахунку та значимістю.

Рис. 2.2. Класифікація критеріїв ІПМ

Значення критеріїв рівня дискусії та спільноти прив’язані до учасника спільноти. За допомогою цих критеріїв виявляють підозрілого учасника спільноти, діяльність якого необхідно перевірити на наявність ІПМ.

А статичні критерії рівня повідомлення безпосередньо вказують на елементи інформаційного наповнення дискусії, які потрібно проаналізувати на наявність ІПМ.

Динамічні критерії використовують для аналізу конкретного акту інформаційної активності. Вони не містять узагальненої інформації про роль та поведінку учасника в онлайн-спільноті.

На відміну від статичних, динамічні критерії не поділяються за структурно-організаційними рівнями. Динамічні критерії використовують для аналізу комунікації на рівні повідомлення. Вони вказують на наявність у повідомленні слідів та наслідків застосування прийому ІПМ. На основі слідів та наслідків застосування ІПМ виявляють використані прийоми ІПМ та стани учасників-жертв ІПМ, відповідно. Ця інформація необхідна для визначення застосованої тактики ІПМ та ідентифікації прецеденту ІПМ.

2.3.2. Формальна модель системи фільтрів для виявлення підозрілих фрагментів дискусій

Поставивши за мету підвищити ефективність алгоритму виявлення ІПМ в онлайн-спільнотах, уникнуто перебору всього інформаційного наповнення дискусії. Цього досягнуто завдяки окресленню областей дискусії, які мають найбільше ознак наявності ІПМ, тобто виділення підозрілих фрагментів дискусії.

Підозрілі фрагменти дискусії — це набори логічно пов’язаних повідомлень, кількісні і якісні характеристики яких та кількісні і якісні характеристики профілів авторів цих повідомлень властиві повідомленням з ІПМ.

Систему критеріїв підозрілих фрагментів дискусії розроблено на основі закономірностей виявлених внаслідок аналізу комунікації у онлайн-спільнотах, їх узагальненої структури та способів презентації учасників. На основі цих критеріїв та формальної моделі онлайн-спільноти (див. розд. 2.1 «Спеціальна модель онлайн-спільноти з погляду виявлення ІПМ») запропоновано систему фільтрів, яка рекомендуватиме для подальшого аналізу на наявність ІПМ підозрілі фрагменти дискусії, тобто фрагменти, які потенційно містять ІПМ.

Систему фільтрів розроблено на основі статичних критеріїв наявності ІПМ (див. розд. 2.3.1 «Статичні та динамічні критерії наявності ІПМ»). Ознаки рівнів спільноти та дискусії створюють узагальнену картину діяльності учасника в дискусіях Інтернету, тоді як критерії рівня повідомлення вказують на ознаки ІПМ в конкретних актах інформаційної діяльності.

Виявляючи маніпуляцію, потрібно враховувати той факт, що кожна інтернет-дискусія має свої правила, стиль спілкування й аудиторію, а також передбачені платформою структуру та засоби спілкування. Від переліченого залежатиме формування набору фільтрів для виявлення ІПМ в конкретній дискусії, вага критеріїв, на яких ґрунтуються обрані фільтри та порогові значення, які не може перевищувати результат аналізу дискусії за допомогою фільтра.

Почнімо з розгляду фільтрів найвищого рівня, які засновані на критеріях, що потребують найменше даних та часу для їх розрахунку, тобто критеріях спільноти.

Тривалість учасницької діяльності $MembershipPeriod$, розраховуємо на основі такої формули:

$$MembershipPeriod = CurrentDate - \mathrm{Re}\,gistrationDate. \quad (2.35)$$

Якщо $MembershipPeriod$ перевищує встановлене експертами порогове значення ($MembershipPeriodThreshold$), то діяльність учасника потребує додаткової перевірки.

Ще один фільтр, який заснований на критеріях рівня спільноти — це вичерпність інформації про учасника. Він розраховується як відношення заповнених учасником полів профілю до всіх наявних полів, які передбачають внесення інформації учасником:

$$PersonalInformationCompleteness_i = \frac{FilledOutFieldsCount}{AllFields}. \quad (2.36)$$

Для прикладу розглянемо фільтр, який ґрунтується на статичному критерії рівня дискусії. $\mathrm{Re}plyRatio$ — це відношення кількості відповідей учасника на повідомлення до усіх повідомлень, які опублікував учасник.

$$ReplyRatio = \frac{ReplyCount_i}{PublishedMessageCount} * 100\%, \quad (2.37)$$

де $\mathrm{Re}plyCount_i$ — це кількість відповідей учасника на повідомлення інших; $PublishedMessageCount$ — це кількість усіх повідомлень, які написав учасник.

У дослідженні [19] визначено, для 84,3% учасників, які розміщували повідомлення з визначеним замовниками змістом $\mathrm{Re}plyRatio < 40\%$, в тоді для користувачів, які розміщували не замовлену інформацію $\mathrm{Re}plyRatio \approx 90\%$. Відповідно на основі цих даних оптимальне порогове значення для спільноти вибирають у діапазоні (30, 50) залежно від особливостей спільноти. Якщо $\mathrm{Re}plyRatio$ є нижчим за порогове значення, то цей фільтр відбирає всі повідомлення цього користувача для подальшої перевірки.

Згідно з дослідженням [19] 60% учасників спільноти, які розміщують проплачені тексти, реагують на коментарі з інтервалами часу 200 с, в той час лише 40% нормальних учасників реагують на повідомлення зі швидкістю близькою до 200 с. Зазвичай, вони виступають не так бурхливо і їхня швидкість публікації повідомлень є меншою. З погляду виявлення ІПМ в

онлайн-спільнотах, це спостереження трактуємо так: аномально швидкі відповіді на повідомлення є або продуктом маніпуляторів, або реципієнтів, які вже «клюнули на гачок», але так чи інакше зменшення інтервалу часу між публікаціями повідомлень свідчить про те, що у фрагменті дискусії потенційно наявна ІПМ.

$$PublishingFrequency = \frac{\sum_{1}^{N^{Messages}} (Timestamp_{i+1} - Timestamp_i)}{N^{Messages}}, \quad (2.38)$$

де $PublishingFrequency_i$ — частота розміщення дописів.

Якщо справджується умова (2.39), то фрагмент дискусії необхідно перевірити на наявність ІПМ.

$$PublishingFrequency < 200\,\text{sec}. \quad (2.39)$$

Як було зазначено вище, фільтри для виявлення підозрілих фрагментів ІПМ засновані на статичних критеріях рівня спільноти, дискусії та повідомлення. Розгляньмо один із фільтрів, які ґрунтуються на критерії рівня повідомлення.

Значна кількість видів інформаційної активності (1.1.3 «Текстові види спілкування в онлайн-спільнотах та їхні характеристики»), пов’язаних повідомленням, є характерною ознакою ІПМ. На основі цього спостереження розроблено фільтр:

$$SignalActivityProvocativeness = Like * LikeWeight + Share * ShareWeight + Comment * CommentWeight. \quad (2.40)$$

Також фільтри можуть базуватися на кількох критеріях, до прикладу наводимо фільтр, який заснований на складеному критерії рівня дискусії.

$$SignalActivityProvocativeness = Like * LikeWeight + Share * ShareWeight + Comment * CommentWeight.$$

На основі кількості дискусій спільноти, в яких учасник розміщує тематично релевантні повідомлення, можна виявити підозрілі профілі.

Розгляньмо три варіанти поведінки учасника спільноти: учасник веде активну інформаторську активність лише в кількох тематично пов’язаних дискусіях, учасник є активним у багатьох дискусіях, учасник є активним у багатьох дискусіях, але всі його пости є одного тематичного спрямування і, тому є нерелевантними у багатьох дискусіях, де вони розміщені. Останній варіант поведінки свідчить про маніпулятивну діяльність. Звичайно, теоретично можливий ще такий варіант: більшість постів, які розміщує учасник, не релевантні до тематики дискусій, але такий тип діяльності не можна розглядати як маніпуляцію, адже така поведінка більше схожа на нескладного бота, і інформаційне наповнення, створене ним, реципієнти серйозно не розглядатимуть.

Тому для виявлення підозрілих фрагментів дискусії на основі активності учасника та релевантності його постів розроблено такий фільтр:

$$\begin{cases} DiscussionActivenessRatio > 50\% \\ IrrelevantMessagesRatio > 50\% \end{cases}, \tag{2.41}$$

де $DiscussionActivenessRatio$ — це відношення кількості дискусій, в яких учасник веде активну діяльність до кількості усіх дискусій, яке відображає відносну активність учасника спільноти; $IrrelevantMessagesRatio$ — це відношення кількості нерелевантних повідомлень, які розмістив учасник, до усіх створених ним повідомлень.

$$DiscussionActiveness = \frac{EngagementDiscussionCount}{CommunityDiscussionTotal}, \tag{2.42}$$

де $EngagementDiscussionCount$ — це кількість дискусій, в яких учасник веде активну інформаторську діяльність; $CommunityDiscussionTotal$ — це загальна кількість усіх дискусій у спільноті.

$$IrrelevantMessagesRatio = \frac{IrrelevantMessages}{MessagesTotal}, \tag{2.43}$$

де $IrrelevantMessages$ — це кількість розміщених учасником повідомлень, які розмістив учасник, нерелевантних до тематики дискусії; $MessagesTotal$ — це загальна кількість усіх повідомлень учасника.

Якщо подвійна умова (2.41) виконується для учасника, то він потенційно веде маніпулятивну діяльність.

Кожен фільтр базується на простому чи складеному критерії наявності ІПМ та характеризується кортежем з такими елементами:

$$Filter_i = \langle CriterionValue_i, Weight_i, ThresholdValue_i \rangle, \quad (2.44)$$

де $CriterionValue_i$ — значення для конкретної дискусії, розраховане відповідно до критерію, на якому заснований фільтр; $Weight_i$ — ваговий показник фільтра, визначений експертами з галузі управління онлайн-спільнотами; $ThresholdValue_i$ — порогове значення критерію встановлене експертами для конкретної дискусії.

Фільтри мають різну важливість з погляду виявлення підозрілих фрагментів дискусії. Саме тому для кожного з критеріїв встановлено вагові показники. Наприклад, $\mathrm{Re}\,plyRatio$ є важливішим з погляду виявлення підозрілих уривків, ніж $MembershipPeriod$.

Сформувавши набір фільтрів для аналізу дискусії, під час встановлення їхніх вагових показників необхідно враховувати такі вимоги (2.45), (2.46). Тобто сума ваг всіх фільтрів має дорівнювати 1, а значення вагового показника для кожного фільтра не може бути від'ємним:

$$\sum_i Weight_i = 1, \quad (2.45)$$

$$Weight_i \geq 0. \quad (2.46)$$

Введемо поняття індикатора потенційної ІПМ (2.47). Якщо значення критерію, на якому заснований фільтр, перебуває в межах встановленого

експертами порогового значення, то ідентифікатору потенційної ІПМ присвоюємо 0, у протилежному випадку – 1

$$PotentialIPMIndicator_i = \begin{cases} 0, CriterionValue_i \in ThresholdValue_i \\ 1, CriterionValue_i \in ThresholdValue_i \end{cases}. \quad (2.47)$$

Експерти встановлюють порогове значення критерію для кожного фільтра на основі комунікаційних і структурних особливостей певної спільноти

$$ThresholdValue_i \in [\min^i, \max^i]. \quad (2.48)$$

Розрахунок результатів роботи системи фільтрів відбувається за такою формулою:

$$IPMSuspiciousness = \sum_{i=1}^{N^{Filter}} PotentialIPMIndicator_i * Weight_i . \quad (2.49)$$

При чому для системи фільтрів також встановлюється порогове значення, при виході за межі якого, фрагмент дискусії ідентифікують як підозрілий. Перевірка, чи не перевищує *IPMSuspiciousness* порогового значення, відбувається динамічно, тобто розраховується не після проходження аналізу за допомогою всіх фільтрів, а після отримання *PotentialIPMIndicator* з нульовим значенням (рис. 2.3).

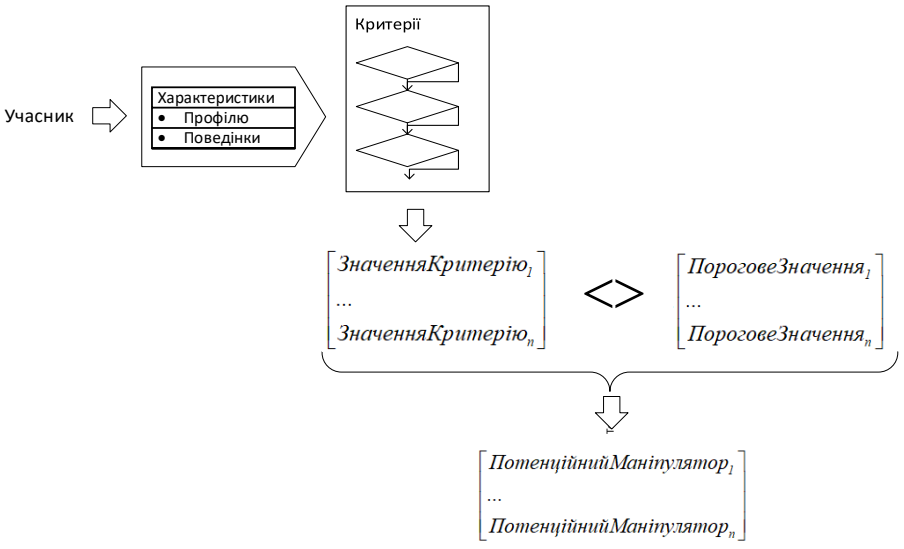


Рис. 2.3. Схема роботи системи фільтрів

Отже, розроблена система фільтрів забезпечує зменшення кількості повідомлень, які доцільно перевірити на наявність ІПМ, а динамічна перевірка перевищення порогового значення встановленого для системи фільтрів, дає змогу автоматично коригувати кількість фільтрів необхідну для ідентифікації конкретного фрагмента дискусії як підозрілого. Таким чином досягнуто підвищення ефективності роботи системи.

2.4. Семантичні змінні

Виявлення прийомів ІПМ відбувається на основі аналізу інформаційного наповнення повідомлень. Ключовими елементами інформаційного наповнення, за допомогою яких виявляють ІПМ, є метаграфічні, невербальні (емотикони і посилання) та вербальні елементи повідомлень.

Вербальне наповнення повідомлення представлене, зокрема, за допомогою семантичних змінних. Семантична змінна — це частина вербального наповнення повідомлення, яка представляє певний концепт та варіює залежно від дискурсу, стилю, теми обговорення. СЗ може бути представлена за допомогою одного слова або словосполучення.

СЗ позначають не лексеми, а поняття. В повідомленнях однакові поняття можуть бути представлені за допомогою різних лексем, а різні поняття за допомогою лексико-семантичних варіантів одної лексеми (різних змістових планів багатозначного слова).

Для кожної СЗ визначена множина значень. Семантична змінна може приймати одне із значення з цієї множини. Це значення може бути словом, а може бути словосполученням.

СЗ поділено на три основні класи семантичних змінних: сутності, предикати та характеристики.

Ці класи не можна ототожнити з певними частинами мови, оскільки СЗ це поняття, але у кожному класі домінує певна частини мови. Клас Сутність становлять більшою мірою іменники або іменникові словосполучення.

Більшість елементів класу Предикати є дієсловами або словосполученнями дієслова з іншими частинами мови. Клас Характеристики становлять прикметники та прислівники. Службові частини мови не є повнозначними, тобто не можуть позначити жодного поняття самостійно, тому не можуть самостійно представляти СЗ, а лише разом із повнозначними частиними мови.

СЗ поділені на групи на основі семантичних характеристик. Кожна група пов'язана з відповідними формами вираження, яким, своєю чергою, відповідають певні морфологічні характеристики.

Кожному класу СЗ притаманна певна структура, вид зв'язків між підкласами та їх елементами, а також типи характеристик, які використовуються для їхнього опису.

Зв'язки між підкласами та елементами класу Сутність ґрунтуються на зв'язках гіперо-гіпонімії, меронімії та синонімії. Перші два види зв'язків задають ієрархію. Останній вид зв'язку можуть утворювати елементи, які перебувають на одному ієрархічному рівні. Ієрархічна структура класу Сутність складається з багатьох підпорядкованих підкласів. Ці підкласи, як було сказано вище, пов'язані між собою зв'язками гіперо-гіпонімії та меронімії.

Класу Предикати властиві зв'язки темпоральної сумісності та темпоральної несумісності. Зв'язки темпоральної сумісності поділюються на два види: «дія - спосіб виконання дії», наприклад, йти — маршувати, говорити — заїкатися; «дія - побічна дія», яка може відбуватись лише поки триває перша дія, наприклад, спати — храпіти, купувати — платити. Зв'язки темпоральної несумісності поділяються на конверсиви та зв'язки «причина-наслідок». Конверсиви позначають певну дію у протилежних напрямках чи відношеннях, наприклад, зав'язати — розв'язати, кинути — зловити. Прикладами предикатів між якими існує зв'язок «причина-наслідок» є взяти — мати, підняти — піднятись.

Між елементами класу Характеристики існують зв'язки комплементарної, контрарної, контрадикторної та векторної антонімії. В

межах одного ієрархічного рівня СЗ цього класу пов’язують зв’язки синонімії. Певні семантичні ознаки цього класу відображаються у мовленні як морфологічні, наприклад, можливість утворювати ступені порівняння. Крім того, характеристикам притаманна морфолого-семантична ознака якісноті, відносності і присвійності, кожну з цих груп у мовленні можна розпізнати за притаманнми наборами суфіксів.

Класи СЗ мають структуру дерева. Кожен із підкласів вирізняється на основі характерних особливостей, і ці особливості притаманні всім його дочірнім підкласам. Інакше кажучи, кожен дочірний підклас СЗ має вужче семантичне значення і чіткіше окреслений набір граматичих ознак.

Висновки до розділу

У розділі представлено розроблені формальні моделі:

1. онлайн-спільноти, яка складається з формальних моделей дискусії, учасника спільноти та повідомлення, кожна з формальних моделей містить характеристики необхідні для аналізу онлайн-спільноти з метою виявлення ІПМ;
2. тактики ІПМ, яка формалізує тактику ІПМ відповідно до особливостей комунікації в онлайн-спільнотах. Тактика ІПМ побудована на основі кусково-лінійних агрегатів, що дало змогу відобразити послідовність зміни психічних станів жертви ІПМ внаслідок впливу маніпулятора;
3. фільтрів для виявлення підозрілих фрагментів дискусії на основі формалізованих статичних та динамічних критеріїв наявності ІПМ. Фільтр дозволяє окреслити область потенційної наявності ІПМ і, отже, підвищує ефективність виявлення ІПМ;
4. формальну модель семантичних змінних, які дозволяють виявляти ІПМ на основі ознак лексичного рівня.

Розділ 3. Методи виявлення інформаційно-психологічної маніпуляції в онлайн-спільнотах

У третьому розділі подано розроблений алгоритм виявлення ІПМ в онлайн-спільноті. Алгоритм складається з чотирьох етапів: підготовчого етапу, етапу виявлення, етапу нейтралізації, етапу формування результатів та рекомендацій. На першому етапі алгоритму використано методи пошуку релевантних дискусій у соціальних середовищах інтернету. На другому етапі використано методи виявлення підозрілих фрагментів дискусії, методи виявлення послідовності психічних станів, методи виявлення прийомів ІПМ у дискусіях, а також інструментарій тональнісного аналізу, засоби обробки тексту. На третьому етапі використано методи визначення автора тексту, а саме мовні, поведінкові ознаки та ознаки профілю. Крім того, застосовано методи визначення впливових профілів і шляхів поширення інформації у спільноті на основі соціального графу онлайн-спільнот.

Основні результати розділу опубліковано у дослідженнях [79, 91, 92, 93].

3.1. Алгоритм виявлення ІПМ в онлайн-спільнотах

Алгоритм виявлення ІПМ в онлайн-спільнотах ґрунтується на формальних моделях онлайн-спільноти і тактики ІПМ (див. розд. 2.1 «Спеціальна модель онлайн-спільноти з погляду виявлення ІПМ» та 2.2 «Формальна модель). Залежно від мети, інструментарію та аналізованих даних виділено чотири етапи алгоритму: підготовки, виявлення, нейтралізації та формування результатів (рис. 3.1).

Етап підготовки полягає у пошуку спільнот, які містять релевантні дискусії, та сортуванні дискусій за наявністю передумов для здійснення ІПМ. Конкретні завдання виявлення ІПМ передбачають обмеження області пошуку, наприклад, може бути задана тематика або локація (онлайн-спільнота, набір

дискусій), за якою треба виявляти ІПМ. Після ідентифікації спільнот, які містять відповідні до обмежень дискусії, необхідно визначити порядок аналізу дискусій. Дискусії сортують за найсприятливішими передумовами для здійснення ІПМ за спадання (див. розд. 2.3.2 «Формальна модель системи фільтрів для виявлення підозрілих фрагментів дискусій»).

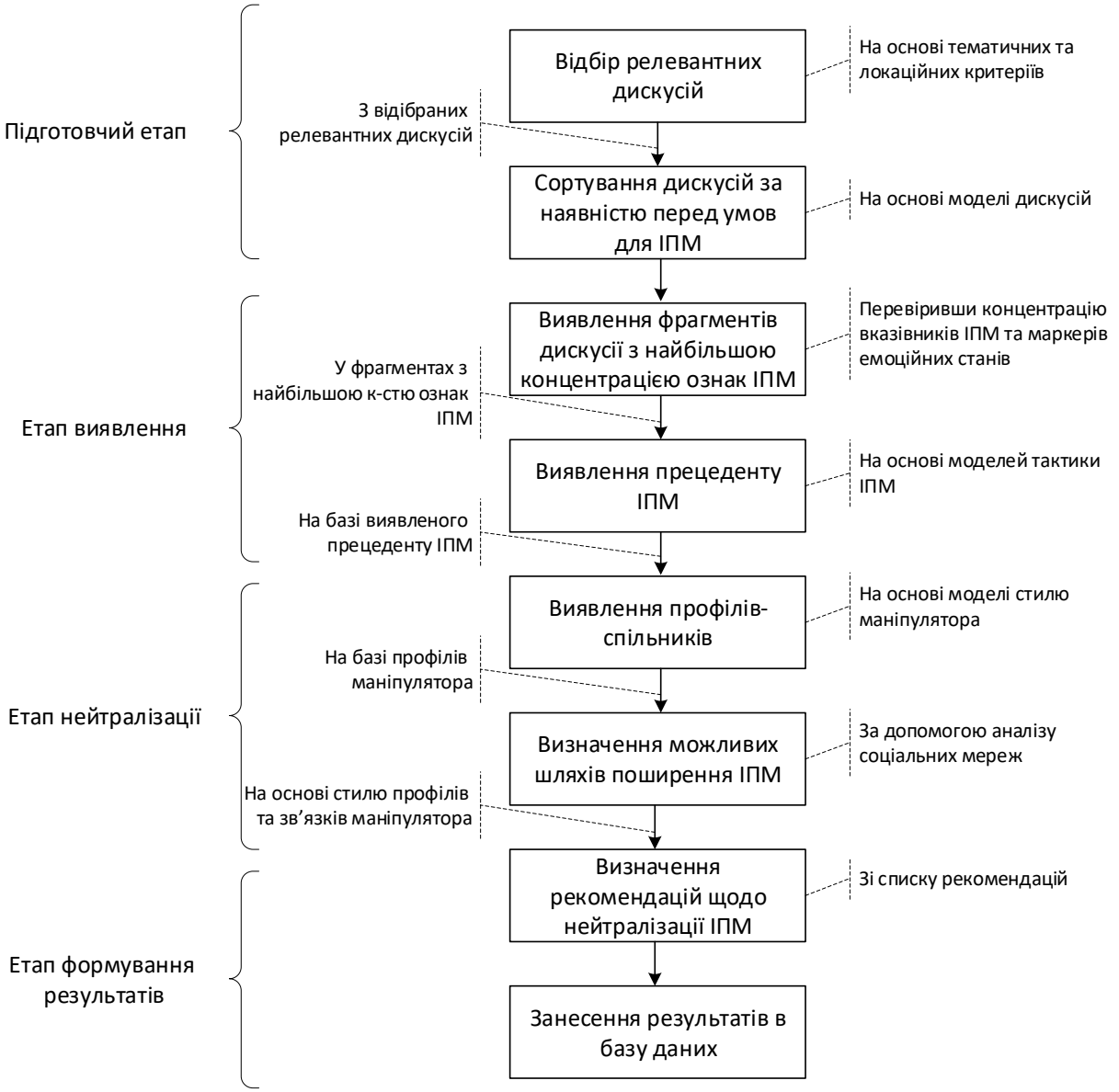


Рис. 3.1. Схема виявлення ІПМ в онлайн-спільнотах

На етапі виявлення визначають фрагменти дискусій, з яких варто почати пошук ІПМ, та проводять аналіз інформаційного наповнення з метою виявити

прецеденти ІПМ. Результатом цього етапу є опис прецеденту ІПМ, а саме суб'єктів ІПМ, застосованої тактики та використаних прийомів.

На третьому етапі, етапі нейтралізації, аналізують дані про суб'єкти ІПМ, з метою виявити угрупування маніпуляторів. Також здійснюється аналіз зв'язків учасників дискусії. Інформацію про учасників, які піддалися впливу ІПМ, використовують для визначення можливих шляхів поширення ІПМ.

На етапі формування результатів, на основі вихідних даних попередніх етапів, надаються поради щодо нейтралізації ІПМ, до черги додаються нові дискусії, які необхідно проаналізувати на наявність ІПМ. На цьому етапі алгоритму вносять до бази даних виявлені прецеденти ІПМ.

Алгоритм виявлення ІПМ в онлайн-спільнотах розроблено для ідентифікації прецедентів ІПМ в онлайн-спільнотах відповідно до окресленої області пошуку та завдання ІПМ та встановлення можливих шляхів поширення ІПМ, що є важливим з погляду нейтралізації ІПМ. Алгоритм передбачає використання та наповнення бази даних ІПМ інформацією про нові виявлені прецеденти ІПМ.

3.1.1. Підготовчий етап

Мета підготовчого етапу — створити список дискусій, які відповідають вказаним у завданні виявлення ІПМ тематичним або локаційним обмеженням та відсортовані за спаданням відповідно до наявності сприятливих передумов до здійснення ІПМ.

