

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)

Кафедра комп'ютерних наук
(повна назва кафедри)

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

магістр

(назва освітнього ступеня)

на тему: Дослідження процесів консолідації інформації користувачів
стрімінгових сервісів

Виконав: студент VI курсу, групи СНм - 61
спеціальності 122 Комп'ютерні науки
(шифр і назва спеціальності)

(підпис)

Кузьо О.О.

(прізвище та ініціали)

Керівник

(підпис)

Боднарчук І.О.

(прізвище та ініціали)

Нормоконтроль

(підпис)

Мацюк О.В.

(прізвище та ініціали)

Завідувач кафедри

(підпис)

Боднарчук І.О.

(прізвище та ініціали)

Рецензент

(підпис)

Цуприк Г.Б

(прізвище та ініціали)

Тернопіль
2022

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)

Кафедра комп'ютерних наук
(повна назва кафедри)

ЗАТВЕРДЖУЮ

Завідувач кафедри

Боднарчук І.О.
(підпис) (прізвище та ініціали)

« » 2022 р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

на здобуття освітнього ступеня Магістр
(назва освітнього ступеня)

за спеціальністю 122 Комп'ютерні науки
(шифр і назва спеціальності)

Студенту Кузьо Олегу Олеговичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження процесів консолідації інформації користувачів стрімінгових сервісів

Керівник роботи Боднарчук Ігор Орестович, к.т.н., доцент.
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від «22» листопада 2022 року № 4/7-948

2. Термін подання студентом завершеної роботи 20 грудня 2022р.

3. Вихідні дані до роботи Наукові публікації про процеси консолідації інформації

4. Зміст роботи (перелік питань, які потрібно розробити)

Вступ. 1 Дослідження основних аспектів роботи. 2 Аналіз методів консолідації інформації користувачів стрімінгових сервісів 3 Практична реалізація обґрунтованих методів і моделей для консолідації інформації. 4 Охорона праці та безпека в надзвичайних ситуаціях. Висновки
Додатки

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)

1 Титульна сторінка 2. Актуальність дослідження 3. Мета, Об'єкт, Предмет дослідження
4. Завдання дослідження 5. Основні аспекти роботи 6. Актуальні методи консолідації інфор-
мації користувачів 7. Практична реалізація методів конслолізації інофрмачії користувачів
8. Алгоритми роботи досліджених методів 9. Висновки 10. Завершальний слайд

6. Консультанти розділів роботи

| Розділ | Прізвище, ініціали та посада консультанта | Підпис, дата | |
|----------------------------------|---|----------------|------------------|
| | | завдання видав | завдання прийняв |
| Охорона праці | Мацюк О.В., доцент | | |
| Безпека в надзвичайних ситуаціях | Клепчик В.М., ст. викладач | | |

7. Дата видачі завдання 14 листопада 2022 р.

КАЛЕНДАРНИЙ ПЛАН

| № з/п | Назва етапів роботи | Термін виконання етапів роботи | Примітка |
|-------|--|--------------------------------|----------|
| 1. | Ознайомлення з завданням до кваліфікаційної роботи | 14.11.2022-15.11.2022 | Виконано |
| 2. | Підбір наукових джерел про процеси консолідації інформації користувачів стрімінгових сервісів | 16.11.2022-20.11.2022 | Виконано |
| 3. | Переклад та опрацювання наукових джерел про процеси консолідації інформації користувачів стрімінгових сервісів | 21.11.2022-23.11.2022 | Виконано |
| 4. | Виконання дослідження щодо процесів консолідації інформації користувачів стрімінгових сервісів | 24.11.2022-27.11.2022 | Виконано |
| 5. | Оформлення розділу «Дослідження основних аспектів роботи» | 28.11.2022-30.11.2022 | Виконано |
| 6. | Оформлення розділу «Аналіз методів консолідації інформації користувачів стрімінгових сервісів» | 01.12.2022-04.12.2022 | Виконано |
| 7. | Оформлення розділу «Практична реалізація методів і моделей для консолідації інформації» | 05.12.2022-07.12.2022 | Виконано |
| 8. | Виконання завдання до підрозділу «Охорона праці» | 08.12.2022-09.12.2022 | Виконано |
| 9. | Виконання завдання до підрозділу «Безпека в надзвичайних ситуаціях» | 10.12.2022-11.12.2022 | Виконано |
| 10. | Оформлення кваліфікаційної роботи | 12.12.2022-13.12.2022 | Виконано |
| 11. | Нормоконтроль | 14.12.2022-15.12.2022 | Виконано |
| 12. | Перевірка на плагіат | 15.12.2022 | Виконано |
| 13. | Попередній захист кваліфікаційної роботи | 16.12.2022 | Виконано |
| 14. | Захист кваліфікаційної роботи | 20.12.2022 | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

Студент

_____ (підпис)

Кузьо О.О.

_____ (прізвище та ініціали)

Керівник роботи

_____ (підпис)

Боднарчук І.О.

_____ (прізвище та ініціали)

АНОТАЦІЯ

Дослідження процесів консолідації інформації користувачів стрімінгових сервісів // Кваліфікаційна робота освітнього рівня «Магістр» // Кузьо Олег Олегович // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно-інформаційних систем і програмної інженерії, кафедра комп'ютерних наук, група СНм-61 // Тернопіль, 2022 // С. 78, рис. – 12, табл. – 1, кресл. – 0, додат. – 1, бібліогр. – 60.

Ключові слова: стрімінг, користувач, диференціація даних, консолідована інформація, OSINT-технології, структури великих даних.

Дипломна робота присвячена дослідженню процесів консолідації інформації користувачів стрімінгових сервісів.

В першому розділі дипломної роботи було проведено предметне дослідження та виділення основних аспектів роботи.

В другому розділі дипломної роботи проведено аналіз методів консолідації інформації користувачів.

В третьому розділі дипломної роботи було проведено практичну реалізацію обґрунтованих методів і моделей для конслодації інформації.

Об'єкт дослідження: процес консолідації інформації користувачів.

Предмет дослідження: моделі та методи диференціації даних консолідованої інформації.

Мета роботи: дослідження використання моделей та методів консолідації інформації на прикладі створення користувацького портрету.

ANNOTATION

Research of Information Consolidation Processes for Users of Streaming Services // Qualification work of educational level "Master" // Kuzo Oleg // Ternopil Ivan Pulyuy National Technical University, Faculty of Computer Information Systems and Software Engineering, Department of Computer Science, SNm-61 group // Ternopil, 2022 // P. 78, fig. - 12, tables - 1, chair - 0, annexes - 1, references - 60.

Keywords: streaming, user, data differentiation, consolidated information, OSINT technologies, big data structures.

The thesis is devoted to the research of the processes of information consolidation of users of streaming services.

In the first section of the thesis, a subject research was conducted and the main aspects of the work were highlighted.

In the second section of the thesis, an analysis of user information consolidation methods was carried out.

In the third section of the thesis, the practical implementation of substantiated methods and models for information consolidation was carried out.

Object of research: the process of user information consolidation.

Subject of research: models and methods of data differentiation of consolidated information.

Purpose: study of the use of models and methods of information consolidation on the example of creating a user portrait.

ПЕРЕЛІК СКОРОЧЕНЬ І ТЕРМІНІВ

Стрімінг – потокова передача даних

СУБД – система управління базами даних

Big Data – великі дані

OSINT – розвідка на основі відкритих джерел

Data Mining – добування даних

ШІ – штучний інтелект

ГПК – груповий портрет користувача

ЗМІСТ

| | |
|--|----|
| ВСТУП..... | 7 |
| 1 ДОСЛІДЖЕННЯ ОСНОВНИХ АСПЕКТІВ РОБОТИ..... | 10 |
| 1.1 Аналіз сфери дослідження..... | 10 |
| 1.2 Концепції та моделі систем консолідації інформації..... | 19 |
| 1.3 Дослідження завдань систем консолідації інформації..... | 26 |
| 1.3 Висновок до розділу 1..... | 29 |
| 2 АНАЛІЗ МЕТОДІВ КОНСОЛІДАЦІЇ ІНФОРМАЦІЇ КОРИСТУВАЧІВ СТРІМІНГОВИХ СЕРВІСІВ..... | 30 |
| 2.1 Метод консолідації інформації користувачів за допомогою оптимізації трансформації структур Big Data | 30 |
| 2.2 Принцип і архітектура розробки системи обробки Big Data..... | 37 |
| 2.3 Використання технології OSINT для формування портрету користувача..... | 42 |
| 2.4 Висновок до розділу 2..... | 46 |
| 3 ПРАКТИЧНА РЕАЛІЗАЦІЯ ОБГРУНТОВАНИХ МЕТОДІВ І МОДЕЛЕЙ ДЛЯ КОНСОЛІДАЦІЇ ІНФОРМАЦІЇ..... | 48 |
| 3.1 Розробка обробки даних користувачів за технологією OSINT..... | 48 |
| 3.2 Алгоритм диференційованої обробки даних користувачів..... | 52 |
| 3.3 Висновки до розділу 3..... | 60 |
| 4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ | 63 |
| 4.1 Організація робочого місця користувача ЕОМ | 63 |
| 4.2 Види та причини виникнення надзвичайних ситуацій техногенного характеру | 66 |
| ВИСНОВКИ..... | 70 |
| СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ..... | 72 |
| ДОДАТКИ | |

ВСТУП

Актуальність роботи. На даний час існує велика кількість різноманітних стрімінгових сервісів. Враховуючи досить складну конкуренцію між розробниками виникає потреба постійно розвивати продукт, для підтримання інтересу аудиторії, покращення досвіду використання та залучення якомога більшої кількості нових користувачів. Враховуючи можливість отримання великої кількості даних від користувачів, виникає потреба певним чином обробити дану інформацію, для використання результатів аналізу цих даних в прийнятті рішень.

У більшості випадків інформація, яка підлягає обробці та необхідна для прийняття управлінських рішень консолідується. В умовах обмежених ресурсів сучасного інформаційного простору необхідно обробляти великий обсяг консолідованої інформації, враховуючи обмеження апаратно-програмних обчислювальних ресурсів інформаційної системи. Рішення повинні прийматися в режимі реального часу, виходячи з особливостей комплексної інформації з метою скорочення часу її обробки, використовується метод диференціації даних, який дозволяє обробляти великі масиви даних.

Проте існуючі моделі та методи консолідації даних не можуть повною мірою отримати належне значення показника ефективності. Таким чином, існує протиріччя між обробкою все більшого обсягу інформації та вимогою швидкого прийняття рішень та існуючими комплексними моделями та методами обробки інформації. Враховуючи вищезазначене, дослідження процесів консолідації інформації є актуальною науково-практичною задачею.

Мета і задачі дослідження. Метою даної роботи є дослідження процесів використання моделей та методів консолідації інформації.

Зазначені цілі зумовили постановку та вирішення таких завдань:

- удосконалення методів консолідації інформації за допомогою оптимізації перетворень структури великих даних;
- аналіз існуючих моделей та методів диференціації консолідованих даних та демонстрація вибору напрямку дослідження на прикладі побудови користувацьких портретів;
- розробка інформаційних моделей для формування профілів користувачів та профілів груп користувачів на основі OSINT-техніки для законного доступу та використання інформації з відкритих джерел;
- аналіз розроблених моделей та методів формування користувацьких портретів та формулювання практичних рекомендацій щодо їх застосування.

Об'єкт дослідження: процес консолідації інформації користувачів.

Предмет дослідження: моделі та методи диференціації даних консолідованої інформації.

Наукова новизна одержаних результатів. Проведено дослідження процесів консолідації інформації на основі персонального і групового портретів користувачів та опис подальшого розвитку моделі процесу визначення важливості параметрів користувацького портрету, що відрізняється від відомих моделей використанням покращених коефіцієнтів значущості, які дозволяють зменшити невизначеність вхідних даних та підвищити ефективність прийняття рішень, зберігаючи рівень об'єктивності оцінювання.

Практичне значення одержаних результатів. Результати проведеної роботи можуть бути використанні розробниками стрімінгових сервісів для полегшення та здешевлення аналізу великої кількості консолідованої інформації отриманої від користувачів, що зпрощує подальший розвиток продукту, підтримання інтересу аудиторії, покращення досвіду використання та залучення якомога більшої кількості нових користувачів.

Апробація результатів магістерської роботи. Результати роботи були представлені на XI Міжнародній науково-практичній конференції

молодих учених та студентів «Актуальні задачі сучасних технологій» Тернопільського національного технічного університету імені Івана Пулюя (м. Тернопіль, 2022 р.)

Публікації. Основні результати кваліфікаційної роботи опубліковано у двох працях конференції (Див. додаток А).

Структура й обсяг кваліфікаційної роботи. Кваліфікаційна робота складається зі вступу, чотирьох розділів, висновків, списку літератури з найменування та додатків. Загальний обсяг кваліфікаційної роботи складає сторінок, з них сторінок основного тексту, який містить рисунків.

1. ДОСЛІДЖЕННЯ ОСНОВНИХ АСПЕКТІВ РОБОТИ

1.1 Аналіз сфери дослідження

Щодня мільйони користувачів на різних стрімінгових сервісах споживають багато різного контенту, кожного дня кількість інформації невпинно росте, розвивається та адаптується. Цифрові технології замінюють аналогові, та прогресують, адже користувачі стають все більше зацікавлені в таких перевагах як доступність, практичність та персоніфікація.

У реаліях ринку розважального контенту очевидний перехід від типової моделі розповсюдження (джерело-провайдер-користувач) до моделі, в якій контент транслюється безпосередньо на пристрій користувача. Проте, не акцентуючи увагу на різних моделях поширення та бізнес стратегії, об'єднує усі стрімінгові платформи факт того, що для успішного функціонування виникає гостра потреба враховувати потреби та бажання користувачів, які хочуть споживати медіаконтент за власними пріоритетами.

Це сприяє стрімкому зростанню кількості користувачів, конкуренції між різними сервісами, та вимагає розробників підтримувати рівень сервісу і постійно розвиватися. Основними перевагами стрімінгових сервісів, є швидкий доступ до контенту з різних платформ та за різних умов, швидкість виходу актуального контенту, відсутність реклами. Також варто не забувати про наявність субтитрів, актуальність та використання технологій для аналізу даних користувачів для створення добірок, рекомендацій та розвитку сервісу.

Стрімінг – потокова передача даних, через мережу, що транслюється від певного провайдера. Цей термін застосовується до різних видів медіаконтенту, що генерується в режимі реального часу, або ж попередньо записаного. Ця технологія використовується в дуже різноманітних сферах в житті людини. Основною перевагою потокового мовлення є те, що дані

безперервно надсилаються користувачу поступово, відразу увесь, отже немає потреби завантажувати локально цілий файл, щоб отримати до нього доступ. Як і будь які дані, контент який передається користувачу розподіляється на пакети даних, кожен пакет вміщує частини файлу який передається, а вже на користувацькому пристрої плеєр у браузері, або ж додаток об'єднує цей потік та інтерпретує його у медіаконтент.

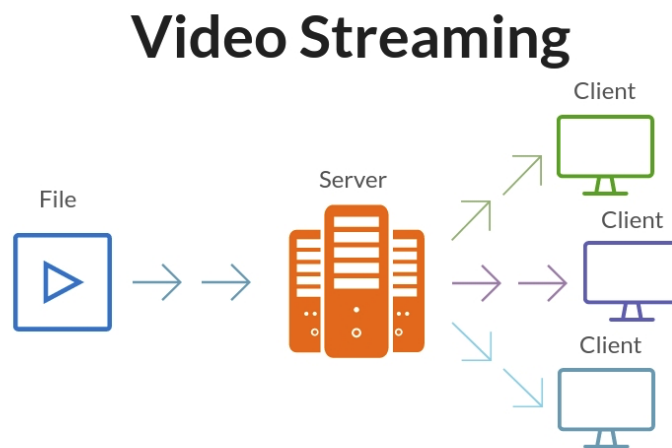


Рис. 1.1. Принцип роботи стрімінгового сервісу

Більшість систем потокового передавання медіа можна розділити на такі великі категорії:

Потокове відео на вимогу (Video on Demand, VOD). Така система має каталог, з якого за допомогою пошукової системи користувач може вибрати фільми, серіали, мультфільми тощо. Сучасні програми часто мають механізми рекомендацій на основі особистих уподобань.

Приклади: Netflix, HBO Max, Disney Plus, Amazon Prime, Hulu, YouTube.

Потокове аудіо. Завдяки платній підписці користувачі можуть слухати музику, не зберігаючи її на своєму пристрої. Приклади: Spotify, Deezer, iTunes, YouTube Music, Apple Music.

Пряма телевізійна трансляція. Телевізор більше не потрібен вдома, але важливі телепрограми можна дивитися онлайн зі смартфона чи ноутбука. Приклади: Megogo, Divan TV, Sweet TV. Дані платформи визначаються як розважальні платформи, оскільки переважно крім каталогового контенту, вони мають права на різні трансляції, як телевізійних каналів так і спортивних подій, подкастів тощо.

Один із найпопулярніших потокових жанрів. Це перегляд відео в реальному часі з одночасним записом і відтворенням. Цей тип найчастіше зустрічається у геймерів (або стримерів, коли транслюється, поки користувач грає в комп'ютерну гру), блогерів (коли просто ведуть розмову в прямому ефірі або відповідають на запитання) і під час трансляції відео спортивної події. Приклади: Twitch, Instagram Live, Facebook Live, YouTube Live.

Розглянемо більш детально існуючі сервіси потокового контенту.

Netflix — американський постачальник медіа-послуг і виробнича компанія, основною сферою якої є надання послуг за передплатою через веб-трансляцію своєї бібліотеки фільмів і телевізійних шоу, включаючи власного виробництва. Хоча основними пропозиціями сервісу з моменту його створення були шоу, серіали та фільми, Netflix починався зовсім інакше як компанія з прокату DVD. Станом на квітень 2020 року у Netflix було 182 мільйони передплатників у всьому світі, у тому числі 69 мільйонів у США. Netflix доступний у всіх країнах, крім материкового Китаю (через місцеві обмеження), Ірану, Сирії, Північної Кореї та окупованих Росією Криму та Донбасу (через санкції США). Netflix є членом Американської кіноасоціації (ААС). Компанія також виробляє та розповсюджує контент по всьому світу[1].

