

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

НУБІП України

Факультет інформаційних технологій

НУБІП України

УДК 004.92:379.823

«ПОПОДЖЕНО»

Декан факультету
інформаційних технологій

«ДОНУСКАЄТЬСЯ ДО ЗАХИСТУ»

Завідувач кафедри комп'ютерних наук

Голуб Б.Л., к.т.н., доцент

Глазнова О.Г., д.т.н., професор

НУБІП України

2021р. 2021р.

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

НУБІП України

на тему Інтелектуальна система трекінгу та аналізу ЗМІ

Спеціальність 122 Комп'ютерні науки

Освітня програма Інформаційні управляючі системи та технології

Орієнтація освітньої програми _____

(освітньо-професійна або освітньо-наукова)

НУБІП України

Гарант освітньої програми

(науковий ступінь та вчене звання)

(підпис)

(ПІБ)

Керівник магістерської кваліфікаційної роботи

НУБІП України

кандидат технічних наук, доцент

(науковий ступінь та вчене звання)

(підпис)

Ткаченко О.М.

(ПІБ)

Виконав

(підпис)

Мусієнко В.М.

(ПІБ студента)

НУБІП України

КНІВ-2021

РЕФЕРАТ

до магістерської роботи Мусієнко В.М.

Інтелектуальна система трекінгу та аналізу ЗМІ

Метою дослідження є аналіз новин для отримання даних про кількість інформаційного шуму в джерелах отримання інформації користувача.

Об'єктом дослідження є телеграм канали, на які підписаний користувач.

Предмет дослідження: необхідність споживача у інформації, що він отримує

Методи та технології, що використовувалися під час дослідження: загальнонаукові методи – емпіричні (опис) та теоретичні (аналіз, пояснення, класифікація).

Завданням дослідження є розробка інтелектуальної системи, яка дозволить фільтрувати інформацію, що надходить до споживача, на основі його побажань та отримання даних для аналізу кількості «шкідливої» інформації.

В результаті дослідження визначено кількість інформації що споживається та її відношення до бажаної кількості споживання інформації. Близько 90% інформації, що ми читаємо – не потрібна для нас.

Кількість сторінок – 33, кількість ілюстрацій – 15 (без додатків), кількість таблиць – 1. До пояснювальної записки додається графічний матеріал – постер.

ABSTRACT

to the master's work Musienko V

Intelligent system of media tracking and analysis

The purpose of the study is to analyze the news to obtain data on the amount of information noise in the sources of user information.

The object of research is the telegram channels to which the user subscribes.

Subject of research: the consumer's need for information he receives.

Methods and technologies used during the research: general scientific methods - empirical (description) and theoretical (analysis, explanation, classification).

The task of the study is to develop an intelligent system that will filter the information coming to the consumer, based on his wishes and obtain data to analyze the amount of "harmful" information.

The study determined the amount of information consumed and its relationship to the desired amount of information consumption. About 90% of the information we read is not needed by us.

Number of pages - 33, number of illustrations - 15 (without appendices), number of tables - 1. Graphie material - poster is attached to the explanatory note.

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

ЗАТВЕРДЖУЮ
Завідувач кафедри
комп'ютерних наук

(назва кафедри)

к.тех.н., доцент

Голуб Б.Л.

(вчене звання і ступінь)

(підпис)
(ініціали і прізвище)

2021

ЗАВДАННЯ

ДО ВИКОНАННЯ МАГІСТЕРСЬКОЇ РОБОТИ СТУДЕНТУ

Мусієнку Валентину Михайловичу

(прізвище, ім'я, по-батькові)

Спеціальність 122 Комп'ютерні науки

(шифр і назва)

Освітньо-професійна програма Інформаційні управляючі системи та технології

Тема магістерської роботи: Інтелектуальна система трекінгу та аналізу ЗМІ

затверджено наказом ректора НУБіП від « 11 » листопада 2021 р. № 1634

Термін подання завершеної роботи на кафедру 10 грудня 2021 року

число) (рік, місяць,

3. Вихідні дані до магістерської роботи:

Групи на канали в соціальній мережі Телеграм для кожного тестованого користувача

4. Перелік питань, що підлягають дослідженню:

1. Дослідити методи збору даних із груп соціальної мережі Телеграм, передачі їх в систему аналізу.

2. Спроекувати та розробити систему аналізу.

3. Дослідити та обґрунтувати аналіз отриманих даних

4. Дослідити кількість даних, які забруднюють інформаційний простір користувача

5. Перелік графічного матеріалу (за потребами): постер

Дата видачі завдання

“ ”

2021 р.

Керівник магістерської роботи

(підпис)

Ткаченко О.М.

(прізвище та ініціали)

Завдання прийняв до виконання

(підпис)

Мусієнко В.М.

(прізвище та ініціали)

НУБІП України

ЗМІСТ
ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ
ВСТУП

6

7

1 АНАЛІЗ ТА ПРОЕКТУВАННЯ

9

1.1 9

НУБІП України

1.2 11

1.3 12

2 16

2.1 16

НУБІП України

2.2 18

3 24

3.1 24

3.2 Результати аналізу

29

ВИСНОВКИ

31

НУБІП України

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

32

НУБІП України

НУБІП України

НУБІП України

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

НУБІП України

- BI - Business intelligence.

- СД – сховище даних.

- БД – база даних.

НУБІП України

- ІС – інформаційна система.

- MS – Microsoft.

- SQL - structured query language, мова структурованих запитів.

- API - application programming interface.

НУБІП України

- UML - Unified Modeling Language.

НУБІП України

НУБІП України

НУБІП України

НУБІП України

НУБІП України

ВСТУП

На сьогоднішній день у світі відбувається гіперактивне зростання інформації в мережі інтернет. Кожен користувач мережі додає свої дані через різні джерела такі як: соціальні мережі, месенджери, блоги, інформаційні сайти і тд. У зв'язку з цим виникає велика кількість інформаційного шуму, і дедалі важче стає отримувати потрібну інформацію, що спричиняє меншій продуктивності роботи та навчання, а також впливає на фізичний та емоційний стан споживачів інформації.

Тому, **метою дослідження** є аналіз новин для отримання даних про кількість інформаційного шуму в джерелах отримання інформації користувача.

Об'єктом дослідження є телеграм канали, на які підписаний користувач.

Предмет дослідження: необхідність споживача у інформації, що він отримує

В рамках даного дослідження створюється система, що аналізує усі групи та канали, на які підписаний користувач месенджера Телеграм та видає йому новини, які він хоче бачити, в залежності від вказаних ним ключових слів. Також збирається інформація про кількість повідомлень, що проходять в зону цікавості користувача та тих, що не входять.

Таким чином, **завданням** дослідження є розробка інтелектуальної системи, яка дозволить фільтрувати інформацію, що надходить до споживача, на основі його побажань та отримання даних для аналізу кількості «шкідливої» інформації.