Підготовчий етап складається з двох підетапів: відбору релевантних дискусій із соціальних середовищ інтернету та сортування дискусій за наявністю передумов для ІПМ.

Результатом виконання першого підетапу є список онлайн-дискусій, поданий у порядку спадання відповідно до релевантності дискусії до обмежень завдання виявлення ІПМ. Результатом виконання другого підетапу є список релевантних дискусій, поданий у порядку спадання відповідно до наявності у

дискусіях сприятливих передумов для здійснення ІПМ. Результат виконання цього підетапу є кінцевим та передається для опрацювання на наступний етап.

На підетапі відбору релевантних дискусій для пошуку у соціальних середовищах інтернету використовуємо агрегатор пошукових систем, засоби пошуку у Facebook та пошук у базі даних виявлених прецедентів ІПМ. Агрегатор пошукових систем збирає результати кількох ГПС та подає їх у відповідному до налаштувань вигляді.

Послідовність дій на підетапі пошуку у соціальних середовищах інтернету за допомогою агрегатора ГПС зображено на рис. 3.2.

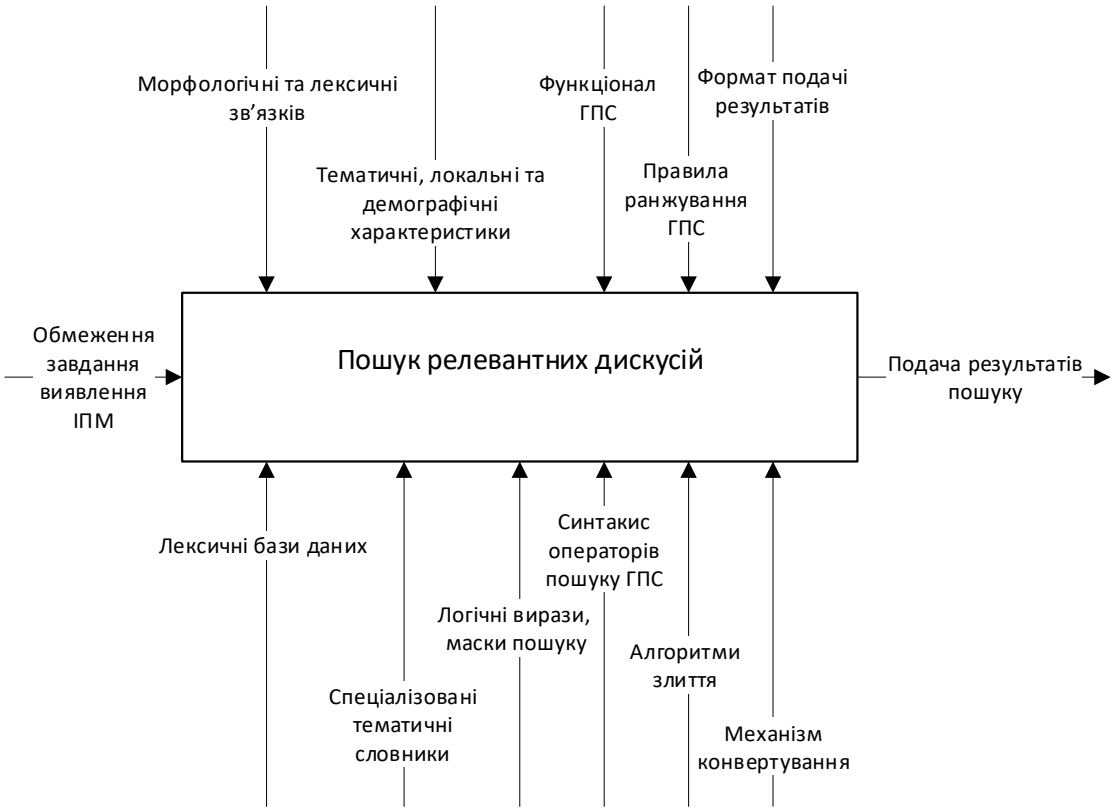


Рис. 3.2. Пошук релевантних дискусій за допомогою агрегатора ГПС

ГПС відрізняються між собою алгоритмами обходу, індексації, ранжування, тому їхні результати пошуку є відрізняються між собою. Використання агрегатора ГПС з метою виявлення онлайн-дискусій є

доцільним, оскільки дозволить підвищити релевантність результатів видачі та збільшити їхню кількість.

Facebook є найпопулярнішою платформою соціальних мереж у багатьох країнах Європи, якій властиве текстове подання інформації, тому необхідно здійснювати пошук і в ньому. Пошук у Facebook потрібно здійснювати спеціальними засобами, враховуючи структуру Facebook як графа.

Крім того, щоб підвищити релевантність результатів пошуку, використовуємо для формування запитів інформацію з бази даних виявлених прецедентів ІПМ.

Відбір релевантних дискусій відбувається на основі морфологічних та лексичних зв’язків між ключовими словами, тематичних, локальних та соціально-демографічних характеристик, заданих у завданні виявлення ІПМ, функціоналу ГПС, правил ранжування ГПС та вимог до подачі результату пошуку агрегатора ГПС

Дії відбуваються за допомогою таких механізмів: лексичних баз даних, спеціалізованих тематичних словників, логічних виразів та масок пошуку, синтаксису операторів пошуку у ГПС, алгоритмів злиття та механізмів конвертування.

Після пошуку релевантних до завдання виявлення ІПМ дискусій переходимо на другий підетап підготовчого етапу, а саме створення черги дискусій за сприятливістю до здійснення ІПМ.

Цей підетап полягає у формуванні черги дискусій, які потрібно аналізувати на наявність ІПМ з урахуванням терміновості аналізу. Сприятливість до здійснення ІПМ і відповідно терміновість аналізу певної дискусії визначається за трьома групами критеріїв: критерії популярності дискусії, підозрілості та вразливості (табл. 3.3).

Таблиця 3.1

Критерії сортування дискусій

	Критерій	**Пояснення**
Популярність	Кількість учасників	Якщо кількість учасників дискусії перевищує встановлене порогове значення, то дискусія є популярною.
	Час публікації останнього допису	Якщо різниця поточного часу та часу публікації останнього допису менша за встановлене порогове значення, то дискусія є популярною.
	Кількість дописів на день	Якщо кількість дописів у дискусії перевищує встановлене порогове значення, то дискусія є популярною.
Підозрілість	Кількість підозрілих профілів	Якщо підозрілих профілів, виявлених у дискусії, перевищує встановлене порогове значення, то дискусія є підозрілою.
	Кількість метаграфеміки	Якщо кількість елементів метаграфеміки, виявлених у дискусії, перевищує встановлене порогове значення, то дискусія є підозрілою.
	Кількість емотиконів	Якщо кількість емотиконів, виявлених у дискусії, перевищує встановлене порогове значення, то дискусія є підозрілою.
Вразливість	Кількість впливових профілів	Якщо впливових профілів, які беруть участь у дискусії, перевищує встановлене порогове значення, то дискусія є вразливою.
	Кількість представників цільової аудиторії	Якщо представників цільової аудиторії, які беруть участь у дискусії, більше за встановлене порогове значення, то дискусія є вразливою.

Критерії популярності та підозрілості розраховуються на основі формальної моделі онлайн-спільноти (див. розд. 2.1 «Спеціальна модель онлайн-спільноти з погляду виявлення ІПМ»). Розрахунок більшості критеріїв цих груп не вимагає складного аналізу і його здійснюємо під час аналізу усіх дикусій.

Розрахунок деяких критеріїв, наприклад, кількості підозрілих акаунтів та більшості критеріїв вразливості, потребує значного збору й аналізу даних. Якщо в базі даних виявлених прецедентів ІПМ наявна інформація про дискусію, то ці критерії варто врахувати під час визначення сприятливості дискусії до здійснення ІПМ. Якщо аналізуємо дискусію на наявність ІПМ

уперше, то розраховувати згадані вище критерії недоцільно, оскільки вони вимагають великих затрат часу.

Підетап сортування дискусій за сприятливістю до здійснення ІПМ є важливим для ефективності алгоритму виявлення ІПМ, оскільки він підвищує оперативність виявлення ІПМ. Лише оперативно виявлену ІПМ, можна вчасно нейтралізувати та запобігти її розповсюдженню.

Результатом виконання підготовчого етапу є список релевантних дискусій, посортований за кількістю ключових слів та їхньою значимістю у дискусії.

3.1.2. Етап виявлення

Етап виявлення передбачає виконання двох ключових завдань: виділення фрагментів дискусії з найбільшою концентрацією ознак ІПМ та виявленні прецедентів застосування тактик ІПМ у дискусії. Відповідно до цих завдань його поділено на підетапи: виявлення фрагментів дискусії з найбільшою концентрацією ознак ІПМ, виявлення прецеденту ІПМ.

Підозрілі фрагменти дискусії виявляємо за допомогою відповідних методів і засобів (див. розд. 3.3 «Методи виявлення підозрілих фрагментів дискусії»).

Пошук прецедентів ІПМ є багатокроковим процесом, в якому застосовують методи контент-аналізу. Виявлення прецедентів ІПМ має такі проміжні цілі, як виявлення станів учасників спільноти (див. розд. 3.4 «Методи виявлення послідовності зміни психічних станів учасників дискусії»), виявлення маркерів на основі вербальних ознак ІПМ, зокрема: діалогічних актів, синтаксичних структур і семантичних змінних, та невербальних маркерів, використаних для текстової реалізації повідомлення (див. розд. 3.5 «Методи виявлення прийомів ІПМ у дискусіях онлайн-спільнот»). Виявлення прецеденту ІПМ ґрунтується на формальній моделі тактики ІПМ, поданій у вигляді кусково-лінійного агрегата (див. розд. 2.2.1 «Формальна модель

тактики ІПМ»). Алгоритим виявлення ІПМ можна поділити на три основні кроки (рис. 3.3).

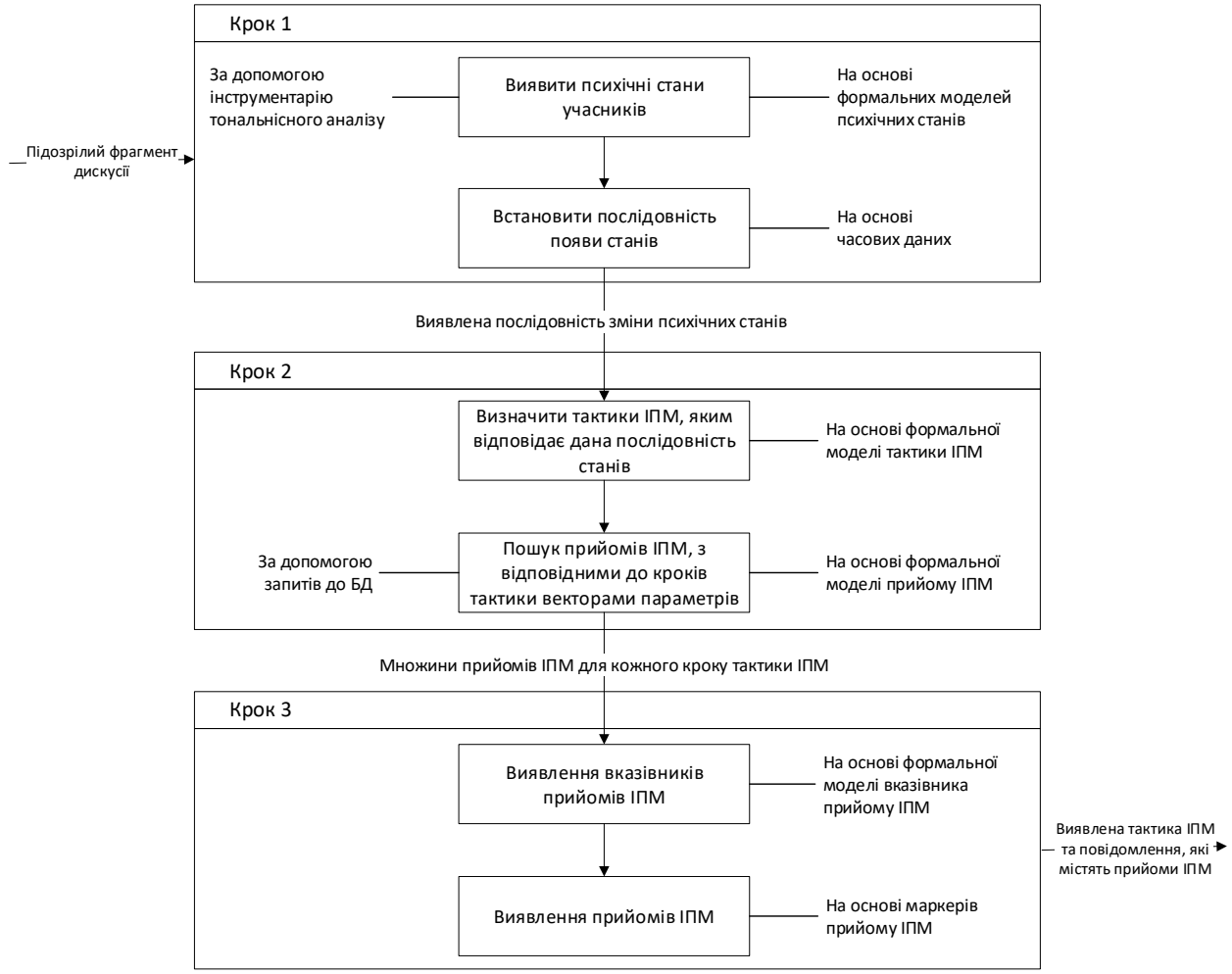


Рис. 3.3. Схема виявлення прецедентів ІПМ

На першому кроці алгоритму виявлення ІПМ, за допомогою методів і засобів виявлення послідовності психічних станів виявляються стани учасників фрагменту дискусії (див. розд. 3.4 «Методи виявлення послідовності зміни психічних станів учасників дискусії»). Кожен стан ідентифікує набір емоцій (див. розд. 1.3.1 «Аналіз наявних класифікацій емоцій і психічних станів людини»). Інформацію про набори емоцій, які визначають стан отримуємо, внаслідок аналізу існуючих схем тактик ІПМ (див. розд. 2.2 «Формальна модель інформаційно-психологічної маніпуляції») та знаходиться у таблиці бази даних. Для виявлення емоцій використовуємо інструментарій

тональнісного аналізу (див. розд. 1.3.2 «Класифікація інструментарію тональнісного аналізу»). Пізніше на основі часових даних встановлюємо послідовність появи психічних станів.

На другому кроці послідовність виявлених станів порівнюють із послідовністю станів у моделях тактик ІПМ поданих за допомогою кусково-лінійного агрегату. Після встановлення тактики ІПМ, яка відоповідає послідовності зміни станів, визначають кроки тактики ІПМ відповідно до формальної моделі тактики ІПМ (див. розд. 2.2.1 Формальна модель тактики ІПМ). Пізніше в базі даних прийомів ІПМ шукають прийоми, вектори параметрів яких збігаються з кроком тактики ІПМ. Після ідентифікації цих прийомів з бази даних беруть вказівники цих прийомів.

На третьому кроці виявляють вказівники цих прийомів у повідомленнях фрагментів дискусії (див. розд. 2.2.2 Формальна модель прийому ІПМ). На цьому кроці застосовують методи і засоби виявлення прийомів ІПМ у дисусіях (див. розд. 3.5 «Методи виявлення прийомів ІПМ у дискусіях онлайн-спільнот») Пошук починають з повідомлень, авторами яких є підозрілі профілі, або повідомлень, які провокують бурхливу активність у дискуії. Це відображається великою кількістю реакцій, наприклад, коментарів та уподобань.

Під час виконання перелічених кроків необхідно дотримуватись наведених нижче правил.

Якщо вказівник прийому ідентифіковано, то необхідно перейти до аналізу наступного кроку даної тактики ІПМ. Якщо вказівники не ідентифіковано, то потрібно шукати вказівник наступного прийому ІПМ, з набору прийомів, які містять вектори, ідентичні з кроком тактики ІПМ.

Якщо ж вказівник жодного з множини прийомів, які відповідають кроку ІПМ, не ідентифіковано в підозрілому фрагменті дискусій, то потрібно перейти до аналізу наступного кроку тактики ІПМ.

Якщо всі кроки ІПМ проаналізовано на наявність відповідних прийомів у підозрілому фрагменті дискусії, проводиться верифікація наявності тактики ІПМ у фрагменті. Тобто, якщо кількість виявлених прийомів перевищує встановлене експертами порогове значення, то тактику ІПМ виявлено. Якщо - ні, то проводиться аналіз тактики ІПМ, формальна модель якої подібна до виявленої в фрагменті схеми переході між станами.

3.1.3. Етап нейтралізації

Мета етапу нейтралізації – запобігти розгортанню та поширенню ІПМ у онлайн-спільноті. Це досягається за рахунок виконання двох завдань:

- виявлення груп профілів, які спільно здійснюють ІПМ, на основі інформації про одиничні маніпулятивні профілі;
- виявлення профілів-жертв та профілів-зомбі на основі соціального графа.

ІПМ здійснюють з одного профіля або один чи кілька маніпуляторів використовують кілька профілів для реалізації ІПМ необхідного спрямування. З метою ефективної нейтралізації ІПМ необхідно виявити всі профілі, які належать до маніпулятивної групи і, таким чином, запобігти їхній подальшій діяльності.

Виявлення груп профілів, які здійснюють ІПМ, відбувається за допомогою пошуку учасників спільноти, в яких характеристики поведінки, стилю чи профілю подібні до характеристик маніпулятора. Профілі, які належать до однієї маніпулятивної групи, можуть мати не всі однакові характеристики одного з видів, тому перевірка лише за допомогою одного виду характеристик є недостатньою.

Аналіз учасників дискусії, в якій було виявлено прецедент ІПМ, на наявність групи маніпулятивних профілів здійснюється за наведеними вище видами характеристик, послідовно, починаючи з характеристик, які не потребують складних обчислень. При цьому кожен вид містить обов’язкові і опціональні характеристики, які застосовуються залежно від вимог завдання

виявлення ІПМ. Деякі з характеристик подано на рис. 3.4 відповідно до груп, до яких вони належать.

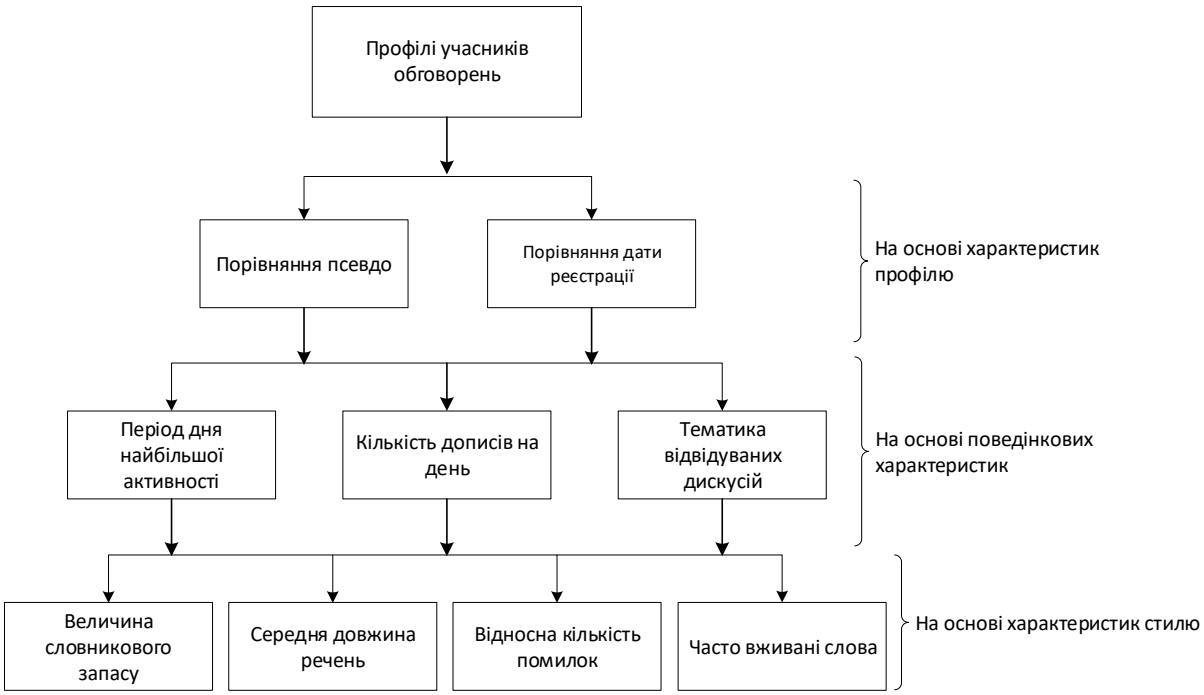


Рис. 3.4. Виявлення профілів-спільників

До характеристик профілю учасника, за допомогою яких можна ідентифікувати профіль маніпулятора-спільника належать нікнейм учасника та дата реєстрації. Порівняння цих двох характеристик із даними профілю маніпулятора не потребує складних обчислень, тому його проводять на першому етапі виявлення груп маніпулятивних профілів.

Зареєстровані в однаковий період учасники чи учасники зі схожими нікнеймами, які належать до даної спільноти, потребують подальших перевірок, наприклад, мовного стилю. Якщо встановлена кількість характеристик, кожна з яких має наперед визначену вагу, співпаде, то маніпулятора-спільника виявлено.

Подібність нікнеймів учасників спільноти встановлюється на основі відстані Дамерау-Левенштейна – міри відмінності двох рядків символів, яка

визначається кількістю операцій вставлення, заміни і транспозиції (перестановки двох сусідніх символів) необхідних для перетворення одного з рядка в інший [94].

Наступні перевірки відбуваються на основі поведінкових характеристик учасника спільноти, наприклад *PublishingFrequency*, *ActivityTime*, *EngagementDiscussionThemes*, (див. розд. 2.1.2 «Формальна модель учасника спільноти»). Якщо маніпулятор та учасник даної спільноти мають схожі характеристики, то виявлено профіль-спільник.

Виявлення профілю-спільника на основі періоду активності відбувається так: графік періодів активності маніпулятора по черзі порівнюється з графіками періодів активності профілів, які проходять перевірку. Якщо 70% періодів активності збігається (рис. 3.5 Б), то результат проходження цієї перевірки буде позитивним, тобто профіль-учасника ідентифіковано як профіль-спільник маніпулятора. В протилежному випадку (рис. 3.5 А), на основі цієї перевірки профіль не становить небезпеки з погляду ІПМ.

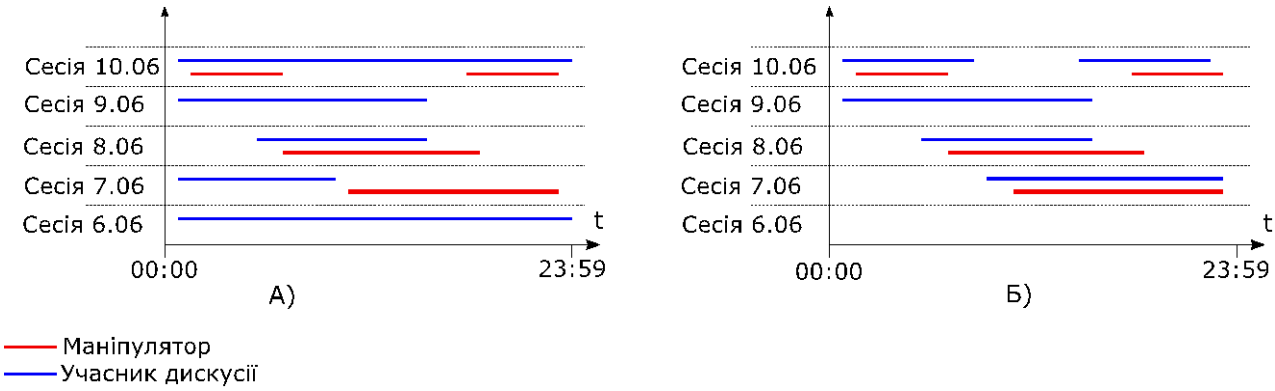


Рис. 3.5. Виявлення профілю-спільника на основі періоду активності

Крім цього, порівнюють кількість щоденних дописів учасників та тематики відвідуваних дискусій.

Найбільше обчислень та опрацювання даних вимагає перевірка мовного стилю учасників. Ця перевірка дає можливість виявити профілі, з яких дописує один маніпулятор або профілі, з яких діють кілька маніпуляторів

дотримуючись інструкцій ІПМ. Ця перевірка забезпечує високу точність результатів, хоч і потребує об’ємних розрахунків.

Необхідні дані для розрахунку характеристик мовного стилю зазначено у формальній моделі повідомлення (див. розд. 2.1.3 «Формальна модель повідомлення»), це, наприклад, кортежі параметрів *FeaturesBasic* та *FeaturesAdditional*.

Для виявлення мовного стилю профілю аналізують усі повідомлення, створені цим профілем. На основі методу k-підпису визначають репрезентативні характеристики мовного стилю [95]. Метод k-підпису полягає у такому: k-підпис – це характеристика, яка з’являється у щонайменше k% повідомлень учасника і відсутня в повідомленнях інших учасників.

Характеристики мовного стилю поділяються на лексичні, символьні, синтаксичні та емотикони (графічні, параграфічні).

Під час аналізу символьних характеристик стилю текст повідомлень розглядають як набір символів, при чому символами вважаються знаки пунктуації, літери, цифри та пробіли (n-грама може складатися з символів, які належать до двох різних слів). Згідно з [97] доцільно розглядати послідовності з 4-х символів.

Визначення мовного стилю на основі ***символьних характеристик*** є особливо ефективним для виявлення авторів, тематично не пов’язаних текстів [98, 99, 100]. Це зумовлено тим, що символьні характеристики дають інформацію про суфікси, афікси, та знаки пунктуації, яким автор надає перевагу, а не тематично-залежних лексичних одиниць.

Символьні ознаки легко видобути з тексту, оскільки не потрібно виявляти слова, тобто враховувати пробіли між символами, регістр, маркери нового рядка [96]. Велика кількість забезпечує велику вибірку, а відсутність необхідності складного опрацювання тексту - простоту реалізації.

Аналіз мовного стилю на основі ***лексичних характеристик*** вимагає опрацювання тексту, наприклад, виділення слів. До цих характеристик

належать: частота появи службових частин мови, частота і довжина слів, багатство словникового запасу [101].

