

ІНТЕЛЕКТУАЛЬНА СИСТЕМА КЛАСТЕРИЗАЦІЇ КОРИСТУВАЧІВ СОЦІАЛЬНИХ МЕРЕЖ НА ОСНОВІ АНАЛІЗУ ТОНАЛЬНОСТІ ДАНИХ

Тарас Батюк, Дмитро Досин

Національний університет “Львівська політехніка”,
кафедра інформаційних систем та мереж, Львів, Україна
E-mail: taras.m.batiuk@lpnu.ua, ORCID: 0000-0001-5797-594X
E-mail: dmytro.h.dosyn@lpnu.ua, ORCID: 0000-0003-4040-4467

© Батюк Т. М., Досин Д. Г., 2023

Головна мета статті – аналіз інтелектуальної системи кластеризації користувачів соціальних мереж на основі аналізу тональності даних. Основна мета цієї інтелектуальної системи – формування загального “образу” користувача системи за допомогою аналізу тональності даних соціальних мереж користувача та їх подальшої кластеризації. Спроектовано інтелектуальну систему, яка з використанням алгоритмів Identity та Access/Refresh JWT токенів забезпечує швидкі та максимально безпечні функції реєстрації, автентифікації та опрацювання різних сеансів користувачів системи. Описано основні підходи до здійснення аналізу тональності користувачьких повідомлень та інших даних різних типів, описано принципи реалізації LSTM рекурентної нейронної мережі, яка є дуже зручною для здійснення аналізу даних, оскільки добре працює і запам’ятовує контекст повідомлень у необхідні проміжки часу, завдяки чому збільшується фактор осмисленості аналізованих даних, відповідно до користувача інтелектуальної системи. Також описано загальні сучасні підходи до кластеризації та найдоцільніший алгоритм кластеризації – *k-means*, оскільки ми кожен раз працюватимемо з невизначеною кількістю даних, яка може істотно змінюватися залежно від кожного окремого користувача, відповідно через це буде змінюватися кількість кластерів і опрацювання даних. Завдяки цьому описано створення загального “образу” користувача інтелектуальної системи на основі комплексного аналізу, що дало змогу здійснювати дослідження користувачів і відображати відповідні результати.

Ключові слова: автентифікація; аналіз тональності даних; кластеризація; LSTM рекурентна нейронна мережа; алгоритм *k-means*.

Вступ

Створення інтелектуальної системи кластеризації користувачів соціальних мереж на основі аналізу тональності повідомлень є актуальним завданням, оскільки сьогодні більшість людей намагаються користуватися тими програмами та застосунками, які найоптимальніше працюють відповідно до вимог пошуку користувачів, тобто дають змогу за мінімальний час із використанням застосунків досягати максимального результату. Тема аналізу даних користувачів актуальна завжди, одна з основних властивостей людей – комунікації та взаємодія з іншими людьми, люди потребують спілкування з давніх давен. Людина – істота соціальна, соціалізація і взаємодія з іншими людьми є важливим аспектом її життя, під час спілкування людина розширяє коло інтересів, дізнається багато нового і цікавого, ділиться своїми знаннями з іншими.

Мета роботи – аналіз наявних систем-аналогів та пошук додаткової інформації в літературі, здійснення системного аналізу поточної інтелектуальної системи, вибір всіх необхідних програмних та апаратних засобів та створення інтелектуальної системи кластеризації користувачів різних соціальних мереж.

Варто зазначити, що ця мета доволі глобальна, та в межах однієї тільки мети важко викласти всю суть потрібного дослідження. Отже, необхідна декомпозиція мети на декілька менших завдань, які необхідно виконати для створення інтелектуальної системи. Розглядаючи алгоритмічну частину, варто відзначити два основні завдання: створення нейронної мережі аналізу тональності даних та реалізація алгоритму кластерного аналізу отриманих даних різного обсягу для вирішення основного завдання системи, а саме аналіз даних щодо кожного користувача, створення певного “образу” користувача, порівняння поточного користувача з іншими користувачами для подальшої взаємодії користувачів зі схожими інтересами та “образами”.

Завдання створення повноцінної нейронної мережі є серйозним, оскільки одним з важливих аспектів інформаційної системи є можлива взаємодія людей. Щоб спілкуватися між собою, люди повинні мати уявлення про тональність спілкування іншого користувача в його соціальних мережах. Щоб це реалізувати, необхідно створити LSTM нейронну мережу для аналізу тональності, яка зможе впоратися з такого типу завданнями, а саме з аналізом тональності користувацьких повідомлень та їх розподілом. Для таких завдань варто використати саме LSTM нейронну мережу, а не CNN, основна їх відмінність полягає у тому, що замість конкатенації вхідного та попереднього прихованого стану маємо різні вагові матриці, які застосовуються до обох перед тим, як передати їх до чотирьох внутрішніх нейронних мереж у комірці LSTM.

Завдання важливе, оскільки, окрім того, що нейронна мережа повинна повноцінно працювати, її робота має бути оптимальною, оскільки дуже істотною складовою сучасної інтелектуальної системи є швидкість відгуку на запити користувача системи. Відповідно потрібно намагатися оптимізувати процеси аналізу тональності тексту.

Також одне з основних завдань – реалізація алгоритму кластерного аналізу *k-means*, який є важливим для поділу отриманих даних про користувача на кластери. Кластеризація *k-means* – простий та елегантний підхід для поділу набору даних на K окремих кластерів, які не перекриваються. Щоб виконати кластеризацію *k-means*, ми повинні спочатку вказати бажану кількість кластерів K , тоді алгоритм *k-means* призначить кожне спостереження рівно одному з K кластерів. Кластеризація – це завдання такого поділу генеральної сукупності або точок даних на кілька груп, щоб точки даних у тих самих групах були більш схожими на інші точки даних у тій самій групі, ніж точки із інших груп. Кластеризація – це певний отриманий алгоритм навчання без учителя.

Звідси випливає ще одне важливе завдання: початкове опрацювання всіх даних, оскільки потрібно здійснити структурний поділ на групи, які надалі проаналізують створені алгоритми аналізу тональності. Також необхідно так реалізувати зберігання та опрацювання даних, щоб покращити роботу з БД, де усі наявні дані будуть зберігатися. Це потрібно, щоб не навантажувати роботу з БД складними запитами і пробувати зберігати використовувані дані локально всередині клієнта користувача і робити запити в БД, коли це справді потрібно, і отримувати лише ті дані, що необхідні в поточний момент. Оптимізація як така важлива, оскільки явно пришвидшить роботу системи і зробить її максимально зручною для користувача.

Оскільки завданням інтелектуальної системи, окрім аналізу тональності тексту, є і аналіз тональності інших даних користувача, то варто відзначити таке важливе завдання, як аналіз фотографій та інших медіафайлів користувача. Це необхідно для остаточної побудови та створення “образу” користувача за допомогою комплексного аналізу всієї доступної інформації. Насамперед варто описати згорткову нейронну мережу, за допомогою якої здійснюватиметься аналіз медіафайлів поточного користувача системи. Основне – проаналізувати обличчя користувача та загальний тон фотографії, для розуміння семантичного контексту. Аналізуючи обличчя користувача, можна дізнатися про його

настрій та додати отриману семантичну інформацію до загального контексту тональності повідомлень і даних користувача. Якщо говорити про збереження користувацьких фотографій, то для того, щоб не перевантажувати БД файлами великих розмірів, варто використати хмарне сховище медіафайлів. У наш час плата за використання хмарного сховища, яке є доволі зручним, невелика, а для доступу до медіафайлів достатньо зберігати в БД лише ідентифікатор та посилання на файл.

Об'єктом дослідження є важливий процес соціалізації користувачів. Одним із основних процесів є саме соціалізація і сьогодні більшість соціальних мереж намагаються максимально спростити й оптимізувати цей процес. Щоб змагатися з іншими популярними соціальними мережами, необхідно дослідити цей об'єкт і зрозуміти, як саме його покращити, для того щоб створювана інтелектуальна система успішно функціонувала і була зручною для користувачів. Адже дуже важливо, щоб і система була достатньо розподіленою для оптимізації всіх істотних для функціонування процесів, і користувачам було приємно нею користуватися і здійснювати соціалізацію так, щоб не виникало бажання якнайшвидше закрити її, а навпаки, її було легко і зручно використовувати протягом виділеного вільного часу.

Предмет дослідження в конкретному випадку – всередині, представляє частину об'єкта дослідження. Це користувач інтелектуальної системи. Оскільки поточна система створюється для кінцевого користувача, який відповідно використовуватиме її залежно від потреб, то потрібно звернути увагу на можливих користувачів систем і проаналізувати, як вони використовують такі системи, що саме людям подобається та що відштовхує у системах цього виду, як краще структурно здійснити реалізацію системи, щоб користувачу все було зрозуміло та щоб він міг просто знаходити інших людей, взаємодіяти з користувачами. Отже, знайдені користувачі повинні мати індекс схожості, що дасть можливість людям зі спільними інтересами шукати один одного та взаємодіяти між собою. Також всі користувачі, очевидно, зможуть створювати особисті публікації на відібрані теми і знаходити людей зі схожими інтересами, із якими можна ці інтереси обговорювати. Якщо оптимально продумати всі потрібні процеси, зростуть шанси, що системою користуватиметься предмет поточного дослідження, а саме користувач цієї системи. Що більше людей буде в системі, то вищі шанси, що відбудеться взаємодія.

Також інтелектуальна система має важливе значення в плані наукової новизни одержаних результатів, адже поки що просто не існує систем-аналогів для взаємодії користувачів, які виконували б повний аналіз тональності всіх користувацьких даних та формували кластери на підставі цього аналізу. Завдяки цьому можна буде зрозуміти, що являє собою користувач і в результаті сформулювати загальний “образ” користувача системи. Отже, вперше буде створено систему, яка дасть змогу максимально зрозуміти “образ”, середовище та інтереси користувача і вирішити, чи потрібно з ним взаємодіяти.

