Наукова робота під шифром “Нейрон”

Всеукраїнський конкурс студентських наукових робіт

«Створення нейронної мережі для аналізу активності користувачів соціальних мереж»

2020 р.

ЗМІСТ

ВСТУП 3
РОЗДІЛ І. АНАЛІЗ ОСОБЛИВОСТЕЙ СОЦИОМЕТРИЧНИХ ДОСЛІДЖЕНЬ У СОЦІАЛЬНИХ МЕРЕЖАХ 5
1.1. Інструменти і завдання аналізу даних в соціальних мережах 5
1.2. Методи аналітики в соціальних мережах 12
РОЗДІЛ ІІ. РЕАЛІЗАЦІЯ ПРОГРАМНОГО ЗАСОБУ ДЛЯ АНАЛІЗУ АКТИВНОСТІ КОРИСТУВАЧІВ СОЦІАЛЬНИХ МЕРЕЖ 14
2.1. Аналітика великих даних 14
2.2. Інструменти аналізу соціальних мереж 17
2.3. Технічне завдання на розробку програмного засобу для аналізу активності користувачів соціальних мереж. 25
ВИСНОВКИ 30
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 31
ДОДАТКИ 37

ВСТУП

Актуальність. Дослідження мереж соціальних медіа розвивалася протягом останніх двох десятиліть в рамках інтер-дисциплінарної наукової галузі досліджень. Мережевий аналіз даних в соціальних мережах з'явився в кінці 20-го століття, коли вперше в історії були зафіксовані величезні соціальні взаємодії, які стали доступні для дослідників. Соціологи внесли теоретичні та концептуальні засади для розуміння структур соціальних мереж і ролі ключових гравців і спільнот в цих комунікаційних мережах на ширшому соціологічному, політичному, економічному і психологічному рівнях. Натуралістичні вчені ввели розуміння структури і динаміки великих мереж (неврологічного, генетичні, статистичні дані), а також методи і алгоритми, необхідні для аналізу соціальних мереж. Мережевий аналіз даних соціальних мереж розквітає в цьому унікальному перетині соціальних і природних наук. Таким чином створюється принципово новий способ проведення соціометричних досліджень із науковим чи екномікопрогнозуючим аспектом. Дослідним коже швидко проаналізувати групу користувачів із метою визначення наявності зазначеної групи користувачів, які будуть відокремлені за групою ознак. Використання нейромережевого підходу у подібній сфері аналізу якісно спрощує та прискорює процес відбору та аналізу емоційного стану користувачів і в той же час даний спектор використання нейронних мереж все ще не достатньо досліджений, через що вважаємо актуальною темою дослідження «Розробка нейронної мережі для аналізу активності користувачів соціальних мереж».

Проведений аналіз літературних джерел показав, що проблеми створення і застосування нейронних мереж досліджували Перлмуттер Б.А., Шофелт Д., Чуча І.І., Уосермен Ф., Ліпман Р.П. та інші.

Об'єкт – нейронні мережі для аналізу даних.

Предмет – програмний засіб для аналізу активності користувачів соціальних мереж.

Мета – розробка програмного засобу для аналізу активності користувачів соціальних мереж.

З**авдання роботи:**

1. Проаналізувати інформаційні джерела з питання використання інформаційних технологій у соціометричному аналізі.
2. Проаналізувати алгоритми та методи аналізу активності користувачів соціальних мереж.
3. Розробити технічне завдання до програмного засобу для аналізу активності користувачів соціальних мереж.
4. Розробити програмний засіб для аналізу активності користувачів соціальних мереж.

У процесі реалізації поставлених завдань були використані такі **методи наукового дослідження**: *теоретичні*: теоретичний аналіз алгоритмів роботи нейронних мереж та програмування з метою вибору методів розробки та проектування програмного засобу; *емпіричні*: проектуванні та моделювання програмного засобу; інтерпретаційні методи, які дають можливість узагальнення і пояснення встановлених фактів і їх взаємозв'язку; *практичні:* розробка програмного засобу для аналізу активності користувачів соціальних мереж.

Практична значущість – полягає у розробці програмного засобу, метою якого буде полегшення процесу аналізу активності користувачів соціальних мереж.

Структура роботи:

Дипломна робота складається зі вступу, 3 розділів, висновків та списку використаних джерел.

РОЗДІЛ І
АНАЛІЗ ОСОБЛИВОСТЕЙ СОЦИОМЕТРИЧНИХ ДОСЛІДЖЕНЬ У СОЦІАЛЬНИХ МЕРЕЖАХ

1.1. Інструменти і завдання аналізу даних в соціальних мережах

Наш мережевий світ з повсюдним створенням даних показує, що мережеві концепції широко поширені в цілому ряді дисциплін. Онлайн соціальна мережа (OSN) - це сучасний тип мережі, історія якої відносно коротка, але бурхлива. Поява масового впровадження онлайн-сайтів соціальних мереж (SNS) призвело до змін в тому, як люди спілкуються і діляться знаннями, як бізнес працює і конкурує, і як політики конкурують і впливають. В області досліджень аналіз OSN майже замінив будь-який традиційний інструмент соціальних наук (опитування, інтерв'ю, анкетування), оголошуючи, таким чином, обчислювальну соціальну науку. У сфері бізнесу аналіз соціальних мереж застосовується для розуміння ринків і спільнот [1, с.681-682], при цьому «соціальне підприємництво» є новою необхідністю для управління знаннями, поліпшеннями, змінами, співпрацею і ризиками. Для розуміння зв'язків, таких як то, як люди пов'язані один з одним за допомогою машин і як в цілому вони створюють фінансовий ринок, уряд, компанія і інші соціальні структури, А. Пентланд і А. Оздаглар [48] недавно створили Центр MIT для зв'язку науки і техніки . Щоб проілюструвати вплив того, як соціальні дані великого розміру змінили наше повсякденне життя, давайте подивимося, як змінився досвід прокату фільмів, і в даний час це сервіс, який використовує величезний масив даних для генерації рекомендованих закінчень [47].

Вражаюче зростання SNS можна розглядати як іскру, яка підірвала епоху великих даних. Це робить доступним безпрецедентний масштаб особистих даних, даних про події та соціальних відносинах, громадських

настроях і поведінці, які, коли вони використовуються і інтерпретуються, мають величезне значення. Нові види додатків виникають при розумному використанні даних OSN, що призводить до нової хвилі продуктивного зростання. OSN є багатим джерелом самовпевненого текстового і мультимедійного контенту, який останнім часом набув величезної популярності, особливо в області моніторингу політичних або маркетингових кампаній. Поширення останніх новин, особливо в Твіттері, проходить набагато швидше, ніж в звичайних ЗМІ. Тому раннє виявлення подій і аналіз соціальних мереж грають згубну роль в управлінні стихійними лихами, епідеміями і терористичними актами. Інформація про соціальні мережі також останнім часом включена в системи рекомендацій. Останні здатні впоратися з проблемами інформаційного перевантаження і фільтрації інформації.

Термін «Соціальна мережа» використовується для опису веб-сервісів, які дозволяють окремим користувачам створювати загальнодоступний чи частково загальнодоступний профіль в домені, щоб вони могли спілкуватися з іншими користувачами в мережі. В теорії мереж соціальна мережа зазвичай моделюється графом, який складається з користувачів або груп, званих вузлами, з'єднаних шаблонами контактів або взаємодій, званими ребрами або зв'язками. Унікальним елементом даних соціальних мереж є те, що вони надають нові можливості для розуміння окремих людей і суспільства, за умови, що люди проявили до них визнання і довіру [5, с.134-136].

Проте, дані соціальних мереж є об'ємними та, в основному, неструктурованими, і їх динамічна природа розвивається надзвичайно швидкими темпами, що ускладнює аналіз даних і вилучення знань. Перейшовши від аналізу окремих маленьких графів і властивостей окремих вузлів до розгляду великомасштабних властивостей графів, потреба в нових інструментах і методах аналізу даних неминуча. Хоча були зроблено багато наукових зусиль і досягнутий прогрес в напрямку певних подзадач для аналізу соціальних мереж, отримання знань з даних, отриманих з соціальних мереж, залишається великою проблемою, головним чином з двох причин.