Найефективнішою ознакою, з погляду виявлення автора повідомлення за допомогою лексичних ознак, є частота появи службових частин мови [102], оскільки службові частини мови тематично незалежні та автори використовують службові частини мови підсвідомо [103].

Перевірка частоти і довжини слів полягає у визначенні найпоширеніших довжин слів та їхньому порівнянні з відповідною характеристикою іншого учасника спільноти.

Багатство словникового запасу визначається як відношення унікально вжитих слів у повідомлення до всіх слів повідомлення:

$$Vocabulary\text{Re}achness = \frac{TermCount}{WordCount}. \qquad (3.1)$$

Для визначення мовного стилю використовують такі ***синтаксичні характеристики***: частота і довжина речень, частота появи різних частин мови та частота і довжина типів словосполучень (іменникових, дієслівних). Більшість синтаксичних характеристик вимагає повного або часткового синтаксичного парсингу речення і, відповідно, більше часу та ресурсів для здійснення. Результативність визначення мовного стилю за допомогою синтаксичних характеристик значно поступається виявленню за допомогою символьних характеристик та службових слів [104], тому у даній системі воно має опціональний характер.

Характеристики на основі емотиконів також використовуються для визначення авторства повідомлення [105], а саме: кількість емотиконів, кількість видів емотиконів в одному реченні, кількість емотиконів зі знаків пунктуації, кількість видів емотиконів зі знаків пунктуації в одному реченні.

$$EmoticonCharacteristics = \left\{ \begin{matrix} EmoticonTokenCounts, EmoticonTypesPerSentence, \\ PunctuationTokenCount, PunctuationTypesPerSentence \end{matrix} \right\}. \qquad (3.2)$$

Ще однією особливою характеристикою, яка використовується для виявлення авторського стилю є ***типові помилки***, тобто орфографічні або синтаксичні помилки, які трапляються більше ніж n разів у повідомленнях певного автора. Ця ознака також потребує детального опрацювання тексту, тому є не обов'язковою, а має рекомендаційний характер.

Виявлення профілів-жертв, профілів-зомбі (див. розд. 2.1.2 «Формальна модель учасника спільноти») та встановлення можливих шляхів поширення ІПМ є необхідним для нейтралізації ІПМ.

Встановлення можливих шляхів поширення ІПМ полягає у виявленні характеристик впливовості учасників спільноти та сили їхнього зв'язку з маніпулятором. Впливовість профілю залежить від кількості та видів зв'язків, які пов'язують його з іншими профілями. Важливість видів зв'язку визначають експерти. Ступінь поширення ІПМ залежить від:

- ролі профілів з точки зору ІПМ;
- впливовості профілів;
- топологічних характеристик профілю як вузла графа.

Можливі ролі профілів учасників ІПМ описані у розділі 2.1.2 «Формальна модель учасника спільноти».

Профіль «Маніпулятор» – це профіль, який провадить тактику ІПМ в онлайн-спільноті.

Профіль «Жертва» – це профіль, на основі повідомлень якого прослідковано зміну психічних станів згідно з тактикою ІПМ.

Профіль «Зомбі» – це профіль, який раніше був жертвою ІПМ, а потім у його повідомленнях почали простежувати пропагування ідей ІПМ.

Нейтральний профіль – це профіль, у комунікативній поведінці якого не можна відстежити змін, спровокованих певною тактикою ІПМ.

Види зв'язків між профілями ґрунтуються на інформаційній діяльності учасників онлайн-спільноти, щоб визначити види зв'язків, необхідно дати

відповідь на питання: хто спілкується в одній дискусії, цитує один одного, коментує, ставить уподобання, стежить.

Можливі види зв'язків між профілями залежать від платформи, на якій реалізована спільнота. Наприклад, для профілю у Facebook беремо до уваги кількість друзів профілю. Якщо спільнота реалізована на форумі, то кількість учасників-послідовників. Учасник-послідовник – це учасник, який бере участь у такій кількості дискусій, у яких присутній конкретний маніпулятор, що вона перевищує встановлене експертами порогове значення.

Більшість платформ надає засоби для таких видів діяльності: цитування, коментування, уподобання і відстежування. Крім того, у Facebook можна реалізувати ще такі види діяльності, як «бути другом» та «розміщувати дописи на стіні іншого учасника».

З метою аналізу профілів та зв'язків між ними необхідно побудувати орієнтований граф, вершинами якого є профілі учасників, а ребрами – зв'язки між учасниками.

З погляду інформаційної діяльності профілів кожне ребро графа описане вектором параметрів:

$$Edge = (Cite, Comment, Like, Follow). \tag{3.3}$$

У випадку аналізу профілів у Facebook описуємо ребро ще додатковим вектором параметрів:

$$AdditionalFacebookParameters = (Friend, PostInTimeline). \tag{3.4}$$

Експерти встановлюють вагу кожного з параметрів ребра, тобто кожного виду діяльності. Переважно в такій послідовності у порядку спадання важливості: Cite, PostOnTimeline, Follow, Comment, Friend, Like. Впливовість профілю обчислюємо за такою формулою:

$$\begin{aligned} Influence = {} & CiteCount * CiteWeight + CommentCount * CommentWeight \\ & + LikeCount * LikeWeight + FollowCount * FollowWeight + FriendCount * FriendWeight \\ & + PostInTimeline * PostInTimelineWeight). \end{aligned} \tag{3.5}$$

Експерти встановлюють кількість ребер на які поширюється інформація в онлайн-спільноті залежно від впливовості профілю, який поширює цю інформацію.

Також під час розрахунку шляху поширення необхідно враховувати характеристику ребер – множинність. Множинність – це кількість видів діяльності, які характеризують ребро (рис. 3.6). Наприклад, якщо ребро має всі параметри не нульові, то воно є потенційним шляхом поширення ІПМ.

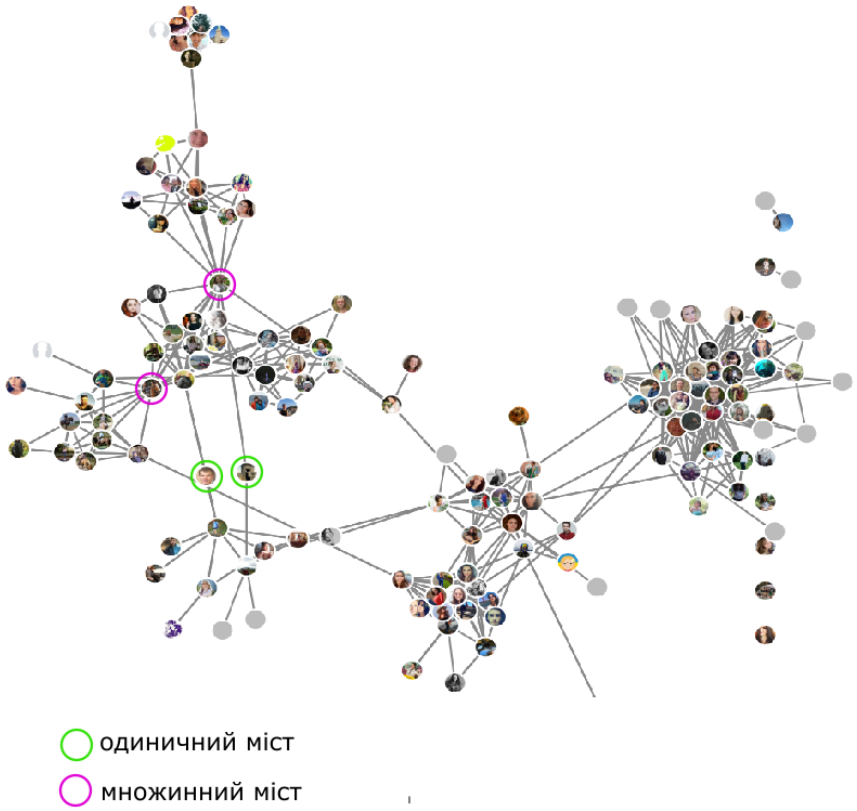


Рис. 3.6. Виявлення можливих шляхів поширення за допомогою аналізу характеристик соціального графа

Аналізуючи будову графа, можна встановити значення вершини за обміном інформації між учасниками спільноти. Важливими за передачею інформації між кластерами, з яких може складатися спільнота, є мости. Міст – це єдиний профіль, який пов’язаний певними видами діяльності з учасниками різних кластерів. Мости бувають двох видів: одиничні і множинні. Одиничні мости – це профілі, які пов’язані зв’язком з одним профілем з кожного

кластера. Множинні – це профілі, які пов’язані зв’язком із багатьма профілями з кожного кластера.

Для виявлення можливих шляхів поширення необхідно визначити, яку роль відіграє впливовий профіль у ІПМ.

Якщо виявлено заражений профіль, то потрібно розрахувати шлях поширення ІПМ, враховуючи характеристики ребер, які виходять із вузла профілю, впливовість даного профілю та наступних профілів, через які перевірятимемо поширення. Тобто, якщо заражено невпливовий профіль, але він перебуває у множинному зв’язку з впливовим профілем, то кількість ребер, на які може поширитись інформація, потрібно розраховувати відносно впливовішого профілю.

Крім того, треба перевірити роль із погляду ІПМ мостів графа і врахувати, що зараження множинних мостів небезпечніше, ніж одиничні.

3.1.4. Етап формування результатів та рекомендацій

На завершальному етапі на основі результатів попереднього етапу даються рекомендації щодо нейтралізації ІПМ та заносять результати в базу даних виявлених прецедентів ІПМ.

Інформацію з бази даних використовують на кожному з етапів алгоритму виявлення маніпуляції. Проміжні та кінцеві результати попередніх аналізів з метою виявлення ІПМ використовуються для пошуку релевантних спільнот, розрахування значень критеріїв підозрілості і вразливості обговорень. База даних містить інформацію про прийоми ІПМ, характеристики маніпулятивних профілів. Крім того, аналізуючи базу даних, можна відслідковувати тенденції маніпуляції, аналізувати помилки алгоритму. Таким чином, ця інформацію використовуватимуть для підвищення ефективності виявлення маніпуляції.

3.2. Методи пошуку релевантних дискусій

Методи пошуку релевантних спільнот необхідні для здійснення першого етапу алгоритму виявлення ІПМ в онлайн-спільнотах (див. розд. 3.1.1 «Підготовчий етап»). Пошук релевантних веб-спільнот за допомогою агрегатора глобальних пошукових систем (агрегатора ГПС) передбачає виконання п’яти послідовних дій. Деталізоване схематичне подання методу, на основі якого побудований засіб пошуку релевантних спільнот, зображене нижче (рис. 3.7).

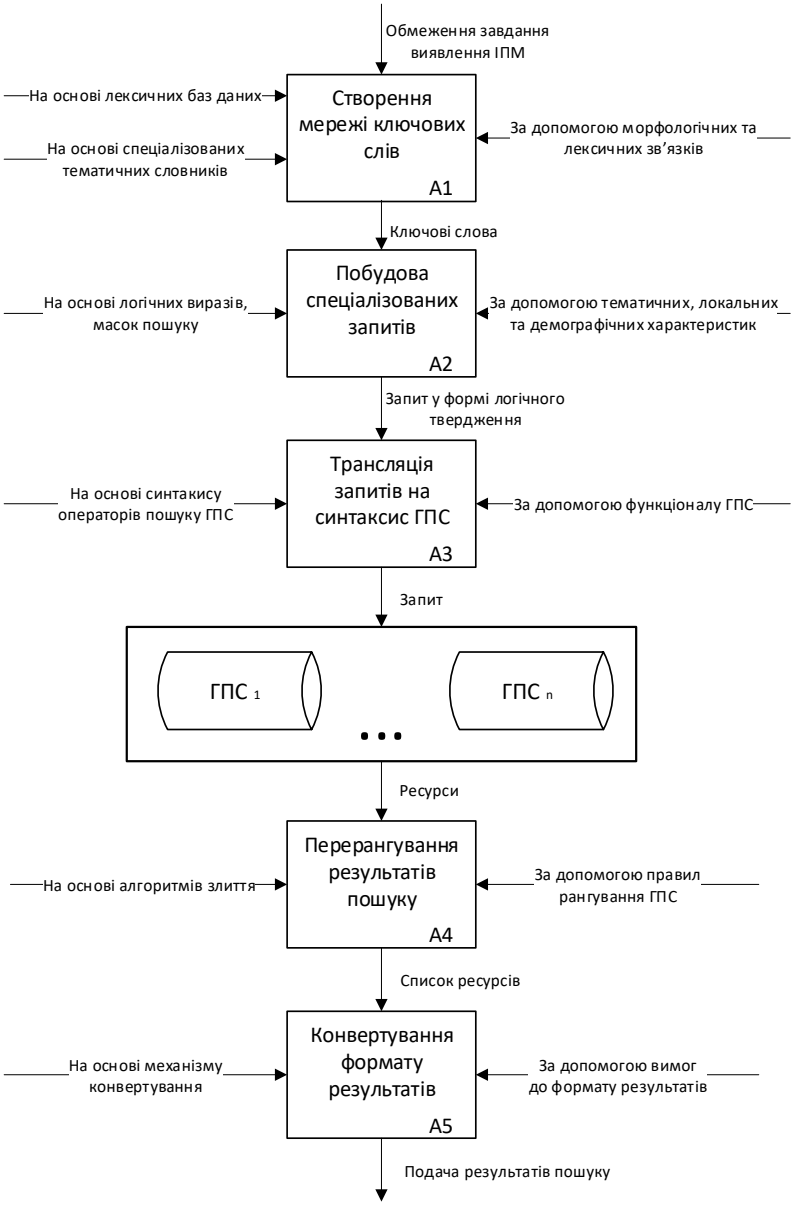


Рис. 3.7. Детальна схема пошуку релевантних дискусій за допомогою агрегатора ГПС

Вхідною інформацією є завдання виявлення ІПМ. Конкретні завдання виявлення ІПМ передбачають обмеження області пошуку, наприклад, може бути задана тематика, локація (онлайн-спільнота, набір дискусій), соціально-демографічні характеристики учасників (стать, вік, професія, захоплення). З метою ідентифікації релевантних онлайн-дискусій необхідно виявляти ІПМ, пов’язані із заданими характеристиками. Соціально-демографічні характеристики учасників виявляємо за допомогою маркерів соціально-демографічних характеристик [106].

Першим кроком є створення мережі ключових слів. Ключові слова до кожного з аспектів тематичної області беремо зі спеціалізованих словників та з лексичних баз даних, наприклад, Wordnet [107]. Під час створення мережі ключових слів керуємось морфологічними та лексичними зв’язками.

На другому кроці для побудови спеціалізованих запитів використовуємо логічні вирази та маски пошуку. Подавши ключові слова, отримані на попередньому етапі за допомогою логічних операторів та масок пошуку, маємо спеціалізовані запити у базовому вигляді, які потребують подальшої трансляції на синтаксис кожної з ГПС, яка входить до складу агрегатора пошукових систем.

Третій крок полягає у перекладі базових спеціалізованих запитів на синтаксис ГПС, з яких складається агрегатор. Деякі ГПС надають можливість використовувати оператори пошуку. Зміст основних операторів пошуку (наприклад, кон’юнкція і диз’юнкція) співпадає, але вони можуть мати різну форму запису. Агрегат ГПС забезпечує трансляцію записів операторів пошуку. У результаті виконання цього кроку запити, подані у відповідному форматі, надходять до ГПС.

На четвертому кроці агрегатор ГПС отримує результати пошуку за допомогою різних ГПС та об’єднує їх у один список, використовуючи алгоритми злиття. Потім переранговує результати пошуку, враховуючи правила рангування кожної ГПС, яка входить до складу агрегатора.

На п’ятому кроці агрегатор отримує список ресурсів із релевантними дискусіями та формує подачу результатів пошуку відповідно до налаштованого формату.

Оскільки сервіс соціальних мереж Facebook є популярною платформою для створення онлайн-спільнот та відрізняється архітектурою від веб-форумів, розглянемо метод пошуку релевантних дискусій за допомогою засобів Facebook. Пошук у Facebook можна поділити на дві стадії: пошук за ключовими словами та пошук за інтересами учасників (рис. 3.8).

Рис. 3.8. Пошук релевантних дискусій у Facebook

Перша стадія передбачає пошук ключових слів серед назв сторінок та груп і пошук ключових слів у постах та коментарях до постів. До того ж, виявлення ключового слова у назві має набагато більшу вагу, ніж у тілі постів чи коментарях до постів. Під час формування мережі ключових слів необхідно враховувати тематичні, локаційні та демографічні обмеження завдання пошуку ІПМ.

На першій стадії необхідно виконати такі кроки:

1. Виявити дискусії, пов’язані з групами, сторінками, постами, персональними профілями, які містять ключові слова у назві.
2. Здійснити пошук ключових слів у коментарях та реакціях на коментарі постів.
3. Відсортувати дискусії за кількістю ключових слів та їхнім розміщенням у структурному елементі поста.

Друга стадія полягає у виявленні користувачів, які опублікували найбільше коментарів у релевантних дискусіях, та аналізі їхніх інтересів щодо інших майданчиків спілкування у Facebook (сторінок, на які вони підписані, поширюють їхню інформацію, коментують і т. ін.). Причому виявляти ступінь активності учасника необхідно відносно до середніх показників активності його діяльності.

На другій стадії необхідно виконати такі кроки:

1. Виявити учасників, які написали найбільше коментарів у дискусіях, виявлених на першій стадії;
2. Виявити майданчики спілкування у Facebook, на яких, виявлені на першому кроці учасники, проявляють значну активність.

Після цього інформація про виявлені майданчики пошуку передається на першу стадію та використовується для пошуку релевантних дискусій.

Отже, використовуючи засіб для пошуку релевантних веб-спільнот (реалізованих на базі веб-форумів) та метод пошуку засобами Facebook, охоплено значний сегмент комунікації в онлайн-спільнотах, що дає змогу виявити велику кількість релевантних до завдання виявлення ІПМ дискусій.

3.3. Методи виявлення підозрілих фрагментів дискусії

Для того, щоб виявити підозрілі фрагменти дискусії, необхідно спершу виявити ознаки прецедентів ІПМ. Кількість і вага ознак ІПМ визначає послідовність аналізу фрагментів дискусії на підетапі виявлення прецеденту ІПМ (див. розд. 3.1.2 «Етап виявлення»).

Система фільтрів для виявлення підозрілих фрагментів ІПМ і засоби виявлення тактик ІПМ ґрунтується на критеріях різної природи. На основі статичних критеріїв виявляють підозрілі профілі та підозрілі фрагменти дискусії (див. розд. 2.3 «Розроблення »). Тоді як засоби виявлення тактик ІПМ базуються виключно на динамічних критеріях.

Аналіз комунікації в онлайн-спільноті з метою виявлення підозрілих фрагментів дискусії починається з найвищого рівня (рис. 3.9), тобто спільноти. Критерії перевірки рівня онлайн-спільноти не потребують складних обчислень, тому за допомогою критеріїв цього рівня можна без значних затрат часу та ресурсів проаналізувати великі об’єми інформації. Якщо фрагмент дискусії був визначений як маніпулятивний наперед встановленим числом фільтрів N^{Filter} цього рівня, то він не передається фільтрам наступних рівнів, а вноситься в базу даних як підозрілий фрагмент дискусії.

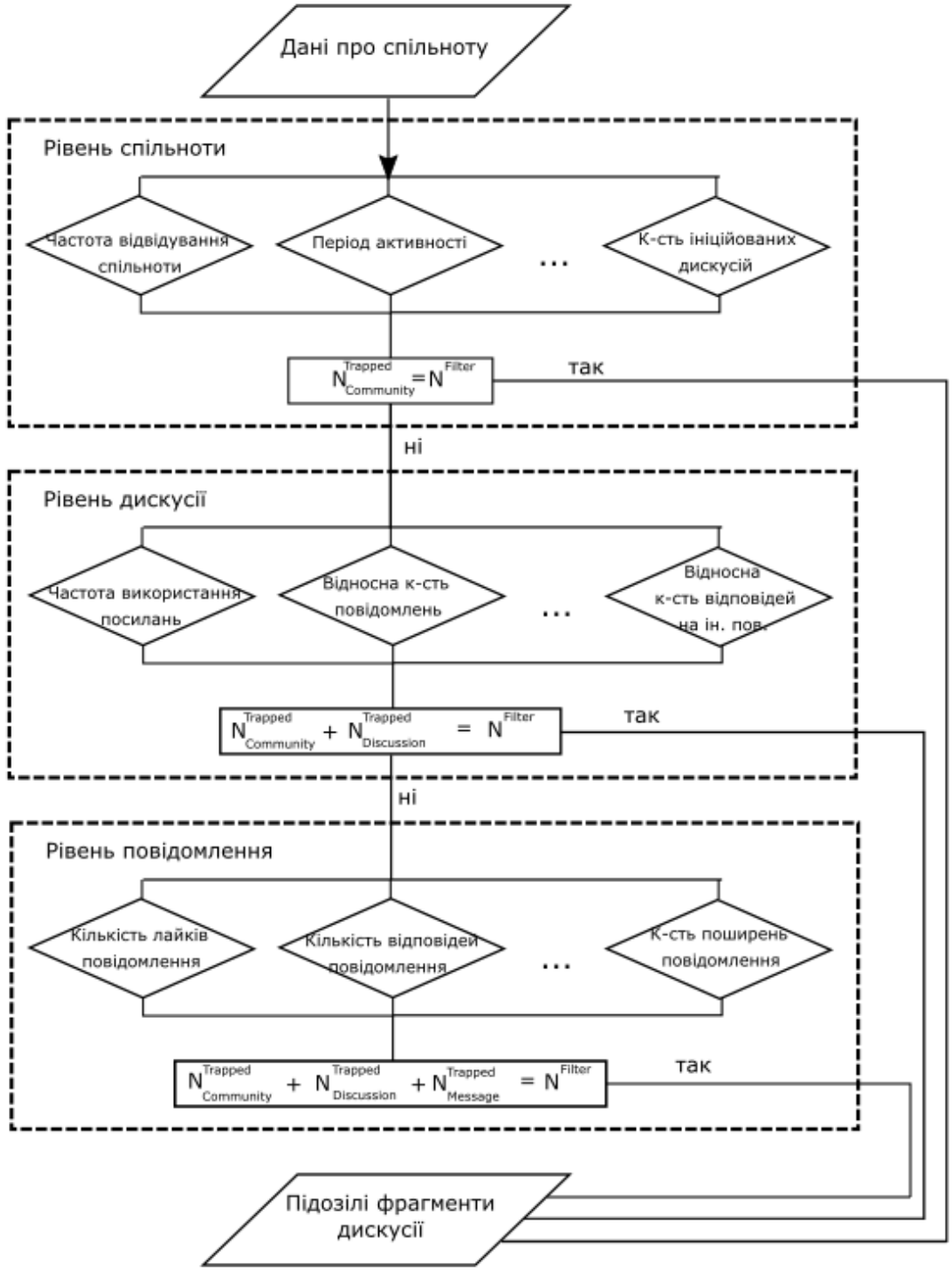


Рис. 3.9. Фільтрування за критеріями організаційно-структурних рівнів

Після структурно-організаційного рівня спільноти проводять аналіз на рівні дискусії. Спершу аналізують фрагменти, число фільтрів попереднього рівня, які їх затримали найближче до N^{Filter}. Якщо кількість фільтрів даного та попереднього рівнів, які затримали фрагменти, дорівнює N^{Filter}, то фрагменти

заносяться в базу даних як ті, що містять ІПМ. Наприклад, кількість фільтрів, які затримали фрагмент на рівні спільноти та на рівні дискусії, дорівнюють встановленому експертами пороговому значенню кількості фільтрів (3.6), то фрагмент дискусії зазначають як підозрілий та передають на наступний підетап для детальнішого аналізу з метою виявлення ІПМ (рис. 3.10).

$$N_{Community}^{Trapped} + N_{Discussion}^{Trapped} = N^{Filter}, \tag{3.6}$$

де $N_{Community}^{Trapped}$ - кількість фільтрів, які затримали фрагмент на рівні спільноти; $N_{Discussion}^{Trapped}$ - кількість фільтрів, які затримали фрагмент на рівні дискусії; N^{Filter} - встановлене експертами порогове значення кількості фільтрів, необхідне для ідентифікації фрагменту як підозрілого.

При сумуванні фільтрів, які затримали фрагменти ІПМ, завжди враховується вага фільтрів (див. розд. 2.3.2 «Формальна модель системи фільтрів для виявлення підозрілих фрагментів дискусій»).

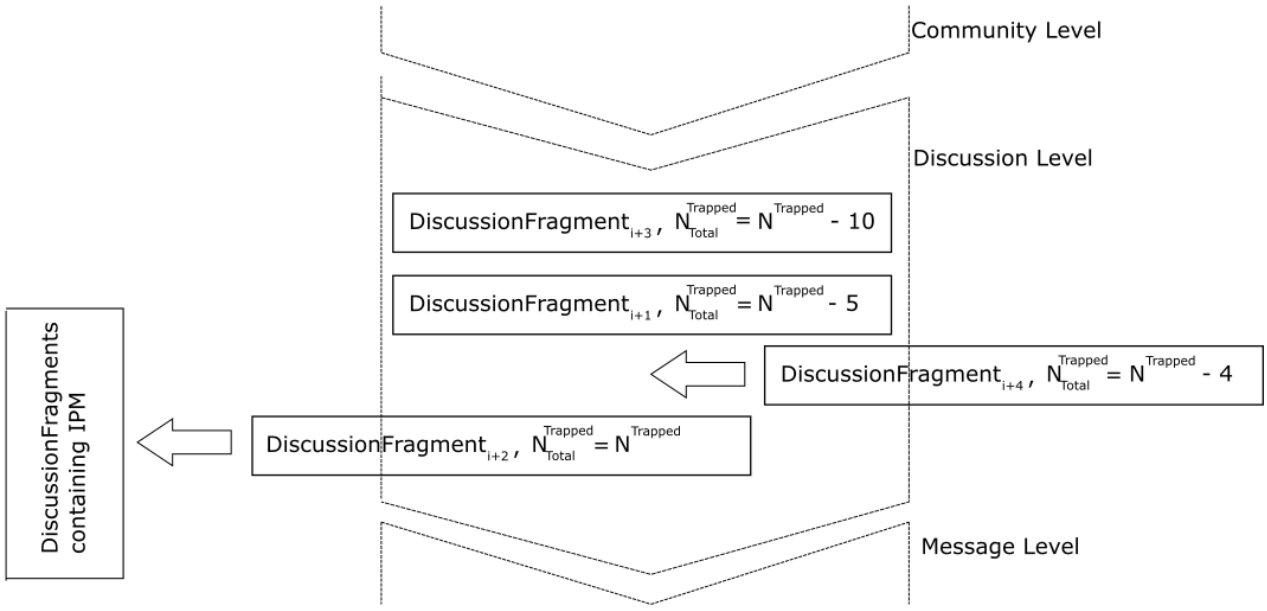


Рис. 3.10. Умова переходу до фільтрування за критеріями наступного рівня

Фільтри кожного наступного рівня вимагають більше даних для обчислення, ніж фільтри попередніх, тому їх доцільно застосовувати лише до фрагментів, які не були затримані фільтрами попередніх рівнів.