Аналіз останніх досліджень та публікацій

У статті [1] висвітлено розроблення інтелектуальної моделі прогнозування продажів транспортних засобів з використанням громадської думки в інтернеті та даних індексу онлайн-пошуку. Автори використали набір даних щомісячних продажів десяти основних брендів автомобілів у Китаї у 2012–2019 рр., а також дані про громадську думку в інтернеті від Sina Weibo та індекс онлайн-пошуку від Baidu. Автори застосували модель часових рядів і алгоритми машинного навчання для аналізу даних і прогнозування майбутніх продажів автомобілів, детально пояснили набір даних і різних моделей і алгоритмів, використаних в аналізі. Результати дослідження свідчать, що громадська думка в інтернеті та дані індексу онлайн-пошуку можуть бути корисними для прогнозування продажів транспортних засобів, а запропонована модель працює краще, ніж традиційні моделі часових рядів. Стаття [2] містить цінну інформацію про використання громадської думки та даних індексу онлайн-пошуку для прогнозування продажів транспортних засобів. Хоча дослідження має деякі недоліки, результати є цікавими та мають важливі наслідки для автомобільної промисловості, для використання онлайн-даних у прогнозованому моделюванні.

У статті [3] автори запропонували методiku оптимізації вебсторінок та підвищення їх рейтингу в пошукових системах на основі контент-аналізу ключових слів вебсторінок. Автори стверджують, що цей підхід може бути ефективним для збільшення видимості та трафіку вебсайтів, що, своєю чергою, може сприяти успіху онлайн-бізнесу. Загалом, стаття містить вичерпний огляд запропонованої методології, зокрема її теоретичні основи, упровадження та результати оцінювання. Автори надають чітко та стисло пояснення методології, що полегшує читачам розуміння та повторення підходу у власній роботі. Одним із потенційних недоліків статті є відсутність емпіричних доказів на підтримку ефективності запропонованої методології.

У статті [4] автори висвітлюють тематичні дослідження, проілюстрували застосування підходу, використавши більше емпіричних доказів, щоб продемонструвати ефективність методології для поліпшення рейтингу вебсайта та залучення трафіку. Недоліком статті є те, що автори зосереджуються насамперед на технічних аспектах методології, нехтуючи іншими факторами, які можуть сприяти успіху онлайн-бізнесу, такими як досвід користувача, брендинг і залучення соціальних мереж. Повніший аналіз цих факторів дав би читачам повніше уявлення про виклики та можливості просування бізнесу в інтернеті. Стаття є корисним внеском у сферу просування бізнесу в інтернеті, пропонуючи практичний та інноваційний підхід до поліпшення рейтингу вебсайтів і збільшення трафіку. З подальшими дослідженнями та вдосконаленнями запропонована методологія може стати цінним інструментом для онлайн-бізнесу. В статті [5] автори аналізують вплив розподілу кількості функцій у навчанні моделей глибокого навчання, досліджують вплив розподілу кількості функцій на цифрові стереотипи в дизайні машинного інтерфейсу (МІ). Дослідники аналізують, як поточні моделі глибокого навчання можуть опрацювати цифрові стереотипи на основі розподілу функцій, використовуваних для навчання.

Стаття [6] містить вичерпний вступ до теми цифрових стереотипів у дизайні НМІ та ролі моделей глибокого навчання у вивченні цих стереотипів. Автори виконують експерименти з використанням різних розподілів кількості ознак і вимірюють отримані цифрові стереотипи в дизайні НМІ. Результати показують, що розподіл кількості функцій істотно впливає на збереження цифрових стереотипів і що існує потреба в різноманітніших наборах даних, щоб зменшити упередженість під час проектування НМІ. Загалом, стаття містить цінну інформацію про роль розподілу кількості ознак у цифрових стереотипах для розроблення НМІ. Методологія дослідження продумана, а результати подано чітко та ефективно. Автори також надають продумані рекомендації щодо зменшення упередженості в проектуванні НМІ за рахунок використання різноманітніших наборів даних.

Автори статті [7] дослідили ставлення споживачів до продакт-плейсменту у фільмах і вплив його на їхню поведінку. Автори провели експеримент із 200 учасниками, яким показали фільм із продакт-плейсментом або без нього, а потім запитали про їх поведінку та ставлення до пропонованих продуктів. Виявилось, що учасники, які зазнали продакт-плейсменту, позитивніше ставились до продуктів і, швидше за все, брали участь у купівлі, пов'язаній з ними. Автори також виявили, що релевантність і ознайомлення із продуктом відіграють роль у визначенні ефективності розміщення продукту. Загалом стаття надає корисну інформацію про вплив продакт-плейсменту на ставлення та поведінку споживачів. Однак дослідження має деякі недоліки, такі як невеликий розмір вибірки та використання одного фільму для експерименту. Крім того, дослідження [8] зосереджено лише на короткострокових ефектах і не враховувало довгострокового впливу продакт-плейсменту на поведінку споживачів. Незважаючи на ці недоліки, отримані результати можуть допомогти маркетологам і рекламодавцям зрозуміти, як оптимізувати продакт-плейсмент для ефективного націлювання на аудиторію, підвищити впізнаваність бренду та продажі.

У статті [9] автори подали вичерпний огляд досліджень настроїв інвесторів за останні кілька десятиліть. Вони виконали бібліометричний аналіз 630 статей, опублікованих у 1988–2020 рр., щоб визначити ключові теми та тенденції в цій галузі. Автори визначили три основні теми дослідження

настроїв інвесторів: вимірювання та моделювання настроїв інвесторів, вплив настроїв інвесторів на фінансові ринки та наслідки настроїв інвесторів. У межах кожної з цих тем визначено підтеми, такі як використання соціальних мереж для вимірювання настроїв інвесторів і роль настроїв інвесторів у ціноутворенні активів.

Стаття [10] містить огляд досліджень настроїв інвесторів і висвітлює поточний стан знань у цій галузі. Ретельний бібліометричний аналіз дає змогу визначити найвпливовіші статті та авторів у цій галузі. У статті також підкреслено необхідність додаткових досліджень передумов і наслідків настроїв інвесторів, які можуть допомогти інвесторам приймати кращі рішення та зменшити волатильність ринку. Одним з недоліків дослідження є те, що воно охоплює лише статті, опубліковані в англомовних журналах, що може обмежити узагальнення результатів. Крім того, дослідження охоплює лише статті, опубліковані до 2020 р., новіші дослідження можуть не входити в аналіз. Загалом у статті подано цінний огляд досліджень настроїв інвесторів і визначено напрями майбутніх досліджень. Бібліометричний аналіз є корисним інструментом для визначення ключових статей і авторів у цій галузі та може допомогти дослідникам та інвесторам дізнатись про останні події у цій галузі.

У статті [11] досліджено вплив електронного радіомовлення (e-WOM) на наміри споживачів щодо купівлі у результаті розширення сприйняття інформації. Дослідження має на меті забезпечити всебічне розуміння факторів, які впливають на прийняття повідомлень e-WOM у рішеннях споживачів про покупку. Автори використали метод опитування, щоб зібрати дані від 388 респондентів у Тайвані, та проаналізували дані за допомогою моделювання структурних рівнянь. Отримані дані свідчать про те, що на намір споживачів прийняти повідомлення e-WOM впливають кілька факторів, серед яких сприйнята корисність, сприйнята легкість використання, довіра, соціальний вплив і особиста інноваційність. Дослідження також виявило, що вплив повідомлень e-WOM на наміри щодо покупки залежить від ставлення споживачів до повідомлень.

Стаття [12] робить вагомий внесок у літературу про e-WOM, розширяючи модель сприйняття інформації та досліджуючи фактори, які впливають на прийняття споживачами повідомлень e-WOM. Отримані результати мають практичне значення для маркетологів, які прагнуть використовувати соціальні мережі для впливу на рішення споживачів про покупку. Однак дослідження обмежене зосередженням на конкретному географічному регіоні – Тайвані. Узагальнення результатів на інші регіони та культури може бути обмеженим. Крім того, дослідники не вивчали вміст повідомлень e-WOM, який може вплинути на їхню ефективність у прийнятті рішень про покупку. Загалом стаття містить цінну інформацію про роль e-WOM у прийнятті споживачами рішень про покупку та пропонує корисні винувки для маркетологів, які прагнуть використовувати соціальні мережі для впливу на поведінку споживачів.

Стаття [13] спрямована на визначення ключових факторів успіху стартапів за допомогою аналізу настроїв відгуків користувачів. Автори стверджують, що традиційні методи визначення факторів успіху часто обмежені доступністю та якістю даних, а аналіз настроїв може забезпечити об'єктивніший і комплексніший підхід. У статті наведено детальну методологію для аналізування настроїв щодо відгуків користувачів, яку використано для аналізу відгуків про стартапи в індустрії замовлення поїздок. Автори виділяють кілька ключових факторів успіху, зокрема ціни, зручність і обслуговування клієнтів. Вони також надають рекомендації для стартапів на основі своїх висновків. Використання аналізу настроїв [14] дає цінну перспективу щодо факторів успіху, які неможливо охопити традиційними методами. Автори також надають корисну інформацію та рекомендації для стартапів на підставі своїх висновків. Одним із недоліків статті є те, що аналіз зосереджений лише на індустрії послуг поїздок, тому незрозуміло, наскільки результати можна узагальнити на інші галузі. Крім того, дослідники аналізують лише відгуки користувачів з однієї платформи, тому результати можуть бути

упередженими щодо користувачів цієї конкретної платформи. Також автори статті [15] зробили цінний внесок у літературу щодо визначення факторів успіху стартапів за допомогою аналізу настроїв. Пропоновані методологія та висновки добре реалізовані та надають корисні відомості та рекомендації для стартапів у галузі послуг автівок.