По-перше, соціальна природа вузлів в соціальних мережах робить дані суб'єктивними для багатьох проблем конфіденційності. Яскравим прикладом є те, що конфіденційні дані користувачів можуть використовуватися адміністратором OSN і комерційними компаніями для визначення переваг користувачів і визначення аудиторії для їх реклами, що призводить до порушення конфіденційності та безпеки користувачів [5, 41]. Насправді, найбільша проблема великих даних - це конфіденційність [48, 23, 41, 32] і багато досліджень про надійне потоці персональних даних; і оскільки соціальні мережі містять особисті дані і в значній мірі впроваджені в повсякденне життя, проводиться багато досліджень про збереження конфіденційності в OSN [44, 4, 5, 25, 40]. Первісне зусилля щодо збереження конфіденційності - це сучасний і креативний системний інструмент рекомендації [40], розроблений для того, щоб допомогти користувачам захистити свої дані в OSN.

По-друге, наука все ще далека від автоматичного аналізу неструктурованих даних людського спілкування, тому що машини ще не здатні розуміти людську мову; і тому соціальна наука великих даних все ще розвивається. Крім того, прислів'я сміття на вході в сміття в минулому була жива і здорова. З-за неформального обміну мовними даними по OSN і галасливого характеру середовища традиційні технології попередньої обробки неадекватні. У цьому сенсі можна сказати, що, коли джерелом даних є соціальні мережі, всі проблеми, пов'язані з великими даними, стають ще більш значущими, а забезпечення якості даних, включаючи збереження конфіденційності, залишається відкритою проблемою для досліджень.

У відповідь на цю хаотично-розвиваючуся науку про соціальні дані і прогнозні знання, це дослідження спрямоване на другу проблему і аналізує феномен даних в соціальних мережах з технічної точки зору. Зокрема було провдено огляд сучасних систем аналізу даних в даній області, з огляду на різні види аналізу, різноманітність методів і функціональних можливостей, пропонованих ними.

З огляду на дуже великий набір даних, основне завдання полягає в тому, щоб з'ясувати, які дані є, і як їх аналізувати [49]. Соціальні мережі зазвичай містять величезну кількість контенту і даних про зв'язки, які можна використовувати для аналізу. Ці типи можуть бути додатково розділені на неструктуровані і структуровані дані відповідно, в залежності від того, організовані вони заздалегідь певним чином (структуровані дані) чи ні (неструктуровані дані). Щоб проілюструвати це на прикладі, засновані на часі події структуровані, тоді як дані подій, засновані на твіти та «лайки», неструктурованого. Структуровані дані в OSN зазвичай мають графічну структуру. У самій базовій структурі вони моделюються за допомогою соціальної мережі, яка представлена у вигляді графа G = (V, E), де V - це набір вузлів або об'єктів (наприклад, люди, організації та продукти), а E - це набір ребер або відносин , які з'єднують вузли через шаблони взаємодій. Дані такого типу вимірюються за допомогою аналізу соціальних мереж, додатки графічної аналітики, яке фокусується на добуванні інтелекту з таких взаємопов'язаних даних. З іншого боку, неструктуровані дані - це дані контенту, спільно використовувані в OSN, також відомої як користувальницький контент (UGC). Вони вважаються джерелом життєвої сили SNS і включають в себе текст, зображення, відео, твіти, огляди продуктів і інші мультимедійні дані, які зазвичай вивчаються за допомогою контент-аналізу [7, 24], методи якого включають, серед іншого, алгоритми структурування даних. Підходи аналізу соціальних мереж і контент-майнінг засновані на міждисциплінарних засадах штучного інтелекту (AI), статистики та суміжних областях. За десятиліття до появи OSN AI дослідники спробували впровадити суперечливе поняття «інтелект» в машини, щоб розуміти, міркувати і дізнаватися про те, як влаштований світ, і, отже, купувати додаткові можливості з простих логічних обчислень [51, 10]. OSN може використовуватися в якості середовища, наділеною машинами, що володіють такими знаннями. Статистика, з іншого боку, включає менш

складні процедури, які роблять упор на статистичні моделі для кращого розуміння процесу генерування даних.

Контент-аналіз в OSN вивчається за допомогою аналізу великих даних, і його основним завданням є отримання інформації з контенту, створеного і переданого в мережі. Аудіо або мовної аналіз слід широкосмугового безперервному розпізнаванню мови або фонетичному підходу для добування інформації з неструктурованих аудіо [24, с.815]; аналіз відеоконтенту включає в себе різні методи для моніторингу, аналізу та вилучення значимої інформації з відеопотоків [24, с.816]; Методи аналізу зображень варіюються від простих до складних в залежності від завдання аналізу, в той час як методи розпізнавання осіб [6] і вилучення настрою [26, с.165] в даних соціальних мереж привертають велику увагу.

Аналіз тексту витягує шаблони з текстових даних за допомогою засобів пошуку інформації, підсумовування тексту і обробки природної мови (NLP). Це часто включається з іншими методами. Аналіз зображень і аналіз тексту розпізнали величезні додатки в OSN, оскільки користувачі часто публікують зображення або окремо, або на додаток до тексту в своїх повідомленнях. Також аналіз відео-контенту включений з аналізом тексту. Наприклад, він може розгорнути уявлення транскриптов відео в пакетах слів, щоб витягти деякі приховані шаблони. В цілому, методи аналізу, такі як виявлення викидів [30, 8] і аналіз настроїв [26, 27] зазвичай використовуються в аналізі відео і зображень. За умови, що більшість фреймворків орієнтовані на текстовий контент, ми в основному зосереджені на методах інтелектуального аналізу тексту. Тим не менше, кілька структур, які витягують інформацію з відео, аудіо та зображень, також аналізуються.

Контентний аналіз і СНС не є взаємовиключними підходами, і далеко не обов’язково повинні співіснувати в аналізі. Інформація про контент в різних частинах мережі часто тісно пов’язана з її структурою [7, с.111] і, отже, об’єднання обох джерел інформації вважається більш ефективним в аналізі. Наприклад, аналіз настроїв може використовувати як дані про

зв’язки, так і текст. Попередні методології аналізу настроїв часто припускали, що тексти незалежні; але в контексті соціальних мереж дані об’єднуються в мережу, і цю функцію не слід випускати з уваги [29, 2]. На додаток до цього, соціальні відносини між користувачами останнім часом розглядаються як рівноцінна інформація в рекомендаційних системах як шаблони контенту, які використовуються користувачами спільно. Адже унікальним елементом даних соціальних мереж є те, що вони розкривають інформацію про взаємодію між користувачами-спільнотами-контентом.

Публічні інтерфейси провайдерів додатків (API) - це стандартна програма вилучення даних соціальних мереж з хмари, і вони, як правило, призначені для заохочення розробки стороннього програмного забезпечення, наприклад, плагіна для WordPress. Одна альтернатива - використовувати комерційні інструменти для видалення, які збирають дані, захищаючи їх необроблені форми і мають деякі додаткові функції фільтрації. А. Каушнік [34, с.17] використовував Sysomos, інструмент соціального моніторингу, для виявлення певних подій. Sysomos також є одним з інструментів, що використовуються на BBC для моніторингу соціальних мереж і веб-сайтів [17, с.2037-2038]. Іншою альтернативою є використання комбінації API і сканера, як це зробили дослідники в [11, с.145-146]. Шукач створений для вилучення інформації, що не автоматизована для вилучення за допомогою API служби. Важливо відзначити, що кожна соціальна платформа має дуже конкретні правила використання відповідних даних, які можна знайти в їх Умовах надання послуг. Хоча більшість SNS надають API, який включає методи для отримання діапазону даних, включаючи друзів, події, групи, вони обмежують кількість транзакцій API в день. Відзначено, що різноманітність даних, зібраних для аналізу, можна розрізнити в явних даних, а саме в інформації, безпосередньо пов’язаної з використанням служби (наприклад, подробиці профілю, кількість друзів тощо), І непрямих даних, які є якою інформацією, яка обробляються автоматично в системі (наприклад, дані браузера, відвідані веб-сайти) або можуть бути виявлені в

результаті дій користувача шляхом аналізу великих і повторюваних взаємодій між користувачами (голосування, обмін, позначки, коментування) [42, 65], Існує аналіз, який використовує неявні дані [37, с.65-67], явні [18] чи обидва одночасно [29, с. 22].