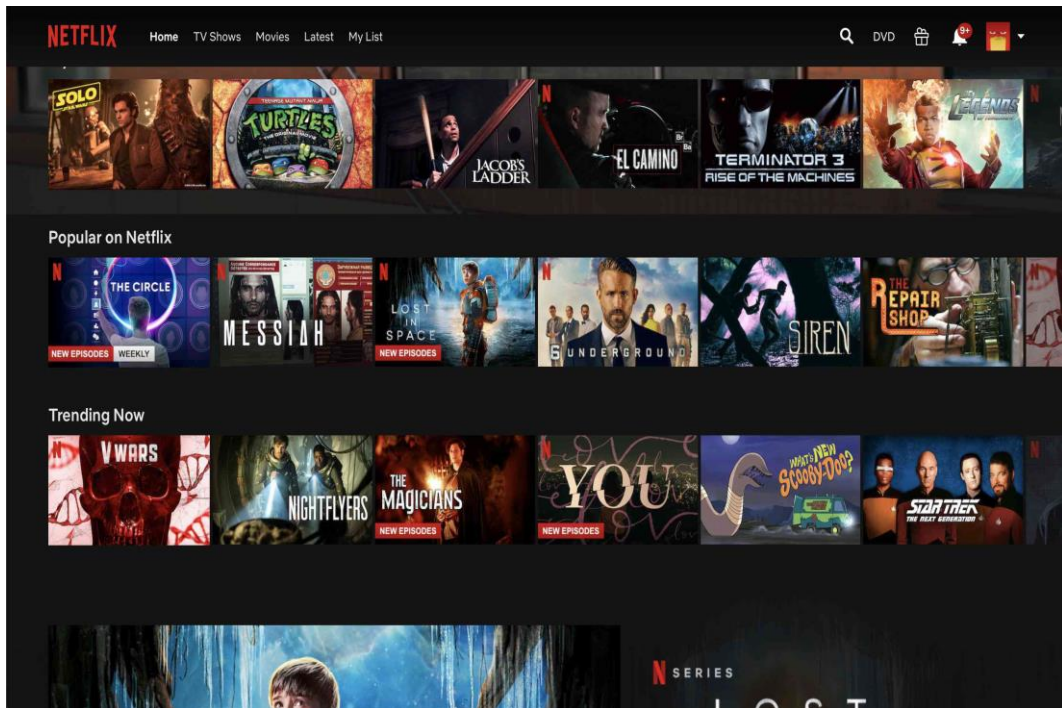


Рис 1.2 Початкова сторінка сервісу Netflix

Той самий алгоритм Cinematch є чи не основним фактором приголомшливого успіху Netflix. За словами засновника стрімінгу, ще в 2007 році він обслуговував кожного передплатника 75 відсотків шоу і фільмів, які йому були дійсно цікаві для перегляду. На даний момент понад 80% переглядів шоу та фільмів на Netflix, здійснюються через систему рекомендацій.

Ні, Netflix не вдавався до чорної магії (мабуть). Система алгоритмів працює на основі штучного інтелекту та машинного навчання з часів продажу DVD. Сервіс аналізує кожне рішення, яке ви приймаєте на платформі: ваші вподобання, кількість переглядів, середній час перегляду та багато інших факторів.

Netflix тепер має спеціальну команду, яка класифікує шоу та фільми не лише загальними жанрами, як-от «комедія» чи «жахи», але й більш детально. Наприклад, «візуально приголомшлива ностальгічна драма» або «недооцінений фільм про романтичну подорож». Netflix використовує технологію адаптивної потокової передачі, яка регулює якість аудіо та відео

на основі швидкості Інтернету користувача, і компанія також пропонує клієнтам можливість самостійно встановлювати якість відео.

Величезне поєднання кодеків і знайомих бітрейтів Netflix означає, що «фільм має бути закодований 120 різними способами, щоб бути доставленим на будь-яку потокову платформу».

YouTube — це найбільший у світі відеохостинг, куди можна завантажувати будь-яке відео (звичайно, з дотриманням авторського права та законодавства). Ви можете конвертувати свої відео та завантажувати їх на комп'ютер за допомогою різних сервісів і програм. YouTube був створений трьома друзями, які раніше працювали в PayPal – Джаведом Карімом, Стівом Ченом і Чадам Герлі. Зараз YouTube є третім за відвідуваністю у світі.

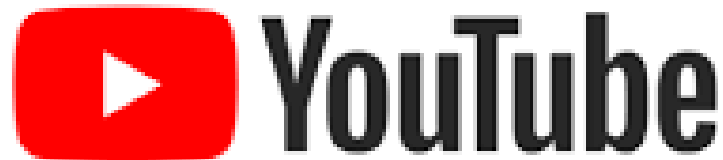


Рис 1.3 Логотип Youtube

YouTube пропонує 8 основних функцій, якими ви можете користуватися будь-коли:

- переглядайте різноманітні відео, деякі з яких будуть відкриті після реєстрації та підтвердження віку;

- завантажте власне відео на відеохостинг. Якщо їх тривалість не перевищує 10 хвилин, первинний перегляд не проводиться;
- використовувати спеціальні сайти або програми для завантаження відео на комп'ютер;
- шанс прокинутися до слави, завантажити відео свого виступу та отримати 1 мільйон переглядів за кілька годин;
- можливість збільшення трафіку шляхом розміщення посилань на ваш сайт, блог та інтернет-магазин в кінці відео;
- можливість отримувати прибутковий дохід за допомогою реклами в Google AdSense;
- можливість отримати корисні навички та дізнатися цікаву інформацію;
- нарешті, ви можете знайти багато друзів і знайомих з усього світу, тому що кількість користувачів YouTube просто вражає.

Spotify — це сервіс потокової передачі музики та подкастів, запущений у Швеції в 2006 році, і зараз доступний у 92 країнах світу з 286 мільйонами активних користувачів щомісяця .

У всьому світі Spotify має найбільше користувачів, при цьому сервіс займає 36 відсотків ринку. Послуга надає вам доступ до десятків мільйонів треків, включаючи подкасти, та безлічі списків відтворення. За допомогою безкоштовної або платної підписки на Spotify Premium, тож ви можете слухати музику покращеної якості офлайн без реклами.

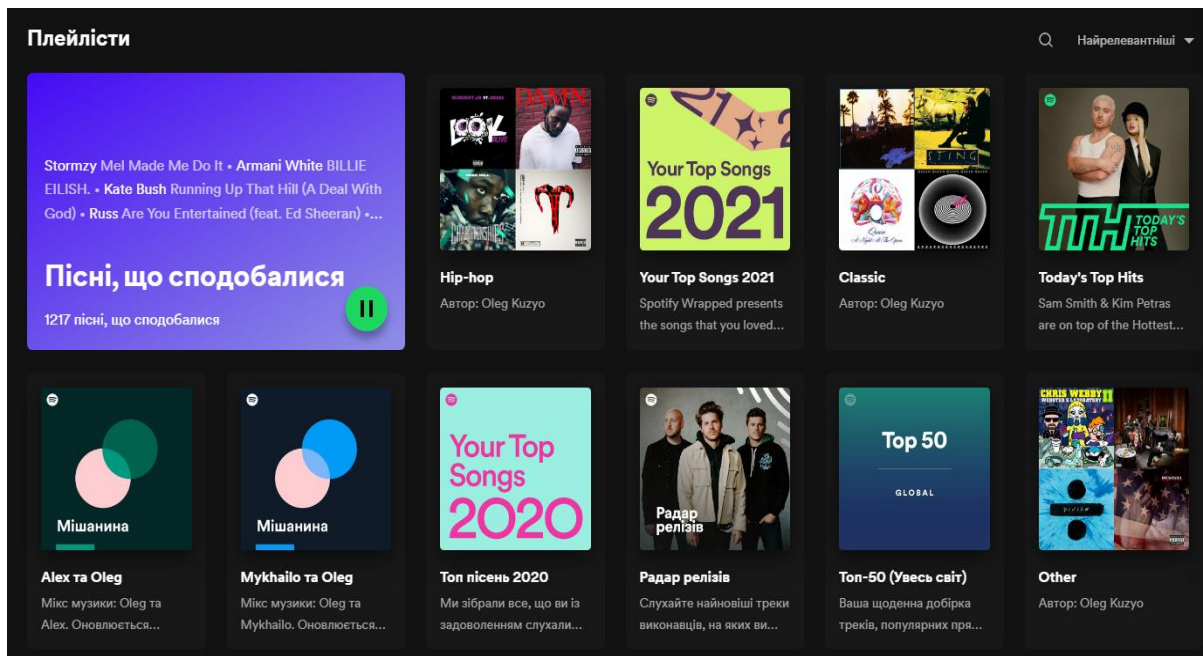


Рис 1.4 Список персонально створених сервісом Spotify плейлистів

Користувачі Spotify високо оцінили персоналізований алгоритм рекомендацій щодо музики на основі смаку сервісу, який працює на основі прослуховування, збереження та лайків. Spotify також відрізняється від інших музичних платформ тим, що має величезний вибір подкастів, деякі з яких ви можете слухати виключно через службу.

MEGOGO — міжнародна платформа OTT для перегляду відео, ТБ та прослуховування аудіо. 55 мільйонів глядачів щомісяця користуються MEGOGO, який володіє понад 25 000 одиниць фільмів та співпрацює з кількома великими студіями, які створюють контент у різних частинах світу. Компанія була заснована в 2011 році, штаб-квартира знаходиться в Києві. Спочатку сервіс був веб-сайтом, потім був створений додаток для телевізорів з функцією Smart TV.

Реалізація програми для пристроїв Android та iOS відбулася восени 2012 року. У 2015 році було запущено власне платне телебачення MEGOGO. Таким чином, до 2021 року компанія запустить 41 канал самостійного виробництва. Весною 2016 року компанія почала розвивати створення

власного оригінального контенту. Влітку 2016 року платформа почала показувати спортивних та культурних заходів.

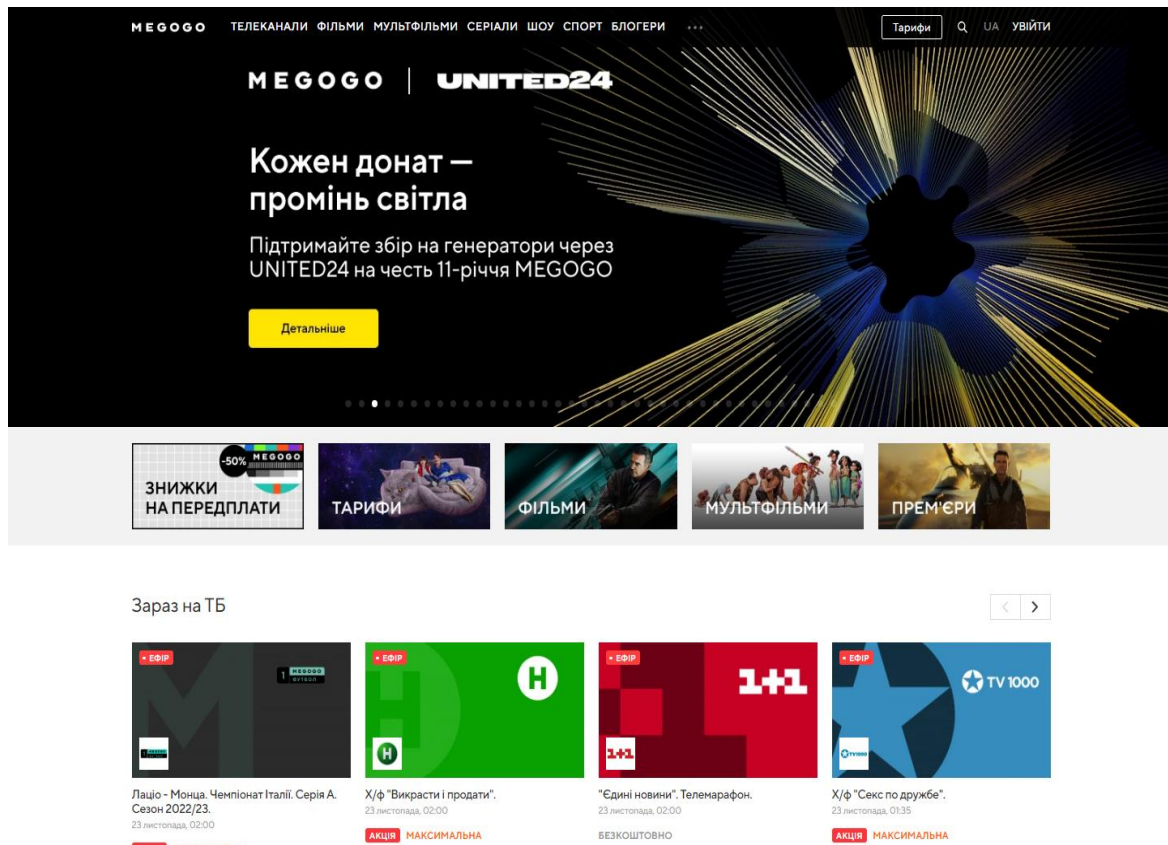


Рис 1.5 Головна сторінка платформи MEGOGO

Взимку 2019 року на платформі з'явився спортивний канал «MEGOGO Футбол», з трансляціями матчів та коментуванням, проект «MEGOGO Audio» де деякий аудіовміст є власного виробництва. З травня 2020 року MEGOGO відкрив аудіоплатформу подкастів, до якої ви можете додати авторський подкаст, відправивши запит Telegram-боту. Літом 2020 року розпочали партнерство з Maincast, кіберспортивні турніри Dota 2 і CS:GO стали доступні на медіаплатформі. В січні У 2021 році компанія отримала права на показ Ліги чемпіонів УЄФА та сезонів чемпіонатів УЄФА до 2024 року.

В розробників даних сервісів є можливість отримувати безліч різних даних від користувачів для подальшого аналізу та використання у власних цілях. Чим більшою популярністю користується продукт, тим більшим стає

потік даних, який розробники отримують. Отже, виникає потреба в консолідації даної інформації для подальшої обробки та використання.

Збір цих даних, дає можливість розробки інформаційної системи, яка їх структурує, аналізує та дає змогу розробникам отримати певні висновки щодо вподобань користувачів, в які періоди дня збільшується активність, популярність жанрів, конкретних виконавців, місцезнаходження користувачів тощо. Отримані дані зберігаються та можуть використовуватися для подальшого використання розробниками в цілях покращення користувацького досвіду, аналізу цих даних для визначення який функціонал потрібен користувачам, формування персональних плейлистів.

Використовуючи даний підхід, розробникам суттєво легше визначати вектори розвитку продукту, отримувати розуміння того, що варто змінити, додати, або ж покращити для збільшення кількості користувачів та проведення аналізу конкурентів.

Також, одним з способів отримання даних від користувачів для подальшого аналізу є зворотній зв'язок від користувачів, за допомогою різного типу опитувань та форм для визначення якості обслуговування, та рівня задоволення клієнта від користування певним функціоналом, або ж враження від переглянутого контенту.

Which of these movies have you seen online video advertising for recently?

| | |
|--|--|
| Probably Interested <input type="checkbox"/> | Definitely Interested <input type="checkbox"/> |
| Definitely Not Interested <input type="checkbox"/> | Probably Not Interested <input type="checkbox"/> |
| None of the above <input type="checkbox"/> | Skip survey <input type="button" value="▶"/> |

YouTube Survey (0:19) ⓘ

Рис 1.6 Приклад опитування на сервісі Youtube

1.2 Концепції та моделі систем консолідації інформації

Термін консолідації інформації визначається процесом об'єднання багатьох обчислювальних потужностей в єдине середовище, призначення якого є виконання певних задач, для прикладу, наукових проблем або ж розрахунків[2].

З часом ця структура накопичує великі обсяги даних, розподілених між обчислювальними вузлами та сховищами. Як правило, програми, що працюють у розподіленому обчислювальному середовищі, посилаються лише на одне з цих джерел даних. Однак виникають труднощі, якщо потрібно отримати доступ до кількох джерел одночасно, адже вони можуть містити дані різного походження і перебувати на деякій відстані одне від одного. Крім того, користувачі, аналізуючи накопичені дані, можуть легко звертатися до єдиного джерела інформації, формувати запити та отримувати результати в тому ж форматі. Тому основною проблемою методів зберігання інформації в

розподілених обчислювальних системах є неоднорідність і віддаленість джерел даних.

Рішення повинні прийматися в режимі реального часу, що вимагає скорочення часу, необхідного для обробки консолідованої інформації. Ці показники можна гарантувати за рахунок підвищення ефективності прийняття управлінських рішень. Удосконалення методів управління обробкою даних може підвищити ефективність прийняття управлінських рішень. Виходячи з особливостей комплексної інформації, з метою скорочення часу її обробки використовується метод диференціації даних, який дозволяє обробляти великі масиви даних. Проте існуючі моделі та методи диференціації даних не можуть повною мірою отримати належне значення показника ефективності. Таким чином, існує протиріччя між обробкою все більшої кількості інформації та вимогою швидкого прийняття рішень та існуючими комплексними моделями та методами обробки інформації[3]. Враховуючи вищезазначене, розробка моделей і методів диференціації даних є актуальним науковим і практичним завданням у процесах консолідації інформації.

Розподілені системи, особливо бази даних, мають тенденцію надавати якусь модель узгодженості. Ці моделі гарантують, що можна очікувати, що система матиме певні властивості (тобто послідовність, доступність або стійкість до збоїв у мережі), якщо виконуються певні умови. У цьому випадку розглядається популярний набір властивостей ACID, який забезпечує надійну обробку транзакцій бази даних порівняно з інверсною моделлю BASE.

