

Кокідько Богдан Сергійович

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського, Київ
ORCID 0009-0003-2669-3458

Шушура Олексій Миколайович

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського, Київ
ORCID 0000-0003-3200-720X

АРХІТЕКТУРА МАСШТАБОВАНОЇ СИСТЕМИ РОЗПІЗНАВАННЯ ПОВЕДІНКОВИХ ПАТЕРНІВ У СОЦІАЛЬНИХ МЕРЕЖАХ З ВИКОРИСТАННЯМ ГРАФОВИХ БАЗ ДАНИХ ТА НЕЧІТКОЇ ЛОГІКИ

Анотація. Предметом дослідження є архітектура масштабованої інформаційної системи для поведінкового аналізу в соціальних мережах. Зі зростанням складності та обсягу контенту, що генерується користувачами, зростає попит на передові фреймворки, здатні розуміти, моделювати та прогнозувати поведінку користувачів у динамічних та великомасштабних цифрових середовищах. Метою роботи є розробка сервісно-орієнтованої архітектури інформаційної системи, яка інтегрує графові бази даних, обробку природної мови та нечітку логіку для вилучення семантичних висновків із соціальних даних, виявлення моделей поведінки та підтримки візуалізації та прийняття рішень у режимі реального часу. Задачі: формування функціональних та архітектурних вимог до поведінково-орієнтованої аналітики для неоднорідних потоків даних, формалізація багатопланового нечіткого графічного представлення соціальних взаємодій, розробка алгоритмів з розпізнавання патернів, що поєднують системи нечіткого висновку та графові ознаки для виявлення тенденцій, аномалій та динаміки спільноти та розробка архітектури інформаційної системи на основі мікросервісів, яка забезпечує масштабованість, модульність та сумісність на всіх етапах обробки. Отримані результати: архітектура системи дозволяє будувати семантично збагачені нечіткі графи, використовуючи кілька нечітких відношень для формального представлення поведінки користувачів та інтенсивності взаємодії. Вона інтегрує NLP та нечітку логіку для перетворення сигналів настроїв та тематичних сигналів на інтерпретовані нечіткі анотації, що зберігаються в графовій структурі. Дослідження демонструє, що інтеграція нечіткої логіки та аналітики графів у модульну систему забезпечує гнучкий, інтерпретований та масштабований поведінковий аналіз у соціальних мережах. Цей підхід розширює можливості виявлення складних, накладаючих поведінкових моделей, враховуючи як явні взаємодії, так і приховані лінгвістичні сигнали.

Ключові слова: інформаційна система, графова база даних, нечітка логіка, соціальна мережа, класифікація, кластеризація, прогнозування, обчислювальна соціальна наука, поведінковий аналіз, аналіз настроїв.

Kokidko Bohdan

National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv
ORCID 0009-0003-2669-3458

Shushura Oleksii

National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv
ORCID 0000-0003-3200-720X

ARCHITECTURE OF A SCALABLE SYSTEM FOR BEHAVIORAL PATTERN RECOGNITION IN SOCIAL NETWORKS USING GRAPH DATABASES AND FUZZY LOGIC

Annotation. The subject of this study is the development of modular and intelligent system architecture for behavioral analysis within social networks. With the increasing complexity and volume of user-generated content, there is a growing demand for advanced frameworks capable of understanding, modeling, and predicting user behavior in dynamic and large-scale digital environments. The purpose of this work is to design a service-oriented architecture that integrates graph databases, natural language processing and fuzzy logic to extract semantic insights from social data, identify behavioral patterns, and support real-time visualization and decision-making. The objectives of the study include: formalizing functional and architectural requirements for behavior-aware analytics on heterogeneous data streams; developing a multi-layer fuzzy graph representation of social interactions; designing pattern recognition algorithms that combine fuzzy inference systems and graph-based features to detect trends, anomalies, and community dynamics; and

implementing a microservice-based system that ensures scalability, modularity, and interoperability across processing stages. Results obtained: The proposed system constructs semantically enriched fuzzy graphs using multiple fuzzy relations to represent user behavior and interaction intensity. It integrates NLP and fuzzy logic to convert sentiment and thematic signals into interpretable fuzzy annotations stored in the graph structure. The study demonstrates that integrating fuzzy logic and graph analytics within a modular system enables flexible, interpretable, and scalable behavioral analysis in social networks. This approach enhances the ability to detect complex, overlapping behavioral patterns by accounting for both explicit interactions and hidden linguistic signals.

Keywords: *information technology, graph databases, fuzzy logic, social networks, computational social science, behavioral analysis, sentiment analysis.*

Постановка проблеми.

Експоненціальне зростання контенту, створеного користувачами, на платформах соціальних мереж призвело до безпрецедентних можливостей та викликів у розумінні людської поведінки в цифрових спільнотах. Традиційні методи аналізу даних часто не справляються з проблемами складно пов'язаної та неструктурованої природи соціальних взаємодій в Інтернеті. У відповідь на це, в наявних дослідженнях було запропоновано використовувати графові бази даних та обробку природної мови (NLP) для подолання цієї аналітичної прогалини. Графові моделі даних за своєю суттю фіксують взаємозв'язки та моделі впливу між користувачами, тоді як NLP сприяє семантичній інтерпретації величезного текстового контенту, починаючи від постів і коментарів та закінчуючи особистими повідомленнями.