Стаття [16] містить огляд поточного стану систем оцінювання мобільних застосунків та їх ефективності. Автори провели опитування рейтингових систем і проаналізували їх ефективність у точному відображенні якості мобільних застосунків. У статті підкреслено важливість рейтингових систем на ринку мобільних застосунків, оскільки вони можуть істотно вплинути на успіх або невдачу програми. Автори в статті [17] визначили кілька проблем, пов'язаних із наявними рейтинговими системами, зокрема можливість упереджених або фальшивих рейтингів, відсутність стандартизації на платформах і вплив інфляції рейтингів. Щоб вирішити ці проблеми, автори пропонують кілька рекомендацій щодо вдосконалення систем оцінювання мобільних застосунків, зокрема розроблення складніших алгоритмів для виявлення фальшивих або упереджених оцінок, встановлення стандартизованої шкали оцінювання для різних платформ і введення до рейтингу більшої кількості контекстної [18] інформації. Загалом стаття містить цінний аналіз поточного стану систем оцінювання мобільних застосунків і проблем, з якими вони стикаються. Рекомендації авторів щодо вдосконалення цих систем можуть мати важливі наслідки як для розробників програм, так і для споживачів.

Стаття [19] є дослідницькою роботою, яка вивчає ефективність різних підходів до виявлення сарказму в даних соціальних мереж. Автори використовують методи машинного навчання, щоб класифікувати твіти як саркастичні чи несаркастичні, порівнюють ефективність моделей, навчених із хештегами, що вказують на сарказм, і без них. Стаття [20] містить чіткий і детальний огляд методології та результатів дослідження. Автори аналізують сарказм та його роль у соціальних мережах, а потім описують збирання даних і методи попереднього опрацювання, які використовують для підготовки даних до аналізу.

Автори в статті [21] порівнюють продуктивність різних моделей машинного навчання, зокрема логістичну регресію, машини опорних векторів і нейронні мережі, у виявленні сарказму як з використанням хештегів сарказму, так і без нього. Результати показують, що використання хештегів сарказму істотно поліпшує продуктивність моделей виявлення сарказму, і що найефективнішою моделлю є нейронна мережа, яка містить як функції хештегу, так і текстові функції. Автори також вказують на недоліки свого дослідження, зокрема той факт, що їхній набір даних складається тільки з англійських твітів, і пропонують напрями майбутніх досліджень. Стаття [22] робить внесок у сферу опрацювання природної мови та допомагає зрозуміти проблеми і можливості виявлення сарказму в даних соціальних мереж. Результати дослідження можуть бути корисними для розроблення інструментів аналізу настроїв, платформ соціальних мереж та інших застосунків, які покладаються на точну інтерпретацію текстових даних.

У статті [23] досліджено можливість використання даних Twitter у реальному часі для прогнозування рухів фондового ринку. Автори запропонували новий метод, який поєднує аналіз настроїв із методами машинного навчання для аналізу твітів, пов'язаних із фондовим ринком. У дослідженні використано набір даних із понад 200 000 твітів, пов'язаних із фондовим ринком, і застосовано різні алгоритми машинного навчання, серед них випадковий ліс, логістичну регресію та машину опорних векторів, щоб передбачити рух фондового ринку.

Автори статті [24] доводять, що запропонований метод аналізу даних перевершує традиційні методи, такі як Buy-and-Hold і Moving Average, з погляду точності. Автори також обговорили недоліки свого дослідження, такі як якість даних Twitter і припущення, зроблені під час аналізу настроїв. Вони припустили, що майбутні дослідження мають бути зосереджені на поліпшенні якості даних і розробленні точніших методів аналізу настроїв. Стаття [25] містить інформацію про потенціал використання даних соціальних мереж для прогнозування рухів фондового ринку. Результати

дослідження мають важливе значення для інвесторів, фінансових аналітиків та інших зацікавлених сторін, які покладаються на точні прогнози руху фондового ринку. Проте необхідні подальші дослідження, щоб усунути недоліки та підвищити точність запропонованого методу. У статті [26] запропоновано новий підхід до аналізу настроїв на основі аспектів (ABSA) з використанням згорткових мереж на графах (GCN) і механізму воріт частини мови (POS). Автори починають із того, що підкреслюють важливість ABSA для розуміння настроїв [27], висловлених щодо різних аспектів продукту чи послуги, що може бути корисним для компаній, які прагнуть поліпшити свої пропозиції. Потім вони пропонують підхід, який передбачає побудову графа вхідного тексту, де вузли представляють слова, а ребра – залежності.

Формулювання цілі статті

Головні завдання системи – аналіз тональності даних користувачів соціальних мереж, їх кластеризація та подальше створення “образу” користувача. Це завдання є важливим та концептуально комплексним, оскільки необхідно здійснити аналіз не тільки текстових даних користувача, а всієї можливої інформації, тобто профілю користувача, його повідомлень, медіафайлів та інших даних. Важливим завданням є опрацювання геолокації користувача для подальшого можливого визначення його суспільної ролі та соціального середовища. Тобто для нас важливим завданням є аналіз тональності всієї наявної інформації про користувача, завдяки чому можна максимально точно створити загальну картину та “образ” користувача системи. В цьому випадку потрібно реалізувати одразу кілька алгоритмів аналізу тональності даних: це аналіз тональності текстових даних відповідно до мови користувача, також аналіз його фото, відео та інших медіафайлів, що потребує використання різних підходів для реалізації декількох нейронних мереж та їх об’єднання у межах однієї системи. Сформувавши результати аналізу тональності, необхідно кластеризувати отримані дані методом *k-means*, завдяки чому можна визначити загальні кластери щодо користувацьких даних і, відповідно, підсумувати кінцевий “образ” користувача реалізованої інтелектуальної системи.

Варто підкреслити, що система повинна складатися з декількох основних кроків. По-перше, це авторизація та автентифікація користувача в сучасних популярних соціальних мережах, таких як Twitter, Facebook та Reddit. Також необхідно зчитати коментарі та публікації користувача за певний проміжок часу, аналіз отриманих та збережених даних потрібно виконувати за допомогою алгоритму аналізу тональності повідомлень і відповідно навченої нейронної мережі. Проаналізувавши повідомлення, необхідно здійснити поділ на кластери за допомогою кластерного аналізу та алгоритму *k-середніх*, після чого сформувати результати у вигляді тексту та діаграм і продемонструвати поточному користувачу.

У сучасних системах багато завдань є розподіленими і виконуються різними системними інтерфейсами. Це ідеально для асинхронної реалізації інтелектуальної системи, оскільки безліч завдань будуть виконуватися на стороні з використанням сторонніх бібліотек, або інтерфейсів, а частину завдань можна буде розподілити за допомогою паралельних обчислень, що дасть змогу додатково поліпшити технологію. Також важливо згадати про реалізацію клієнт-серверної моделі системи, що дасть можливість чітко розподілити роботу серверної програми, яка виконує основні обчислення, опрацювання даних, приймає та відповідає на запити і клієнтської частини, яка є важливою під час взаємодії з користувачем. Правильна реалізація цієї моделі дасть змогу користувачеві зручно використовувати нескладну технологію, завдяки зрозумілому дизайну системи і оптимальній реалізації основного доступного для користувача функціоналу. За допомогою клієнт-програми зможуть вносити зміни та керувати системою користувачі, яким доступно більше можливостей, такі як модератор чи адміністратор. На рис. 1 зображено концептуальну блок-схему роботи описаної системи.

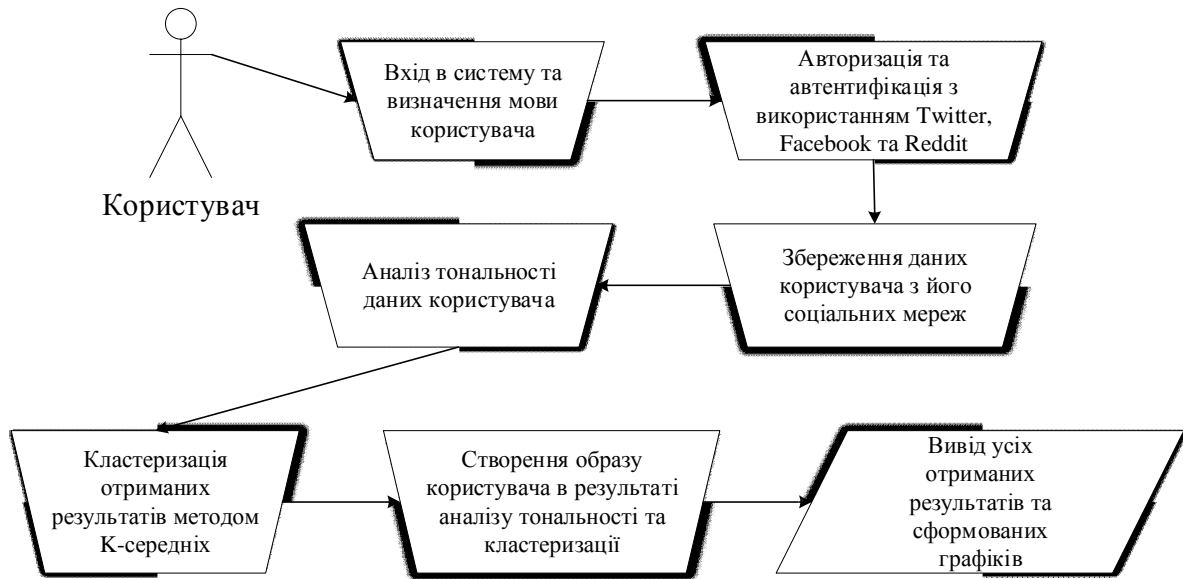


Рис. 1. Концептуальна блок-схема роботи системи

Виклад основного матеріалу

Для того щоб деталізувати структуру створюваної системи, вибрано функціональні діаграми IDEF0. Відповідно для того, щоб узагальнено відобразити головні функції розроблюваної інтелектуальної системи, створено контекстну діаграму, яка містить різні основні вхідні та вихідні дані, а також основні дані для відображення механізмів та контролю. До вхідних даних належать: “Логін користувача” та “Мова користувача”, “База даних”, “Соціальні мережі користувача” і “Запит користувача”. До вихідних – “Образ користувача” та “Оновлена база даних”. Серед основних моментів контролю варто виділити “Рівні доступу” і “Вимоги”, що впливають на поточні процеси інтелектуальної системи. За виконання основного функціоналу системи відповідають механізми системи, а саме: “Алгоритми формування образу користувача”, “Алгоритми кластерного аналізу”, “Алгоритми аналізу тональності повідомлень” та “Методи реєстрації та авторизації”. На рис. 2 зображено вузол створеної контекстної діаграми проектованої інтелектуальної системи.