Зрозуміло, що аналітика - це складний процес, що вимагає досвіду і вміння. Дані, розуміння і вибір відповідних методів і прийомів і інтерпретація результатів аналізу. Інструменти є фундаментальними, щоб допомогти людям виконувати ці завдання. Однак процес відкриття знань став ще більш заплутаним з настанням великої епохи; де постійно з'являються нові інструменти для заміни традиційних і неефективних, і тепер для отримання цінності даних потрібно використовувати гібрид методів [49]. Що стосується області соціальних мереж, існує велика плутанина серед даних через відсутність чіткого визначення класифікації безлічі методів та інструментів, стандартизації процесів і структур аналізу, які зберегти якість даних. Сприяння вищезазначеного розриву в знаннях є метою дослідження даного опитування. Запропоновано підхід до аналізу великих даних в соціальних мережах з використанням інструментів, методів і методів. Зокрема, всі структури розділені з точки зору найбільш поширених методів аналізу в OSN, а саме: аналіз соціальних мереж, визначення теми, аналіз настроїв і спільні рекомендації [24, 18, 15]. До цих практик часто звертаються як за допомогою до аналізу великих даних, так і аналітики соціальних мереж, таких як прогнозування посилань, аналіз впливу і виявлення спільноти. Це опитування важливе з багатьох причин. По-перше, воно забезпечує складну класифікацію великої кількості останніх статей відповідно до методів аналізу даних, інструментами і середовищами для передачі даних в соціальних мережах. Цей ракурс може допомогти дослідникам в даній області вибрати конкретну практику аналізу і вивчити її різні інструменти і методи, які можна використовувати для цілей аналізу.

1.2. Методи аналітики в соціальних мережах

Аналіз думок (або аналіз настроїв) - це спроба скористатися величезною кількістю створеного користувачами текстового і новинного контенту в Інтернеті. Однією з основних характеристик такого контенту є його текстове розлад і високу різноманітність. Тут, натури обробки мови, компутаціонний і аналіз текстів розгорнуті, щоб ідентифікувати і виділити суб'єктивну інфорацію з початкового тексту. Загальна мета полягає в тому, щоб утримати ставлення письменника (або динаміки) по відношенню до якої - то темі або контекстних полярностях документа.

Автоматизований аналіз настроїв цифрових текстів використовує елементи машинного навчання, такі як латентний семантичний аналіз, підтримка векторних машин, мішок з слів моделі і симантики. Простіше кажучи, ці технічні наробки, що використовують три основних напрямки:

1. Обчислювальна статистика - відноситься до обчислювально-інтенсивним статистичних методів, включаючи методи повторної вибірки, методи Монте-Карло з ланцюгом Маркова, локальну регресію, оцінку щільності ядра і аналіз головних компонентів.

2. Навчання нейронної мережі, здатне до автономного величезного придбання та інтеграції знань, витягнутої з досвіду, аналітичного спостереження тощо. Ці субсімвольні системи далі підрозділяються на:

1.1. Навчання з вчителем, таке як дерева регресії, аналіз дискримінантних функцій, машини опорних векторів.

1.2. Навчання без вчителя, таке як самоорганізуються карти, K-Means.

Машинне навчання ставить собі за мету вирішити проблему наявності великих обсягів даних з великим числом змінних і MoNLy використовується в таких областях, як розпізнавання образів (мови, зображень), фінансові алгоритми (кредитні, алгоритмічної торгівлі) прогнозування енергії (навантаження, ціна) і біологія (виявлення пухлини, виявлення ліків).

1. Суб’єктивність настрою - рішення про те, висловлює чи даний текст думка або є фактичною (тобто без вираження).

2. Орієнтація / полярність настрою - рішення, чи є думка в тексті позитивною, нейтральною або негативною.

3. Сила настрою - визначення «сили» думки в тексті: слабка чи сильна.

Популярний підхід полягає в призначенні оцінок орієнтації / полярності (+1, 0, -1) всім словами: позитивну думку (+1), нейтральне думку (0) і негативну думку (-1). Загальна спрямованість / Полярність оцінки тексту є сумою орієнтирів, що становлять оцінки всіх «думок» знайденого слова. Проте, існує різні потенційні проблеми в цьому спрощеному підході, такі як заперечення (немає нічого, я ненавиджу).

Одним з методів оцінки орієнтації / полярності тексту є точкова взаємна інформація (PMI), міра асоціації, яка використовується в теорії інформації і статистиці.

Підводячи підсумок проведеного аналізу, можна стверджувати про наявність великої кількості унікальних особливостей, пов’язаних із використанням нейромережевого підходу у врішенні задач аналізу даних у соціальних мережах, це пов’язано у першу чергу із тим, що метрики соціальних мереж забезпечують унікальне розуміння моделей потоку інформації, а також уваги і отримання. Вони надають контекст для визначення ключових користувачів і їх потенційного впливу на різних рівнях аналізу. При застосуванні аналізу мережі соціальних мереж одиниці аналізу розрізняються і можуть включати в себе вузол, посилання чи всю мережу. Додаткової складності надає потреба використання різних соціальних мереж, кожна з яких має свою унікальну структуру як з точки зору програмної реалізації і рівня конфеденційності даних, так і з точки зору соціальної орієнтовності.

РОЗДІЛ II
РЕАЛІЗАЦІЯ ПРОГРАМНОГО ЗАСОБУ ДЛЯ АНАЛІЗУ АКТИВНОСТІ КОРИСТУВАЧІВ СОЦІАЛЬНИХ МЕРЕЖ

2.1. Аналітика великих даних

Аналіз вмісту вивчає неструктурований контент, згенерований у OSN користувачами. Останні розробки в сферах таких технологій, як хмарні обчислення і аналіз великих даних, сприяють виробленню ідей. Соціальні медіа-сайти мають велику кількість користувачів, розкиданих по всьому світу, які роблять S їх ідеальними кандидатами для адаптації хмари. Аналітика великих даних застосовується в соціальних мережах для вилучення значимої інформації за допомогою аналізу тексту та мультимедіа [24, 33]. Згідно з недавнім оглядом [33], відкритою проблемою в аналітиці великих даних є використання алгоритмів м'яких обчислень, оскільки, хоча вони можуть аналізувати таку складну природу даних, на жаль, до сих пір не багато досліджень присвячені цьому питанню. М'які обчислення є основою обчислювального інтелекту (CI), який, на відміну від систем Al-ba, не вимагає побудови точних моделей для обробки неточної, неповної і невизначеної інформації [51].

Зазвичай використовуються в аналізі текстового контенту підходи можна розділити на лінгвістичні, семантичні, статистичні та гібридні, а саме їх комбінацію. Синтаксичний аналіз - це лінгвістичний підхід, при якому частина мови (POS) є найбільш широко використовуваним методом; метрики косинусного подібності відносяться до статистичних підходів, де термін «частота-зворотна частота документа» (TF- IDF) є найкращим методом; і включення значення слів вказує на семантичні підходи, які в основному здійснюються через бази знань.

Самою основною одиницею мовної структури є слово, воно є основоположним для операцій аналізу, починаючи з навчання моделі

машинного навчання, оцінки документів за запитом, класифікації контенту і кластеризації [52, 53], представлення набору документів у вигляді векторів слів, відомих як вектор космічна модель. Мовні моделі зазвичай використовуються для ранжирування пропозицій і обчислення релевантності на основі інформації про контент. Вони навчаються за допомогою ряду строкових функцій, таких як фонеми, букви або слова. Мовне моделювання - це функція, яка поміщає міру ймовірності в рядки, взяті з деякого словника [52]. Тобто для мовної моделі M над алфавітом E:

Модель мови N-грам є безперервною послідовністю послідовності n рядків тексту або мови, і коли вона має розмір 1, вона називається «уніграммой», розмір 2 є «біграмм» і так далі.