Протягом останнього десятиліття інтеграція цих технологій показала значні перспективи в багатьох сферах. Дослідники розробили гібридні фреймворки для виявлення еволюції спільноти та змін у поведінці, поєднуючи часові знімки графів з моделями NLP на основі трансформаторів. Графи знань також стали надійними інструментами для об'єднання метаданих поведінки із семантичним контентом, покращуючи контекстуальне розуміння дій користувачів. Крім того, зростання графових нейронних мереж (GNN) у поєднанні з аналізом настроїв дозволило прогнозувати поведінку користувачів та динаміку дискурсу в режимі реального часу. Зусилля дослідників також зосереджені на виявленні шкідливої або оманливої поведінки, такої як поширення дезінформації або активність ботів, за допомогою алгоритмів поширення графів, збагачених виявленням лінгвістичної позиції. Ці методи виявилися особливо цінними в модератії, аналізі психічного здоров'я та відстеженні впливу, ілюструючи універсальність та соціальну актуальність моделювання поведінки в соціальних мережах.

Незважаючи на ці досягнення, актуальною задачею є розробка архітектур інформаційних систем, які цілісно включають попередню обробку даних, семантичне збагачення, розпізнавання образів та дієву візуалізацію. Дана робота пропонує модульну, сервісно-орієнтовану архітектуру (SOA), яка використовує графові бази даних, NLP та нечітку логіку для надійного аналізу поведінки в соціальних мережах.

Аналіз останніх досліджень і публікацій.

Моделювання поведінки в соціальних мережах стало основою для міждисциплінарних досліджень, які поєднують методи обробки природної мови (NLP), графової аналітики, нечіткої логіки та архітектур масштабованих систем. В умовах інформаційного переважання особливої актуальності набули інтерпретовані моделі, здатні працювати в реальному часі з неоднозначними та структурно складними даними.

Серед ключових напрямів розвитку домінують підходи, що комбінують глибоке навчання з аналізом соціальних графів. Запропоновані архітектури, які одночасно здійснюють класифікацію контенту і виявлення спільнот, об'єднуючи контекстуальні мовні моделі з топологічними ознаками графів [1]. У дослідженнях глибокого представлення даних виявлено, що інтеграція структурної інформації з вбудованими репрезентаціями покращує видимість прихованих поведінкових шаблонів [2]. Інші праці підтверджують, що включення контексту взаємодії у графову структуру дозволяє точніше прогнозувати наявність шкідливих або токсичних обговорень [3].

Поряд із структурним аналізом, активно розвивається напрям використання нечіткої логіки у поведінковому моделюванні. У відповідних роботах представлено системи, які оперують нечіткими правилами для класифікації або прогнозування поведінки користувачів в умовах лінгвістичної невизначеності [4]. Поєднання семантичного аналізу з нечіткими оцінками дозволяє точніше формувати поведінкові профілі [5]. У контексті смарт-міст запропоновано використання моделей soft computing для моделювання соціальної динаміки на основі геопросторових і часових факторів [6].

Цікавим напрямом досліджень залишаються методи аналізу тексту. Предметно-орієнтовані лексикони та контекстуальні вбудовування перевершують універсальні моделі у завданнях аналізу

громадської думки [7]. Хоча лише мовних ознак недостатньо для надійного виявлення фейкових повідомлень, також необхідним є підкріплення моделями взаємодії користувачів та метаданими мережевої активності [8].

У більш прикладному аспекті продемонстровано можливості графових баз знань для об'єднання текстових даних з реляційними атрибутами в завданнях аналізу політичного або стратегічного контенту [9].

З точки зору архітектури інформаційних систем, увага акцентується на гнучкості, масштабованості та модульності. Сервісно-орієнтовані структури, які поєднують класифікатори, API-інтерфейси та графові сервіси в окремих модулях, дають змогу створювати адаптивні конвеєри для поведінкового аналізу [1]. Об'єднання гібридних системи фільтрації контенту з керованими та некерованими методами, зокрема використання моделей на базі BERT, допомагають з оптимізацією обробки у стрімінгових середовищах [10].

У напрямку прогнозування змін громадської думки під час поширення стратегічних наративів у соціальних мережах було запропоновано модель, що поєднує нейронні мережі з дифузійною моделлю Басса для моделювання динаміки соціально-політичних настроїв. Такий підхід акцентує увагу на важливості інтеграції механізмів прогнозування поведінки з адаптацією контенту, що узгоджується з цілями даного дослідження є підтримкою нечіткого міркування в реальному часі та картографуванням впливу у соціальних графах [11].

Загалом, хоча окремі компоненти: NLP, графові моделі, нечітка логіка, архітектури є достатньо добре досліджені, їхня повна інтеграція в межах єдиної інформаційної системи аналізу соціальних мереж залишається недостатньо реалізованою. У наявних рішеннях переважає фокус на окремі модулі, що знижує здатність таких систем до гнучкої інтерпретації даних у режимі реального часу. Це обґрунтовує потребу в розробці єдиної платформи, яка б об'єднувала зазначені підходи у масштабовану й інтерпретовану поведінкову систему для роботи із соціальними графами.

Мета і задачі дослідження.

Метою цього дослідження є розробка масштабованої та модульної архітектури для аналізу поведінкових моделей у соціальних мережах шляхом інтеграції моделей графових баз даних, NLP та нечіткої логіки в цілісну, сервісно-орієнтовану структуру.

Цілі дослідження:

- сформулювати архітектурні та функціональні вимоги до системи аналітики, що враховує поведінку, на потоках даних з різноманітних соціальних платформ;
- формалізувати графові представлення взаємодій у соціальних мережах з використанням нечіткої логіки, які дозволяють реалізувати семантично насичене та топологічно усвідомлене розпізнавання патернів;
- розробити алгоритми розпізнавання патернів, які використовують як системи нечіткого висновку, так і графові структури для виявлення поведінкових тенденцій, аномалій та висновків на рівні спільноти;
- забезпечити масштабованість, модульність та сумісність інформаційної системи шляхом розробки архітектури на основі мікросервісів з чітко визначеними інтерфейсами для отримання, обробки, аналізу та звітності даних.

Результати дослідження.