Концепція ACID. Традиційно реляційні бази даних були в змозі застосовувати та підкорятися властивостям ACID, набору властивостей, які забезпечують надійну обробку транзакцій бази даних і гарантують, що база даних залишається узгодженою під час одночасного доступу та системних

збоїв. ACID — це аббревіатура, що означає атомарність, цілісність, ізоляція та довговічність.

Атомарність означає, що транзакції бази даних повинні відповідати принципу «все або нічого». Враховуючи, що транзакція — це послідовність інструкцій, які потрібно виконати, для «атомарної» транзакції всі інструкції повинні бути виконані, або якщо одна (чи більше) інструкцій не може бути виконана, вся транзакція завершується невдачею, а база даних повинна залишатися незмінною.

Цілісність, гарантує, що в базу даних записуються лише дійсні дані, у випадку коли база даних переходить з одного узгодженого стану в інший, якщо транзакція вдасться. Якщо під час транзакції через якісь причини з'являються помилки, будь-які внесені зміни автоматично скасовуються (відкочуються), таким чином гарантуючи, що база даних залишається в узгодженому стані.

Ізоляція. Ізоляція вимагає, щоб транзакція, яка виконується і не зафіксувала свої зміни, залишалася ізольованою від будь-якої іншої транзакції. Таким чином, виконання однієї транзакції не повинно впливати на виконання інших паралельних транзакцій. Ізоляція важлива, оскільки стан системи не може бути узгодженим під час виконання транзакції, оскільки транзакція лише гарантує, що система перебуває в узгодженому стані після завершення транзакції. Якщо транзакцію не було розпочато ізольовано, вона зможе отримати доступ до даних із наразі неузгодженої системи.

Довговічність. Довговічність гарантує, що будь-яка транзакція в базі даних є постійною та не може бути втрачена. У разі будь-якого типу збою системи (апаратного чи програмного забезпечення) база даних повинна мати можливість відновлювати підтвержені оновлення, зроблені транзакціями. Багато баз даних досягають довговічності шляхом запису транзакцій у журналі транзакцій, щоб їх можна було повторно зафіксувати, щоб відновити

стан системи до збою. Операція не вважається завершеною, доки вона не буде зареєстрована в журналі.

Модель ACID представляє перевагу узгодженості розподілених баз даних, вона песимістична і гарантує узгодженість наприкінці кожної транзакції. Залежно від вимог предметної області ця суворя відповідність не завжди може бути необхідною. Крім того, цю модель дуже важко реалізувати всередині розподіленої системи без створення вузьких місць[4-16].

Наприклад, двофазний протокол фіксації, який гарантує атомарність розподілених транзакцій, повинен блокувати весь вузол, поки вузол очікує повідомлення. Ця модель також несе багато накладних витрат, оскільки передбачає передачу великої кількості повідомлень. Крім того, якщо ми розглянемо теорему CAP, суворя узгодженість може бути забезпечена лише за рахунок доступності або стійкості до відключень мережі. Якщо ми використовуємо двофазний протокол фіксації, ми можемо забезпечити узгодженість вузла, що призводить до втрати доступності. Оскільки різні програми можуть мати різні вимоги до узгодженості, бази даних NoSQL часто пропонують альтернативні моделі узгодженості.

Концепція BASE, як випливає з назви, є логічним антонімом ACID (гра англійських слів base і acid, що позначають хімічні терміни). ACID більше зосереджується на узгодженості та є песимістичним у тому сенсі, що BASE зосереджується на доступності, забезпечуючи остаточну узгодженість, і в цьому сенсі є оптимістичним. BASE є акронімом для набору властивостей: базова доступність, м'який стан, можлива узгодженість.

Зокрема, якщо розглядати системи інтеграції фінансової інформації більш детально, то слід зазначити, що на сьогоднішній день вони з'являються на ринку у вигляді програмних пакетів програм, корпоративних систем планування ресурсів і стандартних інструментів ETL (англ. – Extract, Transform, Load). - Витяг, конвертація, завантаження)[17].

У контексті аналізу моделей системи консолідації інформації в залежності від використовуваної в системі архітектури зберігання інформації можна виділити наступні методи їх побудови.

1. Міграція структури – це перенесення існуючих аналітичних структур на одну платформу. Такий підхід привабливий для компаній, яким необхідно мінімізувати витрати. Така дія дозволяє зменшити кількість використовуваних серверів, а також розвантажити спеціалістів, які займаються їх обслуговуванням. Слід зазначити, що перенесення структур не передбачає зміни програми чи інтеграції даних. Модель даних, метадані, логіка трансформації та технологія звітності залишаються незмінними, адаптовано лише клієнтські інтерфейси.

2. Проект «з нуля» - розробка системи консолідації інформації не використовуючи готові рішення або розробки.

3. Визначення корпоративного стандарту та реалізація - створення існуючих сховищ даних або сайтів як корпоративного стандарту та оперативна інтеграція стандарту в структури . Такий підхід може бути рекомендований у випадках поглинання однією компанією іншої. Таким чином, сховище даних материнської компанії стає корпоративним стандартом, а сховище даних придбаної компанії вбудовується в нього.

4. Синхронізація. Цей підхід є більш доцільним, якщо компанії потрібно уніфікувати довідкові дані для основних об'єктів даних, таких як клієнти, продукти, постачальники тощо, у великій кількості робочих програм. Досить часто формат даних різниться, та дані зберігаються в окремих операційних програмах. Як результат існує багато копій ідентичних записів, що заважає узгодженості та стандартизації цих даних.

5. Скоординовані сховища даних — це один спосіб консолідації сховищ даних без їх фізичного підключення. Це введення розмірів у конструкцію кожної вітрини, щоб вони відповідали один одному. Цей підхід базується на методології , згідно з якою компанії створюють одну тимчасову

зону розміщення, дані з якої використовуються для заповнення узгоджених сховищ даних. Введення цієї нової проміжної області дозволяє консолідувати надлишкові дані, тим самим зменшуючи витрати.

6. Презентація подання даних. Якщо компанія децентралізована і тільки частина корпоративних користувачів потребує консолідованої інформації, варто створити вітрину даних з усіх існуючих аналітичних структур. Відповідно до цього підходу необхідна корпоративна інформація отримується з існуючих структур за допомогою стандартних засобів ETL. Перевага цього підходу полягає в тому, що немає необхідності змінювати існуючі середовища. Крім того, немає необхідності збільшувати штат спеціалістів і закуповувати засоби розробки, що значно економить кошти.

7. Розподілені запити. Використовуючи даний метод залишаються на місці існуючі неінтегровані аналітичні структури. Однак замість використання ETL і пакетної обробки для створення консолідованого подання, подання створюється за допомогою мови структурованих запитів (SQL)[18].

Водночас слід додатково проаналізувати можливості переміщення готових аналітичних структур на єдину платформу. В даному випадку доцільним є перегляд системи консолідації інформації залежно від моделі уніфікації інформації. У зв'язку з цим можна виділити кілька варіантів побудови таких систем:

- об'єднання систем зберігання даних, розташованих на кількох серверах, в межах одного сервера;
- пряме підключення кількох різнорідних серверів до однієї системи зберігання даних;
- уніфікація системи зберігання на основі архітектури SAN (Storage Area Network)[19].

Особливо з точки зору уніфікації систем зберігання, розташованих на кількох серверах, у межах одного сервера, цей варіант створення систем

консолідації інформації зменшує серйозність адміністративних проблем і економить місце, зайняте обладнанням, завдяки централізації системи зберігання на одному високопродуктивному сервері.

В залежності від того, як моделюється обробка запитів користувачів, то можна виділити наступні актуальні моделі.

1. Модель нечітких множин. Ця модель базується на теорії нечітких множин, яка допускає часткову приналежність елемента множини, на відміну від традиційної теорії множин. У цій моделі весь масив документів описується як набір нечітких наборів термінів. Кожен термін визначає монотонну функцію приналежності до документів даної таблиці. У цій моделі логічні оператори перевизначено для врахування можливості часткової належності до множини.

2. Розширена булева модель, яка є модифікацією булевої моделі. У цій моделі була зроблена спроба зняти обмеження, властиві булевій моделі. Насправді це гібрид булевої моделі та векторної моделі, про які йдеться нижче. Він передбачає, що термін описує зміст документа з певною точністю, яка виражається вагою терміну. Щоб визначити важливість терміну, використовується статистика щодо появи термінів разом із відповідною процедурою нормалізації. Крім того, ваги можна вказати як для термінів документа, так і для термінів запиту.

3. Векторна модель. У цій моделі документи та запити користувачів представлені у вигляді n -вимірних векторів у n -вимірному векторному просторі. Розмірність векторного простору n представляє загальну кількість різних термінів у всіх документах. Все різноманіття словоформ кожного терміну зводиться до якоїсь основи. Офіційні словоформи з низькою інформативністю та високою частотністю не розглядаються.

4. Модель на основі неявного семантичного аналізу. У теорії пошуку інформації ця модель називається неявним семантичним індексуванням. Прихований семантичний аналіз (LSA) — це метод отримання та

представлення контекстно-залежних значень слів шляхом статистичної обробки великих масивів текстових документів. У цій моделі сукупність усіх контекстів, у яких дане слово зустрічається і не зустрічається, встановлює ряд взаємних обмежень, які дозволяють визначити схожість семантичних значень слів і наборів слів один з одним.

5. Імовірнісні моделі. Ці моделі заснували застосовуючи методи теорії ймовірності при пошуку. Вони використовують статистичні показники, які визначають вірогідність відповідності документа запиту користувача. Ці моделі базуються на принципі ймовірнісного ранжування (PRP). Суть цього принципу полягає в ранжуванні документів у наборі за зменшенням вірогідності важливості запиту користувача. Розрахунок цієї ймовірності є важливою часткою моделі, і саме цим відрізняються імовірнісні моделі одна від одної. Особливості імовірнісної моделі полягають у наступному. Модель враховує взаємозалежності та зв'язки термів, визначає їх вагу і форму подібності "запит-документ"[20].

Переважно при об'єднанні даних в системах консолідації інформації використовують дані моделі:

- реляційні;
- ієрархічні;
- мережні;
- об'єктно-орієнтовані.

Сучасна система інтеграції інформації про користувачів відіграє важливу роль у забезпеченні конкурентоспроможності продуктів і може бути використана як компонент системи інтеграції інформації зовнішнього середовища .

1.3 Дослідження завдань систем консолідації інформації

Загалом основними завданнями для систем консолідації інформації є:

- вибір шляхів отримання інформації;
- створення стратегій консолідації інформації;
- якісна оцінка інформації;
- збагачення інформації;
- фільтрація інформації;
- передача та збереження інформації.

При цьому слід враховувати, що в основі роботи будь-якої системи консолідації інформації за етапами формування та використання консолідованої звітності лежать такі завдання:

- збір та структурування вихідної інформації;
- процес інтеграції інформації;
- використання комплексних звітів.

Отже, якщо зупинитися на найважливіших завданнях на поточному рівні застосування систем консолідації інформації користувачів, то серед їх конкретних завдань виділяються:

- збір інформації з багатьох джерел;
- реконструкція клієнтської інформації та приведення структури до уніфікованого вигляду;
- пошук і консолідація однотипних файлів;
- підготовка презентацій даних та розрахунок показників.

Однак, залежно від призначення системи консолідації, її завдання можуть значно відрізнятися[11-14].

У цілому, серед переваг нині існуючих закордонних систем консолідації даних про клієнтів варто виділити такі, як популярність бренду та тісна інтеграція з іншим програмним забезпеченням. В якості недоліків необхідно відзначити високу вартість; порівняно низьку продуктивність і проблеми з кодуваннями.

Більш масштабні системи консолідації інформації про зовнішнє середовище компанії орієнтовані на такі напрями:

- консолідація інформації про сприйняття споживачем запропонованої їй цінності (моніторинг реакції споживача на наданий йому сервіс);

- збір інформації про діяльність контрагентів та конкурентів компанії;

- засоби інтеграції корпоративної інформації.

Якщо проаналізувати існуючі моделі систем інтеграції інформації, то можна отримати декілька їх різновидів:

- система консолідації фінансової інформації;

- залежить від використовуваної в системі архітектури зберігання інформації;

- модель, залежна від комбінованої інформації;

- залежно від способу обробки запитів користувачів;

- система консолідації даних про користувачів;

- залежить від методу аналізу інформації.

При аналізі інформаційної системи зовнішнього середовища підприємству необхідно дослідити модель системи консолідації інформації за методом аналізу інформації[25-27].

Таким чином, залежно від методу аналізу інформації в системі консолідації інформації, можливе застосування наступних методів.

1. Аналіз можливостей (англ. - Opportunity Analysis) може висвітлити ризики та можливості. Використання такого підходу дозволяє зосередити основну увагу на тих заходах, які потрібні вашій компанії для використання нових ефективних методів управління та вдосконалення процесу розробки стратегії.

2. Підхід альтернативного результату (англ. – Alternative Outcomes) даний підхід дає можливість сформулювати декілька обґрунтованих варіантів розвитку подій, а також оцінити довгострокові перспективи розвитку компанії.

3. Аналіз подій (англ. - Event Analysis) такий підхід дає змогу проаналізувати останні події таким чином, щоб визначити отримані переваги над конкурентами в контексті окреслення його загальних і конкретних дій у рамках стратегії поведінки.

4. Аналіз ланцюга (англ.-Linchpin Analysis) традиційно використовується для зміни або повної відмови від базових припущень щодо конкурентів і перегляду власної стратегії. Такий підхід дозволяє змінити усталені стереотипи про конкурентів і об'єктивно оцінити їх поведінку, що призводить до ефективних стратегій поведінки.

5. Аналіз конкуруючих гіпотез (англ. – Analysis of Competing Hypotheses) – можна порівнювати різні аналітичні висновки та пояснення щодо стратегічних дій конкурентів, а також відповідний розвиток деяких гіпотез. Хоча аналіз непотрібних припущень займає багато часу, цей підхід забезпечує важливі результати.

6. Тривале використання методу «копіювання» для стримування і врівноваження конкурентів у стратегічних рамках «попереджувальних ударів»[28].

У цілому, враховуючи вищевикладене, розробка та впровадження системи консолідації інформації користувачів з метою підвищення конкурентоспроможності продукту є доцільною та необхідною[29].

1.4 Висновок до розділу 1

Проведено дослідження стрімінгових систем за такими аспектами:

- різні типи додатків потокового медіа, особливості їх реалізації, спільні риси та ключові моменти роботи;
- проаналізовано найуспішніші існуючі системи потокового передавання медіаконтенту;
- шляхи отримання інформації від користувачів даних сервісів.

Крім того зазначено, що консолідація – це набір методів і процедур, призначених для отримання даних з різних джерел, забезпечення їх інформаційного обсягу та якості на необхідному рівні, щоб вони стали єдиним форматом, який можна завантажити в сховище даних або аналітичну систему.

Проаналізовано та виділено основні завдання інформаційних систем конслодації інформації, визначено основні етапи та існуючі моделі інтеграції таких систем.

У цілому, враховуючи вищевикладене, розробка та впровадження системи консолідації інформації користувачів з метою підвищення конкурентоспроможності підприємства є доцільною та необхідною.

2 АНАЛІЗ МЕТОДІВ КОНСОЛІДАЦІЇ ІНФОРМАЦІЇ КОРИСТУВАЧІВ СТРІМІНГОВИХ СЕРВІСІВ

2.1 Метод консолідації інформації користувачів за допомогою оптимізації трансформації структур Big Data

В сучасних умовах відкриваються можливості для вирішення дуже актуальних завдань, пов'язаних з обробкою великих об'ємів даних від користувачів. Особливо перспективною є інтеграція сучасних інформаційних технологій, пов'язаних із збором та обробкою великих неструктурованих даних (Big Data), інтелектуальним аналізом даних (Data Mining), розподіленими базами даних.

Великі дані (англ. Big Data) – це сучасний напрям у сфері інформаційних технологій, що включає перелік технологій, методик, та інструментів для збору та аналізу великих обсягів різноманітних структурованих і неструктурованих даних з метою отримання ефективних результатів в умовах поширення інформації в мережі багатьох обчислювальних вузлів і їх безперервне оновлення[30]. Також прийнятно використовувати цей термін для даних, оброблених за допомогою цієї технології. Спочатку термін Big Data використовувався в академічних колах для вирішення проблем, пов'язаних із проблемою зростання та різноманітності даних. Перші рішення на основі Big Data з'явилися в другій половині нульових років і розглядалися як альтернатива класичним реляційним СУБД в питаннях бізнес-аналітик[31].

Вже згадана технологія може бути використана у випадках, коли занадто багато даних для обробки традиційними методами, зокрема з використанням реляційного сервера СУБД.

Однак точне кількісне значення інформації, коли дані стають «великими», не визначено. Сфера застосування цієї технології залежить від

можливостей обчислювальної техніки та кількості записів у базі даних, тобто в одних випадках гігабайти даних вже можна вважати великими даними, оскільки комп'ютерна система не може їх обробити вчасно, а в інших випадках петабайти інформації можна прийнятно обробляти за допомогою класичних методів, тому вони не є "великими"[32].

Під аналізом великих даних розуміється аналіз масивів даних у межах можливостей персональних комп'ютерів, а також аналіз у межах можливостей систем керування базами даних. Водночас у першому та другому випадках виникнуть певні труднощі у формуванні та візуалізації масивів даних, тобто необхідність забезпечення злагодженої роботи комп'ютерних програм на великій кількості серверів.

Аналітику Big Data можна охарактеризувати наступними параметрами:

- обсяг, кількість згенерованих даних. Цей показник залежить від того, чи можна вважати масив даних вважати Big Data. Дані традиційно зберігаються на серверах SQL у хмарних середовищах;

- різноманітність, категорія, до якої потрапляють Big Data. Розуміння цієї приналежності дозволяє аналітику найбільш ефективно обробляти інформацію;

- швидкість, з якою дані генеруються або обробляються для досягнення поставлених цілей;

- мінливість, тобто нестабільність на часі даних;

- достовірність, тобто якість зібраних даних, від якої залежить точність аналізу;

- складність, яка полягає у трудомісткості процесу зв'язування та встановлення зв'язків між даними[33].