Однак, згідно з Cambria і White [16, с.90-91], системи НЛП поступово перестануть покладатися занадто сильно на методи, засновані на словах, такі як BoW, і почнуть більш послідовно використовувати семантику для подолання таких проблем, як усунення неоднозначності в розумінні слова.

Семантичні технології широко використовуються в багатьох методах контент-аналізу. Щоб включити семантичні відносини між термінами в модель векторного простору або витягти тільки релевантну інформацію, словники з синонімами, такі як WordNet [45], виявилися корисними. WordNet є найбільш поширеним словником і охоплює семантичні і лексичні відносини між термінами та їх значенням, такими як синонімія, гіпонімія і багатозначність. В [9] дослідники використовували його, щоб знайти синоніми, щоб розширити свій створений вручну суб'єктивний набір слів для аналізу настроїв постів мікроблогів. TwiCal розпізнає події за підтримки словників термінів подій, зібраних з WordNet. В. Контопулос [38] використовував WordNet для доповнення базової семантики таксономії понять і атрибутів синонімами і гіпонімії. Також WordNet може застосовуватися до функцій агрегації, заснованим на ієрархічних моделях, де функції нижнього рівня (наприклад, координати GPS) можуть бути агреговані на більш високий рівень (наприклад, міста). Крім того, WordNet

використовується в схемі шифрування з можливістю пошуку для підтримки персоналізованого пошуку по цікавлять вас моделям [39].

Синтаксичний аналіз витягує токени і включає розширений аналіз пропозицій, термінів і порядку термінів. Він ідентифікує POS і розпізнавання іменованих об'єктів (NER) для створення дерев аналізу залежностей для кожної пропозиції. Методи POS і NER використовують структуру і мовні можливості, засвоєні з великого корпусу анотованого тексту. Мета POS, маркування даних метаданими і інших методів попередньої обробки, полягає в тому, щоб надати неструктурованих даними структуру, створити шаблони і зменшити неоднозначність для подальшого аналізу мови. Інший лінгвістичний підхід НЛП полягає в тому, щоб виконати вимір подібності між кластерними іменними фразами, використовуючи графічне представлення іменованих сутностей наборів документів, які пов'язані відносинами залежностей.

З іншого боку, TF-IDF вимірює значення слів з тексту, ігноруючи структуру пропозицій. Це метрика косинусного подібності (COS), яка використовується в контент-аналізі зазвичай для оцінки значущості слова. TF представляє важливість терміна в документі, а IDF вказує важливість або ступінь відмінності у всій колекції документів. Документи представлені в моделі векторного простору, де кожен документ d представлений вектором TF. TF - входження терміна, що зустрічається в документі: $dtf = (tf_h tf2, tfs, ... tfn)$, де tf - частота і-го терміна документа d. IDF надає великої ваги термінам, що зустрічаються тільки в декількох документах, і визначається як дріб: N / df, де N - загальна кількість документів в колекції, а df - кількість документів, в яких зустрічається термін і. Перераховані деякі з додатків TF-IDF в аналітичних середовищах соціальних мереж: обчислення подібності між питанням і темою [19], алгоритми навчання навчаючому механізму [20], пошук відповідної інформації [9], включення декількох ключових слів [22] і персоналізований [39] ранжируваних пошук. Інший менш використовуваний статистичний підхід полягає у використанні евристичних правил.

2.2. Інструменти аналізу соціальних мереж

Теорія графів є основним видатним підходом в аналізі соціальних мереж, а інструменти аналізу графів важливі для дослідження соціальних структур як аналітично, так і візуально. Розробляються бази даних графів, такі як Neo4j, графічні моделі, за допомогою інструментів глибокого навчання та інтелектуального аналізу графів, такі як Networkit, для ефективної обробки необхідності здобуття знань з мережевих даних. Двома великими обмеженнями, що стосуються СНС і обсягу даних, є обмежена кількість витягнутих даних із соціальних мереж через обмежені транзакцій API і складність обробки даних, що виходять за межі певного розміру мережі, за допомогою метрик графа і візуалізація даних [14, с.224-226].

Видобуток контенту OSN спільно з мережею може бути корисним для ефективної відповіді на підпитання аналізу. Інструменти Майнінгу на основі графів необхідні для того, щоб легко моделювати структуру соціальних мереж і виконувати вищевказані завдання. Існує безліч програмних інструментів, які аналізують властивості вузлів і ребер в мережі. Деякі з інструментів були спочатку розроблені для мережевої візуалізації, і тепер містять процедури аналізу, а інші були спеціально розроблені для інтеграції мережевого аналізу та візуалізації. Однак для ефективного дослідження соціальних мереж необхідна тісна інтеграція статистики та візуалізації соціальних мереж [7, с.89]. Кожен інструмент має свої сильні і слабкі сторони, тому вибір будь-якого інструменту для конкретного завдання все ще залишається проблемою. Порівняльне вивчення інструментів аналізу соціальних мереж вже проводилося раніше [50, 36, 21, 43], але не в рамках підходу, орієнтованого на дані. Крім того, категоризація метрик в таблицях призначена не для експертів в області співтовариства аналітиків графів, а головним чином для фахівців за даними, які хочуть використовувати графічну аналітику при аналізі контенту в соціальних мережах і відповідати

на питання. Програмні додатки з пакетами GUI (наприклад, Pajek) легше вивчати, в той час як пакети, створені для мов сценаріїв / програмування (наприклад, Networkit), є більш складними, потужними і розширюваними.

NodeXL [3] - це безкоштовний шаблон з відкритим вихідним кодом для Microsoft Excel, який спрощує основні завдання мережевого аналізу та візуалізації і підтримує аналіз мереж соціальних мереж для користувачів, які не використовують кодування. Він схожий на Pajek і Gephi з тією різницею, що він може безпосередньо збирати дані з соціальних мереж [43]. Хоча Gephi більш гнучкий з точки зору візуалізації. Однак обчислення метрик мережі в NodeXL може бути повільним, тому важливі дослідження в області удосконалених алгоритмів, розпаралелювання виконання з використанням декількох процесорів і використання спеціалізованих графічних співпроцесорів для прискорення обчислень. Два з майбутніх планів включають такі методи хмарних обчислень для ефективного обчислення мережевих кластерів і для поліпшення показників центральності для орієнтованих або дводольних графів і графів з різними вагами ребер [46]. NodeXL підтримує аналіз текстових даних шляхом вимірювання частоти появи суб'єктивних слів [14].

NetworKit [28], модуль Python, являє собою універсальний інструментарій для високопродуктивного мережевого аналізу з ефективними алгоритмами графів, багато з яких дозволяють виконувати паралельне виконання для швидкої обробки великомасштабних мереж. Його мета полягає в тому, щоб надати інструменти для аналізу великих мереж в діапазоні розмірів від тисяч до мільярдів ребер і має намір бути набагато швидше, ніж основні альтернативи. Простота використання і інтеграція з бібліотеками Python для інтерактивної роботи з даними. Це інструмент, який можна порівняти з пакетами NetworkX і igraph Python, які розглядаються в [50, 36], хоча і з упором на масивні мережі, більш швидке виконання алгоритмів, паралелізм і масштабованість. Зверніть увагу, що функціональні можливості Networkit Не такі всеосяжні, як NetworkX і igraph [2]. Pajek

пропонує можливості аналізу даних і можливості мережевої візуалізації для NetworKit [28].