Актуальність дослідження впливає з необхідності зменшення потоку повідомлень від ЗМІ до користувачів Telegram. Актуальність є критично важливою через швидкий ріст кількості даних, яка призводить до нервових проблем та хвороб.

До складу пояснювальної записки входять: вступ, три розділи, висновки, та джерела інформації. У вступі описуються аспекти, що характеризують предметну область дослідження, коротко охарактеризовано сучасний рівень рішення поставленої задачі і взаємозв'язок з іншими роботами. Вказані мета дослідження, об'єкт, предмет, завдання структура пояснювальної записки.

В першому розділі проводиться аналіз постановки завдання та предметної області, розкривається постановка завдання, наводиться діаграма прецедентів та пояснюється робота системи.

В другому розділі подані методи та технології аналізу, відображено зміст та сутність системи, її інформаційне забезпечення та технічна складова.

Третій розділ присвячено розробці системи аналізу. В ньому вказані на зображенні основні частини системи та їх взаємодія між собою.

У висновках показано оцінку отриманих результатів аналізу та визначено потенційний скоуп використання результатів роботи.

До пояснювальної записки також додається графічний матеріал – постер, на якому відображені основні аспекти роботи.

1 АНАЛІЗ ТА ПРОЕКТУВАННЯ

НУБІП України

Мета процесу аналізу і проектування полягає в розробці технічних та розпорядчих інструкцій, де сформульовані завдання та алгоритм виконання роботи, який буде відповідати вимогам.

НУБІП України

До числа вирішуваних завдань при цьому ставляться:

- розробка архітектури логічної програмної системи;
- вибір методів реалізації та визначення проблем на реалізацію;
- розробка структури організації роботи програмного забезпечення;

НУБІП України

Аналіз відображає чітке розуміння завдання та постановку проблем, які необхідно вирішити в ході роботи.

НУБІП України

На етапі проектування відбувається уточнення вимог аналізу, спрямоване на покращення роботи ПЗ із урахуванням обмежень, що накладаються середовищем реалізації і т.д.

1.1 Постановка завдання

НУБІП України

На етапі постановки завдання потрібно визначити яких цілей необхідно досягти в результаті роботи над дослідженням. На даному етапі формуються точні умови завдання з описом вхідної і вихідної інформації.

НУБІП України

Вхідна інформація по завданню - дані, що надходять на вхід системи і використовуються для її вирішення.

НУБІП України

Вихідна інформація може бути представлена у вигляді документів, кадрів на екрані монітора, інформації в базі даних, вихідного сигналу пристрою управління.

Звернемо увагу на те, що основним завданням дослідження є отримання бажаних повідомлень за допомогою ключових слів, що вводить користувач, а вже потім – аналіз кількості інформаційного шуму.

У якості джерел інформації виступили усі Телеграм-канали аккаунта на якому проводилося тестування (близько 50). Зокрема це найпопулярніші ЗМІ України, а також найпопулярніші київські Канали.

Основними фільтрами відбору виступили:

- Ключові слова.
- Максимальна дата.
- Мінімальна дата.

- Максимальна кількість повідомлень з каналу.

Під час авторизації в профілі мало місце дотримання умов конфіденційності Телеграм. Через це було використано додаткове ПЗ в ролі Google Sheets API, як посередника в авторизації, під час роз'єднання з клієнтом.

Дані для авторизації та аналізу результатів роботи ПЗ також зберігаються в Google Sheets, та є зручними для доступу з мережі інтернет. На основі цих даних можна побудувати порівняльну діаграму, яка відобразить відношення шуканої інформації до отриманої. Самі ж бажані повідомлення відправляються системою в чат до користувача.

Використання Telegram API, Google Sheets API, Google Forms, бібліотеки TLSharp є абсолютно безкоштовними, що дозволяю за мінімальні ресурси побудувати та розробити необхідну систему.

1.2 Аналіз предметної області

Перед розробкою було проаналізовано ринок та не було знайдено жодного аналога. Системи включають свої особливості, але жодна з них не дає доступу до особистих повідомлень користувача.

Серед можливих варіантів аналогів були:

- Awario - система моніторингу ЗМІ та соцмереж, що дозволяє аналізувати згадки цікавих об'єктів у медіа
- BuzzSumo – це система роботи з соціальними медіа, що аналізує мільярди публікацій, постів та повідомлень та створює велику, доступну та надійну медіа-аналітику для бізнесів та організацій будь-якого розміру
- JagaJam інтернет-сервіс комплексного аналізу соцмедіа, що забезпечує вирішення основних завдань інтернет-маркетингу та PR від конкурентної аналітики до оцінки ефективності рекламних кампаній.
- Likealyzer це невеликий та надійний інструмент для моніторингу та аналізу соціальних медіа, що дозволяє досліджувати активність каналів, груп та сторінок у соцмережах.

Необхідною вимогою до виконання роботи є збереження зібраної інформації. Для цього було вирішено використовувати

Google Sheets.

Google Sheets – це програма для роботи з електронними таблицями, яка входить до складу безкоштовного веб-редактора Google Docs Editors, який пропонує Google. Сервіс також включає

Google Docs, Google Slides, Google Drawings, Google Forms, Google Sites і Google Keep. Google Таблиці доступні як веб-додаток, мобільний додаток для Android, iOS, Windows, BlackBerry та як

настільний додаток в ОС Google Chrome. Програма сумісна з форматами файлів Microsoft Excel. Додаток дозволяє користувачам створювати та редагувати файли онлайн, співпрацюючи з іншими користувачами в режимі реального часу. Зміни відстежуються користувачем з історією переглядів, яка містить зміни. Позиція редактора виділена кольором і курсором, що стосується редактора, а система дозволів регулює те, що можуть робити користувач.

1.3 Проектування системи

Проведення аналізу предметної області для подальшого проектування системи є завданням, що формує погляд на процеси, які відбуваються в системі, враховуючи не лише їх структури, а також правила зберігання і обробки.

Діаграма варіантів використання (англ. Use case diagram) в UML - це графічне зображення можливої взаємодії користувача з системою. Діаграма варіантів використання показує різні варіанти використання та різні типи користувачів, яких система має і часто супроводжуватиметься іншими типами діаграм. Варіанти використання представлені колами або еліпсами. Актори часто зображені у вигляді фігурок.

У той час як сам варіант використання може детально розглянути кожну можливість, діаграма варіантів використання може допомогти представити систему на більш високому рівні. Можна сказати, що «діаграми варіантів використання є кресленнями для системи».

Завдяки своїй спрощеній природі діаграми варіантів використання можуть бути хорошим інструментом комунікації для

зацікавлених сторін. Малюнки намагаються імітувати реальний світ і забезпечують уявлення зацікавленої сторони, щоб зрозуміти, як буде розроблена система.