Розглянемо основні методи аналізу великих даних.

Обробка Big Data в даний час зазвичай передбачає впровадження спеціальних програмних комплексів, таких як Hadoop, які дозволяють обробляти великі обсяги даних на основі концепції Map-Reduce.

Наразі Hadoop є «де-факто» стандартом обробки великих даних. Hadoop — це структура, на якій розробляються програми для аналізу та візуалізації великих даних.

Зберігання даних у цій структурі здійснюється за допомогою спеціальної розподіленої файлової системи HDFS (англ. - Hadoop Distributed File System), яка є основою Hadoop і дозволяє зберігати та отримувати доступ до даних одночасно на кількох вузлах кластера. Завдяки цьому, у випадку виходу з ладу одного або декількох вузлів кластера, ризик втрати інформації мінімізується, і кластер функціонує нормально[34].

Алгоритм Map-Reduce використовується для обробки великих даних, і кожен етап Map має завершити свою відповідну роботу до початку Reduce, а вхідні дані потрібно попередньо обробити. Одним із найактуальніших завдань сучасних інформаційних технологій — швидка обробка великих об'ємів даних. Ефективне рішення кожного завдання має можливість пришвидшити приймання рішень на основі попередньо отриманих даних.

Зі збільшенням можливостей збору інформації за допомогою багатьох різних пристроїв розширюється і кількість даних.

Бази даних є незамінним інструментом для розробки програмного забезпечення, орієнтованого на обробку Big Data. Будь-яка діяльність в інформаційному суспільстві базується на зберіганні та обробці даних. З появою та розвитком нових завдань щодо зберігання великих даних зростає необхідність пошуку рішень для їх вирішення. До недавнього часу реляційна модель була ідеальною, але тепер для деяких конкретних проектів більш підходить нова модель зберігання даних, яка називається нереляційною або NoSQL. В даний час існують сотні різних СУБД, кожна зі своїми перевагами і недоліками, придатних для вирішення певних завдань[35].

Щоб мати можливість вибрати найкращий інструмент для конкретного завдання, потрібні інструменти для тестування баз даних і порівняльного аналізу. Щоб зберігати дані, вам потрібно не тільки знати інформацію про кожен тип бази даних, але також потрібно вміти перевіряти їх у своїх завданнях, тестувати їх[36]. Таким чином, можна:

- систематизувати інформацію на одному стенді;
- надавати засоби тестування та звіти про виконані тестові завдання.

На основі отриманих балів для кожної моделі для вивчення рекомендується вибрати найпопулярніші бази даних:

- Oracle - реляційна система управління базами даних;
- Redis - сховище ключ-значення;
- Cassandra - розподілене сховище;
- MongoDB - документно-орієнтована СУБД;
- Neo4j - база даних на основі графів.

Слід зазначити, що найбільшого поширення набула реляційна СУБД Oracle.

Таблиця 2.1 Рейтинг баз даних

| Модель | Реляційна | Ключ-значення | Розподілене сховище | Документно-орієнтована | На основі графів |
|--------|-------------------------|---------------|-------------------------------------|------------------------|------------------|
| 1 | Oracle | Regis | Cassandra | MongoDB | Neo4j |
| 2 | MySQL | Memcached | HBase | Amazon DynamoDB | Titan |
| 3 | Microsoft SQL Server | Riak KV | Microsoft Azure Table Storage | Couchbase | Giraph |
| 4 | PostgreSQL | Yazelcast | Hypertable | CouchDB | InfiniteGraph |
| 5 | DB2 | Enhache | Google Cloud Bigtable | RethinkDB | Dgraph |

Отже, для обраної дослідницької бази даних наводяться детальні характеристики, характеристики та сфера застосування. Для цього необхідно зрозуміти, як у кожному з них подається інформація. Зокрема, у реляційній моделі дані групуються в колекції або таблиці, які зберігаються в структурованому вигляді та зв'язуються за допомогою атрибутів. Бази даних побудовані на створенні зв'язків між ними, які дозволяють об'єднувати різні таблиці один з одним, перетворюючи їх в єдине сховище.

Сховище ключ-значення представляє дані як асоціативний масив. У такій системі немає необхідності будувати схеми даних і строгі типи даних, також немає зв'язків між значеннями. Завдяки цій простоті система є масштабованою і швидкою.

Документно-орієнтована база даних — це система зберігання ієрархії документів, яку можна представити у вигляді дерева. Деревоподібна структура починається з кореневим вузлом та може містити декілька внутрішніх та листових вузлів. Між даними немає строгих зв'язків, але модель дозволяє створювати складні структури. Для моделі графа дані представлені у вигляді вузлів графа, а зв'язки між ними – у вигляді ребер. Він дозволяє створювати щільні зв'язки між об'єктами без створення жорстких конструкцій[37].

Розподілене сховище зберігає дані, групуючи їх за стовпцями. Рядки та стовпці цієї моделі утворюють розріджену матрицю, а зв'язки між даними не такі суворі, як у реляційній моделі. Як ключі можна використовувати як рядки, так і стовпці. Далі потрібно відмітити важливі для тесту параметри. Для цього вам потрібно зрозуміти, які основні вимоги до бази даних і як їх вивчити.

СУБД повинна забезпечувати цілісність даних до пошкодження та пошкодження, пов'язаного з відмовою технічних засобів, системною помилкою та помилкою користувача. Це включає:

- відновлення даних;

- запобігання помилок при видаленні даних;
- коректна робота в багатокористувацькому режимі.

Бази даних не можуть забезпечити цілісність інформації, що зберігається в них, і тому не можуть вважатися надійними.

Попередні параметри визначають надійність і ефективність СУБД, але навіть найшвидша і найбезпечніша база даних марна, якщо дозволені запити не дозволяють виконувати всі необхідні операції з даними для даного завдання[38]. Тому останнім параметром має бути можливість вказати мову запитів, яка використовується для доступу до бази даних.

Загалом, питання оптимального вибору рішень для зберігання даних стає все більш актуальним для розробки нових моделей баз даних. Представлені аналітичні та порівняльні характеристики дозволяють створювати уніфіковані методи тестування та дослідження СУБД, структури представлення даних та мови запитів, що, у свою чергу, дозволить ефективно використовувати ресурси системи.

Зараз використовується велика кількість платформ і систем великих даних. Системи обробки великих даних — це фреймворки, які потрібно зв'язати з іншими фреймворками.

Нижче наведено деталі методів і технік, які використовуються для великих даних у обчислювальних завданнях.

1. Технологія Big Data. Розроблені програмні платформи та технології розподілених обчислень і великих даних підходять для вирішення масштабних завдань у сферах економіки, бізнесу, соціології, державного управління тощо.

Водночас критично важливі зміни в управлінні даними, ІТ-інфраструктурі та людських можливостях.

У світі Big Data сучасні технології дозволяють обробляти та аналізувати великі обсяги даних, у деяких випадках усі дані, пов'язані з тим

чи іншим явищем (не покладаючись на випадкові вибірки) у їх оригінальній формі – структурованій, неструктурованій, потоковій.

2. Управління Big Data. Стратегічна місія лабораторії хмарних технологій – аналіз Big Data. Досвід застосування цих рішень дозволив створити програмні платформи, які разом із спеціальними пакетами додатків для таких платформ можуть забезпечити якісно нові рівні обробки та аналізу даних, які були б принципово неможливі без використання цих технологій.

3. Аналіз Big Data. Нові аналітичні програми висувають такі вимоги до платформ, які обробляють великі дані:

- поєднувати та керувати даними всіх видів, швидкістю та обсягом, надійністю та доступністю;
- здатність виконувати розширену аналітику інформації в необробленому вигляді;
- візуалізація всіх наявних даних для спеціального аналізу;
- наявність середовища проектування для створення нових аналітичних додатків;
- можливість оптимізації навантаження та планування;
- управління та захист.

4. Персонал Big Data. Одним із найважливіших компонентів є підготовка експертів, які можуть розробляти парадигми великих даних:

- вирішити проблему;
- розробка платформ і рішень на основі платформи;
- системи підтримки та управління;
- запропонувати нові методи.

5. Завдання, які стоять перед лабораторією хмарних технологій та великих даних:

- розгортання платформи управління великими даними;
- створення кластерної, хмарної та/або грид-інфраструктур;
- зберігання, передача, обробка та аналіз великих даних;

- навчання хмарним та грід-технологіям (користувачі, адміністратори та розробники);
- адаптувати прикладні пакети (тобто програми) для роботи в цих інфраструктурах;
- надавати користувачам хмарні ресурси та інфраструктуру;
- вибір рішень і методів для зберігання та управління організаційними даними (SQL, NoSQL (сховище типу ключ-значення, масштабоване розподілене сховище, документно-орієнтована СУБД, графова СУБД);
- управління життєвим циклом даних (створення, обробка, аналіз, систематизація, візуалізація, формування звітів, видалення);
- дослідження в галузі аналітики великих даних - аналітичні методи та прогнозні моделі[39-40].

6. Вимоги до завдань замовника. Big Data означають більше, ніж звичайний аналіз великих об'ємів інформації. Проблеми з'являються, коли організації створюють гігантські обсяги і велика частина цих даних відформатована не традиційно для структурованих баз даних. Все це зберігається в багатьох різних сховищах, іноді навіть за межами організації. У результаті організації мають доступ до величезних обсягів даних, але не мають можливості встановити зв'язки між даними та їх обробити.

2.2 Принцип і архітектура розробки системи обробки Big Data

Досліджені під час виконання даної роботи моделі, методи збору та обробки неструктурованих даних цифрових соціальних середовищ, спрямовані, зокрема, на вирішення завдання практичного збільшення ефективності виконання комплексу адміністративних завдань.

Архітектура систем обробки великих даних використовує складні системи, де можна виділити кілька компонентів або рівнів. Як правило, компоненти таких систем організовані на чотирьох рівнях: прийом, збір,

аналіз даних і представлення результатів. Такий поділ багато в чому умовний, оскільки, з одного боку, кожен компонент у свою чергу може бути розділений на підкомпоненти, а з іншого боку, певні функції компонентів можуть бути перерозподілені відповідно до задачі, що вирішується, і використовуваного програмного забезпечення, наприклад, розділити сховище даних на окремий рівень.

Щоб використовувати великі дані, розробники систем створюють моделі даних, пов'язані з реальним світом. Розробка відповідної моделі даних є складним аналітичним завданням, яке виконують системні архітектори та аналітики. Моделі даних дозволяють створювати математичні моделі взаємодії об'єктів реального світу та включають описи структур даних, методів маніпулювання даними та аспектів підтримки цілісності даних.

Для зберігання даних використовуються різні типи розподілених систем. Це можуть бути механізми доступу до файлових систем, баз даних, журналів, загальної віртуальної пам'яті. Більшість систем зберігання зосереджені лише на обробці великих даних і мають надзвичайно обмежену функціональність, що можна пояснити внутрішньою складністю створення ефективних розподілених систем.

Для швидшої обробки даних системи зберігання та обробки даних розпаралелюються в кластери (кластер — це група комп'ютерів, об'єднаних мережею для виконання одного завдання). Однак, згідно з припущеннями Брюера, неможливо гарантувати одночасну узгодженість даних (відсутність протиріч), доступність даних і стійкість системи до розділення окремих вузлів. Це припущення було доведено для транзакцій ACID (Atomic, Consistent, Isolated, Durable) і відоме як теорема CAP (Consistency, Availability, Partition Tolerance).

1. Поглинання даних (англ.-Data Ingestion). Джерела даних мають різні параметри, наприклад:

- частота, з якою дані надходять від джерела;

- кількість розділів даних;
- швидкість передачі даних;
- Типи вхідних даних та їх достовірність.

Для ефективного збору даних необхідно встановити джерела даних.

Це можуть бути:

- зберігання даних;
- постачальник агрегованих даних;

Як правило, для кожного джерела потрібно створити власний компілятор (Data Crawler для збору інформації про мережу та Data Acquisition для вимірювання). Поглинання даних — це початкова підготовка даних із джерел для перетворення даних у загальний формат представлення даних. Цей єдиний формат вибирається на основі прийнятої моделі даних. Виконувати перетворення системи вимірювань, типізацію (набір), перевірку. Маніпулювання даними не впливає істотно на інформацію, що міститься в даних, але може змінити її представлення (наприклад, привести координати в єдину систему координат, а значення - в єдиний вимір).

2. Збір даних (англ. - Data Staging). Фаза збору даних характеризується безпосередньою взаємодією із системою зберігання даних. Створіть пункт збору, де зібрані дані надають локальні метадані та зберігаються або передаються для подальшої обробки. Дані, які з певної причини не потрапляють у пункт збору, ігноруються.

Для структурованих даних перетворення виконується з вихідного формату відповідно до заздалегідь визначеного алгоритму. Це найефективніша процедура, якщо відома структура даних. Однак, якщо дані представлені в двійковій формі, структура і зв'язок між даними втрачаються, і розробка алгоритмів і програмного забезпечення для обробки даних на їх основі стає надзвичайно складною.

Напівструктуровані дані вимагають інтерпретації вхідних даних і використання програмного забезпечення, здатного використовувати мову опису даних.

Значною перевагою напівструктурованих даних є те, що вони часто містять не тільки самі дані, а й метадані у вигляді інформації про зв'язок між даними та методами, за допомогою яких вони були отримані.

Розробка програмного забезпечення для обробки напівструктурованих даних є досить складним завданням. Однак існує велика кількість готових конвертерів, які можуть, наприклад, витягувати дані з формату XML для формування табличного представлення.

Робота з неструктурованими даними вимагає найбільших зусиль. Для їх перетворення в заданий формат може знадобитися створення спеціального програмного забезпечення, складна ручна обробка, розпізнавання та опціональне ручне керування.

На етапі збору перевіряються типи даних і можна виконати основні перевірки цілісності даних. Наприклад, жодна область не містить молекул газу, координати яких не можуть лежати за межами цієї області, а також їхні швидкості не можуть значно перевищувати швидкість звуку.

Робота з неструктурованими даними вимагає найбільших зусиль. Для їх перетворення в заданий формат може знадобитися створення спеціального програмного забезпечення, складна ручна обробка, розпізнавання та опціональне ручне керування[70].

Типи даних перевіряються на етапі збору, і можна виконати основні перевірки цілісності даних. Наприклад, жодна область не містить молекул газу, координати яких лежать поза межами області, ані швидкості яких значно перевищують швидкість звуку. Щоб уникнути помилок при введенні, необхідно перевіряти правильність налаштування одиниці вимірювання. Наприклад, в одному наборі даних висота може бути в кілометрах, а в іншому

наборі даних – у футах. У цьому випадку необхідно перевести висоту в одиницю вимірювання, прийняту використовуваною моделлю.

Після того, як дані зібрані, вони систематизуються та надаються метадані, що зберігаються у відповідних метаданих. З великою кількістю джерел даних може знадобитися керування збором даних, щоб збалансувати кількість інформації з різних джерел.

3. Аналіз даних (англ. - Analysis Layer). На відміну від збору даних, аналіз даних використовує інформацію, що міститься в самих даних. Аналіз можна проводити в режимі реального часу та в пакетному режимі. Під час роботи з великими даними аналіз даних займає багато часу.

Існує багато методів обробки даних: прогнозний аналіз, запити та звіти, реконструкція математичних моделей, трансляція, аналітична обробка тощо. Ці методи використовують певні алгоритми відповідно до поставлених цілей.

Наприклад, аналітична обробка може бути:

- зображення, соціальні мережі, аналіз геолокації;
- розпізнавання ознак, аналіз тексту, статистична обробка;
- аналіз мовлення;
- транскрипція.

Алгоритми аналізу даних, як і алгоритми обробки даних, залежать від моделей даних. У той же час під час аналізу можна використовувати кілька моделей, які визначають загальний формат даних, але моделюють значущі процеси, за допомогою яких ми обробляємо дані різними способами. При аналізі методів штучного інтелекту, особливо нейронних мереж, динамічне навчання моделей виконується на різних наборах даних.

Під час аналізу даних визначте сутності, описані даними, на основі інформації, наявної в даних, і використаної моделі. Суть аналізу полягає в механізмі аналізу, який використовує алгоритми аналізу, керування моделлю

та розпізнавання сутностей для отримання нової та значущої інформації з аналізу[41-42].

Для аналізу даних також використовуються методи нейронних мереж ШІ, які не розглядаються в цій роботі.

4. Представлення результату (рівень споживання). Рівень споживання надає результати аналізу даних. Існує кілька механізмів, які дозволяють використовувати результати аналітики великих даних.

1) Моніторинг метаінформації – підсистема відображення в режимі реального часу, яка відображає основні параметри системи, навантаження на комп'ютер, розподіл завдань у кластерах, розподіл інформації в сховищі, наявність вільного місця в сховищі, надходження вихідних даних, дії користувача та обладнання. невдача тощо

2) Моніторинг даних — це підсистема відображення в реальному часі для отримання, збору й аналізу даних, а також навігації за даними.

3) Генерувати звіти, запитувати дані, представляти дані на інформаційних панелях у візуалізації (англ. - Dashboard), у форматі PDF, інфографіці, зведених таблицях і короткій довідці.

4) Конвертація та експорт даних в інші системи.

2.3 Використання технології OSINT для формування портрету користувача

Сьогодні, враховуючи стрімкий розвиток інформаційних технологій і соціальних процесів, проблема ефективного виконання управлінських завдань і прийняття рішень постає гостро при роботі з великими масивами неструктурованих різномірних даних на основі індивідуальних соціальних профілів. Це пов'язано зі складністю як вихідних даних, так і структури МІС.