Node: A-0

Аналіз тональності повідомлень і кластеризація користувачів

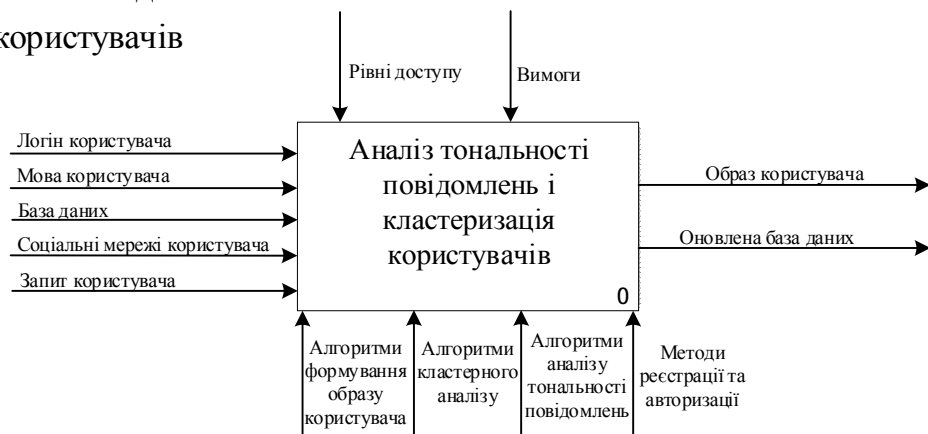


Рис. 2. Контекстна діаграма інтелектуальної системи

Наступним кроком після побудови контекстної діаграми інтелектуальної системи є створення діаграми декомпозиції, яка розкриває внутрішню структуру інтелектуальної системи. Як бачимо,

система складається з чотирьох основних блоків, а саме “Ввійти в систему”, “Проаналізувати тональність повідомлень”, “Здійснити кластеризацію отриманих даних” та “Сформувати образ користувача, вивести і зберегти результати”. Виконавши процес “Ввійти в систему”, ми отримуємо авторизованого користувача, в результаті процесу “Проаналізувати тональність повідомлень” – проаналізовані дані, в результаті процесу “Здійснити кластеризацію отриманих даних” одержуємо результат кластерного аналізу і в результаті процесу “Сформувати образ користувача, вивести і зберегти результати” на виході отримуємо “Образ користувача” і “Оновлена база даних”. Також зі всіма блоками взаємодіють вхідні моменти “Логін користувача” та “Мова користувача”, “База даних”, “Соціальні мережі користувача” і “Запит користувача”, моменти контролю “Рівні доступу” і “Вимоги” та механізми, а саме: “Алгоритми формування образу користувача”, “Алгоритми кластерного аналізу”, “Алгоритми аналізу тональності повідомлень” та “Методи реєстрації та авторизації”. На рис. 3 зображено декомпозицію контекстної діаграми проектованої інтелектуальної системи.

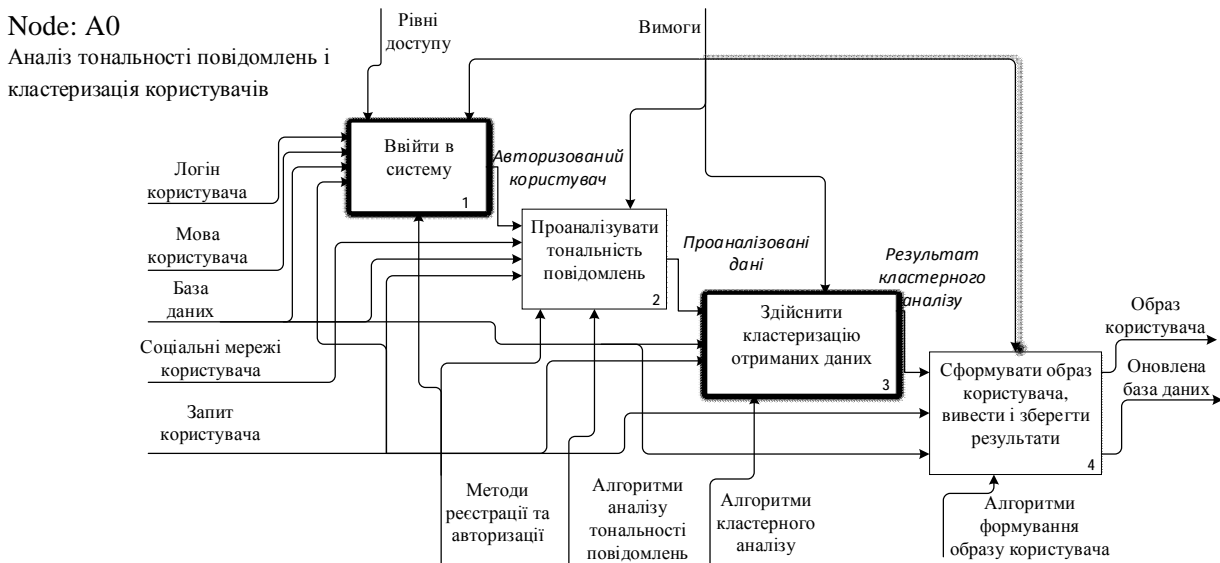


Рис. 3. Декомпозиція інтелектуальної системи

На рис. 4–7 подано декомпозицію основних блоків інтелектуальної системи.

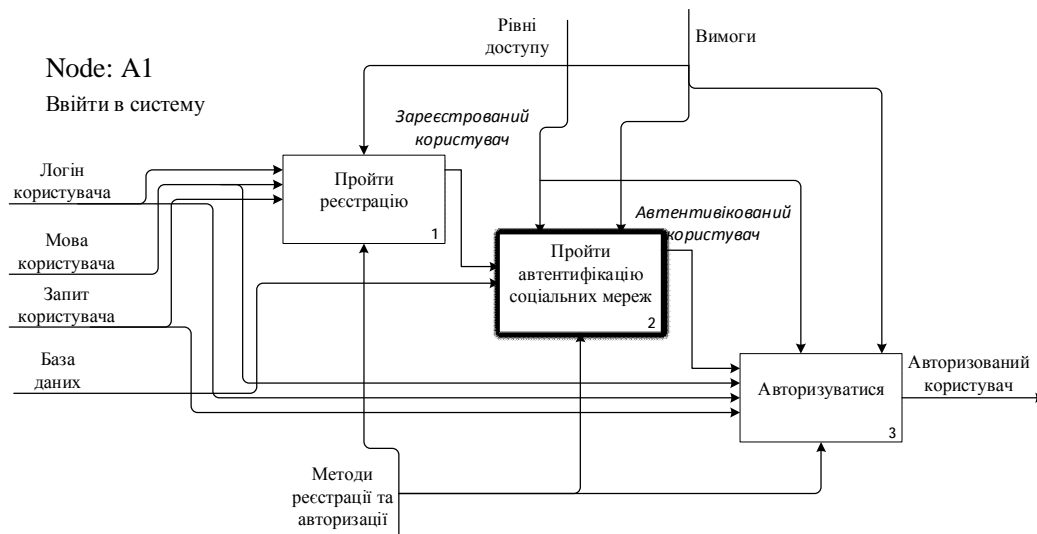


Рис. 4. Декомпозиція блока “Ввійти в систему”

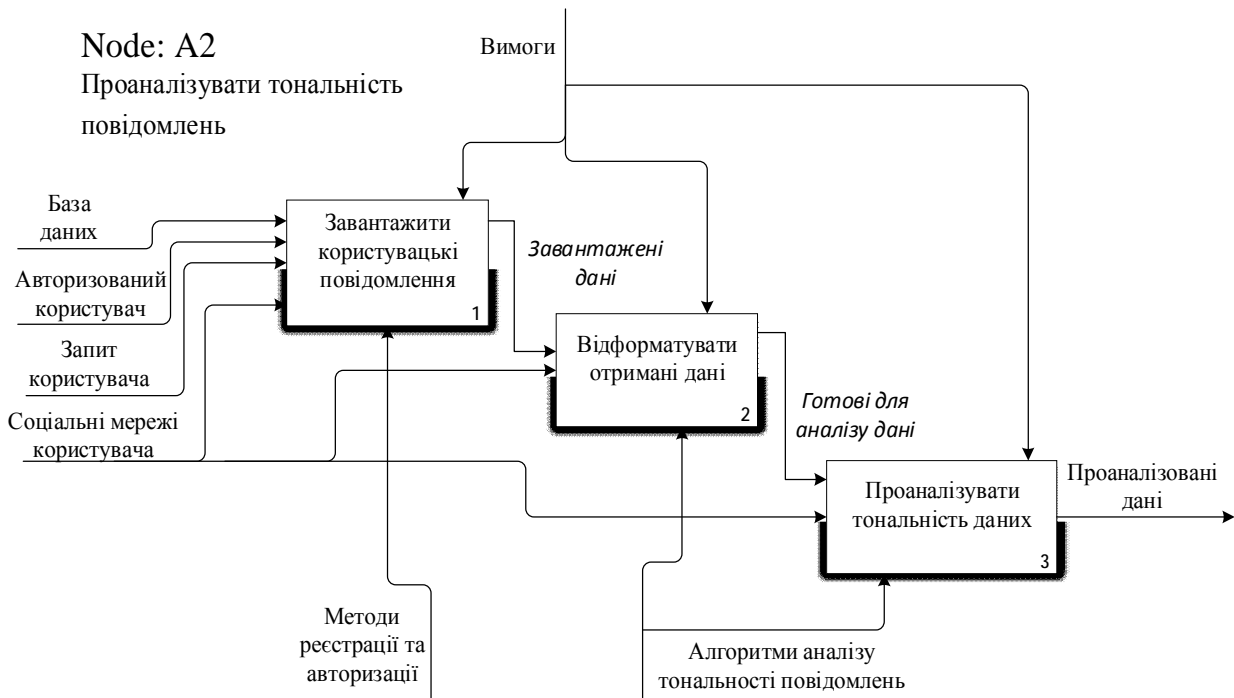


Рис. 5. Декомпозиція блока “Проаналізувати тональність повідомлень”