При роботі з варіантами використання було дотримано

наступних правил:

- кожен прецедент відноситься як мінімум до одного актора;
- кожен прецедент має ініціатора;

• кожен прецедент призводить до відповідного результату.

Діаграма прецедентів для інтелектуальної система трекінгу та аналізу ЗМІ наведена на рис. 1

НУБІП України

НУБІП України

НУБІП України

НУБІП України

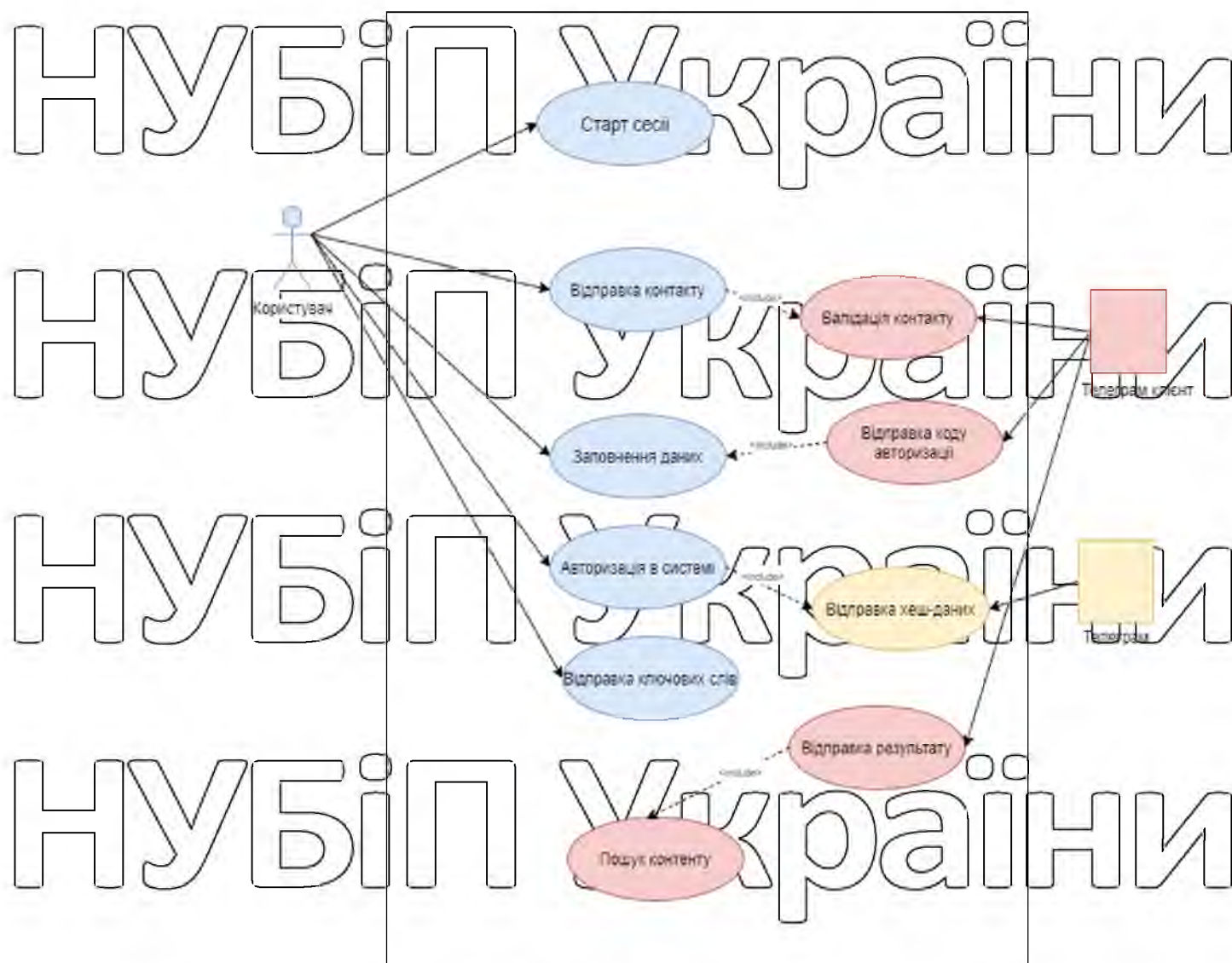


Рис. 1 Діаграма прецедентів

З діаграми зрозуміло, що існує три актори: Користувач,

Telegram клієнт та Telegram. Користувач є основним актором і взаємодія із системою напряму, з використанням телеграм-клієнта.

Сам же телеграм потрібен для авторизації в нього, та отримання

Кожен з прецедентів, вказаних в діаграмі, має свої варіанти

використання, їх детальний опис наведено в таблиці 1.

Опис прецедентів

Таблиця 1

Основний актор	Назва прецеденту	Опис прецеденту
Користувач	Старт сесії	Користувач запускає чат-бот
Користувач	Відправка контакту	Користувач відправляє свій контакт для початку авторизації в системі
Telegram клієнт	Валідація контакту	Telegram клієнт перевіряє чи є контакт контактом
Telegram клієнт	Відправка коду авторизації	Telegram клієнт відправляє код авторизації до Клієнта. В цей час зникає підключення до сесії.
Користувач	Заповнення даних	Користувач вводить отриманий код авторизації в сторонній сервіс (Google Form)
Користувач	Авторизація в системі	Дані зчитуються та відбувається авторизація
Telegram	Відправка хеш даних	Заклучний етап авторизації, після якого система отримує доступ до повідомлень користувача
Користувач	Відправка ключових слів	Користувач відправляє слова, по яких буде відбуватися пошук повідомлень
Telegram клієнт	Пошук контенту	Відбувається пошук повідомлень в системі
Telegram клієнт	Відправка результату	Користувач отримує шукану інформацію

НУВІП України

2 МЕТОДИ ТА ТЕХНОЛОГІЇ АНАЛІЗУ

2.1 Аналіз методів обробки даних

Обробка даних – це перетворення даних у придатну для використання та бажану форму. Це перетворення або «обробка» виконується з використанням попередньо визначеної послідовності

операцій вручну або автоматично. Більшість обробки здійснюється за

допомогою комп'ютерів і, таким чином, відбувається автоматично.

Вихідні або «оброблені» дані можна отримати в різних формах.

Прикладами цих форм є зображення, графік, таблиця, векторний файл, аудіо, діаграми або будь-який інший бажаний формат.

Отримана форма залежить від використовуюваного програмного забезпечення або методу обробки даних. Коли це буде виконано, це називається автоматичною обробкою даних.

Обробка даних в основному синхронізує всі дані, введені в програмне забезпечення, щоб відфільтрувати з нього найбільш корисну інформацію. Це дуже важливе завдання для будь-якої компанії, оскільки воно допомагає їм отримати найбільш релевантний контент для подальшого використання.