Для моделювання та дослідження таких систем нині широко використовується OSINT (англ. – Open Source Intelligence) – концепції,

методи та прийоми легального отримання та використання інформації з відкритих джерел. Наразі найбільш детальні описи OSINT-контенту містяться в збірниках «NATO Open Source Intelligence Handbook» (2006-2017), «NATO Open Source Intelligence Readers» (2006-2017) та «NATO Internet Intelligence Exploitation» (2002) .



Рис. 2.1 Джерела інформації OSINT

На тлі стрімкого розвитку сучасних інформаційних технологій цей вид пошукової діяльності привертає все більше уваги. Різниця між новачком, який шукає інформацію в Інтернеті, і OSINT-продажем цілком зрозуміла: новачок бачить фотографії, ретвіти, групи та сторінки, на які підписана людина чи організація в соцмережі, експерт бачить події, дату, кількість публікацій, можливі причини підписки на певні групи, кола спілкування. Здебільшого людина використовує один або кілька псевдонімів у мережі, а це означає, що, шукаючи знайдений псевдонім, ви також можете знайти інші дії в Інтернеті, наприклад, у соціальних мережах або на форумах, які відвідує людина. Це лише декілька прикладів. Водночас зібрана інформація є основою для подальшої обробки, очищення (шляхом оцінки достовірності джерела

отримання та достовірності інформації), аналітичного узагальнення та інтерпретації кінцевих результатів.

Тому основною ідеєю OSINT є цілеспрямований збір інформації про об'єкти інтересу (Harvesting) для подальшої обробки та багатовекторного контент-аналізу отриманих даних (створення «портрету» особи, виявлення неочевидних фактів). або зв'язків, передбачити його поведінку тощо).

OSINT зручний тим, що:

- набагато менше ризику: ніхто не порушує приватність і закони;
- ця методика дешевша - не потрібно додаткового обладнання та дорогого програмного забезпечення;
- така інформація є легкодоступною (онлайн) і зазвичай завжди актуальна[43].

Є два основні методи збору інформації.

1. Пасивний. У цьому випадку шукачу інформації неможливо розкрити себе і те, що він шукає. Пошук обмежується вмістом веб-сайту суб'єкта, архівною або кешованою інформацією, незахищеними файлами.

2. Активний. Цей підхід рідко використовується в Інтернет-розвідці. Для отримання інформації було досліджено ІТ-інфраструктуру компанії та активно взаємодіяли з комп'ютерами та машинами. Передові методи використовуються для отримання доступу до відкритих портів, сканування вразливостей і серверних веб-додатків. У цьому контексті інформаційний інтелект легко ідентифікувати. Соціальна інженерія також застосовується тут.

Вибір методу збору інформації залежить від зібраної інформації та необхідних даних. Важливо розуміти, що те, що легко отримати, не завжди є законним.

Небезпечною може бути будь-яка загальнодоступна інформація: соціальні мережі, фотографії, дані зі сторонніх профілів і сайтів, публічні документи тощо. Зрештою, у поєднанні з іншими даними, це може сказати хакерам, що вони шукають[44].

Зупинимось на основних напрямках, які регулярно досліджують фахівці з інформаційної безпеки.

1. Метадані файлу. Серед них можна знайти:

- дата створення документа;
- ім'я користувача;
- модель принтера;
- ППЗ встановлений на комп'ютері;
- геолокація.

Наприклад, інформація про встановлені програми та їх версії дозволить відібрати найбільш вразливі програми та відібрати уразливості. У свою чергу, ім'я користувача стає потенційним ім'ям для входу в особисту або корпоративну систему.

2. Конфіденційні документи. Навіть у найрозвиненіших корпораціях і простих державних установах секретні документи можуть випадково стати публічними. Конфіденційна інформація може включати політику створення паролів, ППЗ і використовувані служби.

3. Дані про домени. Існує багато інструментів, які допомагають збирати всі дані із сайту (включно з тими, які не бачать звичайні користувачі), наприклад:

- електронна пошта,
- телефон,
- факс,
- технологія створення сайту,
- сертифікати шифрування для певних доменів.

Після вивчення основного домену необхідно вивчити, як компанія організовує інтернет-ресурси. У цих субдоменах часто знаходяться менш безпечні сайти для тестування нових технологій. Такі субдомени можуть містити деякі важливі документи, залишені на сервері.

4. Серверний мережевий додаток, Інтернет речей. Сервери, маршрутизатори, камери відеоспостереження, веб-камери, онлайн-накопичувачі тощо можна проіндексувати. Ці пристрої містять технічну інформацію, географічне розташування, відкриті порти, запущені служби, доменні імена, пов'язані з пристроєм, інтернет-провайдера, мережеву технологію, на додаток до деяких, до яких можна просто отримати доступ через посилання.

Процес розробки OSINT виглядає наступним чином

1. Оволодіти базовими технологіями, такими як Google Dorks.
2. Знайдіть цікаві способи використання інструментів і методів і напишіть невеликі звіти з візуальними результатами.

3. Максимальна анонімність. Під час дослідження OSINT багато часу йде на те, щоб забезпечити вашу безпеку під час пошуку. Це необхідно для того, щоб компанії або окремі особи не могли визначити, що певна інформація збирається.

Таким чином, деякі такі практики є такими:

- створювати фальшиві профілі;
- за допомогою емулятора Android;
- VPN;
- браузер Tor;

Нижче наведено деякі ключові інструменти OSINT.

1. Shodan – пошукова система пристроїв, підключених до Інтернету, включаючи IoT і веб-додатки. Розділ «Огляд» допоможе вам почати пошук, оскільки тут збираються запити користувачів. Для доступу до розширеного пошуку необхідно зареєструватися. Платні версії надають доступ до більшої кількості пристроїв і необмежену кількість пошуків на день.

2. MaltegoMaltego - програмне забезпечення, яке об'єднує всі дані, допомагає побачити зв'язки та зробити висновки. Результати візуалізуються у

вигляді дерева, яке збирає IP-адреси, електронні листи, телефони, домени тощо в одній системі.

3. Google Dorks – це запити до Google за допомогою спеціальних операторів.

4. Foca — це програма, яка полегшує завантаження, класифікацію та аналіз файлів на віддалених веб-серверах. Для цього він сканує певні домени за допомогою пошукових систем Google, Bing, DuckDuckGo.

5. Spuse - пошукова система технічної інформації веб-сайту. З його допомогою можна знайти різні дані, такі як уразливості, IP-адреси, субдомени та SSL/TLS.

2.4 Висновок до розділу 2

В даному розділі розглянуто інтеграцію сучасних інформаційних технологій, пов'язаних із збором та обробкою великих неструктурованих даних Big Data та використання розподілених баз даних.

Також, описано основні методи аналізу великих даних - впровадження спеціальних програмних комплексів, таких як Hadoop, які дозволяють обробляти великі обсяги даних на основі концепції Map-Reduce.

Охарактеризовано аналітику Big Data наступними параметрами:

- обсяг (кількість згенерованих даних);
- різноманітність;
- швидкість;
- мінливість, тобто нестабільність на часі даних;
- достовірність, тобто якість зібраних даних, від якої залежить точність аналізу;
- складність, яка полягає у трудомісткості процесу зв'язування та встановлення зв'язків між даними.

Отримано висновки щодо використання нереляційних моделей, таких як NoSQL, та системи обробки великих даних — фреймворків, які потрібно зв'язати з іншими фреймворками.

Визначено основну ідею OSINT — цілеспрямований збір інформації про об'єкти інтересу (Harvesting) для подальшої обробки та багатовекторного контент-аналізу отриманих даних, або зв'язків, для передбачення поведінки об'єкту.

Враховуючи результати даних досліджень, виведено такі переваги OSINT:

- набагато менше ризику: ніхто не порушує приватність і закони;
- ця методика дешевша - не потрібно додаткового обладнання та дорогого програмного забезпечення;
- така інформація є легкодоступною (онлайн) і зазвичай завжди актуальна.

3 ПРАКТИЧНА РЕАЛІЗАЦІЯ ОБГРУНТОВАНИХ МЕТОДІВ І МОДЕЛЕЙ ДЛЯ КОНСЛОДАЦІЇ ІНФОРМАЦІЇ

3.1 Розробка обробки даних користувачів за технологією OSINT

Портрет користувача — це інформаційна структура, що описує і атрибути індивідів або груп. Така інформація має характеристики чіткого людино-машинного сприйняття, що забезпечує можливість автоматичної обробки в різних прикладних завданнях.

За визначенням, користувацьке профілювання — це гетерогенна семантична мережа, що складається з персоналізованих даних. На основі інформаційної моделі портрета базується модель середовища та методика вирішення задачі побудови користувацьких портретів[45].

Описи різних явищ важливі для цілісного уявлення про середовище і для портрету користувача: в маркетингових дослідженнях вони часто зустрічаються в багатьох центральних об'єктах. Такі явища мають кілька характеристик:

- час і місце події;
- список учасників для характеристики явища;
- посилання на джерела, що підтверджують факт виникнення явища;
- супутні інші соціальні об'єкти і явища.

Результатом побудови користувацьких портретів є структуровані дані, витягнуті з динамічного контенту та пов'язані з інфографікою під час аналізу. Вони являють собою безмасштабну мережу, яка для подання та подальшого використання аналітиками повинна зберігатися в базі даних, здатній зберігати графіки та слабо структуровану інформацію.

Відповідно інформаційна модель такого портрета буде представлена таким чином:

$$PPU = \{P(X, v), R(Y, v), Q\}, \quad (3.1)$$

де $P = \{P1, \dots, Pm\}$ ознаки набору об'єктів, пов'язаних з особою, що розглядається (тема, подія, інша особа тощо);

$R = \{R1, \dots, Rn\}$ – набір зв'язків (реалізованих у вигляді предикатів) між соціальними об'єктами цього користувача;

$Q = \{Q1, \dots, Q_{m+n}\}$ – набір оцінок типового тону, ознака, на основі якої будується комплексна оцінка настрою зацікавленої особи Q_{avg} , залежить від кількості ознак, є нечітким значенням, $Q \in [-1; 1]$, m, n - кількість та особливості зв'язків та об'єктів і зв'язків відповідно:

$$m - 1 \leq n; -$$

v - коефіцієнт ваги об'єктів і ознак відносин, який є нечіткою величиною;

X, Y - набори атрибутів об'єктів і зв'язків відповідно:

$$\begin{cases} X = AO \times SO \\ Y = AE \times SE \end{cases} \quad (3.2)$$

де $AO = \{AO1, \dots, AO_f\}$ – набір атрибутів (властивостей), пов'язаних із користувачем;

$AO_i = \{AO1, \dots, AO_{fx}\}$ – набір значень властивостей для окремого визначеного об'єкта;

SO - набір синонімів і повторів для визначень соціальних атрибутів;

$AE_i = \{AE1, \dots, AE_g\}$ – набір значень атрибутів для єдиного визначення відносин об'єктів;

SE — це набір синонімів і повторів, визначених суспільними відносинами;

f, g - зв'язок між кількістю ознак об'єкта;

fx, gy – випадки появи властивостей (включаючи синоніми) для об'єктів і зв'язків між ними.

Традиційно користувача не можна розглядати ізольовано від суспільства. Люди організовують формальні та офіційні співтовариства різних напрямків, і кількість таких спільнот, пов'язаних з конкретною людиною, практично необмежена. Проблема вивчення групової діяльності та атрибутів виникає тому, що риси особистості людини можуть змінюватися під впливом певної соціальної групи, що впливає на результати таких досліджень[46]. При аналізі змін характеристик людини виявляється багато закономірностей, які необхідно формалізувати в інформаційному забезпеченні процесу прийняття рішень щодо управлінських завдань[47]. Щоб вирішити цю проблему, необхідно ввести поняття групового портрету користувача (далі – ГПК), який розглядатиме соціальні тенденції спільноти через деякі спільні атрибути, а також міжособистісні взаємодії членів групи та їх моделі поведінки.

Профіль групи включає індивідуальний профіль, поєднаний багатьма закономірностями індивідуальних характеристик, а також притаманними властивостями спільноти, що розглядається. Характеристики групової діяльності можуть бути статичними і динамічними. Різниця полягає в можливій кількості та типі ключових елементів: вони можуть бути ідентифікаторами людей – лідерів спільноти та ідентифікаторами подій або певними об'єктами в реальному світі, які діють як символи – у цьому випадку люди пов'язуються зі схожими ставленнями до спільного символу[48]. З урахуванням цих особливостей кількість внутрішніх зв'язків у ГПК перевищує загальну кількість зв'язків для всіх користувацьких профілів учасників групи. Відповідно, інформаційна модель портрету громади користувачів буде представлена таким чином:

$$GPU = \{PPU, GPU, GP(X, v), GR(Y, v), GQ\}, \quad (3.3)$$

де $PPU = \{PPU1, \dots, PPUt\}$ — профіль користувача членів групи GPU;

t - кількість ГПК в групі GPU;

$GPU, = \{GPU,1, \dots, PPU,tt\}$ – це набір підгруп, які містяться в користувацькій групі GPU;

tt - кількість підгруп у користувацькій групі ГПК;

$GP = \{GP1, \dots, GPM, \}$ – набір функцій для об'єктів, пов'язаних із групою GPU;

$GR = \{GR1, \dots, GRn, \}$ – сукупність прямих зв'язків групи ГПК;

$GQ = \{GQ1, \dots, GQ m,+n, \} \cup \{QGPU1, \dots, QPPU\} \cup \{QGPU, \dots, GQGPUtt\}$ - набір оцінок тонузу характеристик групи та підгруп і осіб у ній, на основі яких робиться загальна оцінка настрою відповідної спільноти:

$$GQ = \frac{\sum_{l=1}^t Qavg1}{t} + \frac{\sum_{ll=1}^{tt} Qavgll}{tt}$$

(3.4)

що представляє собою нечітку величину, причому $GQ \in [-1;1]$;

m, n , – кількість персональних характеристик і зв'язків групи відповідно, причому $m, -1 \leq n$;

v – вагові коефіцієнти об'єктів і зв'язків групи GPU, що являють собою нечіткі величини;

X, Y – множини властивостей персональних характеристик і зв'язків групи відповідно, причому:

$$\begin{cases} X = AO \times SO \\ Y = AE \times SE \end{cases}$$

(3.5)

де $AO = \{AO1, \dots, AO_f\}$ – набір властивостей, пов'язаних з об'єктами, що належать до певної групи;

$AO_i = \{O1, \dots, O_{fx}\}$ - один визначений набір значень властивостей для об'єктів, що належать до даної групи;

SO - набір синонімів і дублікатів, визначених об'єктами, що належать до даної групи у вихідній інформації;

$AE = \{AE1, \dots, AE_g\}$ - набір властивостей, пов'язаних із персональними установками, що належать до групи;

$AE_i = \{E1, \dots, E_{gy}\}$ - один визначений набір значень атрибутів для певних відносин, що належать даній групі;

SE - набір синонімів і повторів, визначених соціальними відносинами приналежності до групи у вихідній інформації;

f, g – кількість атрибутів об'єктів та кількість зв'язків груп GPU;

fx, gy – час появи об'єктів і пов'язаних атрибутів (включаючи синоніми) групи GPU відповідно.

При аналізі групових профілів користувачі легше визначити неявні атрибути об'єкту. Наприклад, якщо одна особа в групі не надає перевагу конкретному продукту, а інші мають приблизно такі ж побажання, можна зробити висновок, що першій особі притаманні ідентичні переваги[49].

Розроблена інформаційна модель індивідуального та групового портрету користувача дозволяє враховувати всі основні атрибути досліджуваних об'єктів, їх тональність і важливість, аналіз для визначення неявних залежностей. Наступним кроком є перехід до розгляду різноманіття елементів цифрового соціального середовища, тому аналіз користувацьких портретів може більш повно охарактеризувати поведінку індивідів та спільноти в цілому.

У результаті проведених досліджень створено інформаційні моделі індивідуального та групового портрету, що дозволяють аналізувати за допомогою OSINT-техніки визначення неявних залежностей.

3.2 Алгоритм диференційованої обробки даних користувачів

Процес побудови профілю користувача ділиться на кілька етапів, під час яких вирішуються важливі підзавдання:

- шукати та збирати попередню інформацію з загальнодоступних джерел Інтернету за допомогою інструментів великих даних;
- попередня фільтрація зібраної інформації для виключення подальшої обробки конфіденційної та явно некорисної інформації третіх осіб;
- аналіз необроблених даних за допомогою комбінації традиційних інструментів і інструментів аналітики великих даних для створення користувацьких портретів окремих осіб або груп;
- представляти результати аналізу в структурованій формі для автоматизованої обробки.

Для виконання перерахованих етапів запропоновано метод збору та аналізу персоналізованих даних, який розглядає набір характеристик із різних загальнодоступних джерел в Інтернеті та відбирає серед них правдиву, неправдиву та суперечливу інформацію. Він реалізує модель для опису взаємодії між елементами цифрового соціального середовища.



Рис 3.1 Структура профілю користувача в загальному вигляді

Однією з важливих відмінностей технології є поділ великих обсягів інформації на постійно структуровані частини та динамічно неструктуровані частини. Метод також пропонує формалізувати елементи користувацького профілю за допомогою розроблених критеріїв суттєвості. Загальна послідовність кроків для цієї методики представлена таким чином:

- ідентифікувати та шукати об'єкти в соціальному середовищі;
- розбити вихідні дані на статичну і динамічну частини;
- фільтрувати вихідну інформацію від конфіденційних даних;
- виключення даних, неправильно віднесених до досліджуваного користувача;
- аналізувати текстові та мультимедійні дані за критеріями суттєвості;
- визначити тональність образу користувача через сукупність тонів його складових елементів;
- візуалізація кінцевих результатів соціальної аналітики.

Процес створення користувацького профілю починається з визначення та збору вихідної інформації від сервісу. У більшості випадків початок роботи з сервісом починається з введення унікальних критеріїв або набору

критеріїв: імені, електронної пошти, віку, ідентифікатора облікового запису, країна реєстрації, дата народження, та проходження опитування на рахунок переваг в контенті (Рис 3.2). Якщо наявна інформація про предмет дослідження недостатня або суперечлива, можна провести додатковий пошук за непрямими ознаками.