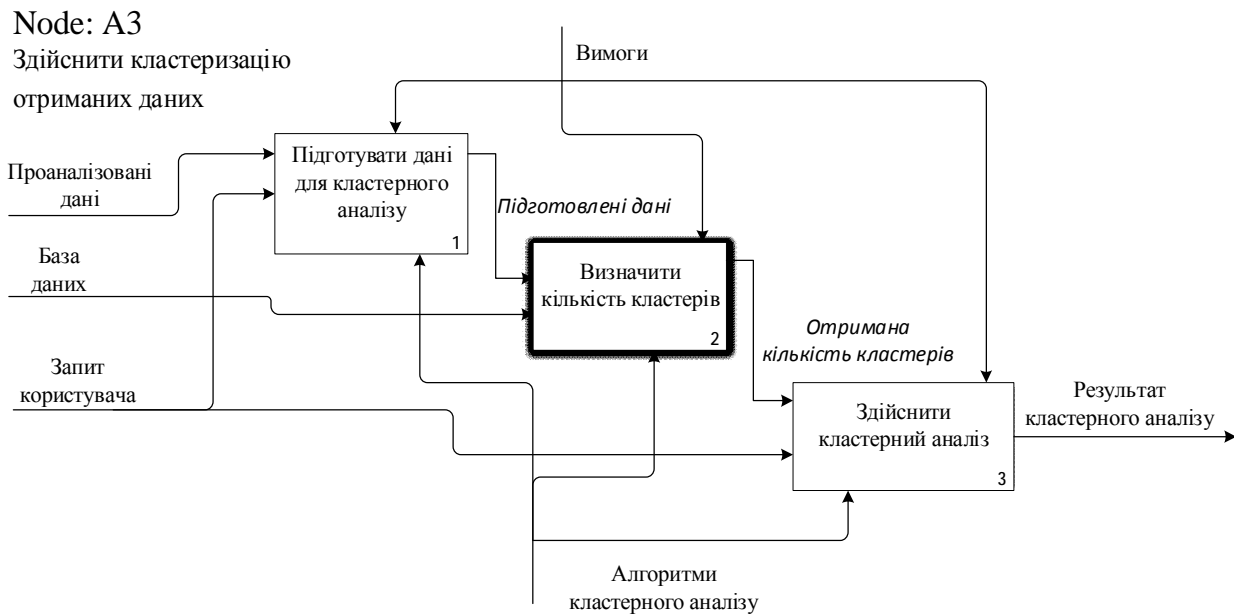


Рис. 6. Декомпозиція блока “Здійснити кластеризацію отриманих даних”

На рис. 8 зображено ієрархію процесів інтелектуальної системи.

Спочатку варто розглянути найузагальненішу діаграму, а саме діаграму варіантів використання, яка визначає загальні вимоги до функціонального забезпечення інтелектуальної системи і формує вихідну концепцію системи для того, щоб згодом деталізувати її в інших формах, які будуть необхідні для поточної системи. Завдяки діаграмі варіантів використання можна загалом висвітлити дійових осіб поточної системи та відобразити їхню взаємодію за допомогою варіантів використання.

Node: A4

Сформувати образ користувача,
вивести і зберегти результати

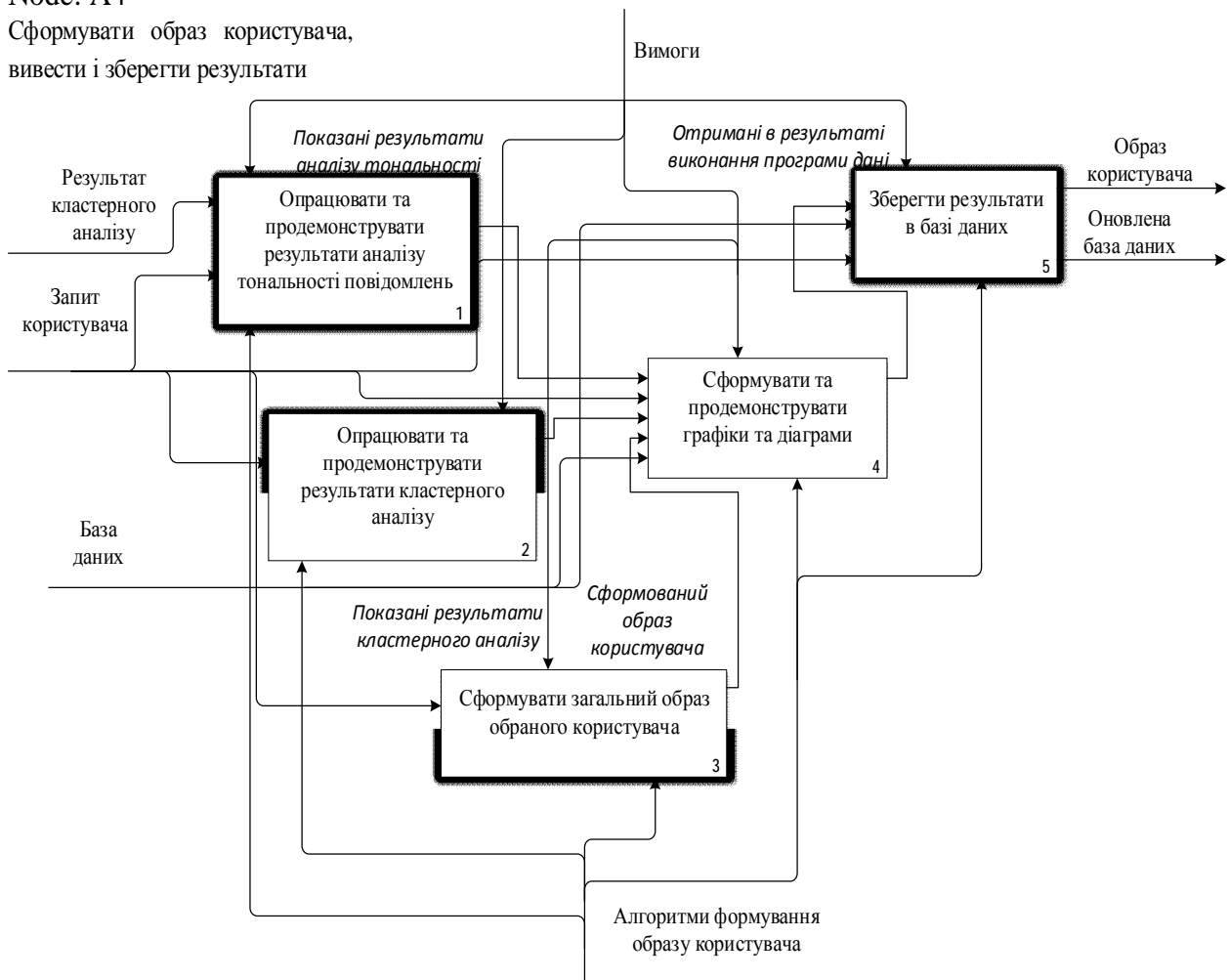


Рис. 7. Декомпозиція блока "Вивести і зберегти результати"

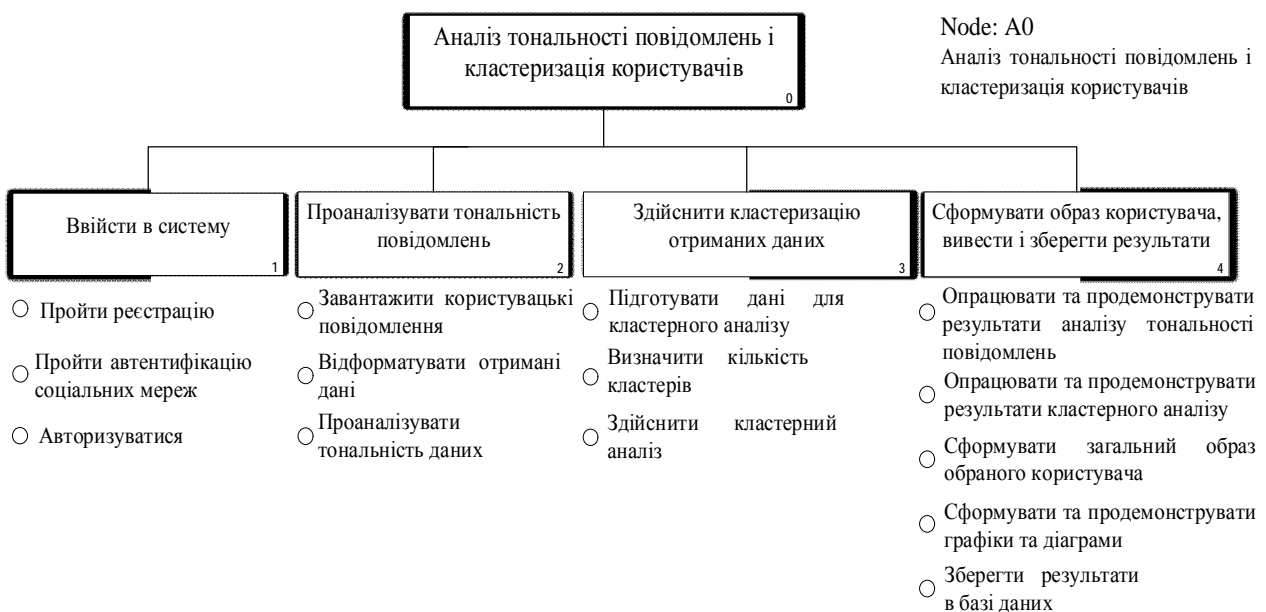


Рис. 8. Ієрархія процесів інтелектуальної системи

Система має одну дійову особу “Користувач”, яка поєднується з варіантами використання “Пройти реєстрацію” і “Авторизуватися” за допомогою відношення асоціації. Якщо розглядати утворену діаграму, то бачимо, що система складається із певних послідовних варіантів використання, які поєднані відношенням включення, а саме “Відформатувати отримані дані”, “Проаналізувати тональність даних”, “Здійснити кластеризацію даних”, “Вивести і зберегти результати”, “Вийти на головну сторінку” та “Вийти з програми”. Також варто звернути увагу на наявний у системі інтерфейс “Форма зберігання даних”, який використовують із варіантами використання “Зберегти параметри у базі даних” та “Зберегти результати в базі даних”. На рис. 9 наведено діаграму варіантів використання.