Послідовне застосування критеріїв наступних рівнів лише до повідомлень, яких не підозрюють у наявності ІПМ, зменшує обсяг інформації, який необхідно проаналізувати на потенційну наявність ІПМ. Окрім того, в межах кожного з рівнів спочатку застосовують фільтри, які є простими, а потім фільтри, які є складеними (див. розд. 2.3.2 «Формальна модель системи фільтрів для виявлення підозрілих фрагментів дискусій»). Це робиться з метою підвищення ефективності роботи системи фільтрів.

Результатом роботи системи фільтрів є фрагменти дискусій, які запідозрено у наявності ІПМ.

3.4. Методи виявлення послідовності зміни психічних станів учасників дискусії

Першим кроком етапу виявлення прецедентів ІПМ є ідентифікація станів учасників дискусії.

Результатом процесу ідентифікації станів учасників дискусії є множина станів. Після виявлення стани порівнюють з наявними у довідковій базі даних тактиками ІПМ, які представлені за допомогою кусково-лінійних агрегатів як послідовності станів та переходи між ними.

Кожному стану відповідає вектор елементів, який представляє характерні для заданого стану емоції. Елементи можуть набувати значення з певного діапазону і, таким чином, відображати інтенсивність певної емоції (див. розд. 2.2.1 «Формальна модель тактики »). Інтенсивність емоцій виявляємо на основі наявності слів-інтенсифікаторів.

Перед початком виявлення емоцій необхідно здійснити об’єктно-орієнтований тональнісний аналіз (entity-centric analysis) відповідно до тематики ІПМ, вказаної в завданні виявлення, на основі мережі ключових слів. Це дасть змогу визначити, якими об’єктами дійсності є афективні прояви учасника, часто тактики ІПМ передбачають спрямування різних емоцій на

різні об’єкти дійсності з метою досягнення мети. Наприклад, гнів на одну компанію та довіра до іншої.

Для порівняння виявлених емоцій з елементами кортежів психічних станів, перебачених тактикою, використовуємо дерево емоцій. Дерево емоцій дає змогу виявляти емоції на різному рівні точності, наприклад, узагальнені емоції-класи чи емоції-листки дерева (рис. 3.11). Дерево емоцій дозволяє враховувати похибку при виявлені суміжних станів, наприклад, якщо виявлено щастя, а згідно з моделлю тактики ІПМ цей стан має характеризувати радість.

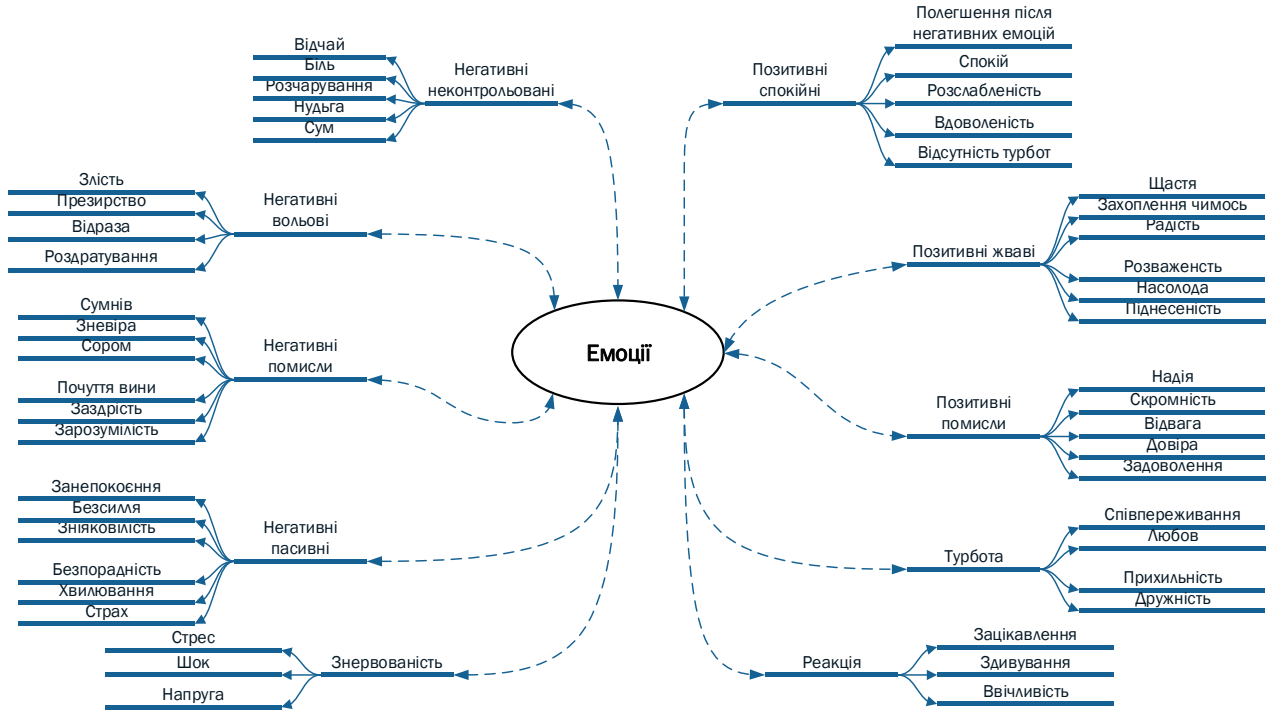


Рис. 3.11. Дерево емоцій

Виявлення станів здійснюється на основі інструментарію аналізу тональності.

Аналіз тональності підозрілого фрагмента дискусії проходить у кілька стадій. Спершу виявляємо емоційно насичені фрагменти дискусій. Проводимо загальний аналіз тональності фрагмента і ділимо повідомлення учасників на позитивні та негативні. Пізніше створюємо мапу настроїв (mood map) та виявляємо кластери повідомлень із найбільшою емоційною концентрацією.

Потім проводимо детальний тональнісний аналіз повідомлень кластерів, щоб якнайточніше виявити конкретні емоції, тобто виявити емоції-листки (рис. 3.11).

Виявлені емоції порівнюють із формальними моделями станів (кожен стан характеризує вектор емоцій). Якщо чітко ідентифікувати 50% емоцій стану та ідентифікувати 40% емоцій, які належать до того самого класу емоцій, що й емоції у формальній моделі, то реципієнт перебуває в згаданому психічному стані.

3.5. Методи виявлення прийомів ІПМ у дискусіях онлайн-спільнот

Виявивши множину тактик ІПМ за психічними станами учасників дискусії, для того, щоб верифікувати наявність ІПМ у дискусії, необхідно виявити прийоми ІПМ, які спричиняють послідовний перехід учасників з одного стану в інший.

Перевірка починається з виявлення вказівників прийомів ІПМ, тобто маркерів, які наявні у найбільшій кількості варіантів реалізації конкретних прийомів.

Виявлення прийомів ІПМ відбувається на основі аналізу інформаційного наповнення повідомлень. Під час аналізу враховуємо параграфічні, невербальні та вербальні елементи повідомлення. Зміст метаграфеміки у повідомленнях визначають на основі таблиць для інтерпретації, значення емотиконів на основі таблиць емотиконів. Семантичні елементи, як найпоширеніші і найбільш значущі, аналізують за допомогою трьох підходів, зокрема: з погляду діалогічних актів, синтаксичної структури та концептуального складу.

Синтаксичні маркери прийомів ІПМ подаємо як характерні синтаксичні структури речень, за допомогою яких реалізовують ці прийоми. Для аналізу мов із вільним порядком слів у реченні надають перевагу граматиці

залежностей перед граматикою із фразовою структурою [108]. З огляду на те, що наша дисертаційна робота орієнтована передусім на аналіз україномовного сегмента онлайн-спільнот, то для формалізації синтаксичних структур використаємо дерева залежностей.

Існують програми для подання англомовних та німецькомовних речень за допомогою дерев залежності, наприклад, displaCy Dependency Visualizer [109]. Завданням розроблення такого модуля для української мови займається Інститут мовознавства ім. О. О. Потебні та лабораторія комп'ютерної лінгвістики Київського національного університету імені Тараса Шевченка [110]. Передбачено, що відповідний модуль входитиме до системи автоматизованого граматичного аналізу українського тексту (АГАТ). На сьогодні модуль ще не розроблено, хоча сформовано певні правила виявлення типів синтаксичної структури, а також типи елементів та типи зв'язку, з яких складаються структури. Правила ґрунтуються суто на морфологічних формах слів. Нижче наведено приклади елементів синтаксичних структур, які використано для подання синтаксичних структур прийомів ІПМ (табл. 3.2).

Таблиця 3.2

Приклади елементів синтаксичних структур

Код	**Елемент синтаксичної структури**	**Приклад**
ГБ	Аналітичний майбутній час	буду робити
ГЗ	Аналітичний наказовий час	хай робить
ГЧ	Умовний спосіб	робив би
ПМ	Складений підмет	один із них
ПС	Складений присудок	почав робити

Також у модулі для синтаксичного аналізу використовують такі синтаксичні зв'язки: координація (КЗ), підрядний ядровий (ПЯ), підрядний ад'юнктивний (ПА), сурядний (СУ).

Як ще один вербальний маркер прийомів ІПМ використовуємо також діалогічні акти (див. розд. 1.4 Аналіз дискусій онлайн-спільнот за допомогою діалогічних актів). Для англійської та німецької мов розроблено системи

маркерів для анотування висловлень за допомогою діалогічних актів. Анотування висловлень за цими системами не потребує експертних знань, тим не менше автоматизоване анотування можливе лише для вищих рівнів ієрархії діалогічних актів.

Ми адаптували систему ДА для аналізу україномовних текстів. Відсутність корпусу україномовних дискусій не дає змоги застосувати алгоритми, використані у дослідженні [65] для розроблення маркерів ДА вищих рівнів ієрархії, хоча адаптована нами система ДА уможливлює виявлення ДА у дискусіях онлайн-спільнот персоналом без відповідної підготовки.

У випадку неоднозначності ідентифікації висловлення як певного ДА рекомендуємо маркувати висловлення як ДА вищого рівня ієрархії. У результаті під час порівняння наявних у повідомленні ДА, за умови збігу інших маркерів ІПМ чи виявлення попередніх кроків тактики ІПМ, виявлені маркери нижчого рівня необхідно увідповіднити маркерам вищого рівня ієрархії.

Семантичні змінні також використовується як вербальні маркери ІПМ. СЗ будуємо за допомогою семантичних мереж та тезаурусів [56].

Для виявлення семантичних змінних використовуємо системи автоматичного стемінгу слів (тобто відкидання флексій). Потім за допомогою концептуальних схем встановлюємо приналежність лексеми до певного поняття. Пізніше порівнюємо наявні поняття зі схемами, які властиві прийомам ІПМ. А також аналізуємо його значення у висловленні відповідно до тематики, вказаної у завданні ІПМ. Існують проекти, які займаються розробленням автоматизованих засобів для опрацювання української мови, наприклад, ВЕСУМ [111, 112].

Подібні механізми і методи роботи з семантичними структурами та елементами використано у програмі для аналізу англомовних текстів FrameNet. FrameNet – це семантичний ресурс для англійської мови, який

складається з концептів чи сценаріїв, названих фреймами. FrameNet має набір з 1000 фреймів, кожен з яких задано лінгвістичними виразами (лексичних одиниць) та структурований відповідно до дійових осіб фрейму (елементами фрейму). На основі FrameNet дослідники розробили Semafor – фреймово-семантичний парсер із відкритим сирцевим кодом. Semafor використовує двоетапну статистичну модель, яка аналізує лексичні одиниці (слова і фрази) у їхніх поняттєвих контекстах та на їх основі робить припущення про фреймово-синтаксичну структуру висловлення.

Отже, на основі трьох видів маркерів ІПМ, а саме вербальних, невербальних та паравербальних, при чому перевагу надаємо вербальним, адже це домінантний спосіб подання інформації у текстових онлайн-спільнотах, виявляємо прийоми ІПМ в онлайн-спільнотах. Для пошуку маркерів у дискусіях онлайн-спільнот повторюємо описану в підрозділі послідовність кроків.

Один із видів посилань, які використовуються в ІПМ, є нерелевантні посилання. Визначення релевантності повідомлення полягає в аналізі його інформаційного наповнення. Тобто визначенні важливості ключових слів повідомлення в інформаційному наповненні ресурсу, на який вказує посилання на рис. 3.12.

Рис. 3.12. Перевірка посилання на релевантність

Визначення показників важливості ключових слів повідомлення у інформаційному наповненні ресурсу здійснюється на основі таких показників: частота вживання слова в первинному ресурсі, на який веде посилання; інверсія частоти, з якою слово трапляється в наборі інтернет-ресурсів, які видала пошукова система внаслідок пошуку за ключовими словами.

Частота вживання слова в первинному ресурсі обчислюється за такою формулою:

$$InnerTermFrequency = \frac{|Word_i|}{|Word|}, \tag{3.7}$$

де $Word_i$ — це слово в інформаційному наповненні ресурсу; $Word = \{Word_i\}_{i=1}^{N^{Word}}$ — множина всіх слів у інформаційному наповненні ресурсу.

Інверсія частоти, з якою слово трапляється в наборі інтернет-ресурсів, які видала пошукова система внаслідок пошуку за ключовими словами, обчислюється за такою формулою:

$$OuterTermFrequency = \log \frac{N^{\mathrm{Re}\,source}}{|(\mathrm{Re}\,source_i \supset Term_i)|}, \tag{3.8}$$

де $N^{\mathrm{Re}\,source}$ — кількість вторинних ресурсів, виявлених унаслідок пошуку за ключовими словами; $|(\mathrm{Re}\,source_i \supset Term_i)|$ — кількість вторинних ресурсів, які містять і-й терм.

На основі наведених нижче ознак можна виявляти посилання на неавторитетні джерела (табл. 3.3). Посилання на неавторитетні джерела, як зазначено у розділі 1.2.2 «Види невербальних та паравербальних засобів, які використовують для реалізації ІПМ», потрібно виявляти на двох рівнях: на користувацькому та на рівні аналізу кодування символів, тобто на front-end та back-end рівнях відповідно.

Таблиця 3.3 Види посилань на неавторитетні ресурси

	Приклад	Приклад реалізації
З back-end пасткою	Містять символи кількох алфавітних систем	https://www.facebook.com/ оо – написні кирилицею
	Містять Unicode символ U+202E	http://www.nationalgeograpmoc[[U+202E]].hic
	Містять візуально подібні символи (наприклад капітелі)	http://www.britishcouncil.org.ua/ u – маленька велика буква
З front-end пасткою	Доменні імена яких написані з помилкою	http://www.kredobonk.com.ua/ о замість а
	Доменні імена яких дуже схожі на оригінал	http://pravda.if.ua/ замість http://www.pravda.com.ua/
	Доменні імена яких містять доменне ім’я оригінального ресурсу	http://www.bbc.com.com/ замість http://www.bbc.com/

Висновки до розділу

У розділі розроблено алгоритм виявлення ІПМ в онлайн-спільнотах, зокрема описано чотири етапи алгоритму:

Підготовчий етап полягає у пошуку релевантних онлайн-спільнот у веб та у Facebook. Це досягнуто за допомогою агрегатора пошукових систем, методу пошуку у Facebook.

Етап виявлення полягає у виявленні фрагментів онлайн-дискусій, які характеризуються частими змінами емоцій учасників, у встановленні за допомогою інструментарію тональнісного аналізу емоцій учасників, на основі емоцій психічних станів. Відповідно до інформації про стани та послідовність їхньої зміни, робиться припущення щодо тактики ІПМ у дискусії. Припущення перевіряють пошуком у дискусії прийомів ІПМ, які можуть переводити учасника у відповідні стани. Якщо прийоми виявлені, то припущення щодо тактики ІПМ підтверджено.

Етап нейтралізації полягає у виявленні груп профілів, з яких відбувається маніпулятивна діяльність на основі визначення характеристик мовного стилю, поведінки та профілю маніпулятора, та порівняння їх з відповідними характеристиками учасника. Крім того, на цьому етапі

встановлюють можливі шляхи поширення ІПМ на основі побудованого соціального графа онлайн-спільноти та визначення його метрик.

Етап формування результатів та рекомендацій полягає у поданні результатів моніторингу, а також рекомендацій щодо нейтралізації актуальних прецедентів ІПМ та запобіганню майбутнім ІПМ.

Розділ 4. Розроблення програмно-алгоритмічного комплексу виявлення ІПМ

У четвертому розділі побудовано програмно-алгоритмічний комплекс виявлення ІПМ на основі розроблених формальних моделей онлайн-спільноти і тактики ІПМ, а також алгоритму виявлення ІПМ в онлайн-спільнотах та методів та засобів, передбачених етапами алгоритму. Програмно-алгоритмічний комплекс розроблено для виявлення ІПМ у певній онлайн-спільноті або щодо певної організації в онлайн-спільнотах. Програмно-алгоритмічний комплекс передбачає виконання завдань виявлення ІПМ з різним ступенем деталізації умов.

Програмно-алгоритмічний комплекс дає змогу:

- задавати параметри завдання виявлення ІПМ;
- відслідковувати результати проміжних етапів та налаштувати процеси та параметри необхідні для виконання дій наступних етапів відповідно до отриманих проміжних результатів;
- уточнити параметри необхідні для виявлення ІПМ відповідно до особливостей комунікацій та структури певної спільноти;
- отримати результати роботи алгоритму у одному з кількох варіантів подання, які відрізняються рівнем деталізації певних ознак чи проміжних етапів.

Споживачами комплексу є відділи інформаційної діяльності організації та адміністративна ланка онлайн-спільнот (адміністратори, модератор, контент-менеджери і т. д.).

Архітектура програмно-алгоритмічного комплексу виявлення ІПМ заснована на методах, описаних у цій дисертаційній роботі. Комплекс передбачає використання БД. Роботу програмно-алгоритмічно комплексу продемонстровано під час використання комплексу і з метою виявлення ІПМ, і

з метою наповнення БД інформацією, необхідною для виявлення ІПМ. Наведено приклади подання тактики ІПМ за формальної моделі ІПМ та приклади подання синтаксичної структури виявлених форм реалізації прийомів ІПМ. Проаналізовано результати роботи програмно-алгоритмічного комплексу і оцінено його ефективність.

Основні результати розділу автор опублікував у дослідженнях [113, 114, 115].

4.1. Загальна схема програмно-алгоритмічного комплексу виявлення ІПМ

Ґрунтуючись на особливостях та закономірностях комунікації в онлайн-спільнотах, схемах традиційних офлайн-тактик маніпуляції, розроблено архітектуру системи виявлення ІПМ.

Програмно-алгоритмічний комплекс виявлення ІПМ використовує як дані, зібрані для вирішення актуального завдання, так і дані, збережені в БД внаслідок виконання попередніх завдань моніторингу. Наприклад, для виявлення дискусій, які потенційно містять ІПМ, використовуємо дані, зібрані під час виконання попередніх завдань виявлення ІПМ (див. розд. 3.1.1 «Підготовчий етап»).

В основі системи лежать формальна модель онлайн-спільноти (див. розд. 2.1 «Спеціальна модель онлайн-спільноти з погляду виявлення ІПМ») та модель тактики ІПМ, подана за допомогою кусково-лінійних агрегатів (див. розд. 2.2.1 «Формальна модель тактики ІПМ»).

Для реалізації системи використано технології глобального пошуку релевантних до поставленого завдання дискусій, для виявлення психічних станів учасників дискусії використано інструментарій тональнісного аналізу, для виявлення прийомів ІПМ використано методи анотування діалогічних актів, метод представлення висловлення за допомогою семантичних змінних, метод подання синтаксичних структур за допомогою дерев залежності.

Крім того, розроблено систему критеріїв для сортування дискусій за сприятливістю до наявності ІПМ (див. розд. 3.1.1 «Підготовчий етап»), систему фільтрів для виявлення підозрілих фрагментів дискусії (див. розд. 2.3 «Розроблення »).

Система виявлення ІПМ аналізує онлайн-спільноту відносно учасників онлайн-спільнот та безпосередньо інформаційне наповнення, створене учасниками, відповідно для реалізації алгоритму використовуємо учасницько-орієнтований та контенто-орієнтований підходи. Дані про учасників використовуємо на початковому та завершальних етапах алгоритму (див. розд. 3.1.1 «Підготовчий етап», 3.1.3 «Етап нейтралізації», 3.1.4 «Етап формування результатів та рекомендацій») з метою окреслення області для моніторингу. Дані, отримані внаслідок аналізу інформаційного-наповнення, створеного учасниками. використовуються для виявлення прецедентів ІПМ (див. розд. 3.1.2 «Етап виявлення»).

Інтеграція цих двох підходів забезпечує ефективність системи, оскільки більш трудо- і часозатратний аналіз, зокрема аналіз інформаційного наповнення, виконується лише для окресленої на попередньому етапі області. Відсіявши профілі учасників, які не є потенційно небезпечними, можна змешити обсяг даних, які потрібно перебрати. Інформаційну діяльність відфільтрованих підозрілих профілів детальніше аналізується на наявність ІПМ.

4.1.1. Архітектура програмно-алгоритмічного комплексу моніторингу онлайн-спільнот

Програмно-архітектурний комплекс моніторингу онлайн-спільнот з метою виявлення ІПМ складається із таких підсистем (рис. 4.1):

- підсистеми пошуку тематичнорелевантних дискусій;
- підсистеми пошуку релевантих дискусій у Facebook,
- підсистеми сортування дискусій;

- підсистеми виявлення прецедентів ІПМ;
- підсистеми нейтралізації;

Комплекс містить дві глобальні бази даних: онлайн-спільнот та виявлених прецедентів ІПМ. Крім зазначених баз даних, деякі підсистеми містить внутрішні бази даних: наприклад, підсистема виявлення прецедентів ІПМ. Внутрішня БД цієї підсистеми містить інформацію про тактики ІПМ, діалогічні акти, семантичні змінні, синтаксичні структури, емоції та стани учасників.

Підсистема пошуку тематичнорелевантних веб-дискусій та **підсистема пошуку релевантних дискусій у Facebook** виконують завдання підготовчого етапу алгоритму моніторингу онлайн-спільнот з метою виявлення ІПМ (див. розд. 3.1.1 «Підготовчий етап»). Зібрана інформація записується в глобальну БД онлайн-спільнот.

Підсистема пошуку релевантних дискусій виконує завдання підготовчого етапу алгоритму виявлення ІПМ в онлайн-спільнотах (див. розд. 3.1.1 «Підготовчий етап»), а саме пошуку дискусій релевантних до завдання виявлення ІПМ. Підсистема складається з агрегатора пошукових систем та містить внутрішні лексичні БД, спеціалізовані тематичні словники, БД шаблонів пошуку.

Підсистема сортування дискусій виконує завдання підготовчого етапу алгоритму виявлення ІПМ в онлайн-спільнотах (див. розд. 3.1.1 «Підготовчий етап»), а саме формування черги дискусій за критеріями популярності, підозрілості та вразливості. Дискусії, які мають найвищі значення цих критеріїв, розміщуються на початку черги. Підсистема сортування дискусій використовує для розрахунку критеріїв інформацію з БД онлайн-спільнот.

Підсистема виявлення прецедентів ІПМ виконує завдання етапу виявлення алгоритму виявлення ІПМ в онлайн-спільнотах (див. розд. 3.1.2

«Етап виявлення»). Результат, тобто виявлені прецеденти ІПМ, підсистема зберігає до глобальної БД онлайн-спільнот.

Підсистема виявлення прецедентів ІПМ складається з таких компонентів:

- компонента виявлення підозрілих фрагментів дискусії;
- компонента виявлення станів учасників;
- компонента виявлення можливих тактик;
- компонента виявлення прийомів ІПМ.

Підсистема опрацьовує чергу релевантних онлайн-спільнот відсортовану за потенційною небезпекою появи ІПМ.

Компонент виявлення підозрілих фрагментів дискусії складається з різнорівневих фільтрів, які виявляють підозрілі фрагменти ІПМ на основі динамічних та статичних критеріїв (див. розд. 2.3 «Розроблення »). Різнорівневість системи забезпечує оперативність виявлення підозрілих фрагментів, оскільки спочатку інформаційне наповнення дискусій проходить перевірку за допомогою фільтрів, які не вимагають складних розрахунків та текстового аналізу. Компонент виявлення підозрілих фрагментів дискусії отримує статичні та динамічні критерії підозрілої діяльності, а також вагові значення необхідні для роботи фільтрів із відповідних таблиць глобальної БД.

Компонент виявлення станів учасників виявляє стани учасників дискусії на основі інформаційного наповнення їхніх повідомлень за допомогою інструментарію тональнісного аналізу. Компонент містить БД станів учасників та емоцій, які характеризують ці стани. Інформацію про виявлені стани компонент передає для зберігання у БД онлайн-спільнот та наступному компоненту підсистеми, тобто компоненту виявлення можливих тактик ІПМ.

Компонент виявлення можливих тактик ІПМ ідентифікує тактики ІПМ на основі послідовностей виявлених станів. Компонент містить внутрішню БД формалізованих за допомогою кусково-лінійних агрегатів тактик ІПМ. Результатом підсистеми є набір можливих тактики ІПМ, які ідентифіковані на основі станів виявлених у певному фрагменті дискусії.

Компонент виявлення прийомів ІПМ перевіряє наявність прийомів, які відповідають ідентифікованій тактиці ІПМ, у фрагменті дискусії. Виявлення прийомів здійснюється на основі текстового аналізу повідомлень за допомогою таких компонент:

- компонента виявлення діалогічних актів;
- компонента виявлення семантичних змінних;
- компонента виявлення синтаксичних структур.