На самому ресурсі вихідна інформація отримується шляхом аналізу активності та пошуку користувачів. Слід зазначити, що збір вихідних даних також може здійснюватися з джерел, не зареєстрованих респондентами, але згаданих у будь-якому контексті.

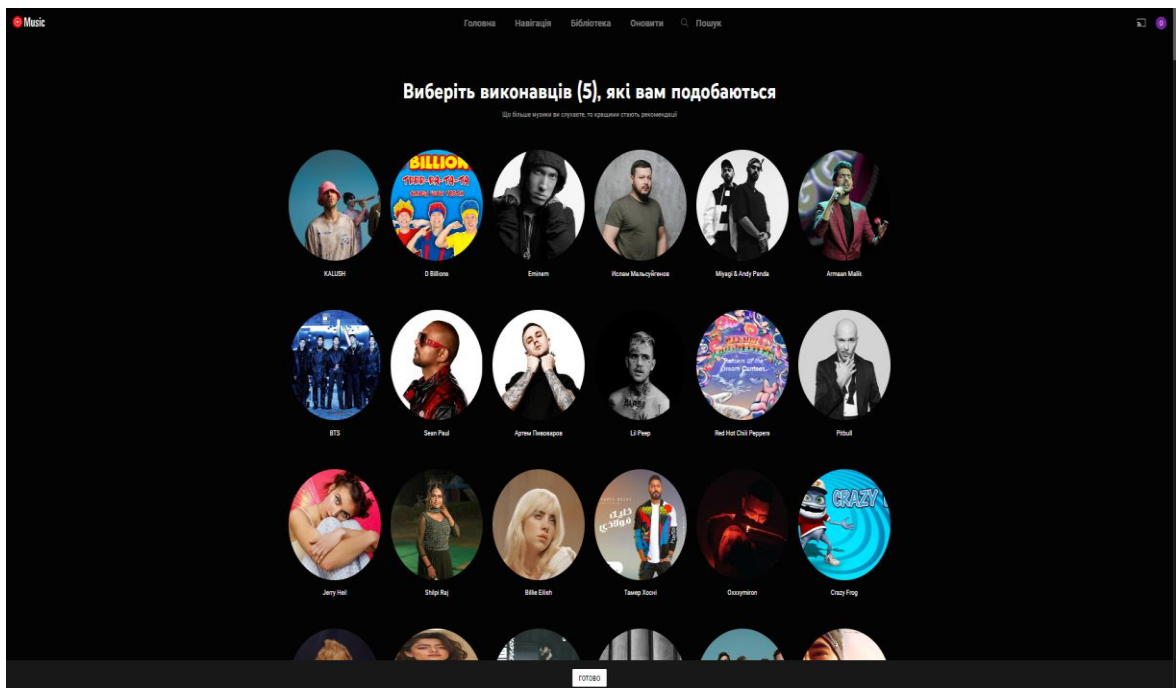


Рис 3.2 Опитування щодо переваг в музиці на Youtube Music

На початку фази збору даних аналітик вводить інформацію та список ключових характеристик досліджуваної особи чи групи в інформаційний граф, представлений у реляційній базі даних. Потім пошукова система автоматично збирає знайдені неструктуровані дані в необробленому вигляді та вказує гіперпосилання на нереляційні сховища. Вихідними даними можуть бути: сторінки, мультимедійні файли, текстові документи, двійкові дані. У

рамках вирішення завдань користувацького аналізу вони класифікуються як динамічний контент, фільтруються та аналізуються з метою вилучення з них структурованої та персоналізованої інформації[50]. Розподілений запис великих даних у сховищі здійснюється фреймворком (Рис 3.3).

Пошукові боти можуть виявляти інформацію про людей, представлену в організованому вигляді, наприклад, у таблиці профілю користувача веб-ресурсу.

Ці дані структуровані та відразу вносяться в інформаційну карту. У процесі фільтрації та аналізу динамічного контенту можна повернутися до етапу збору вихідних даних для уточнення та ідентифікації нововиявлених характеристик об'єкта дослідження.

Отримані на попередньому етапі неструктуровані вихідні дані спочатку непридатні для аналізу через невизначену семантичну повноту їх змісту. Різні соціальні медіа можуть приймати ту саму інформацію, але подавати її з різним рівнем деталізації та спотворювати.



Рис. 3.3. Структурна схема підсистеми збору даних

При автоматизованому зборі даних трапляються випадки коли дані, отримані від користувачів, можуть стати загальнодоступними та потрапити до сховища вихідних даних користувацького профілю на етапі збору. Крім того, можуть виникнути ситуації, коли збір зібраних даних персоналізації (окремо не конфіденційних) формує персональні дані, використання яких порушить недоторканність особистого життя суб'єкта[51]. Підсистема збору даних також може відбирати джерела в Інтернеті, які містять шкідливий вміст. Ці проблеми мають вирішальне значення для можливості впровадження процесу побудови користувацького профілю, і їх потрібно вирішувати, як тільки вони виникають.

Метою етапу фільтрації є усунення технічних труднощів при аналізі та обробці неструктурованих даних, а також вирішення питань конфіденційності та гарантії безпеки використовуваної інформації. Пропонується загальний алгоритм фільтрації вихідних даних користувацьких профілів, який використовується в однойменній підсистемі, структуру якої наведено нижче.

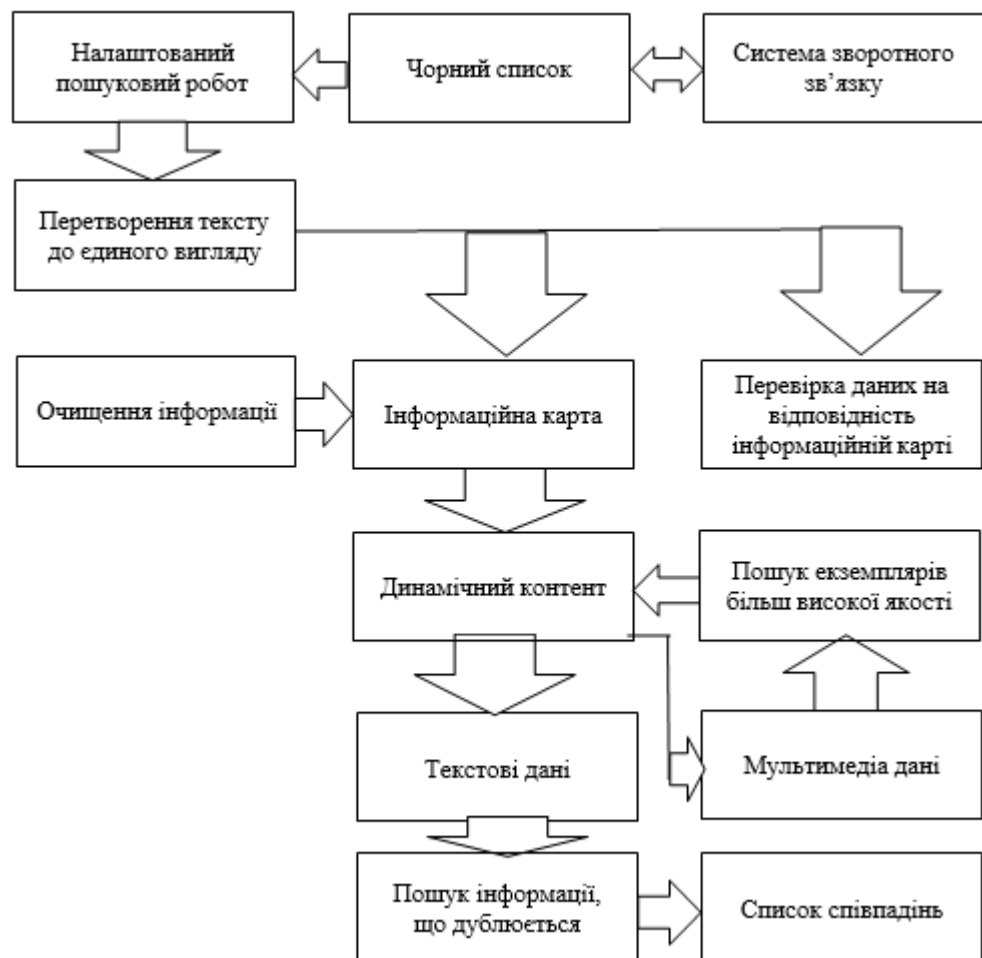


Рис 3.4 Структурна схема підсистеми фільтрації даних

Після фільтрації та ідентифікації аналогів динамічний вміст можна проаналізувати за допомогою поєднання Big Data, Data Mining, машинного навчання та візуальної аналітики.

Наступний крок методу полягає в аналізі всього розмаїття необроблених різнорідних даних і побудові користувацьких профілів у вигляді релевантних структурованих текстових елементів. Окрема проблема – виявлення неявних семантичних залежностей, неоднозначностей, спотворень та інформаційного сміття у вихідній інформації. Сортування початкової інформації профілю передбачає виділення її семантики, а також структурування та розподіл елементів моделі представлення даних користувацького профілю.

Неможливо обробляти мультимедійні дані, як текстові. Для вилучення семантики з нетекстових даних необхідно використовувати спеціальні інструменти. На даний момент існує достатня кількість готових рішень для обробки аудіовізуальної інформації на основі інтелектуального аналізу даних і машинного навчання[52-57].

У рамках завдань користувацького аналізу мультимедійну інформацію можна розглядати як:

- мультимедійні дані, популярні і створені за тематикою - інформація, яка може розповісти про діяльність і переваги людини. Це також включає вміст, захищений авторським правом. Завдання аналізу – порівняти мультимедійні об'єкти з наявними зразками в Інтернеті та користувацьких профілях (наприклад, визначити музичні вподобання за кількома записами);

- вміст, який містить інформацію безпосередньо про відповідну особу. Метою аналізу є отримання необхідної інформації з самого мультимедійного об'єкта.

Перший тип даних аналізується за допомогою контекстного підходу – вміст мультимедійної інформації розбивається на складові та порівнюється між собою. Для другого типу аналізу використовується метод інтерпретації контенту, який, на відміну від контент-методу, дає формальні визначення частинам змісту та вивчає взаємозв'язок між ними.

Тому для ранжування даних у мультимедійному вмісті користувацьких профілів рекомендовано розподіли наступним чином:

- контент розділений за класами;
- контент, призначений для користувача, та інформацію про окремих осіб або спільноти;
- для авторського контенту порівняйте мультимедійні об'єкти з існуючими зразками в Інтернеті та динамічним вмістом користувацьких профілів і витягніть службову інформацію для ідентифікації інформації про діяльність та уподобання відповідних осіб;
- для інформації про осіб провести детальний аналіз контенту з використанням доступних інструментів аналізу мультимедіа, щоб отримати основну інформацію про соціальне середовище з самого мультимедійного об'єкта;
- для всіх типів даних тональний аналіз на основі експертної оцінки самого контенту;
- результати аналізу вносяться в інформаційну карту і прив'язуються до обробленого контенту.

Заключним етапом процесу користувацького профілювання є візуалізація результатів, отриманих на попередньому етапі, у формі, придатній для подальшого використання в різноманітних прикладних завданнях моніторингу та управління соціальними процесами. Остаточне представлення користувацького профілю має бути структурованим, компактним і зручним для обробки людиною-комп'ютером, повністю відображати природу соціальних об'єктів і зв'язок між ними, раціонально використовувати ресурси комп'ютера. У рамках розробленої системи побудови користувацького профілю за виконання перерахованих умов відповідає підсистема, яка представляє результати.

Зважаючи на специфіку структури інформаційної карти та динамічного наповнення користувацького профілю, для візуалізації

результатів використовуються методи теорії графів та мета статистики. Це дає можливість отримати візуальні соціальні графи для експертної обробки, які можна використовувати в різноманітних прикладних дослідженнях[58].

Добре відомі графові алгоритми можуть бути застосовані до результатів, що дозволяє знайти:

- центральний об'єкт;
- найбільш схожий об'єкт;
- зв'язність елементів соціального середовища, найближчих сусідів;
- ступінь незгоди,

та продемонструвати взаємодію між окремими людьми та соціальними групами в різні моменти часу.

На основі аналізу діяльності користувачів, соціальний граф можна перетворити на графік інтересів.

Окрім створення інтерактивних зображень вашого соціального графа, ви також можете брати зразки з баз даних користувацьких профілів, що дає змогу порівнювати їх із довідковими матеріалами, отримувати інформацію про людей із профілів спільнот і будувати статистичні графіки.

3.3 Висновки до розділу 3

За результатами даного розділу, було доведено новий термін портрет користувача — це інформаційна структура, що описує і атрибути індивідів або груп. Також, прийнято визначення користувацького профілювання — це гетерогенна семантична мережа, що складається з персоналізованих даних. На основі інформаційної моделі портрета базується модель середовища та методика вирішення задачі побудови користувацьких портретів.

Розроблена інформаційна модель індивідуального та групового портрету користувача дозволила врахувати всі основні атрибути

досліджуваних об'єктів, їх тональність і важливість, аналіз для визначення неявних залежностей.

У результаті проведених досліджень було створено інформаційні моделі індивідуального та групового портрету, що дозволяють аналізувати за допомогою OSINT-техніки визначення неявних залежностей.

Процес побудови профілю розділено на такі етапи:

- пошук та збір попередньої інформації з загальнодоступних джерел за допомогою інструментів великих даних;
- попередня фільтрація зібраної інформації для виключення подальшої обробки конфіденційної та явно некорисної інформації третіх осіб;
- аналіз необроблених даних за допомогою комбінації традиційних інструментів і інструментів аналітики великих даних для створення користувацьких портретів окремих осіб або груп;
- представлення результатів аналізу в структурованій формі для автоматизованої обробки.

Загальна послідовність кроків для цієї методики представлена таким чином:

- ідентифікація та пошук об'єктів в соціальному середовищі;
- розбиття вихідних даних на статичну і динамічну частини;
- фільтрація вихідної інформації від конфіденційних даних;
- виключення даних, неправильно віднесених до досліджуваного користувача;
- аналіз текстових та мультимедійних даних за критеріями суттєвості;
- візуалізація кінцевих результатів соціальної аналітики.

Також, описано процес створення користувацького портрету. Описано ситуації в яких наявна інформація про предмет дослідження була недостатньою отже, було передбачено можливість проведення додаткового пошуку за непрямыми ознаками.

Визначено такі етапи створення користувачького портрету:

- збір даних аналітиком;
- введення інформації та списку ключових характеристик досліджуваної особи чи групи в інформаційний граф, представлений у реляційній базі даних;
 - розподілення та запис великих даних у сховищі (здійснюється фреймворком);
 - структуризація та внесення даних в інформаційну карту;
 - процес фільтрації даних;
 - візуація отриманого портрету.

4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

4.1 Організація робочого місця користувача ЕОМ

Робоче місце за ЕОМ, що використовується тільки в короткі проміжки часу, може бути організовано при положенні користувача стоячи. Якщо ж користувач постійно завантажений роботою з ЕОМ, більш прийнятною є поза сидячи.

У положенні сидячи основне навантаження припадає на м'язи, що підтримують хребет і голову. При цьому тиск більшої частини ваги тіла припадає на сідниці, не даючи крові проникнути до сідниць. У зв'язку з цим при тривалому сидінні необхідно часто міняти фіксовану робочу позу. Крім того, природний вигин нижньої частини хребта вперед змінюється на вигин назад під час сидіння на роботі, що часто є причиною болю в попереку.

Для фізіологічно правильної робочої пози під час сидіння слідкуйте за тим, щоб частини тіла знаходилися в оптимальному положенні: тіло було прямим, зберігаються природний вигин хребта і кут нахилу тазу[59].

Організаційні вимоги до робочого місця користувача ПК:

Для фізіологічно правильної робочої пози під час сидіння слідкуйте за тим, щоб частини тіла знаходилися в оптимальному положенні: тіло було прямим, зберігався природний вигин хребта і кут нахилу таза. Необхідно уникати сильного нахилу тулуба, повороту голови і крайніх положень суглобів кінцівок.

Вимоги до оформлення робочого місця встановлюються в нормативно-методичних документах відповідно до загальних принципів організації робочого місця. До них належать:

1. Вимоги до дизайну робочого столу, зокрема:

- вимоги до розмірів робочої поверхні обладнання (висота, ширина і глибина робочої поверхні);
- вимоги до простору для ніг (висота, ширина, глибина простору);
- вимоги до параметрів зони розміщення органу керування на пристрої та горизонтальній площині;
- розмір діапазону рухового поля у вертикальній і горизонтальній площинах);
- вимоги до параметрів площі відображення інформації;
- вимоги до взаємного розміщення органів управління та способу відображення інформації.

2. Робочі сидіння (крісла), що відповідають гігієнічним вимогам. Робоче місце та антропометричні параметри, в тому числі:

- вимоги до обробки поверхні і спинки сидіння (нетоксичність, паропроникність, повітропроникність, електрозахист, антистатичність, міцність, морозостійкість, чистота обробки оздоблювального матеріалу);
- вимоги до розміру сидіння (висота, ширина, глибина та кут нахил);
- вимоги до розмірів спинки (висота, ширина постійної поверхні, радіус кривизни спинки та кут нахилу, обмеження регулювання висоти та ширини).

Основним обладнанням на робочому місці користувача комп'ютера є відеомонітор, клавіатура, стільниця, стілець, допоміжним – пюпітр, підставка для ніг, шафа, книжкова полиця тощо. Вимоги до них відображені в нормативних документах: ВСНіПРВЦ; ГОСТ 12.2.032-78; ДСТУ 22269-76 [60].

Під просторовою орієнтацією робочого місця розуміють розташування елементів основного і допоміжного обладнання в певному порядку відносно один одного і працівників. Просторова організація робочого місця визначається в першу чергу розмірами і формою сенсорного і моторного

простору, формою і параметрами елементів робочого місця, просторовим розташуванням елементів відносно працівників.