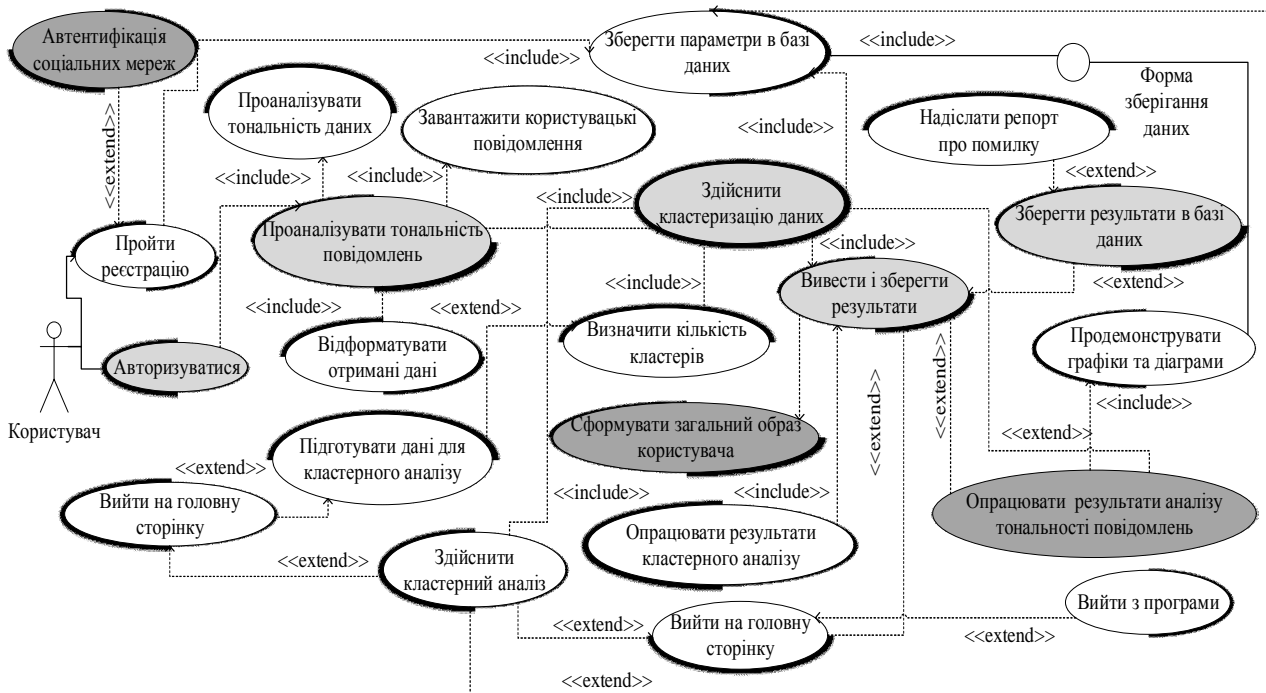


Рис. 9. Діаграма варіантів використання інтелектуальної системи

Діаграму діяльності вважають альтернативою діаграмі станів. Її головною властивістю є те, що вона окремо залежить від часових проміжків, тобто складніша за діаграму станів, а також робить основний акцент не на сам по собі стан, а на стан дії, тобто вся увага зосереджується на дії, яка виконується в межах певного стану. Ця дія може бути розпаралелена та виконана за наявності певних додаткових умов. Діаграма має початковий і кінцевий стани і може використовувати стани всередині системи, або прості, або складені, якщо є необхідність. Особливість створеної діаграми в тому, що є чіткий поділ на об’єкти системи, кожен об’єкт виконує чітку функцію всередині системи і можна побачити явні переходи між станами дій різних системних об’єктів, що порівняно зі звичайною діаграмою станів важливіше для розуміння того, з чого складається система і які функції виконує.

У системі є основні об’єкти – “Користувач”, “Сервер”, “База даних”, вони всі взаємодіють між собою за допомогою станів дій. На діаграмі зображено такі стани дій, як “Ввійти в систему”, у разі виконання умови вибирають “Авторизуватися” чи “Зареєструватися”, “Відправити сформовані дані”, циклічне “Надіслати сформовані дані в БД”, “Завантажити користувацькі повідомлення”, “Відформатувати отримані дані”, “Проаналізувати тональність даних”, “Підготувати дані для кластерного аналізу”, “Визначити кількість кластерів”, “Здійснити кластеризацію даних”, що розпаралелюється

на “Опрацювати результати аналізу тональності повідомлень” та “Зберегти персональну інформацію”, яка за певної умови переходить у “Надіслати сформовані дані в БД”, “Змінити персональну інформацію”, “Завантажити нові користувацькі повідомлення”, “Змінити фактор кластеризації”, “Обробити нові дані”, “Додати дані до вже існуючих”, “Опрацювати результати кластерного аналізу”, “Сформувати отримані дані”. Потім йдуть стани дії “Перевірити отримані з БД дані”, “Сформувати загальний образ вибраного користувача”, “Сформувати та продемонструвати графіки та діаграми”, “Коректно завершити роботу системи” і врешті робота користувача завершується кінцевим станом системи. На діаграмі чітко видно всі три об’єкти системи та розподіл завдань відповідно до кожної з них. Кожен стан дії має свій перехід між об’єктами, завдяки чому дані в JSON форматі відправляються як в тілі запиту, так і як відповідь на надісланий запит, який може бути як і синхронним, так і, здебільшого, асинхронним. На рис. 10 наведено діаграму діяльності.

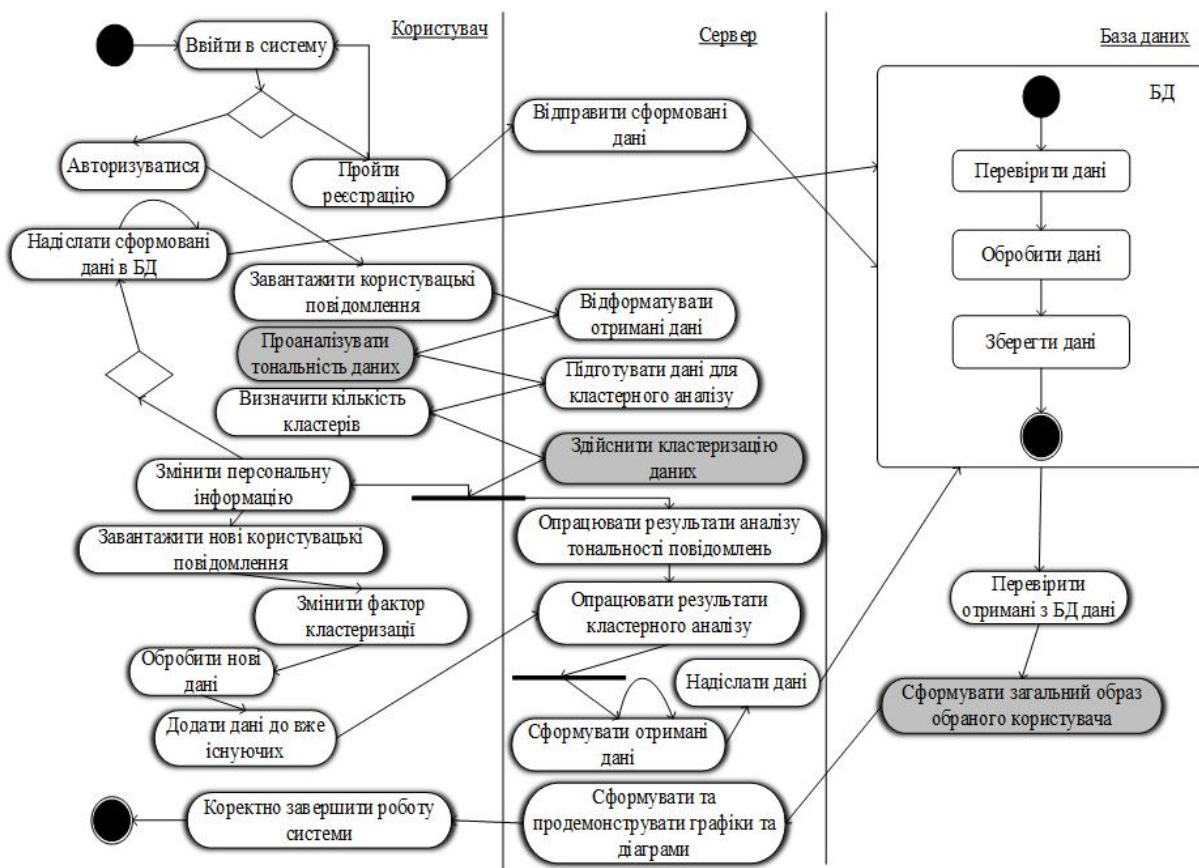


Рис. 10. Діаграма діяльності інтелектуальної системи

На рис. 11 подано діаграму послідовності, яка містить основні об’єкти системи – “Користувач”, “Сервер”, “База даних”, для яких вказано часові особливості передавання і приймання повідомлень.

Отже, маючи діаграми варіантів використання, діяльності та послідовності, можна не лише загалом зрозуміти функції та бізнес-процеси інтелектуальної системи, а й уявити повноцінну структуру системи як об’єктно-орієнтованого програмного забезпечення. Завершивши концептуальне проектування, ми отримали уявлення про всі необхідні системні процеси, активності, дії, компоненти, стани та її акторів, а отже, маємо всі початкові елементи для майбутньої реалізації інтелектуальної системи.