Компонент виявлення прийомів ІПМ містить базу даних маркерів прийомів ІПМ, розроблених на основі діалогічних актів, семантичних змінних та синтаксичних структур висловлень. Результат, тобто ідентифіковану на основі послідовності виявлених прийомів тактику ІПМ, компонент передає у глобальну базу даних виявлених прецедентів ІПМ.

Підсистема нейтралізації ІПМ виконує дії етапу нейтралізації ІПМ алгоритму моніторингу онлайн-спільноти з метою виявлення ІПМ (див. розд.о 3.1.3 «Етап нейтралізації»). Підсистема складається з двох компонентів:

- компонента виявлення стилю маніпулятора;
- компонента виявлення можливих шляхів поширення ІПМ.

Підсистема працює на основі інформації, зібраної в БД, виявлених ІПМ під час моніторингу онлайн-спільноти. Підсистема містить внутрішню базу даних характеристик зв’язків між учасниками спільноти. Цю базу даних використовує компонент виявлення можливих шляхів поширення ІПМ. Результати, тобто підозрілі профілі, виявлені на основі подібностей стилю, та можливі шляхи поширення ІПМ, передаються в базу даних онлайн-спільнот.

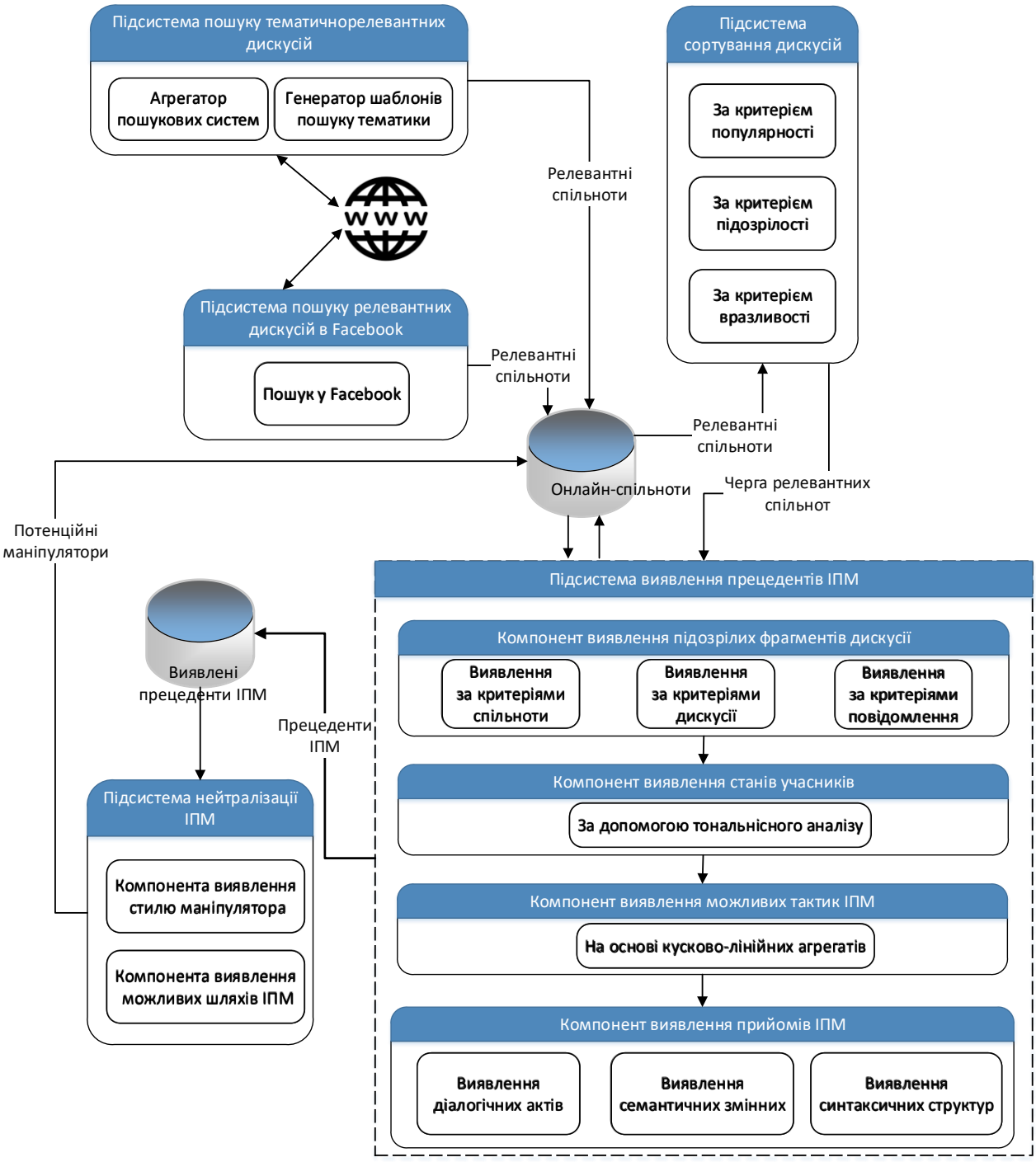


Рис. 4.1. Архітектура програмного комплексу виявлення ІПМ в онлайн-спільнотах

4.1.2. Схеми баз даних

Програмно-алгоритмічний комплекс моніторингу онлайн-спільнот із метою виявлення ІПМ містить дві глобальні БД, а саме онлайн-спільнот та виявлених прецедентів ІПМ. Крім того, підсистеми і компонети комплексу містять внутрішні БД, а саме: БД характеристик зв'язків між учасниками, спеціалізовані тематичні словники, БД шаблонів пошуку, БД тактик ІПМ, БД станів та емоцій учасників.

Бази даних містять таблиці трьох видів: довідкові і фактологічні таблиці та спеціалізовані словники. Довідкові таблиці містять дані, які покладені в основу алгоритму виявлення ІПМ в онлайн-спільнотах, наприклад, типи та компоненти ІПМ, структурні елементи онлайн-спільнот. Фактологічні таблиці містять інформацію про прецедент ІПМ. Спеціалізовані словники містять маркери, за допомогою яких можна ідентифікувати необхідні для виявлення ІПМ об'єкти.

Довідкові таблиці використовуються для співвідношення класів і типів компонентів ІПМ та онлайн-спільнот з їхніми реальними прикладами. Наявність довідкових таблиць забезпечує зручність використання системи виявлення ІПМ.

Спеціалізовані словники містять інформацію, отриману з суміжних із даним дослідженнями, наприклад це словники маркерів демографічних характеристик, словник маркерів висловлень позиції.

4.1.2.1. Інфологічна модель онлайн-спільноти

База даних онлайн-спільнот — це глобальна база даних, що містить вихідну, проміжну та результуючу інформацію, яку використовують під час виконання алгоритму моніторингу онлайн-спільнот з метою виявлення ІПМ.

Інфологічна схема онлайн-спільноти містить такі сутності (рис. 4.2):

Сутність **Community** містить дані про спільноти, які перевіряють на наявність ІПМ. Важливі для процесу моніторингу ознаки сутності описують такі атрибути: **Community ID** — унікальний ідентифікатор екземплярів цієї сутності, тобто онлайн-спільнот, **Community Title** — назва онлайн-спільноти, **Creation Data** — дата створення онлайн-спільноти, **Language** — домінантна мова створюваного користувачами інформаційного наповнення, **Community Type** — тип спільноти за ступенем інтеграції у веб (див. розд. 1.1.1 «Огляд існуючих підходів до класифікації онлайн-спільнот»).

Сутність **Discussion** містить дані про дискусії, що належать до аналізованих спільнот. Цю сутність описують такі атрибути: **Discussion ID** — унікальний ідентифікатор дискусії, **Discussion Title** — назва дискусії, **Discussion Topic** — тематика дискусії, **Discussion Author** — учасник, який створив дискусію, **First Message Date** — дата першого повідомлення в дискусії.

Сутність **Aptitude for IPM** містить дані про сприятливість онлайн-дискусії до здійснення ІПМ. Цю сутність описують такі атрибути: **Criterion Name** — назва критерію, на основі якого оцінюється сприятливість дискусії до здійснення ІПМ, **Criterion Type** — це тип, до якого належить даний критерій (популярність, підозрілість, вразливість), **Criterion Value** — значення критерію для даної дискусії. **Aptitude for IPM** пов’язана з сутністю **Discussion** зв’язком один-до-одного, оскільки показники розраховуються для кожної дискусії і дискусія не може мати кілька значень одного з показників.

Сутність **Member** містить дані про учасників онлайн спільноти. Цю сутність описують такі атрибути: **Member ID** — унікальний ідентифікатор учасника спільноти, **Member Name** — ім’я учасника спільноти, **Email** — емайл учасника спільноти, **IPM Role** — це роль учасника спільноти у прецеденті ІПМ. Атрибут **IPM Role** може набувати наступних значень: нейтральний, маніпулятор, жертва, зомбі (див. розд. 2.1.2 «Формальна модель учасника спільноти»). **Member Name** та **IPM Role** є обов’язковим атрибутом,

тоді як **Email** є опціональним атрибутом, адже не всі спільноти надають інформацію про емейли учасників.

Сутність **Message** містить дані про учасників онлайн спільноти **Message ID** — унікальний ідентифікатор повідомленння дискусії, **Author Name** — ім’я учасника, який є автором повідомлення, **Time Stamp** — дата та час публікування повідомлення**, Message Content** — інформаційне наповнення повідомлення, **Direction Type** — тип інформативної активності, внаслідок якого було створене повідомленя. Атрибут **Direction Type** може набувати двох значень: ініціююче повідомлення та повідомлення-реакція (див. розд. 2.1.3 «Формальна модель повідомлення») **Relevance** — релевантність повідомлення до тематики дискусії, **Weighted Lexis** — наявність значущих видів специфічної лексики, **Rules Violation Record** — інформація про порушені учасником правила спільноти, **Registration Date** — дата реєстрації учасника.

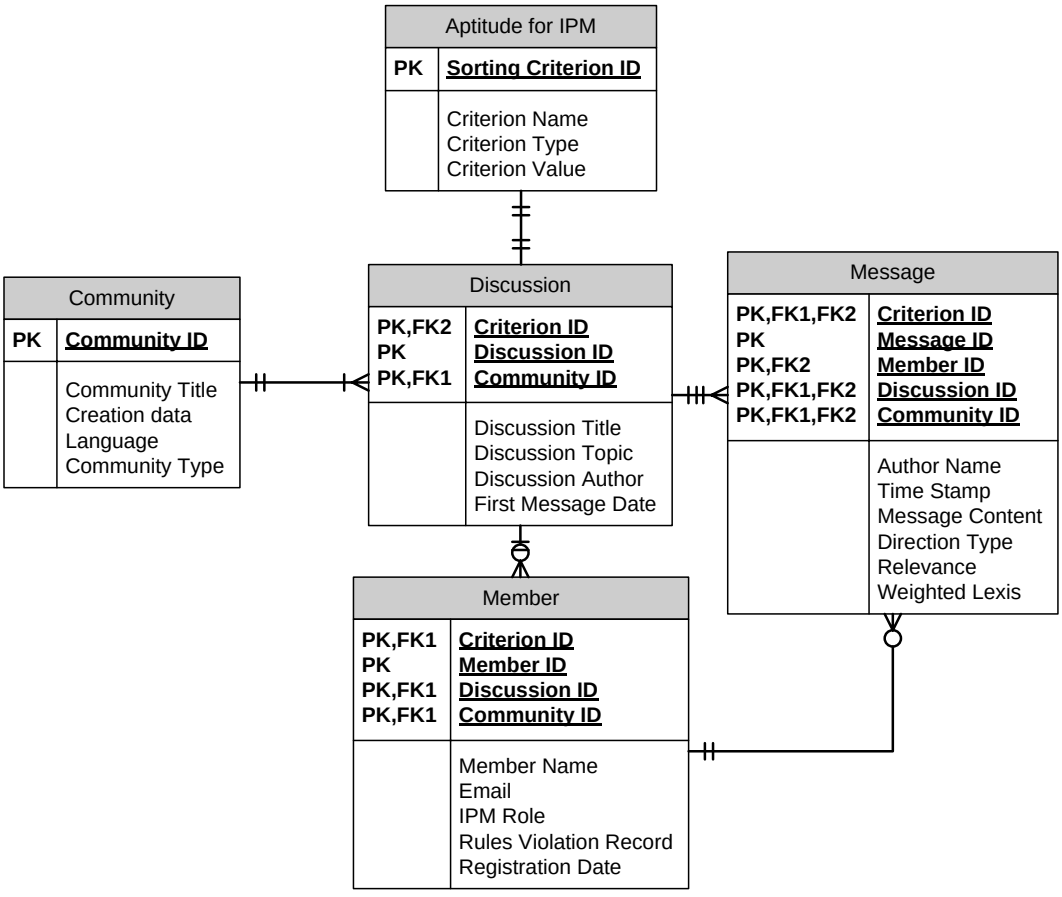


Рис. 4.2. Інфологічна модель онлайн-спільноти

Сутність **Member** пов’язана наведеними нижче сутностями (рис. 4.3):

Сутність **Behavioural Characteristics** містить дані про характерну поведінку учасника в онлайн-спільноті. Сутність **Behavioural Characteristic** описують такі атрибути: **Publishing Frequency** — частота розміщення повідомлень учасником спільноти протягом певного часового періоду, який експерти задають для кожної спільноти окремо (наприклад, 1 місяць), **Reply Ration** — відношення кількості відповідей учасника на повідомлення інших до усіх повідомень опублікованих учасником, **Activity Time** — період активності учасника у онлайн-спільноті**, Self-Centered Activity** — кількість коментарів учасника на власне ініціююче повідомлення до усіх коментарів, **Engagement Discussion Themes** — теми дискусій спільноти, в яких учасник проявив інформаторську активність, **Initiated Discussion Count** — кількість створених учасником дискусій, **Signal Activity Provocativeness** — ступінь сигнальної активності учасника онлайн-спільноти.

Сутність **Profile Features** містить дані, які описують профіль учасника спільноти. Сутність **Profile Features** описують такі атрибути: **Personal Information Completeness** — показник заповненості профілю учасника спільноти, **Friendship Quantity and Quality** — кількість профілів-друзів учасника, які є не фейковими.

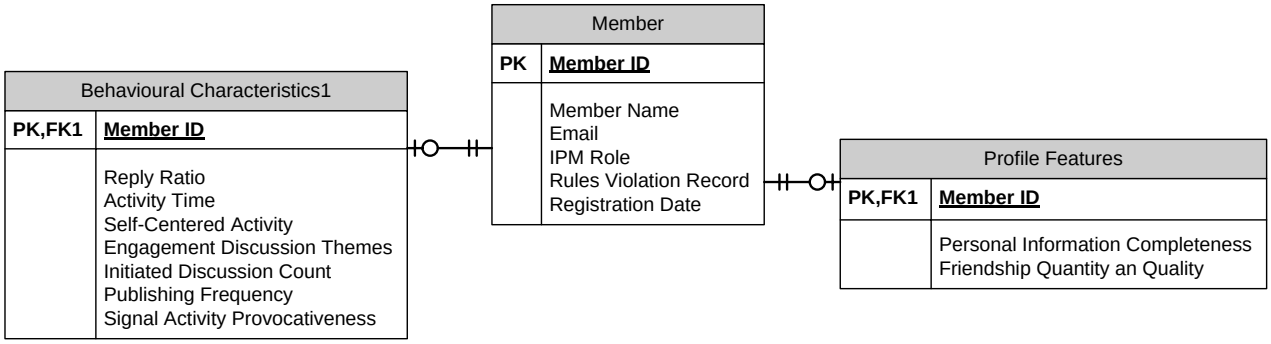


Рис. 4.3. Інфологічна модель учасника онлайн-спільноти

4.1.2.2. Інфологічна модель тактики ІПМ

База даних тактик ІПМ — це внутрішня база даних компонента виявлення прецедентів ІПМ, яка містить довідкову інформацію, необхідну для виявлення застосованої тактики ІПМ.

Інфологічна модель тактики ІПМ складається з таких сутностей (рис. 4.4, рис. 4.5):

Сутність **Tactic Step** містить дані про крок ІПМ, тобто переведення жертви ІПМ з одного стану в наступний. Важливі для процесу моніторингу ознаки сутності **Tactic Step** описують наведені атрибути: **Tactic ID** — унікальний ідентифікатор кроку ІПМ, **Step Title** — назва кроку тактики ІПМ, **Step Parameters** — значення елементів вектору параметрів, які описують крок тактики ІПМ.

Сутність **States** містить дані, які описують стани, в які може переходити жертва ІПМ внаслідок застосування прийомів ІПМ. Сутність **States** описують наступні атрибути: **State ID** — унікальний ідентифікатор стану жертви ІПМ, **States Title** — назва стану жертви ІПМ.

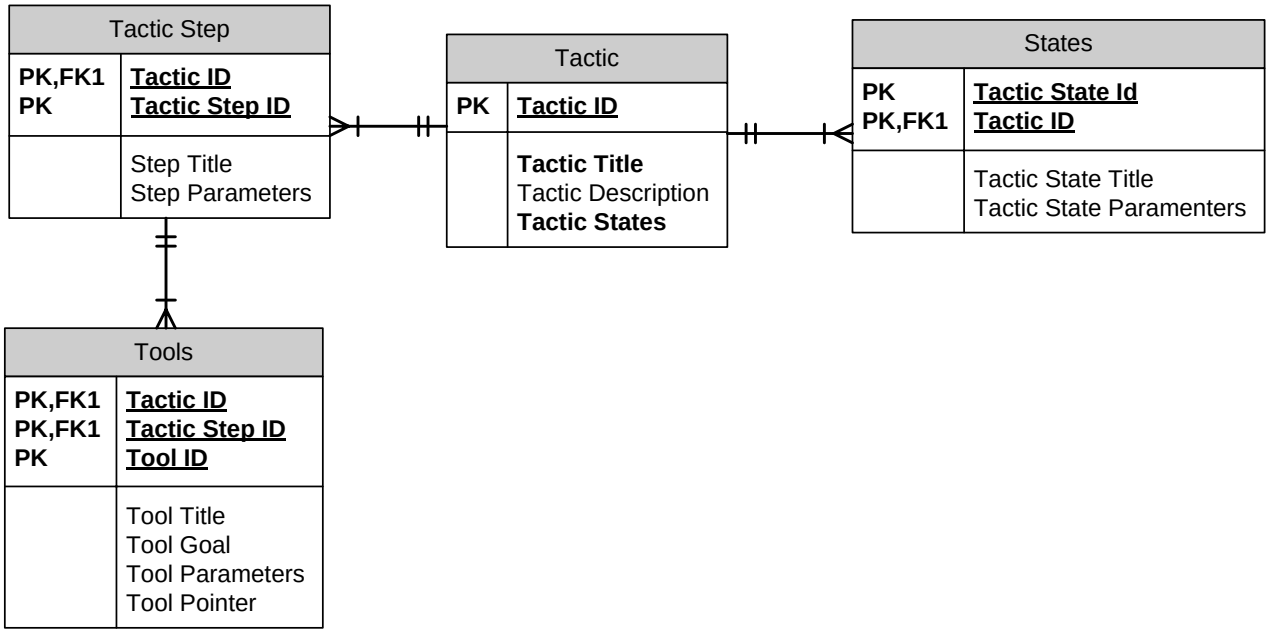


Рис. 4.4. Інфологічна модель тактики ІПМ

Сутність **Tools** містить дані про прийоми, які застосовуються в тактиках ІПМ. У процесі виявлення ІПМ в онлайн-спільнотах прийоми ІПМ ідентифікують на основі певних наперед прописаних ознак. Сутність **Tools** описують такі атрибути: **Tool ID** — унікальний ідентифікатор екземплярів цієї сутності, тобто прийомів ІПМ, **Tool Title** — назва прийому ІПМ, **Tool Goal** — це мета прийому; **Tool Parameters** — це вектор параметрів, який характеризує прийом ІПМ, **Tool Pointer** — це вказівник прийому або найпоширеніші маркери найпоширенішої форми прийому.

Сутність **Syntactic Structure** містить інформацію про синтаксичні структури, за допомогою яких реалізовані прийоми ІПМ, **Syntactic Structure ID** — унікальний ідентифікатор конкретної синтаксичної структури, **Syntactic Structure Type** — тип синтаксичної структури, який відображає основні ознаки будови речення, тобто наявність другорядних членів речення, кількох основ, простої чи складної структури речення (наприклад, поширене, односкладне, повне розповідне, складне з підрядною-сурядністю способу дії і т.д.).

Сутність **Semantic Variable** містить інформацію про семантичні змінні, за допомогою яких здійснено прийом ІПМ. **Semantic Variable ID** — унікальний ідентифікатор конкретної семантичної змінної, **Semantic Variable Class** — клас, до якого належить даний тип семантичної змінної (трьома класами семантичних змінних є концепт, предикат, характеристика) **Semantic Variable Path** — це шлях до конкретної семантичної змінної, який відображає її знаходження в дереві, на основі характерного для цього класу виду зв’язків.

Сутність **Dialog Act** містить інформацію про діалогічні акти, за допомогою яких здійснено прийом ІПМ. **Dialog Act ID** — ідентифікатор діалогічного акту, **Dialog Act Title** — назва діалогічного акту, **Dialog Act Path** — це шлях до конкретного діалогічного акту, який задає його знаходження у дереві ДА.

Сутність **Syntactic Structure Markers** містить маркери синтаксичних структур. **Syntactic Structure Marker ID** — ідентифікатор маркера синтаксичної структури, **Syntactic Structure Marker** — маркер синтаксичної структури.

Сутність **Semantic Variable Markers** містить маркери семантичних змінних. **Semantic Variable Marker ID** — ідентифікатор маркера семантичної змінної, **Semantic Variable Marker** — маркер семантичної змінної.

Сутність **Dialog Act Markers** містить маркери діалогічних актів. **Dialog Act Marker ID** — ідентифікатор маркера діалогічного акту, **Dialog Act Marker** — маркер діалогічного акту.

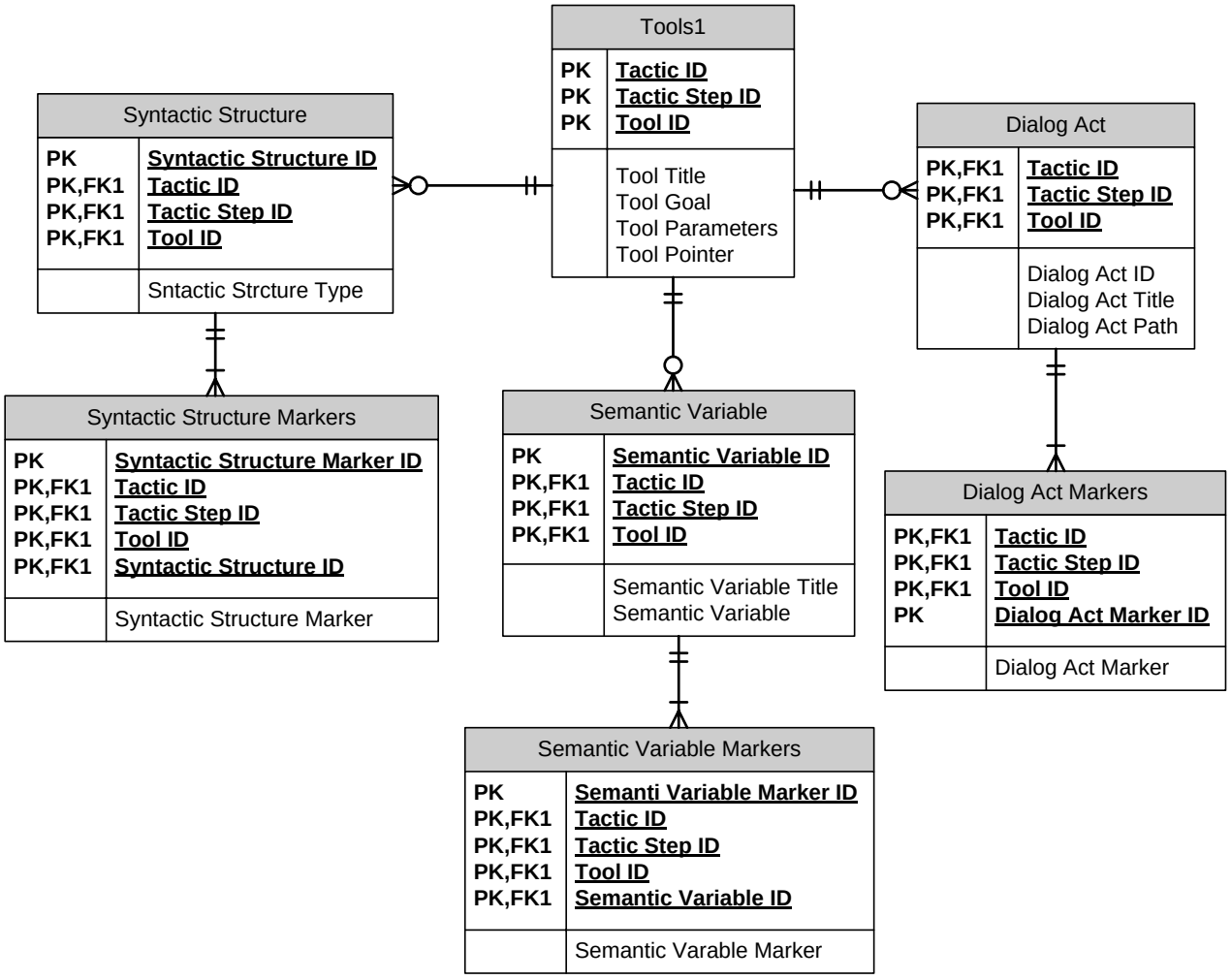


Рис. 4.5. Інфологічна модель прийому ІПМ

4.1.3. Розроблення користувацького інтерфейсу програмно-алгоритмічного комплексу

Графічний користувацький інтерфейс програмно-алгоритмічного комплексу виявлення інформаційно-психологічних маніпуляцій «IPM Detector» розроблений за допомогою Windows Forms та мови програмування C#.