При розміщенні робочого місця персонального комп'ютера відстань між робочими столами з відеомоніторами (у напрямку екрана за одним і за другим відеомоніторами) повинна бути не менше 2,0 м, а з боків від відеомоніторів – не менше 2,0 м. 1,2 м.

Робочі місця з персональними комп'ютерами в приміщеннях з джерелами шкідливих виробничих факторів розміщуються в ізольованих кабінах з організованим повітрообміном. При виконанні творчої роботи, що вимагає великої розумової напруги або зосередженості, робочі місця з персональними комп'ютерами рекомендується відокремлювати одне від одного перегородками висотою 1,5-2,0 м.

З урахуванням розміру буквено-цифрових знаків і символів екран відеомонітора повинен знаходитися на відстані 600-700 мм від очей користувача, але не менше 500 мм.

Висота робочої поверхні для дорослих користувачів повинна регулюватися в межах 680-800 мм, якщо такої можливості немає, висота робочої поверхні повинна бути 725 мм. На підставі розрахунку конструктивних розмірів слід враховувати модульні розміри стільниці ПК: ширина - 800, 1000, 1200 і 1400 мм, глибина - 800 і 1000 мм, висота необмежена 725 мм.

Робочі місця користувачів ПК повинні бути обладнані підставками для ніг шириною не менше 300 мм і глибиною не менше 400 мм, з діапазоном регулювання висоти до 150 мм і опорною поверхнею з кутом нахилу до 20 градусів для стоячи.

Клавіатуру слід розташувати на стільниці на відстані 100-300 мм від краю, який дивиться на користувача, або на спеціальній робочій поверхні з регульованою висотою, окремо від основної стільниці.

4.2 Види та причини виникнення надзвичайних ситуацій техногенного характеру

Щодня у світі фіксуються тисячі інцидентів, у яких порушуються нормальні умови життя та діяльності людей, що може призвести до загибелі людей та значних матеріальних збитків. Такі події називають надзвичайними ситуаціями (НС).

Положення про класифікацію надзвичайних ситуацій за характером виникнення подій, що призвели до їх виникнення на території України, поділяють надзвичайні ситуації на чотири категорії — техногенні, природні, соціально-політичні та воєнні. Кожен тип надзвичайної ситуації поділяється на групи, що містять конкретні види надзвичайної ситуації(Рис 4.1)

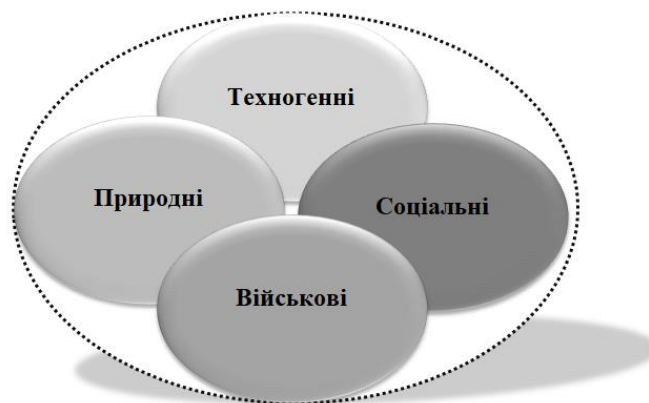


Рис 4.1 - Класифікація надзвичайних ситуацій за характером походження

Розглянемо більш детально надзвичайні ситуації техногенного характеру - це аварії на транспорті (катастрофи), пожежі, неспровоковані вибухи або загрози їх виникнення, шкідливі хімічні речовини, радіоактивність, викиди біологічних речовин (загрози викидів), раптове руйнування споруд і аварії будівель, інженерні мережі та спори покрівлі, життєзабезпечення, дамби, гідродинамічні аварії дамб.

Надзвичайні ситуації техногенного характеру виникають внаслідок раптового виходу з ладу машин, механізмів і агрегатів, що супроводжується серйозним порушенням виробничого процесу, вибухом, утворенням пожежі, радіоактивним, хімічним або біологічним зараженням території, що спричиняє значні матеріальні збитки та враження або смерті.

У міру використання все більших і більших енергетичних потужностей люди змушені концентрувати енергію на невеликих територіях, і найчастіше в межах міст та інших населених пунктів. В результаті забруднення навколишнього середовища, вирубки лісів, все більше людей гинуть від промислових і транспортних аварій.

Аварія - це небезпечна подія техногенного характеру на об'єкті чи окремій території, яка спричинила загибель або загрожує життю і здоров'ю людей і призвела до руйнування будівель, споруд, обладнання та транспортних засобів, порушення виробничих або транспортних процесів або пошкодження. завдають шкоди навколишньому середовищу.

За розміром аварії та завданих збитків вона поділяється на легку, середню, важку та надзвичайно тяжку. Особливо серйозні аварії можуть спричинити величезні збитки з величезними жертвами, тобто катастрофи.

Катастрофи — це великомасштабні аварії, які мають тяжкі наслідки для людей, тварин і рослин і змінюють умови існування. Глобальні катастрофи охоплюють цілі континенти і їх розвиток загрожує виживанню всієї біосфери .

Великі аварії на промислових об'єктах, транспорті тощо, чи то з точки зору кількості збитків, жертв чи характеру наслідків для людей, тварин і рослин, часто дорівнюють або перевищують вплив зброї на масове знищення.

Характер наслідків аварій і катастроф залежить від їх виду, масштабів, населеності території, геолого-географічних умов та інших характеристик.

Основні причини аварій та катастроф:

- конструктивні недоліки, включаючи проектні рішення, які не відповідають вимогам техніки безпеки;
- порушення будівельних норм при будівництві об'єктів та монтажі технічних систем;
- розробка технологічних процесів виробництва без урахування всіх можливих явищ і хімічних реакцій;
- порушення виробничого процесу;
- слабкий контроль над процесом і загальним станом виробництва;
- недотримання правил експлуатації обладнання, машин, механізмів і транспорту;
- недотримання правил зберігання та використання їдких, вибухонебезпечних та пожежонебезпечних речовин;
- фізичне старіння механізмів, конструкцій і матеріалів;
- вихід з ладу обладнання, особливо навігаційного, під час транспортних аварій;
- аварії на прилеглих підприємствах, лініях електропередачі, газопроводах та інженерних мережах;
- несприятливі погодні умови, особливо у разі дорожньо-транспортної пригоди.

Сприяючими чинниками аварій і катастроф є вибухи, пожежі, повені, отруєння людей, обвалення промислових будівель і споруд, ураження людей електричним струмом.

Щорічно в Україні відбувається близько 500 техногенних стихійних лих, які спричиняють близько 400 смертей і 500 травм. Більшість надзвичайних ситуацій є локальними та об'єктними. Випадки загальнодержавного рівня склали близько 1% від загальної кількості ДТП, а регіонального рівня – 4% від загальної кількості ДТП.

Найбільш поширеними надзвичайними ситуаціями є пожежі та вибухи. Найвибухо небезпечнішими є шахти, цукрові заводи,

борошномельні, деревообробні підприємства, для яких характерна велика запиленість. Пил, змішаний з повітрям, з розміром частинок менше 0,5 мм за своїми вибухонебезпечними характеристиками близький до вибуху парів палива або газової суміші. На вугільних шахтах, в тому числі в Україні, регулярно відбуваються вибухи вугільного пилу.

Велика кількість аварій відбувається в системі життєзабезпечення, 12% тепломереж знаходиться в аварійному та аварійному стані, 32% тепломереж повністю вичерпали свій ресурс. 22% від загальної кількості встановлених котелень (електростанцій) експлуатуються більше 20 років, майже 30% каналізаційної мережі є аварійною, а близько 30% мережі комунального водопостачання є аварійною. стан. Наразі закінчився термін експлуатації понад 12 тисяч кілометрів побудованої мережі газопроводів.

Кількість аварій у водопровідній мережі України перевищує відповідний рівень у країнах Європи. Частота пошкоджень водопровідної мережі становить 0,2 на кілометр трубопровідної мережі на рік, а каналізаційної мережі – 0,3 на кілометр на рік, тобто одна аварія трапляється на 5 кілометрів водопроводу та 3 кілометри водопровідної мережі. каналізаційних труб на рік.

Часто надзвичайні ситуації виникають через руйнування будівель і споруд. Сьогодні оцінки технічного стану потребують 115 тис. об'єктів архітектури, визнано непридатними 1800 об'єктів та 700 км інженерної мережі.

Великі несучі металеві та залізобетонні конструкції промисловості, обладнання тощо вичерпали ресурси та потребують заміни. У тому числі 80% енергоблоків електростанцій. Близько 50 000 кілометрів електромережі знаходяться в поганому стані і потребують заміни.

ВИСНОВКИ

Проведено дослідження стрімінгових систем за такими аспектами:

- різні типи додатків потокового медіа, особливості їх реалізації, спільні риси та ключові моменти роботи;
- проаналізовано найуспішніші існуючі системи потокового передавання медіаконтенту;
- шляхи отримання інформації від користувачів даних сервісів.

Крім того зазначено, що консолідація – це набір методів і процедур, призначених для отримання даних з різних джерел, забезпечення їх інформаційного обсягу та якості на необхідному рівні, щоб вони стали єдиним форматом, який можна завантажити в сховище даних або аналітичну систему.

Проаналізовано та виділено основні завдання інформаційних систем конслодації інформації, визначено основні етапи та існуючі моделі інтеграції таких систем.

Також розглянуто процеси інтеграції сучасних інформаційних технологій, пов'язаних із збором та обробкою великих неструктурованих даних Big Data та використання розподілених баз даних, впровадження спеціальних програмних комплексів, таких як Hadoop, які дозволяють обробляти великі обсяги даних на основі концепції Map-Reduce.

Визначено основну ідею OSINT — цілеспрямований збір інформації про об'єкти інтересу для подальшої обробки та багатовекторного контент-аналізу отриманих даних, або зв'язків, для передбачення поведінки об'єкту.

Було введено новий термін портрет користувача — це інформаційна структура, що описує атрибути індивідів або груп. Також, прийнято визначення користувацького профілювання — це гетерогенна семантична мережа, що складається з персоналізованих даних.

Описано процес створення користувацького портрету, відповідно інформаційна модель такого портрета була представлена та доведена

Розроблена інформаційна модель індивідуального та групового портрету користувача дозволила врахувати всі основні атрибути досліджуваних об'єктів, їх тональність і важливість, аналіз для визначення неявних залежностей.

Виведено загальну послідовність кроків для цієї методики та описано процес створення користувацького портрету, визначено такі етапи створення користувацького портрету:

- збір даних аналітиком;
- введення інформації та списку ключових характеристик досліджуваної особи чи групи в інформаційний граф, представлений у реляційній базі даних;
- розподілення та запис великих даних у сховищі (здійснюється фреймворком);
- структуризація та внесення даних в інформаційну карту;
- процес фільтрації даних;
- візуалізація отриманого портрету.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Історія Netflix. Як змінити правила гри та стати головним стрімінгом планети [Електронний ресурс] – Режим доступу: <https://nachasi.com/creative/2020/07/27/netflix-story/>
2. Агальцов В.П. Базы данных. В 2-х т. Т. 2. Распределенные и удаленные базы данных : учебник / В.П. Агальцов. – М. : ИД ФОРУМ, НИЦ ИНФРА-М, 2013. – 272 с.
3. Можаяев М. О., Буслов П. В. Стиснення даних для підвищення ефективності комп'ютерних мереж/ Проблеми інформатизації: тези доповідей восьмої Міжнародної науково-технічної конференції (26–27 листоп. 2020 р.). Черкаси: ЧДТУ; Баку: ВА ЗС АР; Бельсько-Бяла, Польща: УТіГН; Полтава: ПНТУ ім. Юрія Кондратюка, 2020. С.
4. Завадський І.О. Основи баз даних : навч. посіб. / І.О. Завадський. – К. : Видавець І.О. Завадський, 2011. – 192 с.
5. Зарицька О.Л. Базы даних та інформаційні системи : метод. пос. / О.Л. Зарицька. – Житомир : Вид-во ЖДУ ім. І. Франка, 2009. – 132 с.
6. Калитич Г. І. Консолідація інформації, знань і мудрості як проектування і основа гармонійного поступу України / Г. І. Калити // НТІ. – 2008. – № 1. – С. 51.
7. Кеннет Кукієр. Великі дані. Як вони змінюють наші уявлення про світ [Електронний ресурс] / Кеннет Кукієр, Віктор Майєр-Шенбергер. – Режим доступу: <http://www.fundgp.com/ua/events/news/977>.
8. Макарова М. В., Карнаухова Г. В., Запара С. В. Інформатика та комп'ютерна техніка : навч. посіб. / за ред. М. В. Макарової. 3-тє вид., переробл. і допов. Суми : Університетська книга, 2008. – 665 с
9. Можаяев М. О., Буслов П. В. Стиснення даних для підвищення ефективності комп'ютерних мереж/ Проблеми інформатизації: тези

доповідей восьмої Міжнародної науково-технічної конференції (26–27 листоп. 2020 р.). Черкаси: ЧДТУ; Баку: ВА ЗС АР; Бельсько-Бяла, Польща: УТіГН; Полтава: ПНТУ ім. Юрія Кондратюка, 2020. С.

10. Пасічник В. В. Організація баз даних та знань / В.В. Пасічник, В.А. Резніченко. – К.: Видавнича група ВНУ, 2006. – 384 с.

11. Пасічник В. В., Шаховська Н. Б. Сховища даних : навч. посіб. / за ред. В. В. Пасічника. Львів : Магнолія 2006, 2008. – 492 с.

12. Романюк О.Н. Організація баз даних та знань / О.Н. Романюк, Т.О. Савчук. – Вінниця: ВДТУ, 2001.

13. Руденко В. Д. Базы даних в інформаційних системах : навч. посібник / за заг. ред. В. Ю. Бикова. К. : Фенікс, 2010. – 240 с.

14. Сидоренко В.В. Конспект лекцій з предмету СУБД. Навчальний курс [Електронний ресурс] / В.В. Сидоренко. – Режим доступу: <http://dc445.4shared.com/doc/НУВ1Qj5l/preview.html>.

15. Ситник Н. В. Проектування баз і сховищ даних : навч. посібн. – К. : КНЕУ, 2004. – 348 с.

16. Аверченков В. И. Мониторинг и системный анализ информации в сети Интернет : монография / В. И. Аверченков. – М. : Флинта, 2011. – 160 с.

17. Базы данных и знаний. Лабораторный практикум для студентов специальности 1-25 01 12 – Экономическая информатика. – Минск : БГУ, 2015. – 34 с.

18. Буслов П. В. Розробка класифікації систем консолідації комерційної інформації//Вчені записки Таврійського національного університету ім.В.І. Вернацького, серія: Технічні науки, 2018, Т.29 (68)№3. С. 105-108

19. Вандермеер Д. Консолідація / Д. Вандеммер. – М. : Эксмо, 2015. – 384с.

20. Барсегян А. А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / А. А. Барсегян, М. С. Куприянов и др. – СПб. :

БХВ-Петербург, 2009. – 512 с.

21. Додаткові матеріали та програмні засоби з розробки інтелектуальних систем [Електронний ресурс]. – Режим доступу : <http://www.pearsoneduc.com/computing>

22. Завадський І.О. Основи баз даних : навч. посіб. / І.О. Завадський. – К. : Видавець І.О. Завадський, 2011. – 192 с.

23. Зарицька О.Л. Базы даних та інформаційні системи : метод. пос. / О.Л. Зарицька. – Житомир : Вид-во ЖДУ ім. І. Франка, 2009. – 132 с.

24. Калитич Г. І. Консолідація інформації, знань і мудрості як проектування і основа гармонійного поступу України / Г. І. Калити // НТІ. – 2008. – № 1. – С. 51.

25. Гайна Г.А. Основи проектування баз даних. Навч. посібник. - К. : Кондор, 2007. – 208 с.

26. Гужва В.М. Інформаційні системи і технології на підприємствах : навч. посіб. / В.М. Гужва. - К. : КНЕУ, 2001. – 400 с.

27. Дейт К. Дж. Введення в системи баз даних, 8-е видання / К. Дж. Дейт. – М.: «Вільямс», 2005. – 1328 с.

28. Definitive Guide to MongoDB: A complete guide to dealing with Big Data using MongoDB, Third Edition. – Apress, 2015. – 376 с.

29. Великі дані на підприємстві [Електронний ресурс]. – Режим доступу: <http://www.dnipro.net/node/268>

30. M. Chui, V. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. Hung Byers // Big data: The next frontier for innovation, competition, and productivity. – McKinsey Global Institute. - June, 2011. pp. 27-31.