Розробка масштабованої та інтелектуальної системи розпізнавання поведінкових шаблонів у соціальних мережах накладає низку як функціональних, так і не функціональних вимог, зумовлених необхідністю ефективної роботи в складних цифрових екосистемах, що інтенсивно використовують дані. Система призначена для функціонування в середовищах, що характеризуються високошвидкісними потоками даних, семантично різноманітним контентом та постійно мінливими моделями взаємодії з користувачами.

З функціональної точки зору, система повинна підтримувати отримання та обробку даних у режимі реального часу з кількох соціальних платформ, таких як Twitter, Reddit, Telegram тощо. Вона повинна включати можливості обробки природної мови для вилучення настроїв, намірів та контекстного значення з неструктурованого текстового введення. Для ефективного моделювання взаємодії, взаємодії та поведінкових рис користувачів необхідний шар зберігання та аналізу на основі графів. Крім того, система повинна інтегрувати механізми нечіткого мислення для управління невизначеністю, властивою лінгвістичним описам та класифікації поведінки, а також забезпечувати поведінкове профілювання для виявлення та моніторингу таких тенденцій, як поляризація, динаміка

впливу або аномальна активність. Система повинна забезпечувати механізми розпізнавання поведінкових шаблонів шляхом інтеграції систем нечіткого висновку, здатних інтерпретувати неточні або неоднозначні лінгвістичні дані. Це включає визначення та застосування нечітких функцій належності до поведінкових вимірів, таких як вплив, залученість, довіра та мінливість настроїв. Система повинна забезпечувати поведінкове профілювання шляхом логічного виведення над графовими зв'язками та нечіткими атрибутами для виявлення таких шаблонів, як соціальна поляризація, аномальна залученість, поширення дезінформації та нові спільноти.

З не функціональної сторони, архітектура повинна забезпечувати масштабованість шляхом горизонтального масштабування компонентів прийому, обробки та зберігання для розміщення великих та зростаючих наборів даних. Модульність є важливою для того, щоб кожен компонент міг функціонувати незалежно та бути заміненим або оновленим без порушення роботи всієї системи. Взаємодія є ще однією ключовою вимогою, що забезпечує сумісність із зовнішніми API, сторонніми аналітичними службами та різноманітними джерелами даних. Нарешті, система повинна бути розширюваною, що дозволяє майбутню інтеграцію передових модулів штучного інтелекту та нових процесорів даних, таких як зображення або відео-контент, у міру розвитку аналітичних потреб.

Для полегшення семантично значного та топологічно усвідомленого розпізнавання образів спочатку потрібно побудувати графове представлення соціальної мережі, яке інтегрує нечітку логіку у свою структуру. Кожен обліковий запис користувача моделюється як вузол у графі $G = (A, E)$, де A – множина облікових записів, а E – множина ребер, що позначають соціальні взаємодії. На відміну від традиційного бінарного сусідства, ми визначаємо G як нечіткий реляційний граф: ребра несуть ваги та мітки, що відображають силу та тип соціальних зв'язків. Соціальна мережа представлена як нечітка реляційна структура $S = (A, \{F_k\})$, де A — це скінченна множина користувачів, і кожен $F_k \subseteq A \times A$ — це нечітке відношення, що кодує певний тип відношення (якщо кожен F_k були б чіткими, ця структура була б зведена до стандартного графа). Нечіткі відношення виражають нюанси людських зв'язків, що забезпечує різноманітніше лінгвістичне значення, ніж бінарний граф, оскільки кожне ребро може вказувати на те, наскільки сильно пов'язані два користувачі або в якому контексті вони взаємодіють, а не лише на те, чи пов'язані вони.

Для представлення різних аспектів соціальних меж ми вводимо кілька шарів нечітких відношень, використовуючи термінологію з попередньої роботи [12], щоб визначити три ключові нечіткі відношення на A :

FFS (Нечітка сила дружби): Нечіткий зв'язок, що відображає силу або близькість соціального зв'язку між двома обліковими записами. Значення $\mu_{FFS}(a_i, a_j)$ відображає, наскільки «сильна» дружба або зв'язок між користувачами a_i та a_j . Він розраховується шляхом агрегування показників сили зв'язку, таких як близькість взаємодій (особисті повідомлення, емоційний настрій у спілкуванні) та тривалість або довіра у стосунках. Високий рівень μ_{FFS} передбачає міцну, близьку дружбу, тоді як навпаки, низьке значення передбачає слабкий або випадковий зв'язок.

FI (Інтенсивність нечіткої взаємодії): Нечітке відношення, яке фіксує частоту та інтенсивність прямих взаємодій між обліковими записами. Рівень $\mu_{FI}(a_i, a_j)$ є високим коли користувачі a_i та a_j спілкуються або дуже часто взаємодіють одне з одним (часті коментарі, теги тощо) та низько, якщо їхній контакт мінімальний. Щоб визначити це, ми фаззифікуємо кількість або швидкість взаємодій, наприклад, нехай N_{ij} це кількість взаємодій між a_i та a_j в проміжку часу, тоді визначимо $\mu_{FI}(a_i, a_j) = f(N_{ij})$.

FAS (Нечітка подібність спорідненості): Нечіткий зв'язок, що відображає спорідненість між двома обліковими записами з точки зору атрибутів профілю або поведінки. Це вимірює, наскільки два облікові записи схожі, наприклад, за інтересами, членством у групах або моделями активності, що часто корелює з належністю до спільноти або спільними вподобаннями. Ми визначаємо $\mu_{FAS}(a_i, a_j)$ шляхом порівняння характеристик користувачів (теми публікацій, вподобаний контент, демографічні показники тощо) за допомогою нечітких мір подібності. Якщо A_k (a) позначає числовий атрибут облікового запису a (кількість постів на тиждень, інтерес до теми k , тощо), ми можемо визначити нечітку множину для кожного атрибута. Тоді рівень $\mu_{FAS}(a_i, a_j)$ є високим, коли багато з цих нечітких подібностей, специфічних для атрибутів, перетинаються. Наприклад, якщо два користувачі класифікуються як «високоактивні» та мають кілька спільних категорій «інтересів», то їхній зв'язок спорідненості матиме велике значення членства. Ми можемо формалізувати це як: $\mu_{FAS}(a_i, a_j) = \text{Sim}_{fuzzy}(A_k(a_i), A_k(a_j))$, де Sim_{fuzzy} застосовує нечіткий перетин на множинах атрибутів

кожного користувача. Високий FAS може означати, що користувачі схожі або належать до подібних кіл (семантично), навіть якщо вони не часто взаємодіють безпосередньо.