Отримання, опрацювання та аналіз текстової інформації здійснюється за допомогою мови програмування Python 3.6. Для текстового аналізу програмно-алгоритмічний комплекс використовує інструменти бібліотеки NLTK. Для інтеграції модулів програми напсаних на C# і Python використовуємо IronPython.

Програмно-алгоритмічний комплекс має два інтерфейси: для користувача і для адміністратора. Користувацький інтерфейс передбачає використання комплексу для пошуку ІПМ, яка загрожує онлайн-спільноті або пошуку ІПМ, які становлять загрозу для певної особи чи організації. Відповідно під час задання параметрів завдання ІПМ можна вказувати конкретну спільноту і не зазначати тематику, не вказувати спільноти, але окреслити можливу ціль ІПМ.

Розглянемо базові елементи користувацького інтерфейсу. На рис. 4.6 зображено вікно для задання параметрів виявлення ІПМ відповідно до підготовчого етапу алгоритму виявлення ІПМ (див. 3.1.1 «Підготовчий етап»), — це демографічні, локаційні та тематичні характеристики. Для здійснення пошуку комплекс використовує методи і засоби, описані у 3.2 «Методи пошуку релевантних дискусій».

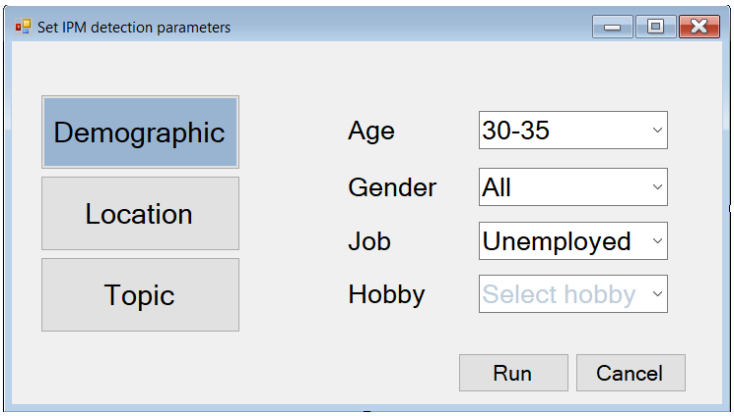


Рис. 4.6. Встановлення параметрів завдання виявлення ІПМ

Наступним кроком є встановлення порогових значень для фільтрів, які виявляють підозрілі фрагменти ІПМ. На рис. 4.7 наведено встановлення порогових значень для фільтрів, що засновані на критеріях рівня онлайн-спільноти (див. розд. 3.3 «Методи виявлення підозрілих фрагментів дискусії»).

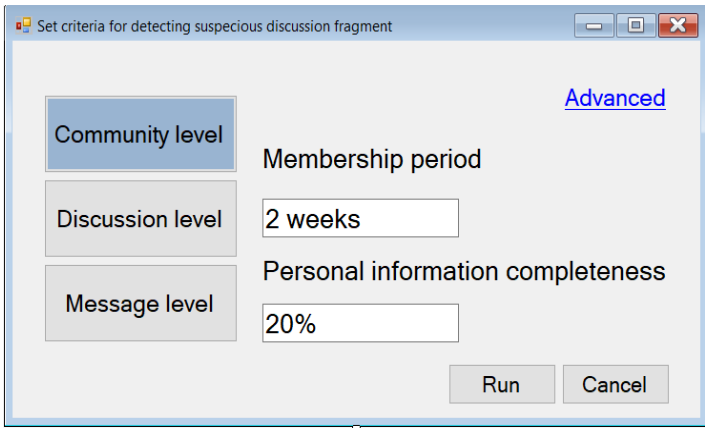


Рис. 4.7. Встановлення порогових значень критеріїв виявлення ІПМ

Задання характеристик стилю дискусії (рис. 4.8) є опціональним, його використовують модератори для моніторингу конкретних онлайн-спільнот.

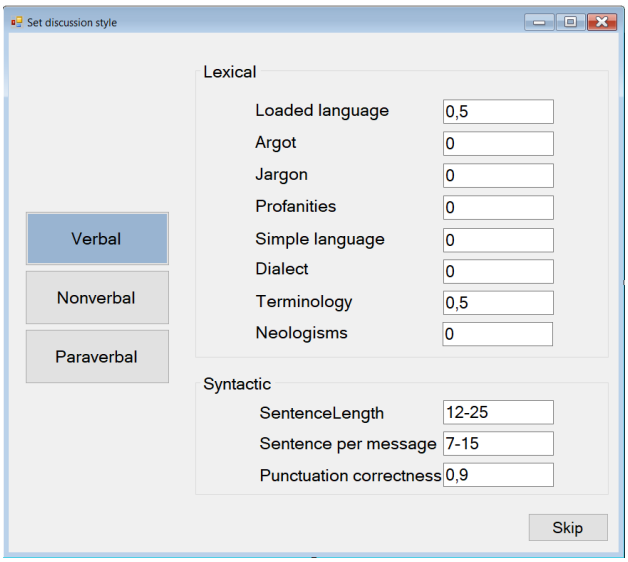


Рис. 4.8. Задання характеристик стилю, властивих для онлайн-спільноти

Згідно з користувацьким інтерфейсом результат виявлення ІПМ можна переглянути в п’яти виглядах (рис. 4.9), а саме: базовий вид, тактика подана в формі кусково-лінійного агрегату та зображена у вигляді діаграми, переглянути учасників дискусії та їхні повідомлення відповідно до їхніх ролей в прецеденті ІПМ, переглянути виявлені підозрілі фрагменти дискусії, переглянути виявлені маркери ІПМ в контексті.

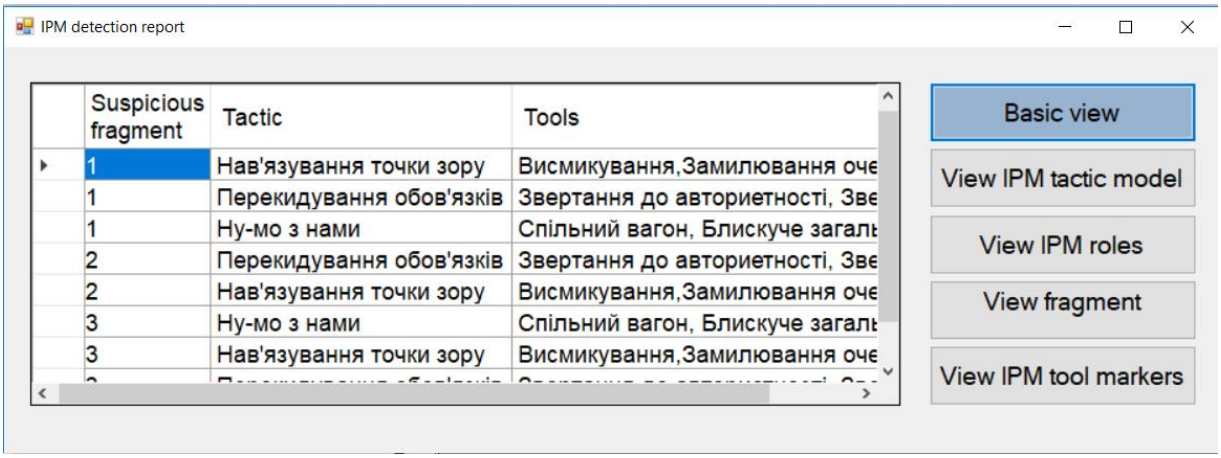


Рис. 4.9. Результат виявлення прецедентів ІПМ

Розглянемо інтерфейс, розроблений для наповнення БД інформацією, необхідною для виявлення ІПМ. Для прикладу наведено інтерфейси для вводу шаблонів тактики ІПМ, а саме послідовностей зміни станів та послідовностей

Наповнення БД шаблонами тактик ІПМ передбачає введення послідовності станів, присутніх в тактиці, та задання їхніх параметрів (рис. 4.10).

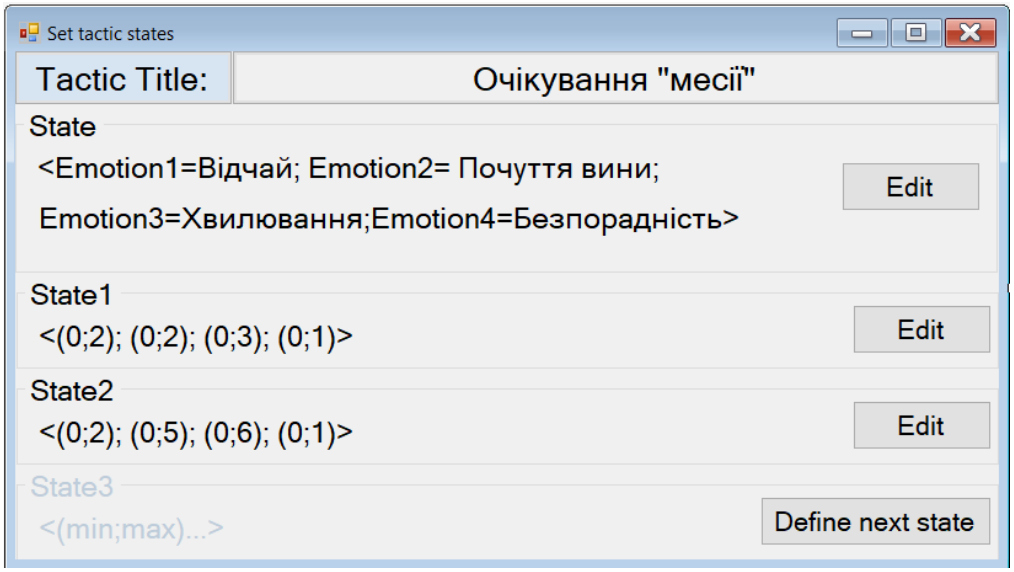


Рис. 4.10. Введення послідовності станів, присутніх в тактиці

Наповнення БД шаблонами тактик ІПМ передбачає задання параметрів станів, присутніх в тактиці (рис. 4.11).

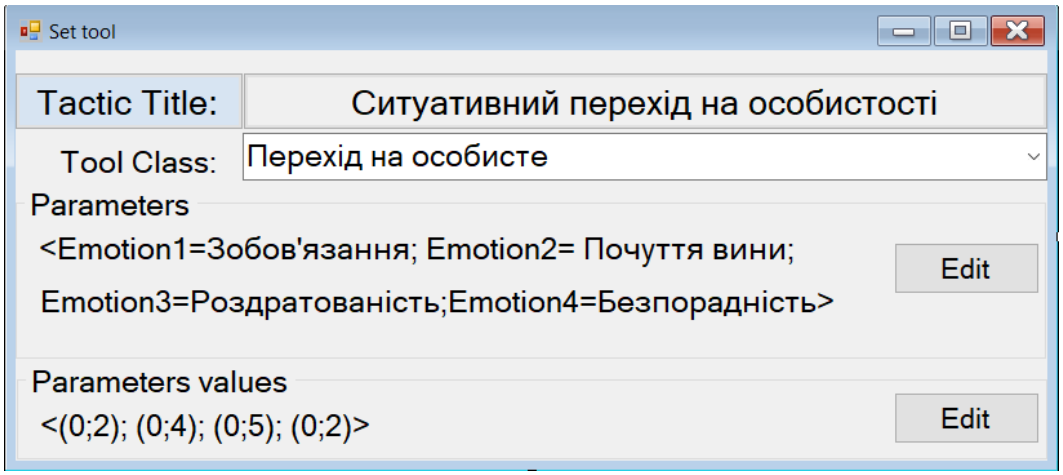


Рис. 4.11. Введення параметрів прийому тактики ІПМ

4.2. Приклади подання тактик ІПМ відповідно до формальної моделі ІПМ

У цьому підрозділі наведено популярні тактики ІПМ, які часто трапляються в дискусіях різного тематичного спрямування. Тактики подано відповідно до формальної моделі ІПМ, описаної у підрозділі 2.2 «Формальна модель інформаційно-психологічної маніпуляції». У такому форматі програмно-алгоритмічний комплекс приймає довідкову інформацію про тактики ІПМ необхідну для виявлення прецедентів.

Тактику «Нав'язування точки зору» подаємо так:

$$ImposingAttitude = \langle TacticState_{IA}, TacticStep_{IA}, ChangeStateFunction_{IA} \rangle. \quad (4.1)$$

Тобто, якщо психічний стан реципієнтів послідовно змінюється відповідно до вказаних нижче станів (4.2), то дискусію необхідно перевірити на наявність прийомів ІПМ:

$$TacticState_{IA} = \{InitialState, Attention, DispositionToSender, Questioning, OpinionAdopted\}. \quad (4.2)$$

Згідно з формальною моделлю тактики ІПМ, кожен стан задано наступним чином:

$$TacticState_0 = \langle InitialState, (0,0,0,0) \rangle, \quad (4.3)$$

$$TacticState_1 = \langle AttentiveRecipient, ([40;50],0,0,0), TacticStep_1 \rangle, \quad (4.4)$$

$$TacticState_2 = \langle DispositionToSender, ([50;60],[40;60],0,0), TacticStep_2 \rangle, \quad (4.5)$$

$$TacticState_3 = \langle Questioning, ([50;60],[40;60],[30;60],0), TacticStep_3 \rangle, \quad (4.6)$$

$$TacticState_4 = \langle OpinionAdopted, ([50;60],[40;60],[30;60],[20;40]), TacticStep_4 \rangle. \quad (4.7)$$

Параметри стану — це встановлені експертами одиниці виміру, які дають змогу оцінити глибину характерних для кожного стану ключових емоцій.

У випадку тактики «Нав'язування точки зору» використано наступні параметри:

$$StateParameters_{IA} = (Interest, Friendliness, Confidence, Trust). \quad (4.8)$$

При чому, вид параметрів однаковий в межах одноієї тактики ІПМ.

Параметри кроку — це встановлені експертами одиниці виміру, які дають змогу оцінити інтенсивність і кількість прийомів необхідну для переведення реципієнта в необхідний стан.

Тактика «Нав'язування точки зору» містить такі кроки:

$$TacticStep_1 = \langle Attract, ([40;50],0,0,0), Tool_{Attract} \rangle, \quad (4.9)$$

$$TacticStep_2 = \langle Befriend, ([10;20],[40;60],0,0), Tool_{Befriend} \rangle, \quad (4.10)$$

$$TacticStep_3 = \langle Assure, ([0;10],[0;20],[30;60],0), Tool_{Assure} \rangle, \quad (4.11)$$

$$TacticStep_4 = \langle IncreaseBelief, ([0;10],[0;20],[0;30],[30;60]), Tool_{IncreaseBelief} \rangle. \quad (4.12)$$

Для виконання кожного з кроків маніпулятор застосовує прийоми з відповідних множин.

$$Tool_{Attract} = \{Anchoring, BreakingNews\} \quad (4.13)$$

$$Tool_{Befriend} = \{AlotInCommon, SandboxFriend, NextOfKin\} \quad (4.14)$$

$$Tool_{Assure} = \begin{Bmatrix} AnnecdotalEvidence, MiddleGround, Bandwagon, CardStacking, KettleLogic, \\ BlackAndWhite, Reification, UnstatedAssumption \end{Bmatrix} \quad (4.15)$$

$$Tool_{IncreaseBelief} = \{BrokenWindowFallacy, ConfirmationBias, RedHerring, ConfirmationBias\} \quad (4.16)$$

Кожен з прийомів подано відповідно до формальної моделі прийому.

Популярну маніпулятивну тактику вирівнювання рейтингу за рахунок антиреклами опонента змодельована за допомогою кусково-лінійного

агрегату. Остання часто використовується для рятування рейтингу скомпрометованої особи за допомогою пошкодження іміджу опонента.

Тактику «Реабілітація іміджу знищуючи опонента» подаємо так:

$$RebuildReputation = \langle TacticState_{RR}, TacticStep_{RR}, ChangeStateFunction_{RR} \rangle. \quad (4.17)$$

Під час маніпулятивної тактики відбілювання замовника за рахунок знищення позитивного іміджу опонента реципієнт перебуває в одному з станів множини (4.18). Кожен стан конкретизований за допомогою такого кортежу параметрів (4.24). Для кожного стану параметри можуть набувати значень із вказаного діапазону. У випадку виходу одного з параметрів за межі встановленого діапазону, система переходить в інший стан.

$$TacticState_{IA} = \{InitialState, Doubt, BlameOpponent, SupportCustomer, ConfidentFollower\}. \quad (4.18)$$

Згідно з формальною моделлю тактики ІПМ, кожен стан задано наступним чином:

$$TacticState_0 = \langle InitialState, (0,0,0) \rangle, \quad (4.19)$$

$$TacticState_1 = \langle Doubt, ([40;50],[0;10],0), TacticStep_1 \rangle, \quad (4.20)$$

$$TacticState_2 = \langle BlameOpponent, ([40;50],[40;60],0), TacticStep_2 \rangle, \quad (4.21)$$

$$TacticState_3 = \langle SupportCustomer, ([10;20],[40;60],[30;60]), TacticStep_3 \rangle, \quad (4.22)$$

$$TacticState_4 = \langle ConfidentFollower, ([0;10],[40;60],[50;60]), TacticStep_4 \rangle. \quad (4.23)$$

Для опису тактики «Реабілітація іміджу знищуючи опонента» використано такі параметри:

$$StateParameters_{RR} = (Conviction, Blame, Support). \quad (4.24)$$

Тактика «Реабілітація іміджу знищуючи опонента» містить такі кроки:

$$TacticStep_1 = \langle CreateSuspicion, ([40;50],[0;10],0), Tool_{CreateSuspicion} \rangle, \quad (4.25)$$

$$TacticStep_2 = \langle Repel, ([0;10],[30;60],0), Tool_{Repel} \rangle, \quad (4.26)$$

$$TacticStep_3 = \langle Appeal, ([-40;-20],[0;20],[30;60]), Tool_{Appeal} \rangle, \quad (4.27)$$

$$TacticStep_4 = \langle IncreaseBelief, ([-10;0],[0;20],[20;30]), Tool_{IncreaseBelief} \rangle. \quad (4.28)$$

Для виконання кожного з кроків маніпулятор застосовує прийоми з відповідних множин.

$$Tool_{CreateSuspicion} = \{CherryPicking, NameCalling, Transfer, GamblersFallacy\}, \quad (4.29)$$

$$Tool_{Repel} = \{PoisoningTheWell, GlitteringGeneralization, RedHerring\}, \quad (4.30)$$

$$Tool_{Appeal} = \{, MisleadingVividness, KettleLogic, BlackAndWhite, Circular\ Reasoning\}, \quad (4.31)$$

$$Tool_{IncreaseBelief} = \{BrokenWindowFallacy, ConfirmationBias, RedHerring, ConfirmationBias\}, \quad (4.32)$$

Кожен з прийомів подано відповідно до формальної моделі прийому.

4.3. Приклади подання синтаксичної структури виявлених форм реалізації прийомів

Синтаксична структура є одним із маркерів для виявлення прийомів ІПМ в онлайн-спільнотах (див. розд. 3.5 «Методи виявлення прийомів ІПМ у дискусіях онлайн-спільнот»). Оскільки прийому ІПМ характерні певні форми реалізації. Подавши форми реалізації за допомогою синтаксичних структур, ми отримали маркери прийомів ІПМ на основі синтаксичної структури. Якщо схожість синтаксичної структури речення і маркеру синтаксичної структури форми реалізації прийому ІПМ відповідає встановленому експертами значенню, то виявлено прийом ІПМ.

Нижче наведено синтаксичні структури характерні для прийомів ІПМ з класу «Перехід на особисте» (див. розд. 1.2.1 «Проекція тактик маніпуляції на комунікацію в онлайн-спільнотах»). Як показано на рис. 4.12 і рис. 4.13

синтаксична структура в межах одного класу прийомів може значно відрізнятись.

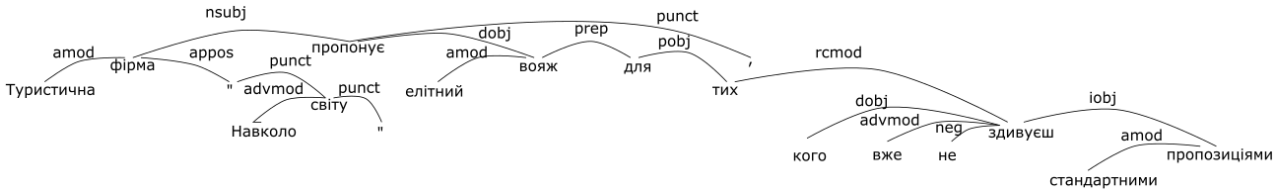


Рис. 4.12. Синтаксична структура прийому з класу «Перехід на особисте» (1)

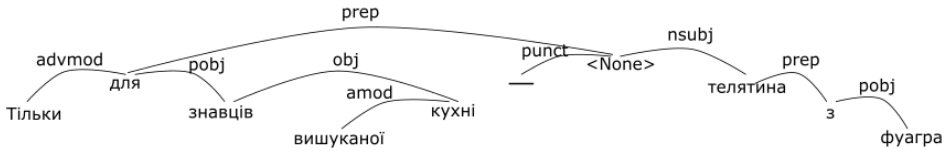


Рис. 4.13. Синтаксична структура прийому з класу «Перехід на особисте» (2)

На рис. 4.14, рис. 4.15 та рис. 4.16 наведено синтаксичні структури виявлених у онлайн-спільноті прийомів класу «Фальшива дилема». Прийоми рис. 4.15 і рис. 4.16 на мають подібну синтаксичну структуру, тому можуть бути виявлені на допомогою однакового маркера синтаксичних структур. Відмінність цих двох синтаксичних структур полягає у наявності багатьох вставних елементів, якщо ж не брати до уваги елементи, які зв’язані з батьківським вузлом зв’яком parataxis, то синтаксична структура цих прийомів буде схожою (рис. 4.17).

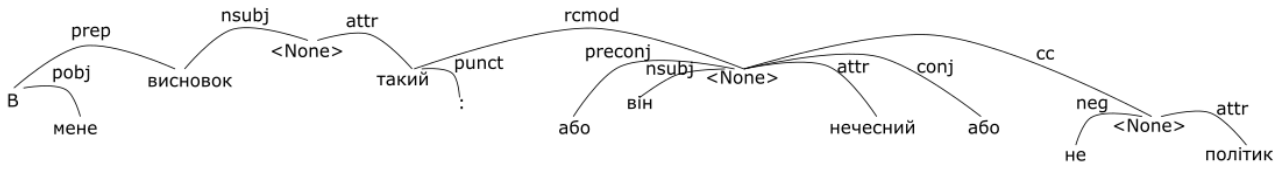


Рис. 4.14. Синтаксична структура прийому з класу «Фальшива дилема» (1)

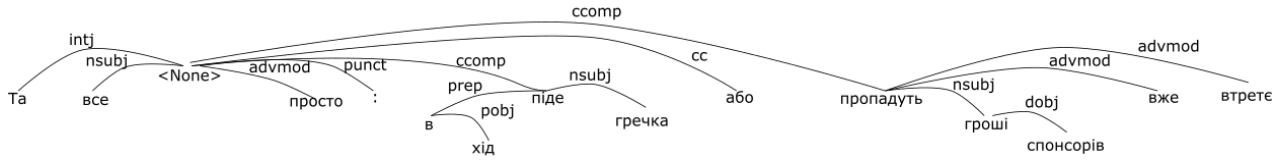


Рис. 4.15. Синтаксична структура прийому з класу «Фальшива дилема» (2)

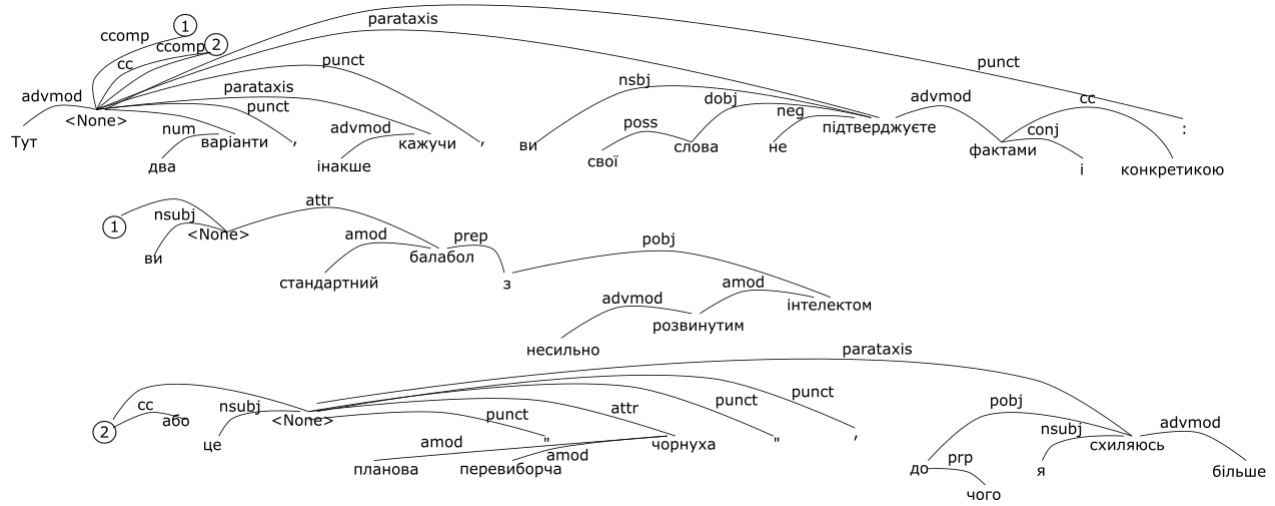


Рис. 4.16. Синтаксична структура прийому з класу «Фальшива дилема» (3)

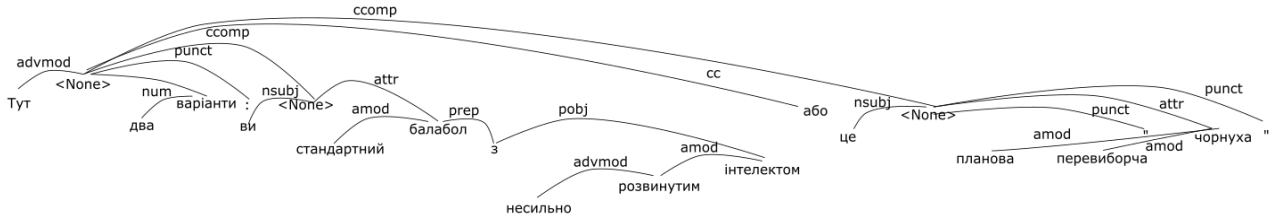


Рис. 4.17. Синтаксична структура прийому з класу «Фальшива дилема» (3) без вставних структур

Синтаксична структура речення подана за допомогою дерев залежності є зручною для виявлення ІПМ, оскільки за зв’язками між елементами можна встановити наявність другорядних членів речення, кількох основ, простої чи складної структури речення (наприклад, поширене, односкладне, повне розповідне, складне з підрядною-сурядністю способу дії), що є важливими ознаками з точки зору виявлення ІПМ.