Кожен важливий сектор, будь то банки, школи, коледжі чи великі компанії, майже всі потребують такої обробки даних. Ця обробка виконується для того, щоб зберегти найбільш витончену інформацію в їхніх системах для подальшого використання. Обробка

вручну займає дуже багато часу і вимагає залучення занадто великої

кількості людей для цього. Це справді неможливе завдання, коли у вас є масові дані.

Сьогодні люди в талуві залежать від потужних та ефективних програмних інструментів, які допомагають обробляти всі ці дані. Це допомагає їм досягти більшої точності та підвищити ефективність.

При належній обробці даних можна сортувати все більше і більше інформації. Це допомагає отримати більш чітке уявлення про матерію та краще її зрозуміти. Це може призвести до кращої продуктивності та більшого прибутку для різних сфер бізнесу.

Обробка даних спрямована на вирішення наступних завдань:

1) упорядкування вихідного матеріалу, перетворення множини даних у цілісну систему відомостей, на основі якої можливий подальший опис та пояснення досліджуваних об'єкта та предмета;

2) виявлення та ліквідація помилок, недоліків, прогалин у відомостях;

3) виявлення прихованих від безпосереднього сприйняття тенденцій, закономірностей та зв'язків;

4) виявлення нових фактів, які очікувалися і були помічені під час емпіричного процесу;

5) з'ясування рівня достовірності, надійності та точності зібраних даних та отримання на їх основі науково обгрунтованих результатів.

Обробка даних має кількісний та якісний аспекти. Кількісна обробка є маніпуляція з вимірними характеристиками об'єкта (об'єктів), що вивчається, з його «об'єктивізованими» у зовнішньому прояві властивостями. Якісна обробка - це спосіб попереднього проникнення в сутність об'єкта шляхом виявлення його невимірюваних властивостей на базі кількісних даних.

Кількісна обробка спрямована здебільшого формальне, зовнішнє вивчення об'єкта, якісна — переважно змістовне, внутрішнє його вивчення. У кількісному дослідженні домінує аналітична

складова пізнання, що відображено і в назвах кількісних методів

обробки емпіричного матеріалу, які містять категорію «аналіз»:

кореляційний аналіз, факторний аналіз тощо. Реалізується кількісна обробка за допомогою математико-статистичних методів.

У якісній обробці домінує синтетична складова пізнання,

причому в цьому синтезі превалує компонент об'єднання та меншою

мірою присутній компонент узагальнення. Узагальнення

прерогатива наступного етапу дослідницького процесу

інтерпретаційного. У фазі якісної обробки даних головне полягає не

в розкритті сутності явища, що вивчається, а поки лише у

відповідному уявленні відомостей про нього, що забезпечує

подальше його теоретичне вивчення. Зазвичай результатом якісної

обробки є інтегроване уявлення про безліч властивостей об'єкта або

безліч об'єктів у формі класифікацій та типологій. Якісна обробка

значною мірою апелює до методів логіки.

В нашому випадку використовується кількісна обробка даних,

що дає змогу порівнювати конкретні результати дослідження та

робити висновки виходячи із них.

2.2 Загальні поняття про джерело даних

Джерело даних (іноді його називають файлом даних) так само

просто, як і звучить: місце, з якого беруться дані. Джерелом можуть

бути будь-які дані будь-якого формату файлу, якщо програма розуміє,

як їх читати.

Різні програми можуть використовувати джерело даних, включаючи додатки для баз даних, такі як Microsoft Access, MS Excel та інші програми електронних таблиць, текстові процесори, такі як Microsoft Word, ваш веб-браузер, офлайн програми тощо. Загальний сценарій, коли мова йде про Microsoft Word з використанням джерела даних полягає в тому, щоб Word здійснив злиття пошти з даних, взятих з документа Excel. Дивіться наш вступ до об'єднання пошти для отримання додаткової інформації.

Важливі факти джерела даних

Файл джерела даних, який використовується в одній програмі для однієї мети, може не мати жодної актуальності в іншій програмі, навіть якщо вони обидва використовують файли джерела даних.

Іншими словами, певне "джерело даних" є суб'єктивним програмі, що використовує дані.

Наприклад, джерелом даних для злиття пошти в Microsoft Word може бути файл CSV, який містить купу контактів, щоб вони могли бути автоматично записані в документ Word для друку конвертів з правильними іменами та адресами. Однак таке джерело даних не дуже корисне в будь-якому іншому контексті.

Приклади джерел даних

Як було сказано вище, джерело даних, яке також називають файлом даних, - це просто сукупність записів, що зберігають дані. Саме ці дані використовуються для заповнення полів злиття в злиттях пошти. Ось чому будь-який текстовий файл може використовуватися як джерело даних, будь то звичайний текстовий файл або власне файл бази даних.

Вони можуть надходити з таких програм, як MS Access, FileMaker тощо. Теоретично будь-яка база даних відкритої бази даних

(ODBC) може використовуватися як джерело даних. Їх також можна створити в електронних таблицях від Excel, Quattro Pro або будь-якої іншої подібної програми. Джерелом даних може бути навіть проста таблиця в текстовому процесорі документа.

Ідея полягає в тому, що джерелом даних може бути будь-який вид документа, якщо він організований для створення структури для приймаючої програми для отримання даних. Наприклад, контакт у адресній книзі може використовуватися в деяких сценаріях, оскільки там стовпчик для імені, адреси, облікового запису електронної пошти тощо.

Іншим видом джерела даних може бути файл, який записує час, коли люди відвідують лікарню. Програма може використовувати джерело даних, щоб агрегувати всі часи реєстрації та відображати їх на веб-сайті або використовувати їх у межах програми, або для відображення вмісту, або для взаємодії з іншим джерелом даних.

Інші джерела даних можуть бути взяті з живого каналу.

Наприклад, програма iTunes може використовувати прямий канал для відтворення Інтернет-радіостанцій. Канал - це джерело даних, а програма iTunes - це те, що відображає його.

Для виконання поставлених завдань використовуються дані із чату користувача Telegram. Процедура їх отримання є дуже складною та має зберігати повну конфіденційність - в ньому головна складність.

Telegram - це безкоштовна, кросплатформна, хмарна система обміну миттєвими повідомленнями. Послуга також забезпечує наскрізне зашифроване відеодзвінки, VoIP, обмін файлами та кілька інших функцій. Він був запущений для iOS 14 серпня 2013 року та Android у жовтні 2013 року. Сервери Telegram розповсюджуються по всьому світу, щоб зменшити часте навантаження на дані з п'ятьма

центрами обробки даних у різних регіонах, а операційний центр розташований в Дубаї в Об'єднаних Арабських Еміратах.