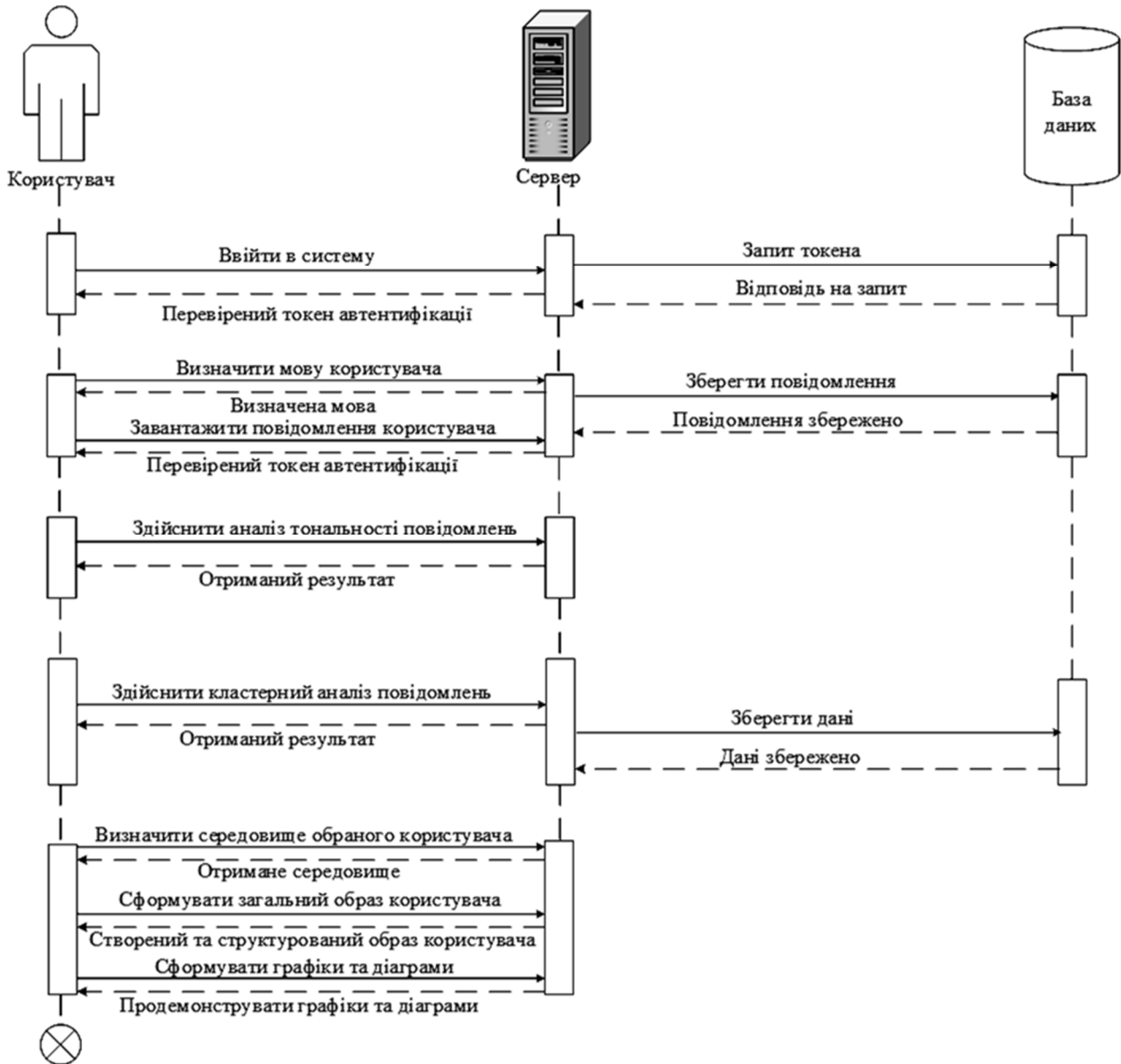


Рис. 11. Діаграма послідовності інтелектуальної системи

Висновки

В ході роботи проаналізовано вже виконані наукові роботи, в яких, відповідно, створено та описано готові рішення щодо роботи із сучасними популярними соціальними мережами, особливості аналізу користувацьких профілів та повідомлень, на основі яких визначено, які методи та засоби, призначені для опрацювання даних, застарілі та неефективні, а які варто використовувати. Також досліджено вже наявні реалізації алгоритмів аналізу тональності даних та особливості кластеризації результатів, проаналізовано нейронні *long short-term memory* (LSTM) мережі та *k-means* алгоритм кластеризації. Проаналізовано алгоритми роботи з користувацькими фотографіями та різні підходи до опрацювання як текстових, так і медіаданих, тобто згорткові нейронні мережі та алгоритми опрацювання природної мови. Під час системного аналізу створено блок-схему, яка концептуально пояснює роботу системи, також побудовано IDEF0 діаграми, які описують функціональну роботу інтелектуальної системи і її основних бізнес-процесів. Сформовано основну ієрархію процесів інтелектуальної системи.

Спроектвано основи подальшої реалізації інтелектуальної системи кластеризації користувачів соціальних мереж на основі аналізу тональності даних за допомогою UML-діаграм. Побудовано три основні об'єктно-орієнтовані діаграми, а саме діаграму варіантів використання, діаграму активності та діаграму послідовності інтелектуальної системи. Висвітлено основні етапи та стани роботи системи, особливості авторизації та автентифікації користувача соціальних мереж, принципи роботи алгоритмів аналізу тональності даних та особливості їх практичної реалізації. В результаті створено та показано особливості реалізації нейронної мережі LSTM для визначення тональності даних та алгоритму кластеризації користувачів *k-means* разом із принципами формування і подання кінцевої інформації користувачу. Отриманий результат є теоретичним та концептуальним базисом, на основі якого здійснюватиметься подальша практична реалізація інтелектуальної системи з урахуванням визначених алгоритмів, методів та засобів.

Список літератури

1. Zhang M., Xu H., Ma N., Pan X. (2022). Intelligent Vehicle Sales Prediction Based on Online Public Opinion and Online Search Index. *Sustainability*, No. 14 (16), 344–361. DOI: <https://doi.org/10.3390/su141610344>.
2. Vysotska V. (2021). Information Technology for Internet Resources Promotion in Search Systems Based on Content Analysis of Web-Page Keywords. *Radio Electronics, Computer Science, Control*, No. 3, 133–151.
3. Antonowicz P., Podpora M., Rut J. (2022). Digital Stereotypes in HMI – The Influence of Feature Quantity Distribution in Deep Learning Models Training. *Sensors*, No. 22 (18), 673–689. DOI: <https://doi.org/10.3390/s22186739>.
4. De-Gregorio F., Sung Y. (2010). Understanding attitudes toward and behaviors in response to product placement. *Journal of Advertising*, No. 39 (1), 83–96. DOI: <http://doi.org/10.2753/JOA0091-3367390106>.
5. Kamath A. N., Shenoy S., Subrahmanya K. N. (2022). An overview of investor sentiment: Identifying themes, trends, and future direction through bibliometric analysis. *Investment Management & Financial Innovations*, No. 19 (3), 229–242. DOI: [https://doi.org/10.21511/imfi.19\(3\).2022.19](https://doi.org/10.21511/imfi.19(3).2022.19).
6. Erkan I. (2016). The influence of e-WOM in social media on consumers' purchase intentions: An extended approach to information adoption. *Computers in Human Behavior*, No. 4, 47–55.
7. Asgari T., Daneshvar A., Chobar A. P., Ebrahimi M., Abrahamyan S. (2022). Identifying key success factors for startups With sentiment analysis using text data mining. *International journal of Engineering Business Management*, No. 14, 435–453. DOI: <https://doi.org/10.1177/18479790221131612>.
8. Gao L. (2014). Online consumer behavior and its relationship to website atmospheric induced flow: Insights into online travel agencies in China. *Journal of Retailing and Consumer Services*, No. 21 (4), 653–655.
9. Abulhajja S., Hattab S., Abdeen A., Etaiwi W. (2022). Mobile Applications Rating Performance: A Survey. *International journal of Interactive Mobile Technologies*, No. 16 (19), 133–146. DOI: <https://doi.org/10.3991/ijim.v16i19.32051>.
10. Guidry J. D., Messner M., Jin Y. (2015). From McDonalds fail to Dominos sucks: An analysis of Instagram images about the 10 largest fast food companies. *Corporate Communications: An International Journal*, No. 20 (3), 344–359.
11. Bagate R. A., Suguna R. (2022). Sarcasm Detection with and without #Sarcasm: Data Science Approach. *International journal of Information Science and Management*, No. 20 (4), 1–15.
12. Salganik M. (2019). Social Research in the Digital Age. *Journal of Interactive Marketing*, No. 2 (9), 345–358.
13. Li Q., Li X., Du Y., Fan Y., Chen X. (2022). A New Sentiment-Enhanced Word Embedding Method for Sentiment Analysis. *Applied Sciences*, No. 12 (20), 712–725. DOI: <https://doi.org/10.3390/app122010236>.
14. Jeff M., Jennifer R., Catherine J., Elke P. (2014). Managing brand presence through social media: The case of UK football clubs. *Internet Research*, No. 24 (2), 181–204.
15. Opila J. (2022). On Employing of Extended Characteristic Surface Model for Forecasting of Demand in Tourism. *Interdisciplinary description of Complex Systems*, No. 20 (5), 621–639. DOI: <https://doi.org/10.7906/indecs.20.5.8>.