Кожне нечітке відношення $F \in \{FFS, FII, FAS\}$ представлене його функцією належності $\mu_F : A \times A \rightarrow [0,1]$. Модель соціальної мережі можна представити як багаточаровий нечіткий граф, де кожен шар F додає різне семантичне пояснення до зв'язків. Зв'язок між a_i та a_j може мати до трьох нечітких ваг (по одній на тип стосунків), наприклад, «Аліса та Боб мають середню силу дружби, високу інтенсивність та низьку спорідненість». Це розрізняє, чому та як користувачі пов'язані, а не розглядає всі зв'язки як рівноцінні.

Для побудови цих зв'язків ми використовуємо ключові характеристики рівня облікового запису та перетворюємо їх на нечіткі змінні, де кожен користувач облікового запису $a \in A$ описується вектором ознак (активність, кількість друзів, обліковий запис, вік тощо). Ми визначаємо нечіткі функції належності для цих ознак, щоб класифікувати користувачів за семантичними термінами. Наприклад, нечітку змінну «Активність» можна визначити для частоти публікацій за допомогою функції належності $\mu_{\text{activity}}(a)$, якщо користувач a публікує більше 100 повідомлень на тиждень, то функцію належності можна встановити на 0,9 або 1, якщо користувач a публікує менше 5 повідомлень, то 0,1 з градаціями між ними для помірної активності. Аналогічно, «Впливова особа» може бути нечіткою множиною на основі кількості підписників або оцінки PageRank на графі, або «Новий обліковий запис» – нечіткою змінною для віку облікового запису. Ці нечіткі змінні для окремих облікових записів враховуються у визначеннях відношень F .

Щоб зіставити користувацький контент з поведінковими або тематичними кластерами (мова ворожечі, реклама, скарга тощо), система обчислює семантичну подібність між користувацьким текстом та попередньо визначеними дескрипторами категорії. Кожен екземпляр тексту вбудовується у багатовимірний векторний простір за допомогою попередньо навчених моделей (вбудовування OpenAI, трансформатори речень BERT тощо). Косинусна подібність між вектором вбудованого тексту txt та посилання на вектор категорій cat обчислюється як [13, 14]:

$$\cos(\mu) = \frac{\sum_{i=1}^n \text{txt}_i * \text{cat}_i}{\sqrt{\sum_{i=1}^n \text{txt}_i^2} * \sqrt{\sum_{i=1}^n \text{cat}_i^2}} \quad (1)$$

де μ ступінь, з якою текст користувача належить або відповідає певній категорії поведінки. Це значення, що варіюється від 0 до 1, інтерпретується як нечіткий ступінь семантичної відповідності з цільовою категорією поведінки. Ці нечіткі семантичні мітки потім поширюються в базу даних графів як ознаки рівня вузлів або рівня зв'язків, формуючи основу для профілювання поведінки.

Для семантичного збагачення графа кожне ребро в графі позначається типом взаємодії. Наприклад, ми розрізняємо, чи є взаємодія прямим повідомленням, коментарем до публікації, «лайком» чи спільним членством у групі. Їх можна розглядати як окремі типи відношень або анотації до існуючих нечітких відношень. Однак, для підтримки ефективності ми часто кодуємо такий семантичний контекст як атрибуту, що подаються в первинні відношення F . Наприклад, FII (взаємодія) зв'язок можна розбити за типом взаємодії: FII_{comment} для частоти коментарів, FII_{post} для інтенсивності публікацій. Це представлення відображає значення взаємодій, тобто хто з ким розмовляє і як.

Графове представлення зберігає топологію соціальних мереж, дозволяючи нам витягувати структурні ознаки, необхідні для розпізнавання образів. Оскільки ребра зважуються за допомогою F , класичні графові метрики адаптовані для врахування цих нечітких ваг. Наприклад, ступінь центральності вузла можна узагальнити шляхом підсумовування нечітких сил зв'язків, а алгоритми виявлення спільноти можна застосувати до зваженої матриці суміжності μ_{FFS} (знайти міцні групи друзів) або μ_{FAS} (знайти спільноти за інтересами). Ці показники роблять представлення топологічно усвідомленим: такі шаблони, як хаби або ізольовані вузли, можна виявити зі структури графа. Збагачуючи кожен вузол цими метриками графа, ми підсумовуємо кожен вузол з його топологічним положенням.

Завдяки визначенню нечіткого графowego представлення ми тепер можемо розробляти та оцінювати алгоритми розпізнавання патернів, які використовують як структурні особливості графів, так і системи нечіткого висновку (FIS) для збору інформації, такої як поведінкові тенденції, аномалії та закономірності на рівні спільноти. FIS може обробляти неточні, шумові дані, використовуючи експертні знання та лінгвістичні правила. Поєднуючи нечітку логіку з аналітикою графів, ми можемо визначити, коли поведінка є незвичайною або лише помірно відрізняється, або коли група є справжньою спільнотою чи вільною приналежністю. Стратегія полягає у вилученні відповідних ознак із збагаченого графа та введенні їх у моделі на основі нечітких правил або гібридні алгоритми для класифікації або кластеризації шаблонів.