31. А.К. Карнаухов, О.О. Кузьо. Консолідація інформації користувачів за допомогою структур Big Data. XI Міжнародна науково-практична конференція молодих учених та студентів «Актуальні задачі сучасних технологій» Тернопільського національного технічного університету імені Івана Пулюя, 2022, с.141

32. Neil Biehn. The Missing V's in Big Data: Viability and Value //Wired. - [Електронний ресурс] - <https://www.wired.com/insights/2013/05/the-missing-vs-in-big-data-viability-and-value/>.
33. Big Data: A Revolution that Will Transform how We Live, Work, and Think. Viktor Mayer-Schönberger, Kenneth Cukier. Houghton Mifflin Harcourt, 2013 - Business & Economics - 242 p.
34. White T. Hadoop: The Definitive Guide, 4th Edition. - O'Reilly Media, Inc., 2015. - 235 p.
35. Буслов П. В. Аналіз особливостей застосування моделей комунікацій у соціальних групах та віртуальних співтовариствах // Ukrainian Scientific Journal of Information Security, 2017, vol. 23, issue 1, p. 39-44
36. Dozorskyi, V., Nykytyuk, V., Dozorska, O., Dediiv, L., Yavorska, E. The Method of Selection and Pre-processing of Electromyographic Signals for Bio-controlled Prosthetic of Hand. 2020 IEEE 15th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT). Volume 1, Lviv-Zbarazh, Ukraine 23-26 September 2020. P. 188-191.
37. Graph-based intelligence analysis. URL: <https://linkurio.us/blog/graph-based-intelligence-analysis>
38. Дейт К. Дж. Введение в системы баз данных, 8-е видання / К. Дж. Дейт. – М.: «Вільямс», 2005. – 1328 с.
39. СУБД MySQL [Електронний ресурс] – Режим доступу: www.mysql.com .
40. СУБД PostgreSQL [Електронний ресурс] – Режим доступу: www.postgresql.org
41. Переваги та недоліки контент-аналізу в порівнянні з опитуванням [Електронний ресурс] – Режим доступу: <https://ukrbukva.net/print:page,1,13554-Preimushstva-i-nedostatki-kontent-analiza-po-sravneniyu-s-oprosom.html>
42. Досвід Cedos: соціальні дослідження для суспільних трансформацій [Електронний ресурс] – Режим доступу:

<https://cedos.org.ua/researches/dosvid-cedos-soczialni-doslidzhennya-dlya-suspilnyh-transformacij/>

43. О.О. Кузьо, В.К. Крилов, Н.Л. Мацюк Використання технології OSINT для формування портрету користувача. XI Міжнародна науково-практична конференція молодих учених та студентів «Актуальні задачі сучасних технологій» Тернопільського національного технічного університету імені Івана Пулюя, 2022, с.140

44. Duda, O., Kusanets, N., Martsenko, S., Nykytyuk, V., Pasichnyk, V. Information technology platform for the selection and analytical processing of information on COVID-19. 2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT), 2021, pp. 231-238.

45. Marinov M. Four-Dimensional Encoding of Character Sequences and Evaluation of their Similarities and Differences. Proceedings of the Technical University of Sofia. 2020. Vol. 70, Issue 2, P. 1-20.

46. Mozhaiev M, Buslov P., Shvedun V. Development of unclear criteria for determining the significance of a composite social profile information / Захист інформації. 2020 Том 22, № 4. С. 119–126.

47. Буслов П.В. Моделі та методи диференціації даних консолідованої інформації для систем підтримки рішень/Черкаський державний технологічний університет – Режим доступу: https://er.chdtu.edu.ua/bitstream/ChSTU/2067/12/dis_%D0%91%D1%83%D1%81%D0%BB%D0%BE%D0%B2%20%D0%9F.%D0%92..pdf

48. Буслов П.В., Зоренко Д.С., Рябуха Ю.М Використання технологій OSINT для отримання пошукової інформації : практичний poradnik. Х. : ІШЮК для СБ України, 2021. 28 с.

49. Виклики аналітики великих даних [Електронний ресурс] – Режим доступу: <https://uk.wiki-code.net/20518111-challenges-of-big-data-analytics>

50. Mozhaiev M, Buslov P. Development of an Information Model for the

Personality's Social Portrait Formation Using OSINT Technology // Proceedings of the Technical University – Sofia, Volume 70, Issue 4, 2020, P 37–48.

51. Захист персональних даних: правове регулювання та практичні аспекти [Електронний ресурс] – Режим доступу: <https://rm.coe.int/handbook-pers-data-protect-2021-web/1680a37a69>

52. Прикладні й теоретичні методи програмування – Режим доступу:

https://elib.lntu.edu.ua/sites/default/files/elib_upload/ЕНП%20ІПЗ%2019.04.16%20Шолом%20Павло%20Степанович/page11.html

53. Мельников О.В. Регулювання сталого розвитку інформаційної сфери України/— Державне навчально-наукова установа «Академія фінансового управління» Міністерства освіти і науки України – Режим доступу: http://afu.kiev.ua/getfile.php?page_id=1006&num=2

54. Шефер О.В. Моделі та методи підвищення якості функціонування бортових радіолокаційних систем/Полтавський національний технічний університет імені Юрія Кондратюка – Режим доступу: http://lib.kart.edu.ua/bitstream/123456789/3820/1/dis_Shefer.pdf

55. Кіріченко Л.О. Моделі та методи оцінювання параметрів самоподібних і мультифрактальних стохастичних процесів/Харківський національний університет радіоелектроніки – Режим доступу: <https://openarchive.nure.ua/server/api/core/bitstreams/1bffc91d-9b3f-4347-a6e6-2f84b717b1b6/content>

56. П'ятницька Г.Т. Класифікація кластерів у системі інформаційного забезпечення стратегії кластеризації – Режим доступу: https://mmi.fem.sumdu.edu.ua/sites/default/files/mmi2015_4_187_208.pdf

57. Григоруک П.М Роль та складові інформаційної технології в процесі прийняття маркетингових рішень – Режим доступу: <http://elar.khmnu.edu.ua/bitstream/123456789/11756/1/026.pdf>

58. Соціальний граф [Електронний ресурс] – Режим доступу:

https://uk.wikipedia.org/wiki/%D0%A1%D0%BE%D1%86%D1%96%D0%B0%D0%BB%D1%8C%D0%BD%D0%B8%D0%B9_%D0%B3%D1%80%D0%B0%D1%84

59. Організація робочого місця користувача ЕОМ [Електронний ресурс] – Режим доступу: http://ni.biz.ua/3/3_16/3_167306_organizatsiya-rabochego-mesta-polzovatelya-evm.html

60. Гігієнічні вимоги до організації і обладнання робочих місць користувачів комп'ютерів [Електронний ресурс] – Режим доступу: <https://cpo.stu.cn.ua/Oksana/posibnik/1010.html>

ДОДАТКИ

Тези конференції

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Тернопільський національний технічний університет імені Івана Пулюя (Україна)
Університет імені П'єра і Марії Кюрі (Франція)
Маріборський університет (Словенія)
Технічний університет у Кошице (Словаччина)
Вільнюський технічний університет ім. Гедимінаса (Литва)
Міжнародний університет цивільної авіації (Марокко)
Наукове товариство ім. Т.Шевченка

АКТУАЛЬНІ ЗАДАЧІ СУЧАСНИХ ТЕХНОЛОГІЙ

Збірник
тез доповідей

**XI Міжнародної науково-практичної
конференції молодих учених та студентів**
7-8 грудня 2022 року



УКРАЇНА
ТЕРНОПІЛЬ – 2022

Матеріали XI Міжнародної науково-практичної конференції молодих учених та студентів
«АКТУАЛЬНІ ЗАДАЧІ СУЧАСНИХ ТЕХНОЛОГІЙ» – Тернопіль, 7-8 грудня 2022 року

| | | |
|-----|--|-----|
| 6. | В.А. Готович, І.Р. Ралік ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ НА ОСНОВІ КЛІЄНТ-СЕРВЕРНОЇ АРХІТЕКТУРИ ДЛЯ ОБЛІКУ РЕАЛІЗАЦІЇ ТОВАРІВ В ТОРГІВЛІ | 126 |
| 7. | В.В. Ковальчук, І.В. Чихіра, О.В. Готосько КЕРУВАННЯ ПРОЦЕСОМ ДОСЛІДЖЕННЯ ВЛАСТИВОСТЕЙ ПОЛІМЕРКОМПОЗИТНИХ ПОКРИТТІВ ПРИ ЗГИНІ | 127 |
| 8. | А. Хом'як МЕТОДИ АНАЛІЗУ ЧАСОВИХ РЯДІВ З ВИКОРИСТАННЯМ РЕКУРЕНТНИХ НЕЙРОННИХ МЕРЕЖ | 128 |
| 9. | Р.С. Гром'як, С.С. Серкіз ВЗАЄМОДІЯ БІЗНЕС ПРОЦЕСІВ З КЛІЄНТАМИ ЗА ДОПОМОГОЮ CRM- СИСТЕМ | 129 |
| 10. | Р.С. Гром'як, С.С. Серкіз CRM-СИСТЕМА ЯК ІНСТРУМЕНТ УДОСКОНАЛЕННЯ ВЗАЄМОВІДНОСИН З КЛІЄНТАМИ | 130 |
| 11. | В.А. Готович, А.В. Мачужак ЗАСТОСУВАННЯ МЕТОДОЛОГІЇ SI/CD ДЛЯ АВТОМАТИЗАЦІЇ ПРОЦЕСІВ ТЕСТУВАННЯ ТА РОЗГОРТАННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ | 131 |
| 12. | І.В. Струтинська, В.О. Мельник РОЛЬ АВТОМАТИЗОВАНИХ СИСТЕМ КЕРУВАННЯ ЗАМОВЛЕННЯМИ | 133 |
| 13. | Є.Б. Яворська, А.С. Каплунова АЛГОРИТМ ПОДАВЛЕННЯ ЗАВАД В ЕЛЕКТРОКАРДІОСИГНАЛАХ | 135 |
| 14. | Є.Б. Яворська, А.О. Карнов ЗАСОБИ БІОМЕТРИЧНОЇ ІДЕНТИФІКАЦІЇ ОСОБИ У СИСТЕМАХ МОНІТОРИНГУ СТАНУ ЗДОРОВ'Я | 136 |
| 15. | О.М. Петрик, В.О. Суховерша, С.В. Марценко ДОСЛІДЖЕННЯ МЕРЕЖЕВИХ АРХІТЕКТУР ДЛЯ КРИТИЧНИХ ІНФРАСТРУКТУР | 137 |
| 16. | О.М. Петрик, В.О. Суховерша, С.В. Марценко ДОСЛІДЖЕННЯ РОЛІ ІОТ-ТЕХНОЛОГІЙ В ПРОМИСЛОВИХ КОМП'ЮТЕРНИХ МЕРЕЖАХ | 138 |
| 17. | І.В. Воробець ВПЛИВ КОМПОНЕНТІВ ЧАСОВИХ РЯДІВ НА ВИБІР МЕТОДУ ПРОГНОЗУВАННЯ | 139 |
| 18. | О.О. Кузьо, В.К. Крилов, Н.Л. Мацюк ВИКОРИСТАННЯ ТЕХНОЛОГІЇ OSINT ДЛЯ ФОРМУВАННЯ ПОРТРЕТУ КОРИСТУВАЧА | 140 |
| 19. | А.К. Карнаухов, О.О. Кузьо КОНСОЛІДАЦІЯ ІНФОРМАЦІЇ КОРИСТУВАЧІВ ЗА ДОПОМОГОЮ СТРУКТУР BIG DATA | 141 |
| 20. | І.Г. Купратий, А.М. Паламар КОМП'ЮТЕРНА СИСТЕМА ДЛЯ ДИСТАНЦІЙНОГО МОНІТОРИНГУ СТАНУ ЗДОРОВ'Я ПАЦІЄНТІВ | 142 |
| 21. | В. Лісовський, А. Зелінський, О. Сороківський АНАЛІЗ ЗАДАЧ МАШИННОГО АНАЛІЗУ ЗОБРАЖЕННЯ ТА СУЧАСНИХ МЕТОДІВ ЇХ ВИРІШЕННЯ | 143 |
| 22. | А. Зелінський, В. Лісовський ПРОЄКТУВАННЯ ХМАРНОГО РІШЕННЯ ДЛЯ АНАЛІЗУ ПОТОКОВИХ ДАНИХ НА ОСНОВІ AWS KINESIS | 145 |

УДК 004.9

О.О. Кузьо, В.К. Крилов, Н.Л. Мацюк

Тернопільський національний технічний університет імені Івана Пулюя, Україна

ВИКОРИСТАННЯ ТЕХНОЛОГІЇ OSINT ДЛЯ ФОРМУВАННЯ ПОРТРЕТУ КОРИСТУВАЧА

Kuzo O.O., Krylov V.K., Matsuk N.L.

USING OSINT TECHNOLOGY TO FORM USER PORTRAIT

Сьогодні, враховуючи стрімкий розвиток інформаційних технологій і соціальних процесів, проблема ефективного виконання управлінських завдань і прийняття рішень постає гостро при роботі з великими масивами неструктурованих різномірних даних на основі індивідуальних користувацьких профілів.

Для моделювання та дослідження таких систем нині широко використовується OSINT (англ. – Open Source Intelligence) – концепції, методи та прийоми легального отримання та використання інформації з відкритих джерел.

Основною ідеєю OSINT є цілеспрямований збір інформації про об'єкти інтересу (Harvesting) для подальшої обробки та багатовекторного контент-аналізу отриманих даних.

OSINT зручний тим, що: - набагато менше ризику: ніхто не порушує приватність і закони; - ця методика дешевша - не потрібно додаткового обладнання та дорогого програмного забезпечення; - така інформація є легкодоступною (онлайн) і зазвичай завжди актуальна.

Є два основні методи збору інформації:

1. Пасивний. У цьому випадку шукачу інформації неможливо розкрити себе і те, що він шукає. Пошук обмежується вмістом веб-сайту суб'єкта, архівною або кешованою інформацією, незахищеними файлами.

2. Активний. Цей підхід рідко використовується в Інтернет-розвідці. Для отримання інформації було досліджено IT-інфраструктуру компанії та активно взаємодіяли з комп'ютерами та машинами. Передові методи використовуються для отримання доступу до відкритих портів, сканування вразливостей і серверних веб-додатків. У цьому контексті інформаційний інтелект легко ідентифікувати. Соціальна інженерія також застосовується тут.

Вибір методу збору інформації залежить від зібраної інформації та необхідних даних. Важливо розуміти, що те, що легко отримати, не завжди є законним.

Процес розробки OSINT виглядає наступним чином: 1. Оволодіння базовими технологіями, такими як Google Dorks. 2. Пошук цікавих способів використання інструментів і методів і написання невеликих звітів з візуальними результатами. 3. Максимальна анонімність. Під час дослідження OSINT багато часу йде на те, щоб забезпечити вашу безпеку під час пошуку. Це необхідно для того, щоб компанії або окремі особи не могли визначити, що певна інформація збирається.

Література

1. Буслов П.В., Зоренко Д.С., Рябуха Ю.М. Використання технологій OSINT для отримання пошукової інформації : практичний poradnik. X : ІПЮК для СБ України, 2021. 28 с

2. Mozhaiev M, Buslov P. Development of an Information Model for the Personality's Social Portrait Formation Using OSINT Technology // Proceedings of the Technical University – Sofia, Volume 70, Issue 4, 2020, P 37-48.

3. White T. Hadoop: The Definitive Guide, 4th Edition. O'Reilly Media, Inc., 2015, 235p.

*Матеріали XI Міжнародної науково-практичної конференції молодих учених та студентів
«АКТУАЛЬНІ ЗАДАЧІ СУЧАСНИХ ТЕХНОЛОГІЙ» – Тернопіль, 7-8 грудня 2022 року*

УДК 004.9

А.К. Карнаухов, О.О. Кузьо

Тернопільський національний технічний університет імені Івана Пулюя, Україна

КОНСОЛІДАЦІЯ ІНФОРМАЦІЇ КОРИСТУВАЧІВ ЗА ДОПОМОГОЮ СТРУКТУР BIG DATA

Karnaikhov A.K., Kuzo O.O.

CONSOLIDATION OF USER INFORMATION USING BIG DATA STRUCTURES

В сучасних умовах відкриваються можливості для вирішення дуже актуальних завдань, пов'язаних з обробкою великих об'ємів даних від користувачів. Особливо перспективною є інтеграція сучасних інформаційних технологій, пов'язаних із збором та обробкою великих неструктурованих даних (Big Data), інтелектуальним аналізом даних (Data Mining), розподіленими базами даних.

Великі дані (Big Data) – це сучасний вектор у сфері інформаційних технологій і систем, що включає перелік технологій, методів, та інструментів для збору та аналізу великих обсягів різноманітних структурованих і неструктурованих даних з метою отримання ефективних результатів в умовах поширення інформації в мережі багатьох обчислювальних вузлів і їх безперервне оновлення.

Аналітику Big Data можна охарактеризувати наступними параметрами: - обсяг, кількість згенерованих даних. Цей показник залежить від того, чи можна вважати масив даних вважати Big Data. Дані традиційно зберігаються на серверах SQL у хмарних середовищах; - різноманітність, категорія, до якої потрапляють Big Data. Розуміння цієї приналежності дозволяє аналітику найбільш ефективно обробляти інформацію; - швидкість, з якою дані генеруються або обробляються для досягнення поставлених цілей; - мінливість, тобто нестабільність на часі даних; - достовірність, тобто якість зібраних даних, від якої залежить точність аналізу; - складність, яка полягає у трудомісткості процесу зв'язування та встановлення зв'язків між даними.

Обробка Big Data в даний час зазвичай передбачає впровадження спеціальних програмних комплексів, таких як Hadoop, які дозволяють обробляти великі обсяги даних на основі концепції Map-Reduce. Hadoop — це структура, на якій розробляються програми для аналізу та візуалізації великих даних. Зберігання даних у цій структурі здійснюється за допомогою спеціальної розподіленої файлової системи HDFS (англ. Hadoop Distributed File System), яка є основою Hadoop і дозволяє зберігати та отримувати доступ до даних одночасно на кількох вузлах кластера.

Алгоритм Map-Reduce використовується для обробки великих масивів даних, і кожен етап Map має завершити свою відповідну роботу до початку Reduce, а вхідні дані потрібно попередньо обробити. Одним із найактуальніших завдань сучасних інформаційних технологій — швидка обробка великих об'ємів даних. Ефективне рішення кожного завдання має можливість пришвидшити приймання рішень на основі попередньо отриманих даних. Зі збільшенням можливостей збору інформації за допомогою багатьох різних пристроїв розширюється і кількість даних.

Література

1. Кеннет Кукієр. Великі дані. Як вони змінюють наші уявлення про світ [Електронний ресурс] / Кеннет Кукієр, Віктор Майєр-Шенбергер.

2. James Manyika. Big data techniques and technologies / J. Manyika,

M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. Hung Byers // Big data: The next frontier for innovation, competition, and productivity. - McKinsey Global Institute. -

3. Neil Biehn. The Missing V's in Big Data: Viability and Value.