4.4. Перевірка результатів

Перевірку результатів програмно-алгоритмічного комплексу, а саме ефективності виявлення ІПМ у певній онлайн-спільноті, ІПМ щодо певної організації у онлайн-спільнотах, а також підвищення рівня захисту суб'єктів та об'єктів комунікації в онлайн-спільнотах від ІПМ, ми проводили на основі зазначених нижче показників.

Точність виявлення ІПМ обчислюємо за формулою:

$$Precision = \frac{IPMFound - FalsePositives}{IPMTotal}, \tag{4.33}$$

де $IPMFound$ — кількість прецедентів ІПМ виявленна програмно-алгоритмічним комплексом «IPM Detector»; $FalsePositives$ — кількість фрагментів дискусії, які комплекс помилково виявив як прецеденти ІПМ; $IPMTotal$ — загальна кількість прецедентів ІПМ, яку виявили експерти у дискусії.

Швидкість виявлення прецедентів ІПМ обчислюємо за формулою:

$$Velocity = \frac{1}{n}\sum_{i=1}^{n}\frac{MessagesInPrecedent_i}{Date_i^{Detection} - Date_i^{LastMessageOf\,Precedent}}, \tag{4.34}$$

де $MessagesInPrecedent$ — середня кількість повідомлень у прецеденті; $Date_i^{Detection}$ — дата виявлення прецеденту ІПМ, $Date_i^{LastMessageOf\,Precedent}$ — дата останнього повідомлення, яке належить до прецеденту.

Точність виявлення ІПМ за допомогою програмно-алгоритміного комплексу становить близько 40%, а швидкість виявлення ІПМ є у 4-и рази більшою ніж швидкість виявлення ІПМ вручну.

Крім того, ефективність виявлення ІПМ відслідковуємо за опосердкованими показниками. Оскільки ІПМ має негативний вплив на суб'єкти комунікації та на онлайн-спільноту загалом, то вважаємо, що ознаки позитивної динаміки розвитку спільноти, за умови відсутності інших заходів для підвищення ефективності, свідчать про зменшення кількості ІПМ в

онлайн-спільноті. Крім того, ІПМ має негативний вплив на інформаційний образ об’єктів комунікації, тобто на обговорюванні бренди, організації, особистості і т.д.. Відсутність необґрунтованих діяльністю об’єкту деструктивних змін інформаційного образу є ознакою захисту від ІПМ.

Ми виділили наступні показники позитивної динаміки розвитку спільноти та поділили їх на кількісні та якісні.

Кількісні показники позитивної динаміки розвитку спільноти:

- збільшення появи нових учасників;
- збільшення відкриття нових дискусій;
- збільшення публікування повідомлень;
- зменшення кількості учасників, які покидають спільноту.

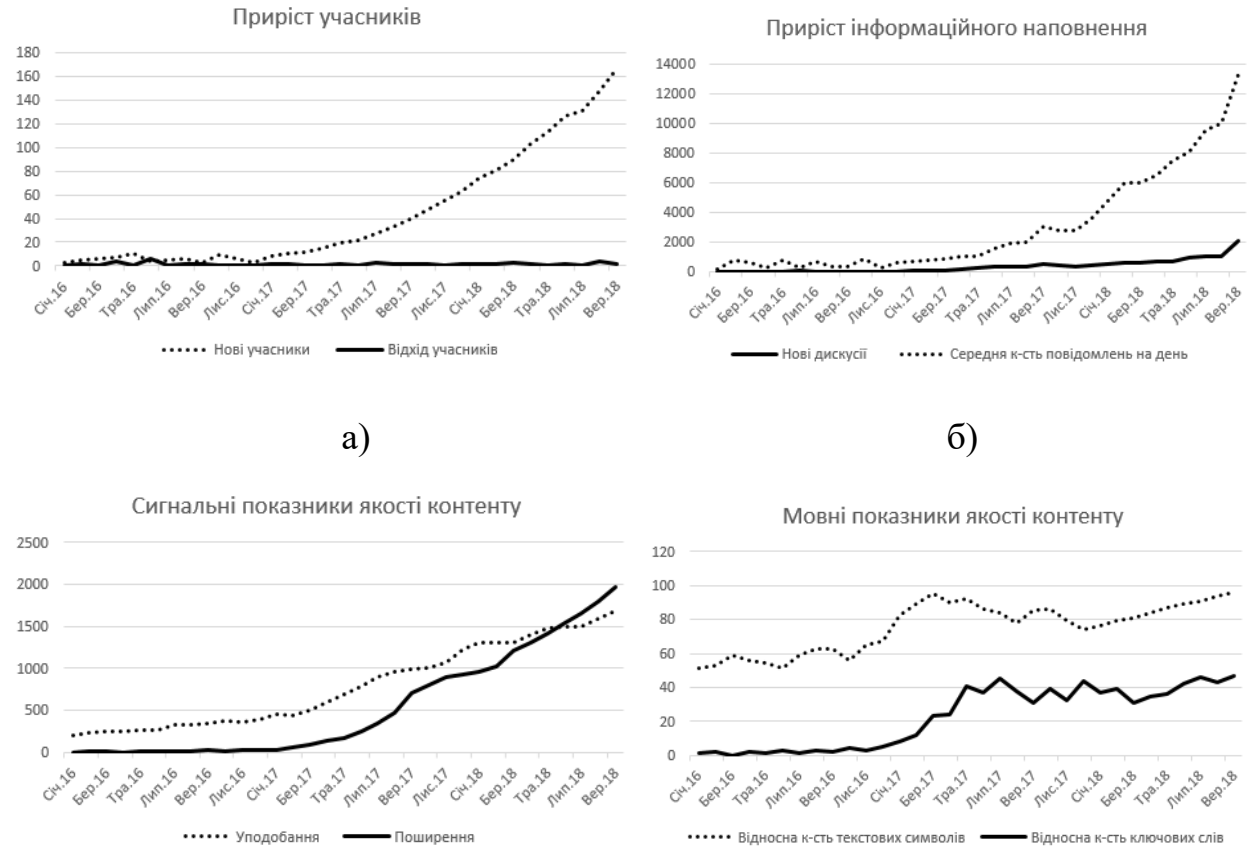
Якісні показники, або показники якості, генерованої користувачами інформації поділяємо на:

1. сигнальні:
    a. підвищення користувацької оцінки повідомлень (рейтинг, уподобання і т.д.);
    b. збільшення розповсюдження інформації користувачами (поширення, зовнішні посилання);
2. мовні:
    a. збільшення кількості текстових символів відносно усіх символів повідомленння;
    b. збільшення кількості ключових слів (за статистичним показником TF-IDF)
3. поведінкові:
    a. зменшення кількості видалених повідомлень;
    b. зменшення кількості повідомлень, які містять агресію (флейм, хейт, холівар)

Рівень захисту від ІПМ ми оцінювали на основі відсутність деструктивних змін інформаційного образу об’єкту, які не є наслідком його діяльності. Для цього ми використовували такі опосерковані показник:

- кількість поширених повідомлень з ІПМ;
- кількість реакцій на повідомлення, які містять ІПМ;
- кількість згадок про об’єкт у запланованій тональності;
- кількість згадок про об’єкт в непередбачуваних відповідальними за інформаційний образ контекстах.

Ефективність методів і засобів виявлення ІПМ в онлайн-спільнотах перевірена для Facebook та веб-реалізованих спільнот. Нижче наведені значення показників ефективності для Facebook спільноти. Методи і засоби виявлення ІПМ були застосовані в період від січня 2017. На графіках зображено показники зібрані впродовж періоду з січня 2016 до вересня 2018 року (рис. 4.18).



а) б)

в) г)

ґ)

Рис. 4.18. Показники позитивної динаміки розвитку спільноти

Показники рівня захисту інформаційного образу організації від ІПМ впродовж періоду з січня 2016 до вересня 2018 року наведені на рис. 4.19. Методи і засоби виявлення ІПМ були застосовані в період від січня 2017.

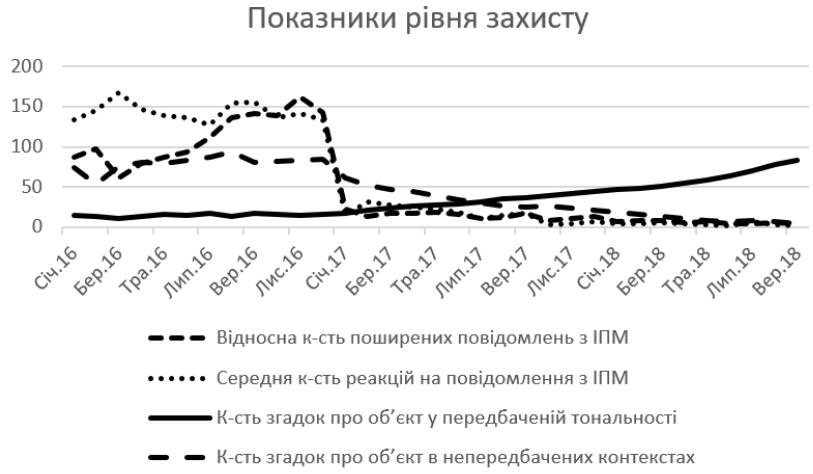


Рис. 4.19. Показники рівня захисту інформаційного образу організації

Зміни значення показників позитивної динаміки розвитку онлайн-спільноти та показникві захищеності інформаційного образу об’єкта внаслідок застосування програмно-алгоритмічного комплексу виявлення ІПМ зображено на рис. 4.20.

а) б)

Рис. 4.20. Зміни значень показників внаслідок застосування програмно-алгоритмічного комплексу

Ефективність розроблених методів та засобів для виявлення ІПМ підтверджена результатами їх застосування в онлайн-спільноті «Молодіжного Націоналістичного Конгресу (МНК) — Львів» (з метою виявлення ІПМ у спільноті) та в онлайн-спільноті Львівської Молодіжної Крайової Скаутської Організації «Білі Горвати» (з метою виявлення ІПМ у онлайн-спільнотах щодо їхньої діяльності).

Висновки до розділу

Розроблений програмно-алгоритмічний комплекс виявлення ІПМ дозволяє виявляти ІПМ у певній онлайн-спільноті або щодо певної організації у онлайн-спільнотах. Він є необхідним для адміністративної ланки онлайн-спільнот та відповідальних за інформаційну діяльність організацій та установ, оскільки дає змогу зменшити вибірку онлайн-спільнот, які необхідно проаналізувати на наявність прецедентів ІПМ вручну.

Програмно-алгоритмічний комплекс виявлення ІПМ побудований на основі методів і засобів, описаних у даній дисертаційній роботі. Комплекс складається з підсистеми пошуку тематичнорелевантних дискусій, підсистеми пошуку релевантих дискусій в Facebook, підсистеми сортування дискусій, підсистеми виявлення прецедентів ІПМ, підсистеми нейтралізації. Підситеми

містять внутрішні БД, крім того система має загальну БД, якою користуються всі підсистеми безпосередньо.

Програмно-алгоритмічний комплекс можна використовувати з метою виявлення ІПМ та з метою наповнення БД інформацією необхідною для виявлення ІПМ.

Висновки

У дисертаційній роботі розв’язано важливе науково-прикладне завдання розроблення науково обґрунтованих комп'ютерно-лінгвістичних методів і засобів виявлення інформаційно-психологічної маніпуляції в онлайн-спільнотах шляхом аналізу текстового інформаційного наповнення онлайн-спільнот.

Основні наукові та практичні результати роботи полягають у такому:

1. Проаналізовано характерні особливості та відмінності онлайн-спільнот від традиційного офлайн-середовища спілкування та від інших інтернет-засобів комунікації та проведено огляд існуючих засобів, необхідних для опрацювання інформаційного наповнення онлайн-спільноти;
2. Побудовано формальну модель тактики ІПМ на основі кусково-лінійних агрегатів, що дало змогу виявляти прецеденти ІПМ в онлайн-спільнотах на основі зміни станів учасників;
3. Розроблено систему лінгвістичних маркерів прийомів інформаційно-психологічної маніпуляції на основі діалогічних актів, семантичних змінних та синтаксичних структур, що дало змогу виявляти прийоми ІПМ в онлайн-спільноті без експертного аналізу;
4. На основі запропонованих у роботі критеріїв, розроблено систему фільтрів, де застосування наступних фільтрів залежить від результатів попередніх, що дає змогу збільшити ефективність та швидкість виявлення підозрілого фрагмента дискусії;
5. Розроблено алгоритм виявлення прецедентів ІПМ, який заснований на формальній моделі тактики ІПМ, заданій за допомогою кусково-лінійного агрегату та передбачає верифікацію виявленої тактики, на основі лінгвістичних маркерів прийому ІПМ, що дало змогу збільшити точність результатів виконання алгоритму;
6. Знайшли подальший розвиток характеристики мовного стилю, поведінки і профілю учасників онлайн-спільноти та подання онлайн-спільноти у вигляді

соціального графа, що дало змогу виявляти профілі-спільники та визначати можливі шляхи розповсюдження ІПМ;

7. Розроблено та впроваджено програмно-алгоритмічний комплекс виявлення ІПМ в онлайн-спільнотах, що дало змогу підвищити рівень захисту.

Достовірність отриманих результатів підтверджено теоретичним обґрунтуванням та практичним застосуванням для виявлення ІПМ у онлайн-спільноті «Молодіжного Націоналістичного Конгресу (МНК) – Львів» та для виявлення ІПМ щодо діяльності Львівської молодіжної крайової скаутської організації «Білі Горвати».

Література

1. Голуб З. Д., Пелещишин А. М. Види спілкування в онлайн-спільнотах та їхні характеристики. *Інформація, комунікація, суспільство 2018* : матеріали 7-ої міжнар. наук. конф. Чинадієво, 2018. С. 45–46.
2. Голуб З. Д. Огляд класифікацій онлайн спільнот. *Інформаційна діяльність, документознавство, бібліотекознавство: історія, сучасність, перспективи»* : матеріали ІІІ Всеукр. наук.-практ. конф. Київ, 2017. С. 17–19.
3. Huminskyi R.V., Peleshchyshyn A.M., Holub (Lazurak) Z.D. Suggestions for informational influence on a virtual community // International Journal of Computer Science and Business Informatics. 2015. Vol. 15, No. 1. P.47-65. Available at: http://ijcsbi.org/index.php/ijcsbi/article/view/512/147
4. Rheingold G. The virtual community. URL: http://www.rheingold.com/vc/boo k/intro.html (Last accessed 15.01.2017).
5. Wei Z. Research on the english online learning community based on SNS. International Conference on Education, Management and Computing Technology (ICEMCT 2015) (June 13-14, 2015, Tianjin, China). Tianjin, 2015. P. 1841–1844.
6. Пелещишин А. М., Кравець Р. Б., Сєров Ю. О. Аналіз існуючих типів віртуальних спільнот у мережі Інтернет та побудова моделі віртуальної спільноти на основі Веб-форуму. *Інформаційні системи та мережі: Вісник Національного університету «Львівська політехніка».* 2011. № 699. С. 212–221.
7. Anwar T., Abulaish T. Modeling a Web Forum Ecosystem into an Enriched Social Graph. URL: http://www.abulaish.com/uploads/MSM12E.pdf (Last accessed: 13.10.2018).
8. Kim A. J. Community building on the Web: secret strategies for successful online community. Boston, 2006. 380 p.

9. Kaplan A. M., Haenlein M. Social media: back to the roots and back to the future. *Journal of Systems and Information Technology*. Vol. 14. P. 101–104.
10. Baxter H. An introduction to online communities. URL: http://www.providersedge.com/docs/km_articles/an_introduction_to_online_communities.pdf (Last accessed: 17.10.2018).
11. Fu T., Abbasi A., Chen H. A hybrid approach to web forum Interactional Coherence analysis URL: https://pdfs.semanticscholar.org/0b4b/7f7da50e34d4935a0a3d660560f69049f165.pdf (Last accessed: 11.04.2018).
12. Jauman A. Online community management: grow and develop an active audience on social media / National Institute for Social Media, Saint Mary's University of Minnesota, The Loft Literary Center, 2017. 110 p.
13. Nash M. C. Cohesion and reference in English chatroom discourse. Proceedings of the 38th Annual Hawaii International Conference on System Sciences. (6 Jan 2005, Washington). Washington, 2005. P. 108–122.
14. Василенко В. С., Матов О. Я. Теорія інформації та кодування. Київ, 2014. 439 с.
15. Birch A. 30 covert emotional manipulation tactics: how manipulators take control in personal relationships. Createspace independent publishing platform, 2015. 66 p.
16. Mentory J. Covert emotional manipulation exposed!: The underhanded mind control tactics that all manipulators use to take control in personal relationships. CreateSpace Independent Publishing Platform, 2015. 232 p.
17. Почепцов Г. Сучасні інформаційні війни. Київ, 2015. 498 с.
18. Зеленін В. По той бік правди: НЛП як зброя інформаційно-пропагандистської війни. Київ, 2015. 384 с.
19. Chen C. Battling the internet water army: detection of hidden paid posters. 13 Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining(25–28 August, 2013, Niagara, Ontario,

Canada) New York, 2013. P. 116–120., URL: https://arxiv.org/pdf/1111.4297.pdf (Last accessed: 08.02.2018).

20. Дишлева С.М. Адвербіальна дистрибуція лексико-семантичних груп українських дієслів: автореф. дис. ...канд. філол. наук: 10.02.01. НПУ ім. М.П. Драгоманова. – К., 2008. – 18 с.
21. The impact of brands on consumer purchase intentions. *Asian Journal of Business Management.* 2012. Vol. 4(2). P. 105–110, URL: https://pdfs.semanticscholar.org/a1e3/6a36b80e7ef78e2318547784675b44b8656a.pdf (Last accessed 15.10.2017).
22. Kailani C, Ciobotar N. Experiential marketing: an efficient tool to leverage marketing communication impact on consumer behaviour. *International Conference on Marketing and Business Development Journal*. 2015. Vol I(1). URL: http://www.mbd.ase.ro/RePEc/aes/icmbdj/2015/ICMBDJ_V1_2015_123.pdf . (Last accessed: 02.10.2018).
23. Durlacher J. Emoticon : roman. Amsterdam, 2004. 436 p.
24. English oxford living dictionaries. URL: https://en.oxforddictionaries.com/definition/emoticon (Last accessed: 23.02.2018).
25. Шубина Н. Л. Невербальная семиотика печатного текста как область лингвистического знания.URL: https://lib.herzen.spb.ru/text/shubina_97_184_192.pdf (дата звернення: 25.10.2018).
26. Стернин И. А. Введение в речевое воздействие. Воронеж, 2001. 252 с.
27. Доценко Е. Л. Психология манипуляции: феномены, механизмы и защита. Москва, 2001. 344 с.
28. Berne E. Games people play. The psychology of human relationships. URL: http://rrt2.neostrada.pl/mioduszewska/course_2643_reading_3.pdf (Last accessed: 23.10.2018).
29. Шейнов В. П. Скрытое управление человеком. Минск, 2002. 848 с.
30. Propaganda Analysis: A Bulletin to Help the Intelligent Citizens Detect and Analyze Propaganda. Publications of the Institute for Propaganda Analysis.

Vol. 1. 1938. URL: https://archive.org/details/IPAVol1/page/n14 (Last accessed: 23.10.2017).

31. Bandler R., Grinder J. Frogs into Princes: Neuro Linguistic Programming. Colorado, 1979. 194 p.
32. Gorina E.V. Functions of paragraphemic tools on the internet. *Russian Linguistic Bulletin*. 2015. №4 (4), P. 11–13. URL: http://rulb.org/en/article/funkcii-paragrafemnyx-sredstv-v-internete/ (Last accessed: 11.07.2018).
33. Emojipedia. URL: https://emojipedia.org/ (Last accessed: 05.12.2018).
34. Emoji Charts. URL: https://unicode.org/emoji/charts/full-emoji-list.html_(Last accessed: 02.10.2018).
35. Хэмп Э. Словарь американской лингвистической терминологии / пер. с англ. и доп. В. В. Иванова : под ред. В. А. Звегинцева. Москва, 1964. 264 с.
36. Вашунина И. В. Взаимовлияние вербальных и невербальных (иконических) составляющих при восприятии креолизованного текста : автореф. дис. ... докт. филол. наук : 10.02.19. Москва, 2009. 20 с.
37. Шубина Н. Л. Пунктуация в коммуникативно-прагматическом аспекте и ее место в семиотической системе русского текста: дис. … д-ра філол. наук. Санкт-Петербург, 1999. 455 с.
38. Григорьева Т. М. Параграфемные явления в современном русском письме. *Язык, культура, коммуникация: аспекты взаимодействия* : науч.-метод. бюл. / под ред. И. В. Пекарской. Вып. 1. Абакан, 2003. С. 68–76.
39. Варій М. Й. Психологія : навч. пос. Київ, 2009. 288 с.
40. Щербатых Ю. В., Мосина А. Н. Дифференцировка психических состояний и других психологических феноменов. *Пихология психических состояний: теория и практика* : материалы I Всерос. науч.-практ. конф. Казань, 2008. Ч. II. С. 526–528.

41. Scherer K. R. Emotion as a multicomponent process: a model and some cross-cultural data. *Review of Personality and Social Psychologi.* 1984. № 5. P. 37–63.
42. Plutchik R. Emotions and life: perspectives from psychology, biology and evolution. Washington, 2002. 592 p.
43. Petta P., Pelachaud C., Cowie R. Emotion-oriented system: the humaine handbook. Berlin, 2011. 794 p.
44. Potts C. Sentiment symposium tutorial: language and cognition. URL: http://sentiment.christopherpotts.net/lingcog.html#tag:reactions (Last accessed: 24.11.2018).
45. New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems. Vol. 28 (2). 2013. P. 15–21. URL: https://sentic.net/new-avenues-in-opinion-mining-and-sentiment-analysis.pdf (Last accessed: 14.06.2018).
46. Combining social network analysis and sentiment analysis to explore the potential for online radicalisation / CLARITY: Centre for Sensor Web Technologies and 2School of Law and Government, Dublin City University, Glasnevin, Dublin, Ireland URL: https://core.ac.uk/download/pdf/147597985.pdf (Last accessed: 08.05.2018).
47. Jackson R. Writing the war on terrorism: language, politics and counter-terrorism. Manchester, 2012. 242 p.
48. A Multi-criteria Recommender System Exploiting Aspect-based Sentiment Analysis of Users' Reviews. URL: https://dl.acm.org/citation.cfm?id=3109905 (Last accessed: 18.10.2018).
49. Peslak A. Sentiment analysis and opinion mining: current state of the art and review of Google and Yahoo search engines' privacy policies. *Journal of Information Systems Applied Research.* 2017. Vol. 10(3). P. 38–47.
50. A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation. URL: https://liris.cnrs.fr/Documents/Liris-6508.pdf (Last accessed: 12.10.2018).

51. Vilares D., Alonso M. A., Gómez-Rodriguez C. Supervised sentiment analysis in multilingual environments. *Information processing and management: an international journal archive.* 2017. Vol. 53(3). P. 595–607. URL: http://www.grupolys.org/biblioteca/VilAloGom2017a.pdf (Last accessed: 15.10.2018).
52. Chen Q., Sokolova M. Word2Vec and Doc2Vec in unsupervised sentiment analysis of clinical discharge summaries. URL: https://arxiv.org/abs/1805.00352 (Last accessed: 02.07.2018).
53. Musto C, Semeraro G, Polignano M. A comparison of lexicon-based approaches for sentiment analysis of microblog posts. URL: http://ceur-ws.org/Vol-1314/paper-06.pdf (Last accessed: 20.10.2018).
54. Ranganath R., Jurafsky D., McFarland D. It's not you, it's me: detecting flirting and its misperception in speed-dates. URL: http://www.aclweb.org/anthology/D09-1035 (Last accessed: 07.05.2018).
55. Nguyen T. T., Chang K., Hui S. C. Supervised term weighting for sentiment analysis. *Proceedings of 2011 IEEE International Conference on Intelligence and Security Informatics* (*10-12 July, 2011. Beijing, China)*. Beijing, 2011. URL: https://ieeexplore.ieee.org/document/5984056 (Last accessed: 28.02.2018).
56. Baccianella S., Esuli A., F. Sebastiani F. SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. URL: http://nmis.isti.cnr.it/sebastiani/Publications/LREC10.pdf (Last access:17.09.2018).
57. EffectCheck Because your words matter. URL. https://effectcheck.com/about/ (Last accessed: 15.10.2018).
58. Canvs. URL: https://www.canvs.ai/ (Last accessed: 08.12.2018).
59. The statistical difference between 1-star and 5-star reviews on Yelp. URL: https://minimaxir.com/2014/09/one-star-five-stars/ (Last accessed 27.03.2017).

60. Thomsen J. F. How sentiment analysis can add value to numeric ratings in review systems. URL: https://deemly.co/blog/sentiment-analysis-value-numeric-ratings-review-systems/ (Last accessed: 15.01.2018).
61. Zarisheva E., Scheffler T. Dialog act annotation for Twitter conversations. *Proceedings of the SIGDIAL Conference (Prague, Czech Republic, 2-4 September 2015).* Prague, *2015*, P. 114–123.
62. Серль Дж. Классификация иллокутивных актов. *Новое в зарубежной лингвистике: вып 17.* Москва, 1986.
63. Austin J. L. How to do things with words. Oxford university press, 1975. 168 p.
64. DIT++ Taxonomy of dialogue acts. URL. https://dit.uvt.nl/ (Last accessed: 28.08.2018).
65. Zhang R., Gao D., Li W. What are tweeters doing: recognizing speech acts in Twitter. *Analyzing microtext: papers from the 2011 AAAI Workshop (WS-11-05).* P. 86–91.
66. Mayfield E., Adamson D., Rosé C. P. Hierarchical conversation structure prediction in multi-party chat. URL: http://www.aclweb.org/anthology/W12-1607 (Last accessed: 05.12.2018).
67. Cerisara C. Multi-task dialog act and sentiment recognition on Mastodon. URL: http://aclweb.org/anthology/C18-1063 (Last accessed: 08.04.2018).
68. Surendran D., Levow G. Dialog act tagging with support vector machines and hidden markov models. URL: https://faculty.washington.edu/levow/papers/IS06_da.pdf (Last accessed: 18.07.2018).
69. Wilks Y. Artificial companions as a new kind of interface to the future internet. *Oxford internet institute, research repor*t *(13 October, 2006).* 2006. 19 p.
70. Evaluation methods for topic models. URL: http://dirichlet.net/pdf/wallach09evaluation.pdf (Last accessed: 25.09.2018).
71. Forbes-Riley K., Litman D. J. Using bigrams to identify relationships between student certainness states and tutor responses in a spoken dialogue corpus.