16. Kudeshia C., Sikdar P., Mittal A. (2016). Spreading love through fan page liking: A perspective on small scale entrepreneurs. *Computers in Human Behavior*, No. 8 (19), 257–270. DOI: <http://doi.org/10.1016/j.chb.2015.08.003>.
17. Albahli S., Irtaza A., Nazir T., Mehmood A., Ali A., Waleed Albattah W. (2022). A Machine Learning Method for Prediction of Stock Market Using Real-Time Twitter Data. *Electronics*, No. 11 (20), 341–363. DOI: <https://doi.org/10.3390/electronics11203414>.
18. Mousavijad M. (2017). The effect of socialization factors on decision making of teenagers consumers in schools. *Journal of School Administration*, No. 5 (1), 217–234.
19. Kim D., Kim Y., Jeong Y.-S. (2022). Graph Convolutional Networks with POS Gate for Aspect-Based Sentiment Analysis. *Applied Sciences*, No. 12 (19), 101–134. DOI: <https://doi.org/10.3390/app121910134>.
20. Parry M. E., Kawakami T., Kishiya K. (2012). The effect of personal and virtual word-of-mouth on technology acceptance. *Journal of Product Innovation Management*, No. 29 (6), 952–966. DOI: <http://doi.org/10.1111/j.1540-5885.2012.00972.x>.
21. Karyukin V., Mutanov G., Mamykova Z., Nassimova G., Torekul S., Sundetova Z., Negri M. (2022). On the development of an information system for monitoring user opinion and its role for the public. *Journal of Big Data*, No. 9 (1), 119–145. DOI: <https://doi.org/10.1186/s40537-022-00660-w>.
22. Murphy S. T. (2011). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology*, No. 8 (3), 723–739. DOI: <http://doi.org/10.1037/0022-3514.64.5.723>.
23. Wang Y., Guo J., Yuan C., Li B. (2022) Sentiment Analysis of Twitter Data. *Applied Sciences*, No. 12 (8), 157–189. DOI: <https://doi.org/10.3390/app122211775>
24. Schmäh M., Wilke T., Rossmann A. (2017). Electronic word of mouth: A systematic literature analysis. *Digital Enterprise Computing*, 147–158.
25. Wang Y., Chen Z., Fu C. (2022). Synergy Masks of Domain Attribute Model DaBERT: Emotional Tracking on Time-Varying Virtual Space Communication. *Sensors*, No. 22 (21), 450–471. DOI: <https://doi.org/10.3390/s22218450>.
26. Park J., Ciampaglia G. L., Ferrara F. (2016). Style in the age of Instagram: Predicting success within the fashion industry using social media. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, No. 22 (8), 64–72. DOI: <http://doi.org/10.1145/2818048.2820065>.
27. Abbas A. F., Jusoh A., Mas'od A., Alsharif A. H., Ali J. (2022). Bibliometrix analysis of information sharing in social media. *Cogent Business & Management*, No. 9 (1), 521–543. DOI: <https://doi.org/10.1080/23311975.2021.2016556>.

References

1. Zhang M., Xu H., Ma N., Pan X. (2022). Intelligent Vehicle Sales Prediction Based on Online Public Opinion and Online Search Index. *Sustainability*, No. 14 (16), 344–361. DOI: <https://doi.org/10.3390/su141610344>.
2. Vysotska V. (2021). Information Technology for Internet Resources Promotion in Search Systems Based on Content Analysis of Web-Page Keywords. *Radio Electronics, Computer Science, Control*, No. 3, 133–151.
3. Antonowicz P., Podpora M., Rut J. (2022). Digital Stereotypes in HMI – The Influence of Feature Quantity Distribution in Deep Learning Models Training. *Sensors*, No. 22 (18), 673–689. DOI: <https://doi.org/10.3390/s22186739>.
4. De-Gregorio F., Sung Y. (2010). Understanding attitudes toward and behaviors in response to product placement. *Journal of Advertising*, No. 39 (1), 83–96. DOI: <http://doi.org/10.2753/JOA0091-3367390106>.
5. Kamath A. N., Shenoy S., Subrahmanya K. N. (2022). An overview of investor sentiment: Identifying themes, trends, and future direction through bibliometric analysis. *Investment Management & Financial Innovations*, No. 19 (3), 229–242. DOI: [https://doi.org/10.21511/imfi.19\(3\).2022.19](https://doi.org/10.21511/imfi.19(3).2022.19).
6. Erkan I. (2016). The influence of e-WOM in social media on consumers' purchase intentions: An extended approach to information adoption. *Computers in Human Behavior*, No. 4, 47–55.
7. Asgari T., Daneshvar A., Chobar A. P., Ebrahimi M., Abrahamyan S. (2022). Identifying key success factors for startups With sentiment analysis using text data mining. *International journal of Engineering Business Management*, No. 14, 435–453. DOI: <https://doi.org/10.1177/18479790221131612>.

8. Gao L. (2014). Online consumer behavior and its relationship to website atmospheric induced flow: Insights into online travel agencies in China. *Journal of Retailing and Consumer Services*, No. 21 (4), 653–655.
9. Abulhajja S., Hattab S., Abdeen A., Etaiwi W. (2022). Mobile Applications Rating Performance: A Survey. *International journal of Interactive Mobile Technologies*, No. 16 (19), 133–146. DOI: <https://doi.org/10.3991/ijim.v16i19.32051>.
10. Guidry J. D., Messner M., Jin Y. (2015). From McDonalds fail to Dominos sucks: An analysis of Instagram images about the 10 largest fast food companies. *Corporate Communications: An International Journal*, No. 20 (3), 344–359.
11. Bagate R. A., Suguna R. (2022). Sarcasm Detection with and without #Sarcasm: Data Science Approach. *International journal of Information Science and Management*, No. 20 (4), 1–15.
12. Salganik M. (2019). Social Research in the Digital Age. *Journal of Interactive Marketing*, No. 2 (9), 345–358.
13. Li Q., Li X., Du Y., Fan Y., Chen X. (2022). A New Sentiment-Enhanced Word Embedding Method for Sentiment Analysis. *Applied Sciences*, No. 12 (20), 712–725. DOI: <https://doi.org/10.3390/app122010236>.
14. Jeff M., Jennifer R., Catherine J., Elke P. (2014). Managing brand presence through social media: The case of UK football clubs. *Internet Research*, No. 24 (2), 181–204.
15. Opiła J. (2022). On Employing of Extended Characteristic Surface Model for Forecasting of Demand in Tourism. *Interdisciplinary description of Complex Systems*, No. 20 (5), 621–639. DOI: <https://doi.org/10.7906/indecs.20.5.8>.
16. Kudeshia C., Sikdar P., Mittal A. (2016). Spreading love through fan page liking: A perspective on small scale entrepreneurs. *Computers in Human Behavior*, No. 8 (19), 257–270. DOI: <http://doi.org/10.1016/j.chb.2015.08.003>.
17. Albahli S., Irtaza A., Nazir T., Mehmood A., Ali A., Waleed Albattah W. (2022). A Machine Learning Method for Prediction of Stock Market Using Real-Time Twitter Data. *Electronics*, No. 11 (20), 341–363. DOI: <https://doi.org/10.3390/electronics11203414>.
18. Mousavijad M. (2017). The effect of socialization factors on decision making of teenagers consumers in schools. *Journal of School Administration*, No. 5 (1), 217–234.
19. Kim D., Kim Y., Jeong Y.-S. (2022). Graph Convolutional Networks with POS Gate for Aspect-Based Sentiment Analysis. *Applied Sciences*, No. 12 (19), 101–134. DOI: <https://doi.org/10.3390/app121910134>.
20. Parry M. E., Kawakami T., Kishiya K. (2012). The effect of personal and virtual word-of-mouth on technology acceptance. *Journal of Product Innovation Management*, No. 29 (6), 952–966. DOI: <http://doi.org/10.1111/j.1540-5885.2012.00972.x>.
21. Karyukin V., Mutanov G., Mamykova Z., Nassimova G., Torekul S., Sundetova Z., Negri M. (2022). On the development of an information system for monitoring user opinion and its role for the public. *Journal of Big Data*, No. 9 (1), 119–145. DOI: <https://doi.org/10.1186/s40537-022-00660-w>.
22. Murphy S. T. (2011). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology*, No. 8 (3), 723–739. DOI: <http://doi.org/10.1037/0022-3514.64.5.723>.
23. Wang Y., Guo J., Yuan C., Li B. (2022) Sentiment Analysis of Twitter Data. *Applied Sciences*, No. 12 (8), 157–189. DOI: <https://doi.org/10.3390/app122211775>
24. Schmäh M., Wilke T., Rossmann A. (2017). Electronic word of mouth: A systematic literature analysis. *Digital Enterprise Computing*, 147–158.
25. Wang Y., Chen Z., Fu C. (2022). Synergy Masks of Domain Attribute Model DaBERT: Emotional Tracking on Time-Varying Virtual Space Communication. *Sensors*, No. 22 (21), 450–471. DOI: <https://doi.org/10.3390/s22218450>.
26. Park J., Ciampaglia G. L., Ferrara F. (2016). Style in the age of Instagram: Predicting success within the fashion industry using social media. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, No. 22 (8), 64–72. DOI: <http://doi.org/10.1145/2818048.2820065>.
27. Abbas A. F., Jusoh A., Mas'od A., Alsharif A. H., Ali J. (2022). Bibliometrix analysis of information sharing in social media. *Cogent Business & Management*, No. 9 (1), 521–543. DOI: <https://doi.org/10.1080/23311975.2021.2016556>.

**INTELLIGENT SYSTEM FOR CLUSTERING USERS OF SOCIAL NETWORKS BASED
ON THE MESSAGE SENTIMENT ANALYSIS****Taras Batiuk, Dmytro Dosyn**

Lviv Polytechnic National University,
Information Systems and Networks Department, Lviv, Ukraine
E-mail: taras.m.batiuk@lpnu.ua, ORCID: 0000-0001-5797-594X
E-mail: dmytro.h.dosyn@lpnu.ua, ORCID: 0000-0003-4040-4467

© Batiuk T., Dosyn D., 2023

The main objective of this article is the analysis of the intelligent system for clustering users of social networks based on the messages sentiment analysis. The main goal of this intelligent system is to form a general image of the user of the system by analyzing the sentiment of the data of the user's social networks and their subsequent clustering. An intelligent system was designed, which, using the Identity and Access/Refresh JWT token algorithms, provides fast and maximally secure registration, authentication and processing of various system user sessions. The main approaches to the sentiment analysis of user messages and other data of various types are described, the principles of LSTM implementation of a recurrent neural network are described, which is very convenient for data analysis, because it works well and remembers the context of messages in the necessary time intervals, which increases the meaningfulness factor of the data analyzed according to the user of the intelligent system. General modern approaches to clustering and the most suitable clustering algorithm *k-means* is also described, since we will work with an undetermined amount of data each time, which can change significantly according to each individual user, the number of clusters and data processing will change because of this. Due to this, as a result of the work, the creation of a general image of the system user was described thanks to its comprehensive analysis, which made it possible to analyze users and display the corresponding results.

Key words: authentication; data sentiment analysis; clustering; LSTM recurrent neural network; *k-means* algorithm.