Кожен обліковий запис та його локальне оточення описуються набором ознак, отриманих з графічного представлення, до яких належать топологічні метрики (нечіткий ступінь, нечітка кластеризація, центральність тощо) та семантичні атрибути (нечіткий рівень активності облікового запису, теми контенту тощо). Також враховуються ознаки спільноти, такі як ступінь членства в різних спільнотах. Для динамічного розпізнавання шаблонів, такого як виявлення трендів, ми відстежуємо ці ознаки з часом (щоденні/тижневі/щомісячні знімки мережі), щоб зафіксувати часову еволюцію. Замість того, щоб застосовувати чіткі пороги до цих ознак, ми перетворюємо їх на нечіткі змінні, наприклад, ознаку, таку як «ступінь центральності = 50» у нечітких термінах можна інтерпретувати як «Високий ступінь» з коефіцієнтом належності 0,9 та «Середній ступінь» з коефіцієнтом приналежності 0,4. Зі зміною в часі, як «+30% збільшення кількості взаємодій цього тижня», можна перетворити на лінгвістичні терміни, такі як «Помірне зростання».

FIS використовує набір лінгвістичних правил виду «ЯКЩО умова, ТОДІ результат», де умови – це нечіткі твердження про ознаки, а результати відповідають класам шаблонів або оцінкам. Нечітка база правил кодує експертні знання та гіпотези про соціальну поведінку. Наприклад, ми можемо визначити такі правила, як: «ЯКЩО інтенсивність взаємодії користувача висока ТА його спорідненість з відомою спільнотою низька, ТО ймовірність аномалії висока» (аутсайдер або фальшивий обліковий запис) або «ЯКЩО внутрішня щільність FFS спільноти висока ТА її зовнішня зв'язність низька, ТО спільнота є згуртованою» (щільна група). Антецедент кожного правила збирає кілька умов ознак за допомогою логічних операторів (І, АБО), а результат (ТО) призначає нечіткий вихід. FIS типу Мамдані є одним з найбільш широко використовуваних завдяки своїй здатності моделювати складні людські міркування. Використовуючи підхід нечіткого висновку, ми можемо отримувати градуйовані оцінки, такі як призначення 70% ймовірності аномалії, що краще відображає невизначеність у соціальній поведінці, на відміну від бінарної класифікації. У цих рамках ми можемо зосередитися на трьох різних категоріях розпізнавання патернів: виявлення аномалій, ідентифікація тенденцій та розуміння спільноти.

Виявлення аномалій: щоб виявити користувачів або взаємодію, які суттєво відхиляються від норми. Для кожного облікового запису ми обчислюємо оцінку аномалії на основі нечітких правил, які поєднують показники відхилення та незвичайні моделі поведінки. Наприклад, щоб зафіксувати аномалії взаємодії: «ЯКЩО користувач має низький показник FFS з іншими (слабка дружба), АЛЕ дуже високий FPI (дуже часта взаємодія), ТОДІ аномалія є середньою», оскільки справжня часта взаємодія зазвичай супроводжується сильними зв'язками. Інший приклад правила може бути спрямований на ботів або спам-акаунти: «ЯКЩО активність облікового запису висока ТА ступінь центральності низький (мало друзів), ТОДІ аномалія висока», що означає, що користувач багато публікує, але має мало зв'язків (типова для ботів модель). Нечітка система об'єднує такі правила, що призводить до оцінки аномалії або лінгвістичного результату, такого як «помірна аномалія». Система надає пояснення для кожного виявлення з точки зору запущених правил, що сприяє довірі до системи виявлення. Якщо обліковий запис, який раптово взаємодіє через дві інакше не пов'язані спільноти, може бути позначений як аномалія мосту, що вказує на можливе втручання інформаційного брокера або вторгнення.

Визначення тренду: Щоб виявити нові тенденції в поведінці користувачів або активності спільноти, ми відстежуємо еволюцію ознак нечіткого графу з часом і застосовуємо розпізнавання нечітких шаблонів до даних часових рядів. Для кожного періоду часу (день, тиждень, місяць тощо) ми обчислюємо сукупні показники, такі як середній FPI у спільноті або загальна кількість активних користувачів у групі. Потім фазифікувати ці показники та їх часові зміни. Наприклад, збільшення частоти публікацій у спільноті на 15% можна позначити як «помірне зростання», тоді як збільшення на 50% буде «швидким зростанням». Правила для визначення тенденцій можна визначити так: «ЯКЩО рівень появи високий ТА темпи зростання високі, ТО це зростаюча тенденція», для поведінки користувачів: «ЯКЩО залученість була низькою ТА залученість значно зросла ТА центральність мережі швидко зростає, ТО користувач є законодавцем тренду». Застосовуючи такі правила по всій мережі, цей алгоритм може виявляти такі явища, як тенденції контенту, вірусні користувачі, зниження залученості. Результатом може бути список трендових тем або спільнот, відсортованих за показником достовірності.