Pajek [12] - це загальний інструмент аналізу графів для аналізу і візуалізації великих мереж. Він надає чудовий діапазон показників крім процедур аналізу соціальних мереж, таких як різні схеми поділу, кліки, кластери, компоненти і багато інших функцій. Цей інструмент існує на ринку вже 20 років і розширив його можливості, виправдовуючи широке використання як в наукових дослідженнях, так і в таких відомих компаніях, як Deutsche Bundesbank і Volkswagen. Тим не менш, він працює тільки на платформі Windows і відносно слабкий в візуалізації. Pajek-XXL - це спеціальне видання Pajek для аналізу величезних мереж.

Statnet [31] являє собою набір програмних пакетів, таких як ergm і network, для статистичного мережевого аналізу в мові програмування R, який реалізує останні досягнення в області статистичного моделювання випадкових мереж. Він залежить від набору цих базових пакетів для забезпечення його основних функціональних можливостей для статичного і динамічного моделювання мережі і використовується з командного рядка R або з недавнього GUI для менш досвідчених користувачів. Відмінність статнета від інших інструментів полягає в тому, що він зосереджений на статистичному моделюванні мережевих даних. Він використовується для оцінки моделі, оцінки моделі і на основі моделі мережеве моделювання, таке як моделі прихованого простору і кластера. Всі моделі засновані на алгоритмі Монте-Карло з центральної ланцюгом Маркова, який може легко обробляти мережі, що складаються з декількох тисяч вузлів і більш.

Gephi [13, с.98-99] є автономним програмним забезпеченням, яке вивчає кореляцію властивостей вузлів і структури мережі за допомогою візуальних шаблонів і підтримує класичні алгоритми інтелектуального аналізу даних аналізу соціальних мереж [35]. Gephi дозволяє дуже легко графічно представити «зв'язність» (ступінь), «вплив» (центральність між) і членство в спільноті окремих людей в мережі.

У таблиці 2.1 представлено порівняння п’яти інструментів мережевого аналізу, заснованих на характеристиках платформи і найосновніших потребах аналізу в залежності від навичок користувача. Таблиця 2.1 показує, що Networkit, Statnet і Pajek можуть використовуватися для більш складного аналізу, і між трьома, простішими для вивчення, знаходиться Pajek, але більш оновленим є Networkit. Statnet пропонує можливість статистичного аналізу моделі мережі. Навпаки, Gephi можна використовувати, коли потрібно приваблива і потужна візуалізація мережі. І останнє, але не менш важливе: NodeXL можна використовувати для аналізу в соціальних мережах, підтримує стандартні функції аналізу і візуалізації. Відзначено, що, коли осередок містить два значення, такі як «M L» означає, що інструменти тощо ovides зацікавлені в метричній шкалі від середнього до низького. Всі значення засновані на літературі, яка вказана в першому ряду таблиці.

Таблиця 2.1

Порівняння інструментів SNA

Program	**Pajek**	**Gephi**	**NodeXL**	**NetworKit**	**Statnet**
ОС	Windows	Windows, Mac OS, Linux	Windows	All	All
Ліцензія	Free	CDDL GNU Free	Microsoft	MIT	GPL
Версія	4.09	0.9.1	332	4.0.1	2016.9
Мова програмування	GUI	GUI	GUI	Python	R
Розмір	L	H	M	H	H
Час обчислення	M	M	H	L	M
Важкість використання	M-L	M-H	M-H	L	L
Якість графіків	L	H	M	L	M
Можливість аналізу	H	L	M	H	H

У таблиці 2.2 представлено порівняння аналітичних можливостей відповідно до критеріїв, згаданими раніше. Ми вибираємо для вивчення

метрики і алгоритми, які використовувалися в поширених методах аналізу OSN, які були проаналізовані. У Таблиці 2.2 різні алгоритми по-різному забарвлені в залежності від того, до якого методу аналізу вони належать. Центральність і описові аналітичні можливості, а також основні алгоритми Майнінг посилань підтримуються всіма інструментами, тоді як контент-аналіз мізерний. Це розумно, оскільки ці інструменти використовуються для маніпулювання і статистичного аналізу графіків, а не для аналізу мультимедійного мережевого контенту.

Таблиця 2.2

Порівняння аналітичних можливостей інструментів SNA

Program	Pajek	Gephi	NodeXL	NetworKit	Statnet
Density	+	+	+	+	+
Clique	+	+	+	+	+
Flow	+	-	-	+	+
Network Diameter	+	+	+	+	+
Geodesic distance	+	+	+	+	+
Census	Triad	Triad, Dyad	Triad, Dyad	Triad, Dyad	Triad, Dyad
Power Law	+	+	+	+	+
Connected Components	-	+	+	+	+
Degree	+	+	+	+	+
Betweenness	+	+	+	+	+
Closeness	+	+	+	+	+
Eigenvector	+	+	+	+	+
Coefficient	+	+	+	+	+
PageRank	-	+	+	+	+
HITS	+	+	+	+	-
Community Detection	+	+	+	+	+

Аналіз показує велику кількість різних за структурою систем і тому виникає потреба у визначені мови програмування, для подальшої роботи:

C ++ (підтримує багато парадигм (multi-paradigm): об’єктно-орієнтована, узагальнена, процедурна, метапрограмування)

За: STL (стандартна бібліотека шаблонів) - багато стандартних типів даних і алгоритмів. Велика "свобода" - можна реалізувати одні і ті ж речі по-різному. Хороша продуктивність скомпільованої коду. Гарна підтримка C ++ сьогодні.

Проти: Відсутність BigInteger і BigDecimal (вони є в бібліотеках Java і C#). Можуть бути різні помилки, викликані нерозумінням між компілятором і програмістом. Ви можете знайти багато тем про це, але це не проблема мови. Але через дуже великий свободи може бути складніше писати і налагоджувати програми на C ++.

Java (об’єктно-орієнтована, структурна, імперативна)

За: більш строгий синтаксис, ніж в C ++ - більш просте читання коду - швидка і проста налагодження. Підказки про помилки і невикористовуваних коді. Дуже багато бібліотек різного типу. Складальник сміття. Нові можливості в останніх версіях Яви (пр .: варіації циклу for).

Проти: Повільна робота програм (в 3-4 рази повільніше ніж C / C ++), довгий (постійно довгий) код, але набір коду швидкий, тому що присутня автодоповнення.

C # (підтримує багато парадигм (multi-paradigm): об’єктно-орієнтовану, узагальнену, процедурне програмування)

За: Швидше ніж Java. Стандартні бібліотеки C #: в останній версії .NET присутні, як і в Java, класи для роботи з довгою арифметикою, але тепер ви можете використовувати їх як змінні базових типів: c = a + b, і т.п.

Проти: Остання версія .NET все ще не доступна на більшості змагань з програмування.

Python (об’єктно-орієнтована, імперативний, функціональний, аспектно-орієнтований)

За: Python - це мова широкого призначення, на ньому пишуть практично будь-які типи програм, за винятком програм реального часу. Не випадково, пітон - це офіційна мова # 3 в Google.

Python відмінно підходить для рішення не дуже складних завдань завдяки стислості запису і наявності вбудованих засобів:

– Вбудована довга арифметика (як целочисленная, так і дрібна)
– Вбудовані list, set, dict, tuple (aka struct)
– Бібліотека для роботи з регулярними виразами re
– Функція sorted () для будь-яких послідовностей
– Зручні рядкові операції
– Зручні конструктори списків
– Функції sum (), max (), min (), здатні обробляти списки тощо.

Проти:

– Низька швидкість виконання програм (в середньому програш в 6 разів у порівнянні з C ++) і особливо повільний введення-виведення
– Мало зручних IDE

Беручі до уваги подальше орієнтування на мобільну розробку, вважаємо доцільним використання мови програмування Java. Проведемо більш детальний аналіз особливостей цієї мови.

Java – об’єктно-орієнтована мова програмування, розроблений компанією Sun Microsystems (надалі придбаної компанією Oracle). Програми Java зазвичай транслюються в спеціальний байт-код, тому вони можуть працювати на будь-який віртуальній Java-машині незалежно від комп’ютерної архітектури.