URL: https://pdfs.semanticscholar.org/06b2/1fe4dc7e0cc297e73eef0c05231445f8c634.pdf (Last accessed: 21.04.2018).

72. Learning to tutor like a tutor: ranking questions in context. *Journal of graph theory archive.* New York. 1996. Vol. 21(3). P. 335–342. URL: https://dl.acm.org/citation.cfm?id=2345840.2345900 (Last accessed: 17.10.2018).
73. Ezen-Can A., Boyer K. E. Unsupervised classification of student dialogue acts with query-likelihood clustering. URL: https://pdfs.semanticscholar.org/c02a/ea22a2ddda77a07305413711a6cb97f3034b.pdf (Last accessed: 23.03.2018).
74. Detecting deception in synchronous computer-mediated communication using speech act profiling. *IEEE International Conference on Security Intelligence and Informatic May 19-20, Atlanta, Georgia)* Georgia, 2005. URL: https://link.springer.com/chapter/10.1007/11427995_45 (Last accessed: 08.05.2018).
75. Towards an ISO standard for dialogue act annotation. URL: http://ict.usc.edu/pubs/towards%20an%20iso%20standard.pdf (Last accessed: 25.12.2017).
76. Clark A., Popescu-Belis A. Multi-level dialogue act tags. URL. http://aclweb.org/anthology/W04-2328 (Last access: 15.08.2018).
77. Голуб З. Д. Формалізація прийомів інформаційно-психологічної маніпуляції. *Вісник Національного технічного університету «ХПІ». Серія: Нові рішення в сучасних технологіях*. Харків, 2017. № 32 (1254). С. 55–61.
78. Голуб З. Д. Система критеріїв для виявлення фрагментів онлайн-дискусій з підозрою на наявність інформаційно-психологічної маніпуляції. *Вісник Національного технічного університету «ХПІ». Серія: Нові рішення в сучасних технологіях.* Харків, 2018. № 9 (1285). С. 106–111.

79. Pełeszczyszyn A., Holub Z. Development of the system for detecting manipulation in online discussions. *Advances in Intelligent Systems and Computing (AISC)*. Lviv, 2017. Vol. 543. P. 111–117.
80. Голуб З. Структура словника маркерів лексичних змінних для виявлення інформаційно-психологічних маніпуляцій. *Вісник Хмельницького національного університету. Серія: Технічні науки,* Хмельницький, 2018. – № 2 (259). С. 264–268.
81. Голуб З. Розроблення формальних моделей для автоматизації виявлення інформаційно-психологічної маніпуляції. *Управління розвитком складних систем* : зб. наук. пр. Київ. нац. ун-ту буд-ва і архіт. Київ, 2018. №34. С. 85–91.
82. Holub Z. The algorithm for detecting online discussion fragments containing information and psychological manipulation. *Regional interuniversity compendium of scientific works «System technologies».* Dnipro, 2017. № 6 (113). P. 85–91.
83. Peleschyshyn A., Holub I., Holub Z. Methods of real-time detecting manipulation in online communities. *11th International Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT)*(September 2016, Lviv, Ukraine). Lviv, 2016. P. 15–17.
84. Peleschyshyn A., Holub Z., I. Holub I. Formal model and key features of an online community fundamental for detecting informational and psychological manipulation. *11th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)(September 2017, Lviv, Ukraine*), Lviv, 2017. P. 101–104.
85. Гумінський Р. Методи і засоби виявлення інформаційних загроз віртуальних спільнот в інтернет середовищі соціальних мереж. URL: http://er.nau.edu.ua:8080/handle/NAU/17400 (дата звернення: 20.10.2018).

86. Sidorov G., Gómez-Adorno H., Markov I., Pinto D., Loya N. Computing text similarity using Tree Edit Distance http://www.cic.ipn.mx/~sidorov/sngrams_ted_2015.pdf (Last accessed: 18.05.2018).
87. Хрящева, Н. Ю. Особенности психических состояний в условиях изоляции. Экспериментальная и прикладная психология. 1983. Вып. 10. С. 83–89.
88. Чалдини Р. Психология влияния. Убеждай, воздействуй, защищайся. Питер, 2016. 336 с.
89. Аврамчук Е. Ф., Вавилов А. А., Емельянов С. В. Технология системного моделирования. Москва, 1988. 520 с.
90. Dialogue act modeling for automatic tagging and recognition of conversational speech. URL: http://www.aclweb.org/anthology/J00-3003 (Last accessed: 18.07.2018)
91. Голуб З. Розроблення алгоритму виявлення шкідливих інформаційно-психологічних маніпуляцій в онлайн-спільнотах ВНЗ. *Вісник Національного університету «Львівська політехніка». Серія: Інформатизація вищого навчального закладу.* 2017. № 879. С. 33–41.
92. Peleschyshyn A. The preliminary stage of the algorithm for detecting information and psychological manipulation in online communities. *13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)(September 2018, Lviv, Ukraine*). Lviv, 2018. P. 30–33.
93. Голуб З. Д., Пелещишин А. М. Виявлення прийомів ІПМ реалізованих за допомогою посилань. *Інформація, комунікація, суспільство 2017* : матеріали 6-ої міжнар. наук. конф. Львів-Славське, 2017. С. 65–66.
94. Brill E., Moore R. C. An improved error model for noisy channel spelling correction. URL: http://www.aclweb.org/anthology/P00-1037 (Last accessed: 10.10.2018).

95. Authorship attribution of micro-messages. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (18-21 October, 2013, Seattle, Washington. Washington, 2013. P. 1880–1891. URL: http://mlai.cs.huji.ac.il/publications/Ari_Rappoport/files/Authorship%20Attribution%20of%20Micro-Messages.pdf (Last accessed: 24.10.2018).
96. Layton R., Watters P., Dazeley R. Authorship attribution for twitter in 140 characters or less. *Proceedings of the 2010 Second Cybercrime and Trustworthy Computing Workshop* (*July 19–20, 2010, Washington).* Washington, 2010. P. 1–8.
97. Jankowska M. Author style analysis in text documents based on character a word n-grams. Submitted in partial fulfillment of the requirements for the degree of doctor of philosophy at Dalhousie University, Halifax, 2017. 129 p. URL:https://dalspace.library.dal.ca/xmlui/bitstream/handle/10222/72872/Jankowska-Magdalena-PhD-CSCI-April-2017.pdf?sequence=4&isAllowed=y (Last accessed: 25.11.2018).
98. Cross-topic authorship attribution: will out-of-topic data help? Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers (August 23–29, 2014,Dublin, Ireland,). Dublin, 2014. P. 1228–1237. URL: http://www.aclweb.org/anthology/C14-1116 (Last accessed: 12.05.2018).
99. Stamatatos E. On the robustness of authorship attribution based on character N-gram features. URL: https://pdfs.semanticscholar.org/ae14/0da1a86c26533a765d10a87ec47b1f294b53.pdf (Last accessed: 25.10.2018).
100. Layton R., Watters P., Dazeley R. Authorship analysis of aliases: does topic influence accuracy? *Natural Language Engineering*. 2015. Vol. 21(4). P. 497–518.

101. Boutwell S. R. Authorship attribution of short messages using multimodal features. URL: https://apps.dtic.mil/dtic/tr/fulltext/u2/a543909.pdf (Last accessed: 14.08.2018).
102. Mosteller F., Wallace D. L. Inference in an authorship problem: a comparative study of discrimination methods applied to the authorship of the disputed federalist papers. *Journal of the American Statistical Association*. Chicago, 1963. Vol.58. № (302). P. 275–309.
103. Kestemont M. Function words in authorship attribution. from black magic to theory? *Proceedings of the 3rd workshop on computational linguistics for literature* (*April 27, 2014, Gothenburg, Sweden*). Gothenburg, 2014. P. 59–66. URL: https://pdfs.semanticscholar.org/9474/c53bba5db83555012b35b581b8ee62b74dbf.pdf (Last accessed: 15.08.2018).
104. Stamatatos E. A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology*. 2009. № 60(3). P. 538–556.
105. Lin J. Automatic author profiling of online chat logs. *Calhoun: The NPS institutional archive.* Monterey. 2007. №3. URL: https://core.ac.uk/download/pdf/36697283.pdf (Last accessed: 25.01.2018).
106. Федушко С. С., Білущак Г. І. Формування системи лінгво-комунікативних індикаторів соціально-демографічних характеристик web-учасників. *Управління розвитком складних систем.* Київ, 2014. № 18. С. 112–123.
107. WordNet. A lexical database of English. URL: https://wordnet.princeton.edu/ (Last accessed 30.09.2018).
108. Choi J. D., Palmer M. Guidelines for the clear style constituent to dependency conversion. URL: http://www.mathcs.emory.edu/~choi/doc/cu-2012-choi.pdf (Last access: 15.10.2018).
109. displaCy Dependency Visualizer. URL: https://explosion.ai/demos/displacy?text=real%20men%20use%20mac&model=en_core_web_sm&cpu=1&cph=0 (Last accessed: 15.10.2018).

110. Дарчук Н. П. Автоматичний синтаксичний аналіз текстів корпусу української мови. *Українське мовознавство*. Київ, 2013. № 43. С. 11–19.
111. Рисін А., Старко В., Чаплинський Д. Словник ВЕСУМ та інші пов'язані засоби NLP для української мови. URL: https://r2u.org.ua/articles/vesum (режим доступу: 16.10.2018).
112. Великий електронний словник української мови (ВЕСУМ). URL: https://r2u.org.ua/forum/viewtopic.php?f=4&t=7848 (дата звернення: 25.11.2018).
113. Пелещишин А., Голуб З. Виявлення маніпуляції щодо потенційних покупців в онлайн спільноті. *Інформаційне суспільство: тенденції регіонального розвитку (ISRDT-2016)* : матер. міжнар. наук.-практ. конф. Львів, 2016. С. 58–59.
114. Голуб З. Д. Пелещишин А. М. Моделювання маніпулятивної тактики за допомогою кусково-лінійного агрегата. *Інформація, комунікація, суспільство 2016* : матеріали 5-ої міжнар. наук.-практ. конф. Львів-Славське, 2016. С. 80–81.
115. Korzh R., Peleschyshyn A., Holub Z. Analysis of integrity and coverage completeness of the informational image of a higher education institution. *Modern problems of radio engineering, telecommunications and computer science (Feb. 23–26, 2016, Lviv-Slavske, Ukraine*) : in: proceedings of the XIIIth International Conference. TCSET`2016. Lviv-Slavske, 2016. P. 825–827.

Додаток А. Акти використання результатів дисертаційного дослідження

"ЗАТВЕРДЖУЮ"
Голова
Львівська Молодіжна Крайова
Скаутська Організація "Білі Горвати"

Сенькавський Л.О.
«13» 09 2018 р.

Акт
про використання результатів дисертаційних досліджень
Лазурак Зоряни Дмитрівни
«Методи і засоби виявлення інформаційно-психологічної маніпуляції в онлайн-спільнотах»

Цей акт складений про те, що результати дисертаційної роботи Лазурак З. Д. «Методи і засоби виявлення інформаційно-психологічної маніпуляції в онлайн-спільнотах», а саме:

- формальна модель тактики та прийому інформаційно-психологічної маніпуляції;
- система критеріїв для виявлення підозрілих ознак ІПМ відповідно до організаційно-структурних рівнів онлайн-спільноти;
- алгоритм пошуку релевантних дискусій щодо ключових проблем діяльності спільноти та метод сортування дискусій за сприятливістю умов до здійснення ІПМ;
- метод виявлення прецедентів застосування прийомів ІПМ на основі вербальних, невербальних та паравербальних ознак;
- метод ідентифікації профілів, які провадять спільну інформаційно-психологічну маніпуляцію;
- алгоритм ідентифікації можливих шляхів поширення інформаційно-психологічної маніпуляції

є використаними для ряду завдань підтримання позитивного інформаційного образу ЛМКСО «Білі Горвати» в онлайн-спільнотах та підвищення ефективності функціонування спільноти ЛМКСО «Білі Горвати» в Facebook, зокрема:

- підтримання позитивного іміджу організації за рахунок своєчасного виявлення та недопущення подальшого поширення інформаційно-психологічних маніпуляцій щодо ЛМКСО «Білі Горвати»;
- захисту учасників спільноти від інформаційно-психологічних маніпуляцій.

Використання результатів дисертаційної роботи Лазурак З.Д. дозволило:

- підвищити ефективності роботи модераторів онлайн-спільноти у напрямку захисту учасників спільноти від негативних інформаційних впливів;
- збільшення кількості учасників спільноти за рахунок зростання якості інформаційного-наповнення спільноти;
- підвищення рівня захисту інформаційного образу ЛМКСО «Білі Горвати» за рахунок оперативного виявлення інформаційно-психологічної маніпуляції у текстових онлайн-спільнотах.

Члени комісії

Голова ЛМКСО «Білі Горвати» — Сеньковський А. О.

Керівник напряму інформаційної підтримки ЛМКСО «Білі Горвати» — Моторний Т. В.

Модератор Facebook сторінки ЛМКСО «Білі Горвати» — Пузаєнко О.В.

"ЗАТВЕРДЖУЮ"
Голова
Львівського осередку Молодіжного
Націоналістичного Конгресу (МНК)
Дацків О.Р.
«21» 09 2018 р.

Акт
про використання результатів дисертаційних досліджень
Лазурак Зоряни Дмитрівни
«Методи і засоби виявлення інформаційно-психологічної маніпуляції в онлайн-спільнотах»

Цей акт складений про те, що результати дисертаційної роботи Лазурак З. Д. у напрямку розробки методів і засобі виявлення інформаційно-психологічної маніпуляції в онлайн-спільнотах, а саме:

- метод виявлення підозрілих уривків дискусій на основі системи фільтрів зі статичних та динамічних критеріїв;
- метод виявлення тактики ІПМ, формалізованих як кусково-лінійні агрегати;
- метод виявлення прийомів ІПМ на основі вказівників прийому ІПМ та системи вербальних, невербальних та паравербальних маркерів прийому ІПМ;
- метод пошуку релевантних дискусій та сортування їх за сприятливістю передмов до здійснення ІПМ.
- методи визначення можливих шляхів поширення ІПМ

використано для підвищення ефективності функціонування віртуальної спільноти «Молодіжний Націоналістичний Конгрес (МНК) – Львів» в соціальній мережі Facebook. Своєчасне виявлення ІПМ у дискусіях спільноти

«Молодіжний Націоналістичний Конгрес (МНК) – Львів» має вагоме значення для функціонування спільноти, як майданчика для обміну інформацією та реалізації об'єктивних обговорень.

Впровадження результатів дисертаційної роботи Лазурак З.Д. забезпечило:

- збільшення кількості своєчасно виявлених прецедентів ІПМ;
- попередження поширення ІПМ;
- підвищення якості створюваного учасниками інформаційного наповнення дискусії;
- збільшення числа дискусій;
- збільшення кількості учасників за рахунок зменшення відходу користувачів спільноти та появи нових учасників;

Члени комісії

Голова (адміністратор Facebook спільноти)		Дацків О. Р.
Відповідальний за зв'язки з громадськістю (модератор Facebook спільноти)		Паламар Х. З.

"ЗАТВЕРДЖУЮ"
Проректор з наукової роботи
Національного університету
"Львівська політехніка"
_____ проф. І.І. Чухрай
"___" _____ 2018 р.

А К Т

про використання результатів дисертаційної роботи "Методи і засоби виявлення інформаційно-психологічної маніпуляції в онлайн-спільнотах" аспіранта кафедри соціальних комунікацій та інформаційної діяльності Лазурак Зоряни Дмитрівни, представленої на здобуття наукового ступеня кандидата технічних наук, при виконанні науково-дослідних робіт Національного університету "Львівська політехніка"

Ми, що нижче підписались, начальник НДЧ, к.т.н., доц. Жук Л.В. та члени комісії: завідувач відділу науково-організаційного супроводу наукових досліджень, к.т.н. Лазько Г.В., завідувач планово-фінансового відділу Чулой Т.М. та завідувач кафедри соціальних комунікацій та інформаційної діяльності д.т.н., проф. Пелещишин А.М. цим актом підтверджуємо, що результати дисертаційної роботи аспіранта кафедри соціальних комунікацій та інформаційної діяльності Лазурак З.Д. використано під час виконання науково-дослідної роботи кафедри соціальних комунікацій та інформаційної діяльності Національного університету «Львівська політехніка»: "Лінгвістичне забезпечення консолідації відкритих інформаційних ресурсів" (№ державної реєстрації 0113U005274).

Лазурак З.Д. запропонувала та описала новий підхід до опрацювання та консолідації інформаційного наповнення онлайн-спільнот за допомогою лінгвістичних засобів. Дослідила специфіку анотування дискусій онлайн-спільнот на основі діалогічних актів, запропонувала набір синтаксичних структур, значущих з точки зору виявлення інформаційно-психологічних маніпуляцій (ІПМ), ввела поняття семантичної змінної як одного з вербальних маркерів прийомів ІПМ. Розроблені за допомогою лінгвістичних засобів маркери прийомів ІПМ дають змогу збільшити кількість виявлених прецедентів ІПМ.

Начальник НДЧ
к.т.н., доцент — Л.В. Жук

Члени комісії:
Зав. відділу НОСНД,
к.т.н. — Г.В. Лазько

Заст. Начальника ПФВ — Т.М. Чулой

Зав. кафедри СКІД
д.т.н., професор — А.М. Пелещишин

Додаток Б. Список досліджуваних онлайн-спільнот

	Назва онлайн-спільноти	**Адреса онлайн-спільноти**
1.	«Каретний двір»	https://www.facebook.com/lvivfilmcenter/?ref=br_rs
2.	"Українська правда"	https://www.facebook.com/ukrpravda/
3.	Видавництво Старого Лева	https://www.facebook.com/starlev/
4.	ТСН	https://www.facebook.com/tsn.ua/
5.	Будинок іграшок	https://www.facebook.com/Ihrashkyua/
6.	Українська молодь - Христові	https://www.facebook.com/umhlviv/
7.	Dushka	https://www.facebook.com/dushka.care/
8.	Vogue Ukraine	https://www.facebook.com/VogueUkraine/
9.	Кафедра СКІД	Кафедра СКІД
10.	НУ «Львівська політехніка»	https://www.facebook.com/LvivskaPolitehnika/
11.	Видавництво «Свічадо»	Свічадо видавництво
12.	Людоньки, порадьте!	https://www.facebook.com/groups/poradte/
13.	Аудиторія	https://www.facebook.com/audytoriya/
14.	Varta1 - Варта1 ГО"Варта1"	https://www.facebook.com/groups/govarta1/
15.	ASUS	https://www.facebook.com/ASUS.Ukraine/
16.	Телеканал ZIK	https://www.facebook.com/telekanalZIK/
17.	ГО «Білі Горвати»	https://www.facebook.com/bilihorvaty/
18.	Радіо Вголос	https://www.facebook.com/radiovgolos/
19.	МНК - Львів	https://www.facebook.com/groups/mnklviv/
20.	«Дівочі посиденьи»	http://posydenky.lvivport.com/
21.	Інтернет-магазин «Розетка»	https://rozetka.com.ua
22.	Український патріотичний форум «Політика»	http://uapolitics.com/
23.	«Рідна Україна»	https://www.facebook.com/ridna.uk
24.	Zero Waste Lviv	https://www.facebook.com/zerowastelviv/

Додаток В. Список публікацій здобувача за темою дисертації та відомості про апробацію результатів дисертації

1. Huminskyi R.V., Peleshchyshyn A.M., Holub (Lazurak) Z.D. Suggestions for informational influence on a virtual community // International Journal of Computer Science and Business Informatics. 2015. Vol. 15, No. 1. P.47-65. Available at: http://ijcsbi.org/index.php/ijcsbi/article/view/512/147
2. Holub (Lazurak) Z. The algorithm for detecting online discussion fragments containing information and psychological manipulation // Regional interuniversity compendium of scientific works «System technologies» [Системные технологии]. Dnipro, 2017. № 6 (113). P. 85-91.
3. Голуб (Лазурак) З.Д. Формалізація прийомів інформаційно-психологічної маніпуляції // Вісник Національного технічного університету «ХПІ». Збірник наукових праць. Серія: Нові рішення в сучасних технологіях. Харків: НТУ «ХПІ», 2017. № 32 (1254). С. 55-61.
4. Голуб (Лазурак) З.Д. Система критеріїв для виявлення фрагментів онлайн-дискусій з підозрою на наявність інформаційно-психологічної маніпуляції // Вісник Національного технічного університету «ХПІ». Збірник наукових праць. Серія: Нові рішення в сучасних технологіях. Харків: НТУ «ХПІ», 2018. 9 (1285). С. 106-111.
5. Голуб (Лазурак) З.Д. Структура словника маркерів лексичних змінних для виявлення інформаційно-психологічних маніпуляцій // Вісник Хмельницького національного університету. Серія: Технічні науки. 2017. № 2 (259). С. 264-268.
6. Pełeszczyszyn A., Holub (Lazurak) Z. Development of the system for detecting manipulation in online discussions // Advances in Intelligent Systems and Computing (AISC). 2017. Vol. 543. P. 111-117. Available at: https://link.springer.com/chapter/10.1007/978-3-319-48923-0_15
7. Голуб (Лазурак) З.Д. Розроблення алгоритму виявлення шкідливих інформаційно-психологічних маніпуляцій в онлайн-спільнотах ВНЗ // Інформатизація вищого навчального закладу: Вісник Національного університету «Львівська політехніка». Львів, 2017. № 879. С. 33-41.
8. Голуб (Лазурак) З.Д. Розроблення формальних моделей для автоматизації виявлення інформаційно-психологічної маніпуляції // Управління розвитком складних систем: зб. наук. пр. / Київський нац. університет будівництва і архітектури. Вип. 34. Київ, 2018. С. 85-91.
9. Peleschyshyn A., Holub I., Holub (Lazurak) Z. Methods of real-time detecting manipulation in online communities // Proceedings of the XIth International Scientific and Technical Conference on Computer Science and Information Technologies (CSIT 2016). Lviv Polytechnic Publishing House, 2016. P. 15-17.

10. Peleschyshyn A., Holub I., Holub (Lazurak) Z. Formal model and key features of an online community fundamental for detecting informational and psychological manipulation // Proceedings of the XIIth International Scientific and Technical Conference on Computer Science and Information Technologies (CSIT 2017). Lviv Polytechnic Publishing House, 2017. P. 101-104.
11. Peleschyshyn A., Holub I., Holub (Lazurak) Z. The preliminary stage of the algorithm for detecting information and psychological manipulation in online communities // Proceedings of the XIIIth International Scientific and Technical Conference on Computer Science and Information Technologies (CSIT 2018). Lviv Polytechnic Publishing House, 2018. P. 30-33.
12. Korzh R., Peleschyshyn A., Holub (Lazurak) Z., Analysis of integrity and coverage completeness of the informational image of a higher education institution // Proceedings of the XIIIth International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET`2016), Lviv-Slavske. Lviv, 2016. P. 825-827.
13. Пелещищин А., Голуб (Лазурак) З. Виявлення маніпуляції щодо потенційних покупців в онлайн спільноті // Матеріали Міжнародної науково-практичної конференції «Інформаційне суспільство: тенденції регіонального розвитку» (ISRDT-2016). Львів: Редакція ”УП” , 2016. С. 58-59.
14. Голуб (Лазурак) З.Д. Огляд класифікацій онлайн спільнот // Інформаційна діяльність, документознавство, бібліотекознавство: історія, сучасність, перспективи : матеріали ІІІ Всеукр. наук.-практ. конф., Київ, 25–26 квіт. 2017 р. Київ: НАКККіМ, 2017. С. 17-19.
15. Голуб (Лазурак) З., Пелещишин А. Моделювання маніпулятивної тактики за допомогою кусково-лінійного агрегата // Матеріали 5-ї міжнародної науково-практичної конференції «Інформація, комунікація, суспільство − 2016». Львів, 2016. С. 80-81.
16. Голуб (Лазурак) З., Пелещишин А. Виявлення прийомів ІПМ реалізованих за допомогою посилань // Матеріали 6-ї міжнародної наукової конференції «Інформація, комунікація, суспільство − 2017». – Львів, 2017. – С. 65-66.
17. Голуб (Лазурак) З., Пелещишин А. Види спілкування в онлайн-спільнотах та їхні характеристики // Матеріали 7-ї міжнародної наукової конференції «Інформація, комунікація, суспільство − 2018». – Чинадієво, 2018. – С. 45-46.

Апробація результатів дисертації. Основні результати дисертаційного дослідження неодноразово доповідалися на міжнародних та всеукраїнських наукових конференціях, зокрема на: 5–7 Міжнародних наукових конференціях «Інформація, комунікація, суспільство» (Львів, 2016–2018); XIII Міжнародній конференції «Сучасні проблеми радіоелектроніки, телекомунікацій, комп'ютерної інженерії» TCSET'2016 (Львів, 2016); XI–XIII Міжнародних науково-

технічних конференціях «Комп’ютерні науки та інформаційні технології» (Львів, 2016–2018); ІІІ Всеукраїнській науково-практичній конференції «Інформаційна діяльність, документознавство, бібліотекознавство: історія, сучасність, перспективи» (Київ, 2017); Міжнародній науково-практичній конференції «Інформаційне суспільство: тенденції регіонального розвитку (Львів, 2016). Результати дисертаційних досліджень регулярно доповідалися на наукових семінарах кафедри соціальних комунікацій та інформаційної діяльності Національного університету «Львівська політехніка» (2016–2018).