Для настільних і мобільних платформ доступні різні клієнтські програми, включаючи офіційні програми для Android, iOS, Windows, macOS і Linux (хоча для реєстрації потрібен пристрій iOS або Android і робочий номер телефону). Також є дві офіційні програми-блізюки Telegram – WebK і WebZ – і численні неофіційні клієнти, які використовують протокол Telegram. Усі офіційні компоненти Telegram є відкритим вихідним кодом, за винятком сервера із закритим вихідним кодом і закритим кодом.

Telegram забезпечує наскрізне зашифровані голосові та відеодзвінки та додаткові наскрізні зашифровані «секретні» чати.

Хмарні чати та групи шифруються між додатком і сервером, тому провайдери та інші треті сторони в мережі не можуть отримати доступ до даних, але сервер Telegram може.

Користувачі можуть надсилати текстові та голосові повідомлення, здійснювати голосові та відеодзвінки, а також ділитися необмеженою кількістю зображень, документів (2 ГБ на файл), розташуванням користувачів, анімованими наклейками, контактами та аудіофайлами. У січні 2021 року Telegram перевищив 500 мільйонів активних користувачів щомісяця. У січні 2021 року це був найбільш завантажуваний додаток у світі: станом на кінець серпня 2021 року його завантажили 1 мільярд у всьому світі.

За допомогою бота можна отримати зв'язок з джерелом даних.

Боти — це сторонні програми, які працюють всередині Telegram.

Користувачі можуть взаємодіяти з ботами, надсилаючи їм повідомлення, команди та вбудовані запити. Боти керуються за допомогою HTTP-запитів до вбудованого API бота.

НУБІП УКРАЇНИ

• Що можна робити з ботами?

- Переписуватися із ботом
- Отримуйте персоналізовані сповіщення та новини. Бот

може діяти як розумний нагадувач, надсилаючи вам відповідний вміст, щойно його публікують.

НУБІП УКРАЇНИ

- Інтегрувати з іншими сервісами. Бот може збагатити чати Telegram вмістом із зовнішніх сервісів.

- Бот може пропонувати платні послуги або працювати як віртуальна вітрина.

НУБІП УКРАЇНИ

- Бот може надавати вам сповіщення, прогноз погоди, переклади, форматування чи інші послуги.

- Бот для розмітки, бот для наклеювання, бот для голосування, бот для лайків

НУБІП УКРАЇНИ

• Як працюють боти?

По суті, Telegram Bots — це спеціальні облікові записи, для налаштування яких не потрібен додатковий номер телефону.

Користувачі можуть взаємодіяти з ботами двома способами:

НУБІП УКРАЇНИ

- Надсилати повідомлення та команди ботам, відкриваючи з ними чат або додаючи їх до групи.

- Надсилати запити безпосередньо з поля введення, ввівши @username бота та запит. Це дозволяє надсилати вміст від

вбудованих ботів безпосередньо в будь-який чат, групу чи канал.

НУБІП УКРАЇНИ

Повідомлення, команди та запити, надіслані користувачами, передаються програмному забезпеченню, запущеному на власних серверах. Сервер-посередник Телеграму обробляє за усе шифрування

НУБІП УКРАЇНИ

та зв'язок із Telegram API. Спілкування з цим сервером відбувається через простий HTTPS-інтерфейс, який пропонує спрощену версію API Telegram.

● Особливості ботів
○ Боти не мають онлайн-статусу та часових позначок останнього відвідування, замість цього інтерфейс показує мітку «бот».

○ Боти мають обмежене хмарне сховище — старі повідомлення можуть бути видалені сервером незабаром після їх обробки.
○ Боти не можуть ініціювати розмови з користувачами.

Користувач повинен або додати їх до групи, або спочатку

надіслати їм повідомлення.
○ Імена користувачів ботів завжди закінчуються на «бот» (наприклад, @TriviaBot, @GitHub_bot)
○ Після додавання до групи боти за замовчуванням отримують

не всі повідомлення

НУБІП України

НУБІП України

НУБІП України

НУБІП України

3 РОЗРОБКА СИСТЕМИ ТА АНАЛІЗ

3.1 Опис взаємодії вузлів системи

Щоб система працювала необхідно сформувати чітку взаємодію між вузлами її системи, що включають в себе вищеописані частини. Основні складові системи складаються із продуктів Telegram та Google, що дає впевненість та стабільність у роботі внутрішніх процесів системи.

При початку роботи системи потрібно написати боту певну команду. На рис. 2 зображено як почати роботу бота. Команда «/start» привсдить бота в робочий стан. Після цього, бот створює з'єднання з Телеграм та показує створену кнопку, яка дозволяє поділитися власним контактом для продовження роботи(рис. 3).




Рис. 2 Початок роботи з ботом

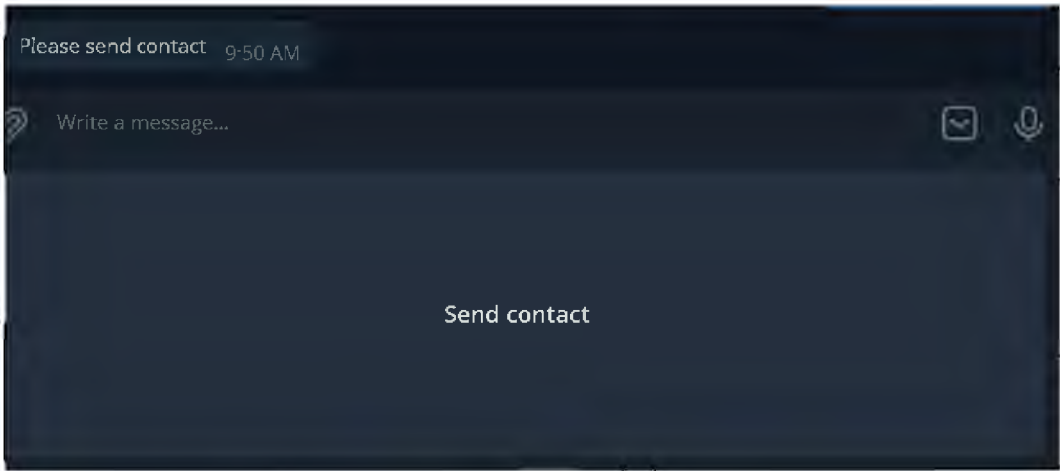


Рис. 3 Кнопка відправки контакту

НУБІП України

Після натиснення на кнопку, Telegram запитає чи ми дійсно хочемо поділитися контактом(рис. 4), та після підтвердження

розпочнеться етап отримання доступу до даних через авторизацію.