Аналітика спільноти: Щоб отримати уявлення про спільноти та групову динаміку в мережі, ми використовуємо алгоритми виявлення спільнот на зваженому графі. По-перше, нам потрібно ідентифікувати спільноти для аналізу. Це досягається шляхом застосування нечіткої кластеризації до графа, яка повертає значення членства $\mu_{C_k}(a_i)$ для кожного акаунту a_i в кожній спільноті C_k . Цей підхід обробляє випадки, коли користувач може перебувати в спільноті 0,8 у складі «Спортивні

вболівальники» та 0,5 у складі «Інженери», демонструючи сильну приналежність до спортивної групи, але також певну залученість до інженерів. Враховуючи ці перекриваючі спільноти, ми застосовуємо нечіткий висновок для оцінки показників рівня спільноти. Наприклад, ми визначаємо згуртованість спільноти як нечітку сукупність зв'язків членів спільноти, пов'язаних зі спортивними фанатами (FFS), у межах цієї спільноти. Якщо більшість пар у групі мають високу силу дружби, то згуртованість позначається як «Висока». Аналогічно, для взаємодії між спільнотами ми можемо розглядати FII (індекс внутрішньої щільності) через межі спільноти, наприклад: «ЯКЩО спільнота А має високу внутрішню щільність та низький зовнішній FII, ТО А є вузькою спільнотою». Щоб ідентифікувати користувачів-містків, які мають значне членство в кількох спільнотах, ми можемо мати правило типу: «ЯКЩО користувач має середньо-високий рівень членства в спільноті X та спільноті Y ТА пряма зв'язність X та X є низькою, ТО користувач є потенційним містком». Щоб знайти приховані поведінкові спільноти, дозволяючи вузлам користувачів бути членами різних поведінкових кластерів з різним рівнем членства, система використовує алгоритм нечіткої кластеризації. Цільова функція мінімізується до [15]:

$$J_m = \sum_{i=1}^N \sum_{j=1}^B u_{ij}^m \|a_i - cc_j\|^2, \quad (2)$$

де N визначає кількість користувачів, B кількість поведінкових кластерів, a_i вектор ознак користувача i , $u_{ij} \in [0,1]$ ступінь належності користувача i в кластері j , $m \in (1, \infty)$ коефіцієнт нечіткості. Центр кластера cc_j та належність u_{ij} оновлюються до [15]:

$$cc_j = \frac{\sum_{i=1}^N u_{ij}^m a_i}{\sum_{i=1}^N u_{ij}^m}, \quad (3)$$

Результати включають класифікацію спільнот (згуртовані, ізольовані, взаємопов'язані тощо) та список ключових користувачів (центр впливу в межах спільноти, мости між спільнотами тощо). За допомогою цих закономірностей можна визначити, чи дві спільноти зливаються з часом через певну подію або сплеск спільних інтересів.

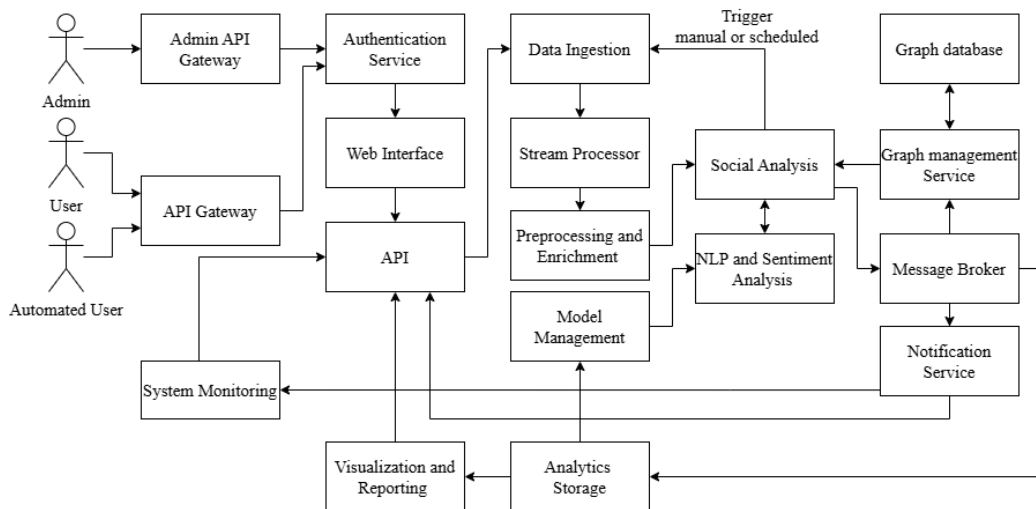


Рис. 1. Схема архітектури системи

Запропонована архітектура інформаційної системи для аналізу соціальних мереж базується на принципах SOA, що підтримується мікросервісною декомпозицією та слабким зв'язком функціональних компонентів. Цей архітектурний вибір зумовлений необхідністю керувати високошвидкісними, великооб'ємними та семантично різноманітними потоками даних, типовими для сучасних соціальних мереж. Делегуючи обов'язки незалежним сервісам, система забезпечує модульну розробку, ізоляцію помилок та безперешкодне горизонтальне масштабування. Кожен функціональний блок, прийом даних, попередня обробка, NLP, побудова графів, застосування нечіткої логіки, поведінкове мислення та візуалізація, реалізовано як окремий мікросервіс, що сприяє цілеспрямованій розробці та гнучкому розгортанню. Зв'язок між компонентами є асинхронним та спирається на легкі протоколи обміну повідомленнями, такі як RESTful API або брокери повідомлень, такі як Kafka або RabbitMQ, що дозволяє системі ефективно працювати в умовах обмежень реального часу, залишаючись стійкою до затримок та збоїв. Архітектура підтримує як пакетний, так і потоковий режими обробки, що дозволяє проводити історичний аналіз разом із моніторингом поведінки в режимі реального часу. В основі системи лежить нечітка семантика, де лінгвістичні дескриптори, такі як

«висока залученість» або «помірна довіра», перетворюються на нечіткі значення членства та поширюються через модель графових даних. Ці семантичні збагачення зберігаються в розширеній графовій базі даних, такої як Neo4j або ArangoDB, де кожен вузол і ребро анотуються нечіткими вагами, що представляють ступені інтенсивності поведінки, довіри або впливу [16]. Для забезпечення прозорості та інтерпретації всі поведінкові висновки ґрунтуються на семантичних мітках та нечітких наборах правил, що дозволяє генерувати такі результати, як «Користувач А, ймовірно, має незначний вплив у спільноті Х», з лінгвістичною ясністю. Для підтримки гнучкості та незалежності від домену система спирається на настроювані набори правил, вивчені моделі подібності на основі вбудовування, а також використання стратегій кешування та паралелізації для оптимізації продуктивності в операціях нечіткої кластеризації та запитів [17, 18]. На рисунку 1 показано запропоновану архітектуру системи.