Android – операційна система для смартфонів, інтернет-планшетів, електронних книг, цифрових програвачів, наручних годинників, ігрових приставок, нетбуків, окулярів Google, телевізорів та інших пристроїв. Основана на ядрі Linux і власної реалізації віртуальної машини Java від Google. Android дозволяє створювати Java-додатки, що керують пристроєм

через розроблені Google бібліотеки.

SDK – комплект засобів розробки, який дозволяє фахівцям з програмного забезпечення створювати додатки для певного пакету програм, програмного забезпечення базових засобів розробки, апаратної платформи, комп'ютерної системи, ігрових консолей, операційних систем і інших платформ.

Android Studio – це інтегроване середовище розробки для роботи з платформою Android. Для розробки додатків потрібно встановити Android SDK та Java SDK.

JSON – простий спосіб зберігати та передавати структуровані дані. За допомогою простого синтаксису можливо легко зберігати все, що завгодно, починаючи від одного числа до рядків, масивів і об'єктів, в простому тексті. Також можна пов'язувати між собою масиви та об'єкти, створюючи складні структури даних.

Після створення рядка JSON, її легко відправити іншому додатку або в інше місце мережі, оскільки вона являє собою простий текст.

JSON має наступні переваги:

− компактний розмір;

− його пропозиції легко читаються і складаються як людиною, так і комп'ютером;

− його легко перетворити в структуру даних для більшості мов програмування (числа, рядки, логічні змінні, масиви й так далі);

− багато мови програмування мають функції та бібліотеки для читання і створення структур JSON.

Назва JSON означає JavaScript Object Notation (представлення об'єктів JavaScript). Як і представляє ім'я, він заснований на способі визначення об'єктів (дуже схоже на створення асоціативних масивів в інших мовах) і масивів.

2.3. Технічне завдання на розробку програмного засобу для аналізу активності користувачів соціальних мереж.

Метою проекту є розробка програмного модуля, який допоможе керівникам проектів різного типу для аналізу активності користувачів соціальних мереж.

Програмний засіб (сервер-клієнт), що становить портрет користувача на основі даних із соціальних мереж.

— клієнт - Android додаток, мова Java або Kotlin

— сервер - Python або Php

Соціальні мережі, до яких звертатиметься сервер для отримання інформації:

— Facebook (профіль, групи, стіна)

— Інстаграм (фото)

Основні кроки роботи програмного засобу:

1. Користувач вибирає соціальну мережу
2. Введення id профілю, за яким він хоче побудувати портрет
3. На сервері виконується збір інформації за профілем
4. До нейромережі подаються зображення, текст
5. Підсумовуються результати і на основі критеріїв виставляється оцінка профілю.

Для розробки інтерфейсу використовувати Java, розробку інтерфейсу проводити у середовищі Android Studio. Програмний засіб має задовольняти таким вимогам:

— мінімальна версія android: Android 5.0 (API level 21)

— максимальна версія android: Android 9 (API level 28)

— Мова програмування: Java

— Середовище розробки: Android Studio

Бібліотеки в Android додатку:

— https://github.com/VKCOM/vk-android-sdk

— https://github.com/square/retrofit (для мережевих запитів)

— https://developer.android.com/topic/libraries/support-library/packages

Бібліотеки в серверному додатку:

— http://sparkjava.com (micro framework for creating web applications)

— https://cloud.google.com/vision/

За для покращення процесу роботи із програмним засобом, він має відповідати таким вимогам: даний програмний продукт повинен бути легким, з точки зору навантажень на ПК користувача і не займати великої кількості місця на жорсткому диску (до 500 МВ). Використання програмного продукту не повинно вимагати установки, а так само не повинно вимагати установки додаткового ПЗ.

Це технічне завдання визначає вимоги та порядок Розробка програмного засобу прогнозування вступу абітурієнтів до закладів вищої освіти.

Мета – розробка програмного засобу для аналізу активності користувачів соціальних мереж.

Програмний засіб призначений для застосування соціальними службами для аналізу активності користувачів соціальних мереж з метою визначення їх емоційного стану.

Завдання, поставлене перед розробником проекту, полягає в тому, що програма повинна складатись з сторінки Android-додатку.

Аналіз проводиться тільки для обраної соціальної мережі, в мобільному додатку основних вікна: вибір, введення ід, налаштування, процесинг, результат.

Зміст, основні вимоги до програмного засобу:

– наявність елементів інтерфейсу, які дозволяють розпочати й відповідно завершити роботу з програмою;

– інтуїтивно зрозумілий і зручний у користуванні інтерфейс;

– дизайн навчально-контролюючої програми має створювати умови для зручного перегляду та сприймання інформації.

– зміст, обсяг та глибина закладеної інформації повинна відповідати рівню знань середнього користувача;

Системні вимоги:

– автономність програми, тобто її запуск не має передбачати наявності на комп'ютері кінцевого користувача спеціальних програмних засобів, окрім операційної системи;

– розміри програми не перевищують 500 Мб й програма являє собою .kфайл, що встановлюється на комп'ютер за допомогою копіювання .exe-файлу та не вимагає спеціальної інсталяції.

– Початок роботи з програмою проходить після запуску .apk-файлу.

Вимоги до функціональності ПЗ:

– надання можливості вибору соціальної мережі;

– надання можливості підгруження та доповнення бази знань;

– розробка можливості, у разі необхідності, виведення повідомлення про помилку у виконання завдання;

– розробка можливості наочного відображення результатів роботи (картинки, графіки, текстові повідомлення тощо);

– закриття програми через пункт меню та за допомогою кнопки керування;

– можливість масштабування інтерфейсу;

Вимоги до дизайну:

— зручне й швидке читання текстової інформації. Текстова інформація (текстові назви) має бути набрана зручними у читанні літерами;

— виважене кольорове візуальне рішення, яке буде акцентованим — реалізоване за принципами двоколірної протилежної гармонії.

— самі засоби відображення результатів повинні бути наочними, активізувати увагу, викликати інтерес й зосередженість на основних об’єктах.

Взаємодія людини з комп’ютером має бути виконано у вигляді системи меню та елементів керування для вибору режиму прогнозування (рис.2.1).

Рис. 2.1 Зображення дизайну сторінки вибору соціальної мережі

Повинен бути використаний діалоговий, інтерактивний режим. Програмний продукт повинен забезпечувати стандартний інтерфейс діалогу з користувачем. Вся необхідна інформація для введення вхідних даних повинна відображатися на екрані монітора(рис.2.2).

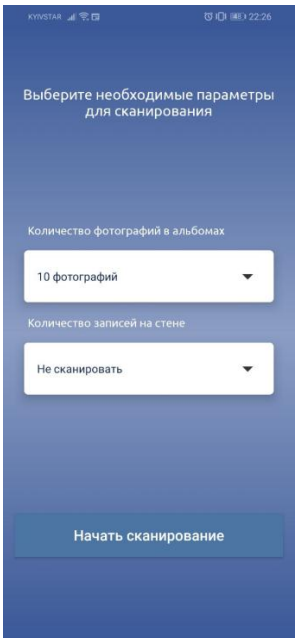

Рис. 2.2 Зображення дизайну сторінки вибору даних

У випадках виникнення виняткових ситуацій необхідно виводити на екран відповідні повідомлення і пропонувати оператору можливі засоби їх подолання.

Необхідно виконання наступних вимог:

- наявність користувальницького інтерфейсу;
- наочність результатів роботи.

Результати, що очікуються: програмний засіб, що розробляється, призначений для застосування соціальними та психологічними службами, ринком праці, керівниками відділів та іншими для полегшення процесу аналізу активності користувачів з метою оцінки їх емоційного стану та вподовань.

Програмний засіб може бути застосований у виробничому процесі керівниками підприємств для визначення рівня емоційного стану працівників.