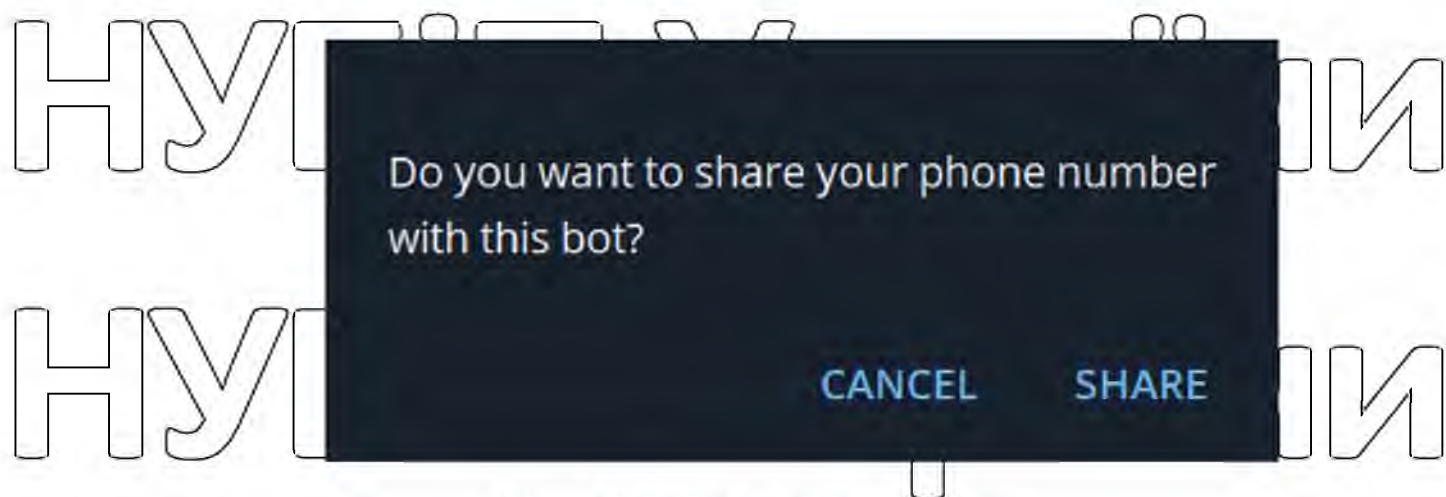


Рис. 4 Підтвердження контакту

НУБІП України

Бот від мені Телеграму відправляє код входження в обліковий запис та просить зачекати(рис. 5, рис.6). Після чого автоматично

відкривається вікно Гугл форми, в яке необхідно ввести отриманий

код(рис. 7).



Рис. 5 Відповідь бота

НУБІП України

Login code: 62958. Do not give this code to anyone, even if they say they are from Telegram!

This code can be used to log in to your Telegram account. We never ask it for anything else.

If you didn't request this code by trying to log in on another device, simply ignore this message.

9:54 AM

Рис. 6 Код авторизації від Телеграм

Authorization

 musienkovalentin07@gmail.com (не зв'язано)
ЗМІНИТИ ОБЛІКОВИЙ ЗАПИС

 Чернетку збережено

*Обов'язкове поле

Code: *

62958

Надіслати

Очистити форму

Рис. 7 Гугл форма авторизації в систему

Після цього код запищеться в Google Sheet (рис. 8), та буде отриманий ботом і використаний для входу в обліковий запис

Телеграм сповістить про це повідомленням, як на рис.9.

03.11.2021 21:09:21	1
03.11.2021 21:09:47	34
03.11.2021 21:09:55	77
04.11.2021 22:18:20	201
07.11.2021 12:06:21	38639
27.11.2021 15:40:20	68688
27.11.2021 15:59:16	91475
28.11.2021 12:41:53	78508
29.11.2021 20:07:25	77540
29.11.2021 20:48:47	72290
30.11.2021 09:55:24	62958

Рис. 8 Дані Google Sheet

New login. Dear Валентин, we detected a login into your account from a new device on 30/11/2021 at 08:03:10 UTC.

Device: NewsTracker, 1.0.0, PC, Web, Win 10.0

Location: Kyiv, Ukraine (IP = 46.164.144.210)

If this wasn't you, you can terminate that session in **Settings > Devices** (or **Privacy & Security > Active Sessions**).

If you think that somebody logged in to your account against your will, you can enable **Two-Step Verification** in **Privacy and Security** settings.

10:03 AM

Рис. 9 Сповіщення про вхід в систему

Після входу бот запропонує ввести ключові слова через пробіл, для подальшого пошуку їх в повідомленнях каналів, на які підписаний користувач (рис.10).

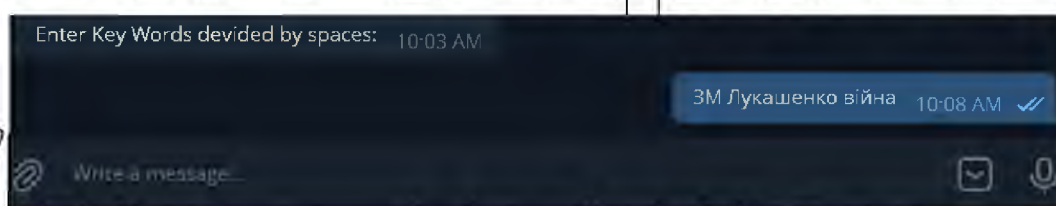


Рис. 10 Введення ключових слів

Бот розпочне аналіз і пошук ключових слів у чатах груп за останню добу. Назви каналів він буде виводити в консоль на сервері (рис. 11), а повідомлення відправляти користувачу (рис. 12).

```

Типичная Украина??
Аренда квартир Киев ? Official Chat
Бруталний Футбол ??
ТСН
ВУТЕОЛ УКРАЇНИ
КОНТР-МИТНИЦЯ-ЧАТ
Новости Украины | ЧАС
НОВИНА ???
Discover
Гуртожиток #7 НУБ?П
Метер дуєт
MyShearTrip - подорож?
Скільки это стоит?
МКСО | Мужская кухня
Планета Футбола
Гіл Мар
Junior двадцет доу.ua
Інструктор
Gripudream
NET learning community
Починаю Киев

```

Рис. 11 Список каналів, що аналізуються

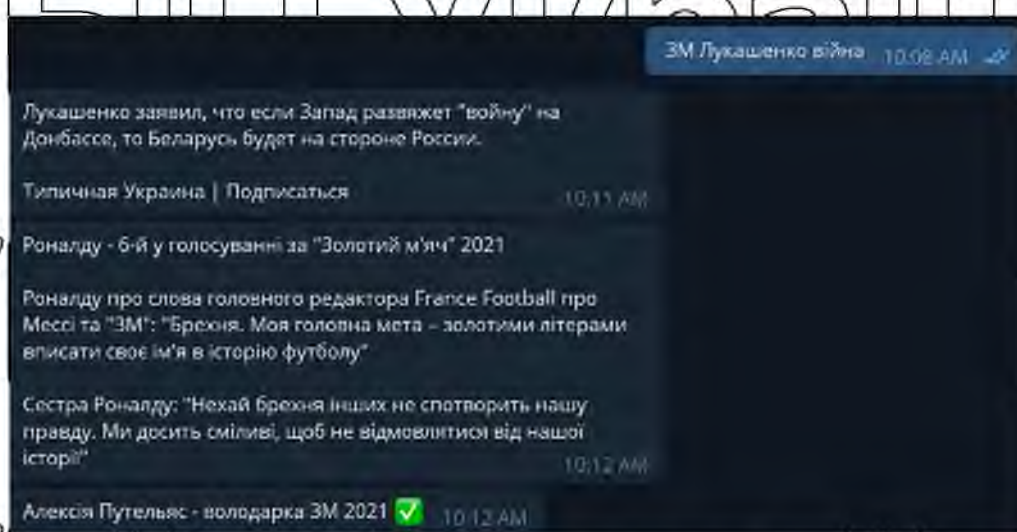


Рис. 12 Повідомлення від бота

Після виконання пошуку бот записує необхідні дані в Google Sheets, в розрізі яких можна провести аналіз (рис. 13).