Рівень **User Interface** показує дві сутності користувачів, що представляють різні ролі в системі. Користувач — це люди, такі як аналітики або модератори, які запитують систему, отримують доступ до панелей інструментів та інтерпретують поведінкові висновки. Автоматизований користувач представляє автоматизовані компоненти, такі як служби отримання даних або зовнішні системи, що взаємодіють через API. Адміністратор відповідає за операції на рівні системи, такі як керування політиками доступу, моніторинг стану системи, оновлення моделей NLP або нечітких міркувань через модуль керування моделями та контроль за загальним виконанням робочого процесу.

Рівень **Data Ingestion** відповідає за збір даних, згенерованих користувачами, з кількох джерел соціальних мереж через API та сервіси потокової передачі в режимі реального часу. Ці дані можуть включати структурований та неструктурований контент, такий як публікації, коментарі, особисті повідомлення, профілі користувачів, позначки часу та метадані взаємодії (лайки, поширення, підписки тощо). Компонент обробки даних забезпечує надійність та своєчасність даних, підтримуючи як пакетний, так і потоковий режим обробки.

Після отримання дані передаються через **Stream Processor**, який обробляє виявлення подій у режимі реального часу, фільтрацію та пересилання до модулів нижче. Далі йде модуль **Preprocessing and Enrichment Module**, де дані очищуються, токенізуються та збагачуються контекстними метаданими. Для однакової обробки всі вхідні записи перетворюються в канонічну схему JSON, яку можна інкапсулювати в наступні поля, показані на рисунку 2.

```
{
  "user_id": "A123",
  "timestamp": "2025-02-10T08:15:30Z",
  "text": "This is absolutely amazing!",
  "interaction_type": "post"
}
```

Рис. 2. Приклад канонічної схеми JSON

Інструменти перетворення використовуються для стандартизації різних форматів (CSV, XML, TXT) у структуровані об'єкти JSON. Цей етап включає такі завдання, як виявлення мови, розпізнавання сутностей та вилучення ознак.

NLP and Sentiment Analysis відіграє ключову роль в інтерпретації поведінки користувачів, витягуючи семантичні та емоційні ідеї з текстових даних. Він використовує такі методи NLP, як тематичне моделювання, аналіз настроїв та розпізнавання намірів. На цьому етапі нечітка логіка інтегрується як основний аналітичний метод для обробки лінгвістичної нечіткості та неоднозначності, що дозволяє отримувати градієнтні оцінки настроїв (позитивні, сильно негативні тощо) замість бінарних виходів. Цей процес виконується за допомогою BERT-кодувальників на основі трансформаторів, залежно від цільової мови та предметної області. Оброблені та проаналізовані дані потім структуруються в сутності та зв'язки, формуючи граф соціальної взаємодії.

Модуль **Social Analysis** інтерпретує граф для виявлення моделей поведінки, структури спільноти та поширення впливу. Використовуючи правила на основі нечіткої логіки, система може відображати невизначені поведінкові тенденції, такі як ступінь впливу або ймовірність членства в групі.

Graph Management and Database керує динамічним зберіганням та запитам до графів взаємодії з користувачем за допомогою механізму баз даних графів. **Graph Management Service** пропонує оптимізоване зберігання та проходження складних зв'язків, що дозволяє відстежувати поведінку в часі та через мережеві вузли. Вхідні дані записуються після попередньої обробки та збагачення, тоді як аналітичні модулі та користувачі отримують інформацію для міркувань, візуалізації та підтримки рішень. Вона підтримує інтеграцію з передовими алгоритмами графів для рекомендацій, виявлення аномалій та тенденцій.

Модуль **Model Management** контролює розгортання та життєвий цикл моделей машинного навчання та нечіткої логіки, що використовуються в аналізі. Він займається навчанням моделей, керуванням версіями, перевіркою та оновленнями. Він також відповідає за налаштування нечітких правил та функцій належності на основі зворотного зв'язку та продуктивності системи, забезпечуючи адаптивність до змін у поведінці користувачів.

Для підтримки активності в режимі реального часу **Notification Service** надає сповіщення та аналітику зовнішнім системам або кінцевим користувачам на основі попередньо визначених правил та нечітких порогових значень. Одночасно модуль моніторингу системи забезпечує справність служби, відстежує показники продуктивності та надає зворотний зв'язок для перенавчання моделі або налаштування системи.

Результати аналізу та поведінкові показники зберігаються в **Analytics Storage** для довгострокової звітності. Інтерфейс **Visualization and Reporting** надає інтерактивні панелі інструментів для адміністраторів та аналітиків, щоб вони могли досліджувати тенденції даних, еволюцію настроїв, кластери поведінки та результати виявлення аномалій.

Висновки і перспективи подальших досліджень.

У цьому дослідженні запропоновано масштабовану та модульну архітектуру інформаційної системи для розпізнавання поведінки в соціальних мережах шляхом інтеграції графових баз даних, обробки природної мови та нечіткої логіки в єдину аналітичну структуру. Робота була мотивована обмеженнями існуючих систем поведінкового аналізу, яким часто бракує семантичної глибини та управління невизначеностями. В результаті було розроблено сервісно-орієнтовану систему для роботи з гетерогенними, високошвидкісними потоками соціальних даних, що дозволяє гнучке розгортання на рівнях прийому, обробки, міркування та візуалізації.