ВИСНОВКИ

У дослідженні було визначено, що нейроні мережі є незамінними при аналізі даних, зокрема, для попереднього аналізу або відбору, виявлення «втрати фактів» або дуже критичних помилок людини, яка приймає рішення. В наш час є доцільним використовувати методи нейронних мереж в задачах з неповною інформацією або інформацією з великою кількістю шумів, особливо в задачах, де рішення можна знайти інтуїтивно, але при цьому традиційні математичні моделі не дають бажаного результату. Методи нейронних мереж можуть використовуватися незалежно від інших або ж служити одним із найкращих доповнень до традиційних методів статистичного аналізу, більшість з яких пов’язані з побудовою моделей, заснованих на тих чи інших припущеннях і теоретичних висновках. Метою роботи була розробка програмного засобу для аналізу активності користувачів соціальних мереж, за для досягнення мети було поставлено та виконано ряд завдань роботи:

1. Під час роботи було вивчено літературу та особливості використання штучних нейронних мереж, після чого було проведено аналіз можливостей нейронних мереж та засобів їх розробки.

2. Проведений аналіз аналогів та наукових наробок показав необхідність звернення уваги на особливості алгоритмів прогнозування у сфері соціології.

3. Розроблене технічне завдання, надало змогу скорегувати функціональні, графічні та системні вимоги до програмного засобу та визначило струтуру користувцького інтерфейсу. Результатом роботи став програмний засіб аналізу активності користувачів соціальних мереж.

У подальшій роботі планується доробка то розвиток програмного засобу з метою розширення його функціоналу та зменшення рівня похибки для більш точного відображення результатів аналізу активності користувачів соціальних мереж.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. A. Abbas, L. Zhang, and S. U. Khan, «A survey on context-aware recommender systems based on computational intelligence techniques» Computing, vol. 97, no. 7, pp. 667-690, Jul.2015.

2. A. Cuesta, D. F. Barrero, and M. D. R-Moreno, «A framework for massive twitter data extraction and analysis» Malaysian J. Comput. Sci., vol.27, no. 1, p. 50, 2014.

3. C. L. Staudt, A. Sazonovs, and H. Meyerhenke, «NetworKit: A Tool Suite for Large-scale Complex Network Analysis» p. 21, 2014.

4. C. Lin and Y. He, «Joint sentiment/topic model for sentiment analysis» in Proceeding of the 18th ACM conference on Information and knowledge management - CIKM ‘09, 2009, p. 375.

5. C. Metz, «AI is transforming Google search. The rest of the web is next» WIRED, 2016, p. 246.

6. Chandrasekar C, Kowsalya N Implementation of MapReduce Algorithm and Nutch Distributed File System in Nutch. Int J Comput Appl 1:6–11, 2011.

7. D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mane, «Concrete Problems in AI Safety» 2016, p. 124.

8. D. B. Kurka, A. Godoy, and F. J. Von Zuben, «Online Social Network Analysis: A Survey of Research Applications in Computer Science» 2015, p. 369.

9. D. Hansen, B. Shneiderman, and M. A. Smith, Analyzing Social Media Networks with NodeXL: Insights from a Connected World. Morgan Kaufmann, 2010, p. 49.

10. D. Li, C. Chen, Q. Lv, L. Shang, Y. Zhao, T. Lu, and N. Gu, «An algorithm for efficient privacy-preserving item-based collaborative filtering» Futur. Gener. Comput. Syst., vol. 55, pp. 311-320, 2016.

11. D. T. Nguyen and J. E. Jung, «Real-time event detection for online behavioral analysis of big social data» Futur. Gener. Comput. Syst., 2016, p. 162.

12. E. Cambria, A. Hussain, C. Havasi, and C. Eckl, «Sentic Computing: Exploitation of Common Sense for the Development of Emotion-Sensitive Systems» in Proceedings of the Second international conference on Development of Multimodal Interfaces: active Listening and Synchrony, Springer-Verlag, 2010, pp. 148-156.

13. E. Davis and G. Marcus, «Commonsense reasoning and commonsense knowledge in artificial intelligence» Commun. ACM, vol. 58, no. 9, pp. 92-103, 2015.

14. EY, «Big Data-Changing the way businesses compete and operate». P. Chaudhary, S. Gupta, B. B. Gupta, «Auditing Defense against XSS Worms in Online Social Network- Based Web Applications» in Handbook of Research on Modern Cryptographic Solutions for Computer and Cyber Security, vol. 36, no. 5, IGI Global, 1AD, pp. 216-245, 2014.

15. F. Atefeh and W. Khreich, «A Survey of Techniques for Event Detection in Twitter", Comput. Intell., vol. 31, no. 1, pp. 132-164, 2015.

16. F. Bonchi, C. Castillo, A. Gionis, and A. Jaimes, «Social Network Analysis and Mining for Business Applications» ACM Trans. Intell. Syst. Technol. Artic., vol. 2, no. 22, pp. 89-92, 2011.

17. F. Jiang, S. Rho, B.-W. Chen, X. Du, and D. Zhao, «Face hallucination and recognition in social network services» J. Supercomput., vol. 71, no. 6, pp. 2035-2049, 2015.

18. Gephi Network Statistics. URL: http://web.ecs.syr.edu/~pjmcswee/gephi.pdf.

19. H. Chen, X. Cui, and H. Jin, «Top-k followee recommendation over microblogging systems by exploiting diverse information sources» Futur. Gener. Comput. Syst., vol. 55, pp. 534-543, 2016.

20. H. Mackay, «Information and the transformation of sociology: interactivity and social media monitoring» Commun. Capital. Crit., vol. 11, no. 1, pp. 117-126, 2013.

21. J. He and W. W. Chu, «A Social Network-Based Recommender System» Springer US, 2010, pp. 47-74.

22. K. Ghazinour, S. Matwin, and M. Sokolova, «Yourprivacyprotector, A recommender system for privacy settings in social networks» Int. J. Secur. Priv. Trust Manag., vol. 2, no. 4, 2016, pp. 26-39.

23. K. N. Vavliakis, F. A. Tzima, and P. A. Mitkas, «Event Detection via LDA for the MediaEval2012 SED Task.” MediaEval 2012 Workshop, 2012, pp. 89-94.

24. K. Schouten and F. Frasincar, «Survey on Aspect-Level Sentiment Analysis» IEEE Trans. Knowl. Data Eng., vol. 28, no. 3, pp. 813-830, 2016.

25. K. Thiel, T. Kötter, B. Michael, R. Silipo, and P. Winters, «Creating Usable Customer Intelligence from Social Media Data: Network Analytics meets Text Mining» 2012, pp. 256-269.

26. Kobayashi M, Takeda K. Information retrieval on the web. ACM Comput Surv CSUR vol. 32, no. 2, pp. 144–173, 2010.

27. Lj. Sheela, «A Review of Sentiment Analysis in Twitter Data Using Hadoop» Int. J. Database Theory Appl., vol. 9, no. 1, pp. 77-86, 2016.

28. M. A. Smith, B. Shneiderman, N. Milic-Frayling, E. Mendes Rodrigues, V. Barash, C. Dunne, T. Capone, A. Perer, and E. Gleave, «Analyzing (social media) networks with NodeXL» in Proceedings of the fourth international conference on Communities and technologies - C&T ‘09, 2009, p. 255.

29. M. Fire, L. Tenenboim-Chekina, R. Puzis, O. Lesser, L. Rokach, and Y. Elovici, «Computationally efficient link prediction in a variety of social networks» ACM Trans. Intell. Syst. Technol., vol. 5, no. 1, pp. 1-25, 2013.

30. M. Mathioudakis and N. Koudas, «TwitterMonitor» in Proceedings of the 2010 international conference on Management of data - SIGMOD ‘10, 2010, p. 1155.

31. MapReduce. What is MapReduce? URL:http://www.mapreduce .org/what-is-mapreduce.php.

32. Mejova Y. Sentiment analysis: an overview URL: http://www.academia.edu/291678/ Sentiment_Analysis_An_Overview.

33. MPQA. 2005. URL: https://mpqa.cs.pitt.edu/

34. Murphy KP. Machine learning: a probabilistic perspective. In: Chapter 1: Introduction. MIT Press, pp 1–26, 2012.

35. N. Siddique and H. Adeli, «Introduction to Computational Intelligence» in Computational Intelligence: Synergies of Fuzzy Logic, Neural Networks and Evolutionary Computing, Oxford, UK: John Wiley & Sons Ltd, 2013, pp. 1-17.