challid	Number of messages	Found message amount	keywords
527734159	1575	17	1
527734159	1679	5	3
527734159	1502	2	3
527734159	1503	3	2
527734159	1514	1	2
527734159	1524	3	2
527734159	1544	166	2

Рис. 13 Дані про проведений пошук

3.2 Результати аналізу

Основним джерелом аналізу є діаграми в Google Sheets на основі даних, що отримуємо від роботи системи

Отримавши дані від роботи системи можна отримати декілька цікавих чисел.

На рис. 14, можна побачити кількість повідомлень, по яких відбувається пошук. Так як цифри дуже схожі, можна зробити висновок що кількість інформації в каналах щодня фактично однакова.

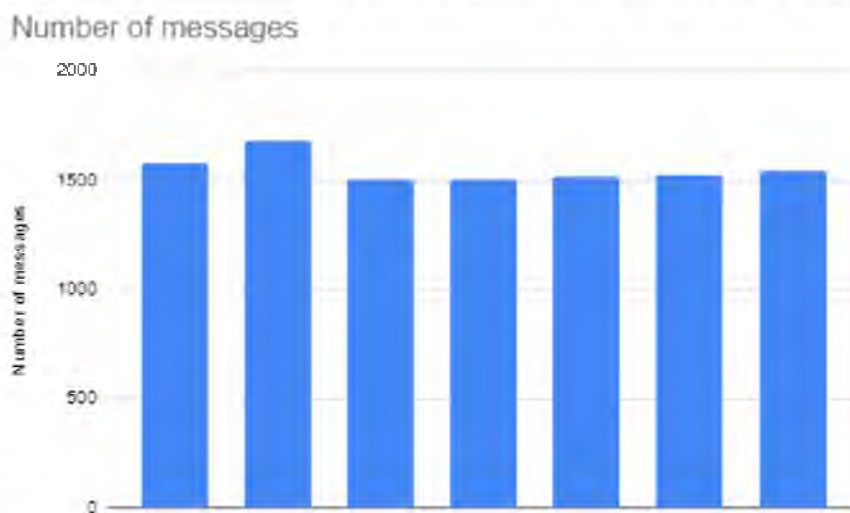


Рис. 14 Кількість повідомлень

На рис. 15 можна побачити що кількість слів дуже корелюється із кількістю знайдених повідомлень, а більше залежить від самих слів.

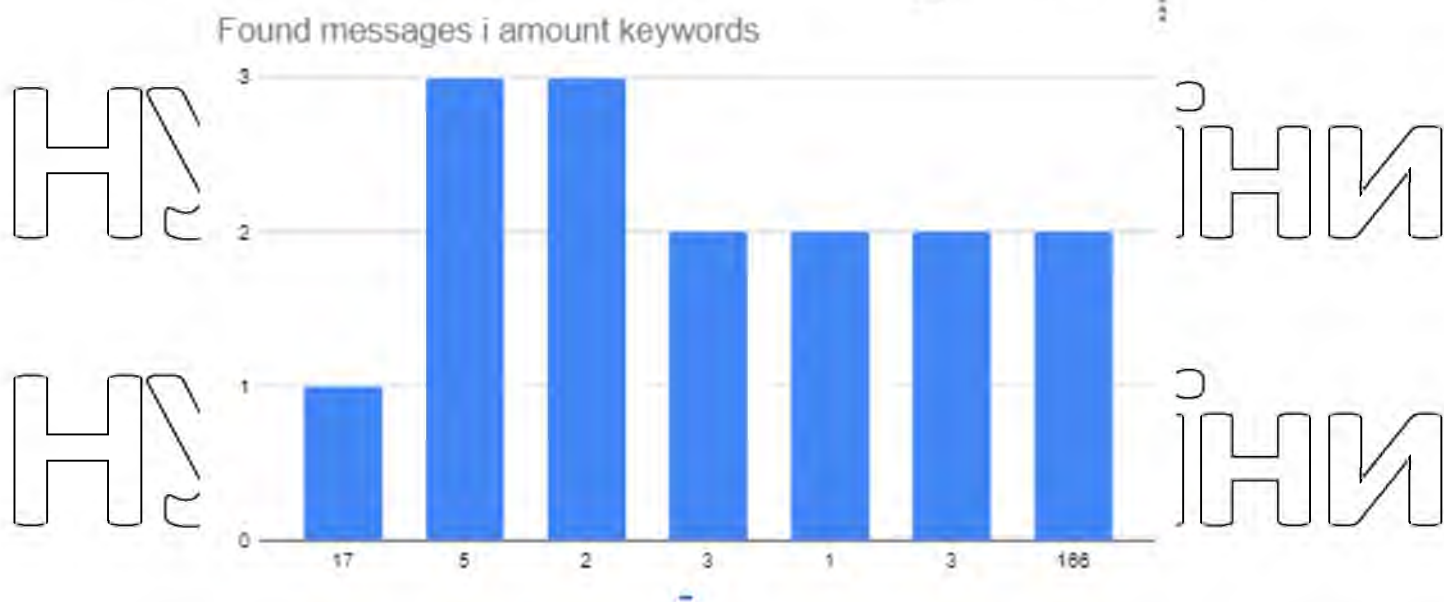


Рис. 15 Кількість ключових слів до кількості знайдених повідомлень. Головна діаграма відображає інформацію, що лише 2.5% інформації, користувачу була необхідна. Це зображено на рис. 16.