Суть цієї роботи полягає в поєднанні нечіткої логіки та графових структур даних для кращого розуміння поведінки користувачів у соціальних мережах. Такий підхід дозволяє гнучкіше оцінювати дії користувачів, враховуючи як прямі взаємодії (публікації, коментарі, лайки тощо), так і приховані сигнали в мові (тон, настрої тощо). Система включає кілька ключових частин: очищення та підготовку даних, обробку природної мови, групування поведінки за допомогою нечіткої кластеризації та гнучке міркування на основі правил.

Подальші дослідження можуть бути зосереджені на прототипуванні системи з використанням даних реального часу, розробці адаптивного навчання нечітких правил та інтеграції моделей великих мов для покращення семантичного збагачення та забезпечення адаптивної еволюції нечітких правил для відображення норм поведінки.

Список використаної літератури

1. Mohsan A., Mehdi H., Kashif K., Jin Y., Saqib H., Muhammad K. Social media content classification and community detection using deep learning and graph analytics. *Technological Forecasting and Social Change*. 2023. Vol. 188. DOI: <https://doi.org/10.1016/j.techfore.2022.122252>.
2. Qiaoyu T., Ninghao L., Xia H. Deep Representation Learning for Social Network Analysis. *Frontiers in Big Data*. 2019. Vol. 2. DOI: <https://doi.org/10.3389/fdata.2019.00002>
3. Liam H., Lukasz G., Robin C. Predicting Hateful Discussions on Reddit using Graph Transformer Networks and Communal Context. *2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. 2022. P. 9-17. DOI: <https://doi.org/10.1109/WI-IAT55865.2022.00012>
4. Alrashed S.A., Abraham A. User Behaviour Classification and Prediction Using Fuzzy Rule Based System and Linear Regression. *Journal of Information Assurance and Security*. 2017. Vol. 12. pp. 86-93. URL: <https://www.researchgate.net/publication/319645234>
5. Chaabi Y., Khadija L., Mounia B. Semantic Analysis of Conversations and Fuzzy Logic for the Identification of Behavioral Profiles on Facebook Social Network. *International Journal of Emerging Research in Management & Technology*. 2019. Vol. 14. No. 7. DOI: <https://doi.org/10.3991/IJET.V14I07.8832>
6. Ruan L., Li C., Zhang Y., Wang H. Soft computing model based financial aware spatiotemporal social network analysis and visualization for smart cities. *Computers, Environment and Urban Systems*. 2019. Vol. 77. DOI: <https://doi.org/10.1016/j.compenvurbsys.2018.07.002>
7. Liam H., Lukasz G., Robin C. Predicting Hateful Discussions on Reddit using Graph Transformer Networks and Communal Context. *2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence*

and Intelligent Agent Technology (WI-IAT). 2022. P. 9-17. DOI: <https://doi.org/10.1109/WI-IAT55865.2022.00012>

8. Oliveira N.R., Pisa P.S., Lopez M.A., Medeiros D.S.V., Mattos D.M.F. Identifying Fake News on Social Networks Based on Natural Language Processing: Trends and Challenges. *Information*. 2021. Vol. 12. P. 38. DOI: <https://doi.org/10.3390/info12010038>

9. Peng W., Zesong L., Zeyuan L., Fang X. A Government Policy Analysis Platform Based on Knowledge Graph. *2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD)*. 2019. P. 208-214. DOI: <https://doi.org/10.1109/ICAIBD.2019.8836979>

10. Poliarush O., Krepych S., Spivak I. Hybrid approach for data filtering and machine learning inside content management system. *Advanced Information Systems*. 2023. Vol.7. No. 4. DOI: <https://doi.org/10.20998/2522-9052.2023.4.09>

11. Romaniuk R., Voitko O., Parkhuts L., Rakhimov V., Kostiak M. Models for predicting changes in public opinion during the implementation of the narrative in social media. *Advanced Information Systems*. 2025. Vol.9. No. 1. DOI: <https://doi.org/10.20998/2522-9052.2025.1.12>

12. Кокідько Б., Шушура О. Використання графових баз даних та нечіткої логіки для аналізу та моделювання соціальних мереж. *Сучасні проблеми наукового забезпечення енергетики*. 2024. 23–26 квітня. Київ. КПІ ім. Ігоря Сікорського. С. 233–236. URL: https://iate.kpi.ua/uploads/p_144_85001196.pdf

13. Majid H.A., Sabrina T., Nazlia O., Nor S.S. Short text clustering algorithms, application and challenges: A survey. *Applied Sciences*. 2023. Vol. 13. DOI: <https://doi.org/10.3390/app13010342>

14. Кокідько Б., Шушура О. Моделювання соціальних мереж за допомогою теорії графів та нечіткої логіки. *Зв'язок*. 2025. Т. 5. DOI: <https://doi.org/10.31673/2412-9070.2024.051035>

15. Кокідько Б., Шушура О. Графові бази даних та нечітка логіка в адаптивних моделях соціальних мереж. *Сучасні інформаційні системи та технології*. 2024. 29 листопада. Хмельницький. С. 31–34.

16. Кокідько Б., Шушура О. Розробка нечіткої графової бази даних. *Сучасні проблеми наукового забезпечення в енергетиці*. 2025. 22–25 квітня. Київ. . КПІ ім. Ігоря Сікорського.

17. Es-sabery F., Es-sabery I., Qadir J., Sainz-de-Abajo B., Garcia-Zapirain B. A hybrid Hadoop-based sentiment analysis classifier for tweets associated with COVID-19 utilizing two machine learning algorithms: CNN, and fuzzy C4.5. *Journal of Big Data*. 2024. Vol. 11. DOI: <https://doi.org/10.1186/s40537-024-01014-4>

18. Qimin Z., Yingcang M., Zhiwei X., Xiaofei Y. A comprehensive framework for designing and learning fuzzy cognitive maps at the granular level. *Applied Soft Computing*. 2024. Vol. 158. DOI: <https://doi.org/10.1016/j.asoc.2024.111601>