36. O. Bowcott, «UK-US surveillance regime was unlawful ‘for seven years’ | UK news | The Guardian» The Guardian, 2015. URL: http://www.theguardian.com/uk- news/2015/feb/06/gchq-mass-internet-surveillance-unlawful-court-nsa.

37. R. Kaushik, S. Apoorva Chandra, D. Mallya, J. N. V. K. Chaitanya, and S. S. Kamath, «Sociopedia: An Interactive System for Event Detection and Trend Analysis for Twitter Data» Springer India, 2016, pp. 63-70.

38. S. Heymann and B. Le Grand, «Visual Analysis of Complex Networks for Business Intelligence with Gephi» in 2013 17th International Conference on Information Visualisation, 2013, pp.307-312.

39. S. Kaisler, F. Armour, J. A. Espinosa, and W. Money, «Big Data: Issues and Challenges Moving Forward» in 2013 46th Hawaii International Conference on System Sciences, 2013, pp. 995-1004.

40. S. Mithun, «Exploiting Rhetorical Relations in Blog Summarization» in Advances in Artificial Intelligence, Springer Berlin Heidelberg, 2010, pp. 388-392.

41. S. Mithun, C. Mulligan, G. Lapalme, N. Bouguila, S. Bergler, G. Butler, and L. Kosseim, «Exploiting rhetorical relations in blog summarization”, Concordia University, 2012, pp. 17-18,

42. S. Poria, E. Cambria, G. Winterstein, and G.-B. Huang, «Sentic patterns: Dependency-based rules for concept-level sentiment analysis» Knowledge-Based Syst., vol. 69, pp. 45-63, 2014.

43. S. Poria, E. Cambria, N. Howard, G.-B. Huang, and A. Hussain, «Fusing audio, visual and textual clues for sentiment analysis from multimodal content» Neurocomputing, vol. 174, pp. 50-59, 2016.

44. S. Rill, D. Reinel, J. Scheidt, and R. V. Zicari, «PoliTwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis» Knowledge-Based Syst., vol. 69, pp.24-33, 2014.

45. SentiWordNet. URL: http://ontotext.fbk.eu/sentiwn.html

46. T. Zhu, Y. Ren, W. Zhou, J. Rong, and P. Xiong, «An effective privacy preserving algorithm for neighborhood-based collaborative filtering» Futur. Gener. Comput. Syst., vol. 36, pp. 142-155, 2014.

47. Thomson Reuters. Thomson Reuters news analytics. URL: http://thomsonreuters.com/products/financial-risk/01_255/News_Analytics_-_Product_Brochure-_Oct_2010_1_.pdf.

48. Thomson Reuters machine readable news. URL: http://thomsonreuters.com/products/financial-risk/01_255/TR_MRN_Overview_10Jan2012.pdf.

49. Thomson Reuters MarketPsych Indices. URL: http://thomsonreuters.com/products/financial-risk/01_255/TRMI_flyer_2012.pdf.

50. Thomson Reuters news analytics for internet news and social media. URL: URL:http://thomsonreuters.com/business-unit/financial/eurozone/112408/news_ analytics_and_social_media.

51. W. Chang, J. Wu, L. M. Aiello, A. Barrat, R. Schifanella, C. Cattuto, B. Markines, F. Menczer, A. Blum, K. Ligett, A. Roth, C. Dwork, S. Kisilevich, L. Rokach, Y. Elovici, B. Shapira, Y. Li, M. Chen, Q. Li, W. Zhang, J. Lin, A. Machanavajjhala, D. Kifer, J. Gehrke, M. Venkitasubramaniam, P. Samarati, L. Sweeney, and L. Sweeney, «A New View of Privacy in Social Networks” in

Handbook of Research on Modern Cryptographic Solutions for Computer and Cyber Security, vol. 6, no. 2, IGI Global, 1AD, pp. 28-51.

52. word2vec. 2013. URL: https://habr.com/ru/post/446530/

53. WordNet Affect. 2009. URL: https://wordnet.princeton.edu/

ДОДАТКИ

Інструкція користувача для роботи з програмним засобом

Робота з програмним засобом починаэться з запуску .apk файлу, піля його встановлення із магазину програмного забеспечення (наприклад PlayMarket).

Стартове вікно надає можливість користувачеві вибору соціальної мережі, яка буде надалі використовуватись під час проведення аналізу соціально-емоційного стану користувача (рис.1).

Рис. 1. Вікно вибору соціальної мережі

Після вибору соціальної мережі, вікривається відповідне вікно введення ідентифікаційних даних користувача у відповідному до вибраної соціальної мережі вигляді. Окрім введення нового ідентифікаційного номеру користувача, надається можливість вибору

ідентифікаційного номеру з існуючого списку контактів у соціальній мережі.

Після введення ідентифікаційного номеру користувача, програмний засіб надає можливість вибору якості та глибини дослідження, для цього користувач має визначити кількість використовуємих у аналізі фотографій з профілю обраного користувача, та кількість використовуємих у аналізі записів, які користувач залишав на своїй особистій сторінці (рис.2).

Після підтвердження (натискання на кнопку «Далее») починається процес сбору інформації та проведення аналізу визначених ресурсів. Процес роботи відображатиметься у вікні робочого програмного засобу, як список виконаних задач та задач, які знаходяться у процесі виконання (рис.3).

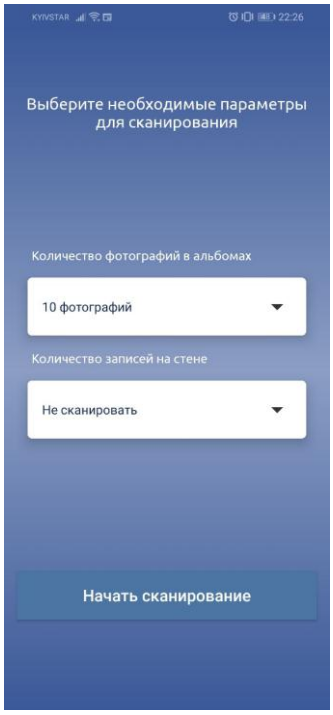

Рис. 2. Вікно визначення кількості використовуємих у аналізі фотографій та кількості використовуємих у аналізі записів

Рис. 3. Вікно відображення процесу аналізу

По завершенню аналізу зібраних даних, відкривається вікно результатів дослідження із емоційним описом особистості досліджуваного користувача. Серед опису визначається загальний емоційний фон, вади, загальні звички та інтереси (рис.4).

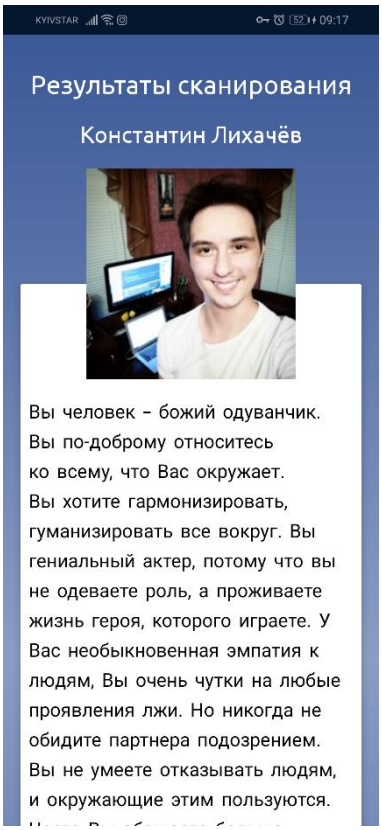

Рис. 4. Вікно відображення результатів аналізу

По завершенню ознайомлення із результатами аналізу, користувачеві надається змога вийти із програмного засобу, вийти до загального меню, чи змінити ідентифікаційний номер досліджуваної людини.