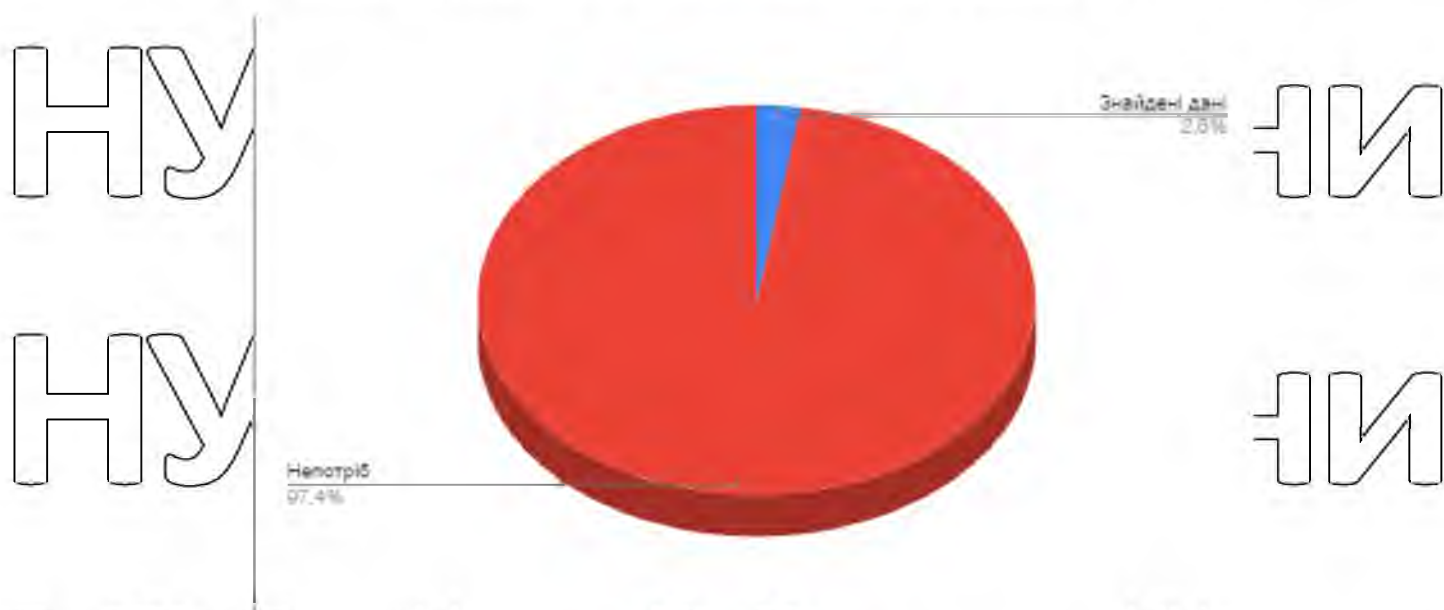


Рис. 16 Відношення шуканої інформації до не потрібної

ВИСНОВКИ

В ході виконання дипломної роботи магістра було інтелектуальну систему для отримання лише необхідної інформації із

ЗМІ користувачем. Проаналізовано кількість інформаційного шуму

та зроблено наступні висновки:

- Близько 90% інформації не є потрібною її споживачу.
- Використання розробленої системи дозволяє зберігати її користувачу більше 50% часу, що вона витрачає на отримання інформації про новини.

- Система дозволяє миттєво отримувати бажану інформацію та сповіщує про неї, що є дуже актуально в наш час.

- Близько 90% людей не усвідомлює цю статистику.

Для роботи були використані наступні інструменти:

- Telegram API
- Google Sheets API
- Google Form

- TlSharp бібліотека

Створені звіти відповідають постановці завдання, а розроблене ПЗ працює та використовується. Результати підтвердили очікування

про проблему у надлишку інформації в соціальних мережах на ЗМІ,

що є важливим кроком для усвідомлення споживання інформації та її поширення.

На мою думку, кожна людина повинна зменшувати свій інформаційний слід в мережі, поширювати якомога менше інформації, тим самим зменшувати забрудненість інтернету. А також

споживати лише необхідний контент, для збереження власного психічного та фізичного здоров'я.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Web Development [Електронний ресурс]: Принципы SOLID. – Режим доступу: <https://medium.com/webbdev/solid-4ffc018077da>

2. W. Ofner [Електронний ресурс]: Repository and Unit of Work Pattern. – Режим доступу: <https://www.programmingwithwolfgang.com/repository-and-unit-of-work-pattern>

3. Nabr [Електронний ресурс]: Особенности работы CLR в .NET framework. – Режим доступу: <https://nabr.com/ru/post/90426>

4. Metanit [Електронний ресурс]: Business Logic Layer. – Режим доступу: <https://metanit.com/sharp/mvc5/23.8.php>

5. Microsoft [Електронний ресурс]: Entity Framework documentation. – Режим доступу: <https://docs.microsoft.com/en-us/ef/>

6. Telegram [Електронний ресурс]: Telegram Bot API. – Режим доступу: <https://core.telegram.org/bots/api>

7. 7. Проектирование информационных систем. Часть 1. Этапы разработки проекта: стратегия и анализ / Лилия Козленко // Компьютер Пресс. – 2011. Режим доступу - <http://www.interface.ru/home.asp?artId=2805>

8. Современные методы анализа данных / Рубаков С.В. Corpus Technologies // Кластерные технологии в исследованиях. 12с. Режим доступу -

<https://riep.ru/upload/iblock/031/031173bb40e099800b248497db44cb88.pdf>

9. Robert Wrembel, Christian Koncilia. Data warehouses and OLAP:

concepts, architectures, and solutions. / Wrembel R., Koncinski Ch. // IRM Press. - 2007. - PP. 1-26. Режим доступа - https://books.google.ru/books?id=XFivorxzDm8C&printsec=frontcover&dq=%22Data+warehouses+and+OLAP:+concepts,+architectures,+and+solutions%22&source=bl&ots=Ss83EvzwXx&sig=qJxPFxXYo7fqKAfUk3bBiMaE_Z0&hl=ru&oi=vrSTS5qeBpXGnAO_Lmt2XCw&sa=X&oi=book_result&ct=result#v=onepage&q=&f=false

10. Бизнес-аналитика: сегодня и завтра / Ольга Мельник //

Intelligent Enterprise: №2 (212), февраль 2010 года. - 09.03.2010. Режим доступа - <https://www.iemag.ru/analytics/detail.php?ID=20110>

11. Ian H. Witten, Eibe Frank and Mark A. Hall. Data Mining: Practical

Machine Learning Tools and Techniques. — 3rd Edition. Morgan Kaufmann, 2011. — P. 664. — ISBN 9780123748560. Режим доступа - <https://www.elsevier.com/books/data-mining-practical-machine-learning-tools-and-techniques/witten/978-0-12-374856-0>

12. Branson D. 8 Major Advantages of Using MySQL [Электронный

ресурс] / Tony Branson // Datamation - Режим доступа до ресурсу: <http://www.datamation.com/storage/8-major-advantages-of-using-mysql.html>.

13. SQL Server Business intelligence [Электронный ресурс]

Режим доступа: <https://www.microsoft.com/ru-ru/sql-server/sql-business-intelligence>

14. Мюллер Р. Дж. Базы данных и UML. Проектирование [Текст] /

Р. Дж. Мюллер - М.: ЛОРИ, - 2002. - 420с.