

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ

Кваліфікаційна наукова праця
на правах рукопису

БУСЛОВ Павло Володимирович

УДК 004.04:004.67:004.77

ДИСЕРТАЦІЯ

**МОДЕЛІ ТА МЕТОДИ ДИФЕРЕНЦІАЦІЇ ДАНИХ КОНСОЛІДОВАНОЇ
ІНФОРМАЦІЇ ДЛЯ СИСТЕМ ПІДТРИМКИ РІШЕНЬ**

05.13.06 – інформаційні технології

Подається на здобуття наукового ступеня кандидата технічних наук

Дисертація містить результати власних досліджень. Використання ідей,
результатів і текстів інших авторів мають посилання на відповідне джерело

_____ П.В. Буслов

Науковий керівник: МОЖАЄВ Олександр Олександрович,
доктор технічних наук, професор

Харків – 2021

АНОТАЦІЯ

Буслов П.В. Моделі та методи диференціації даних консолідованої інформації для систем підтримки рішень. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології. – Черкаський державний технологічний університет, Черкаси, 2021.

Дисертаційна робота присвячена вирішенню науково-практичної задачі підвищення оперативності прийняття управлінських рішень шляхом використання моделей та методів диференціації даних консолідованої інформації.

Для досягнення мети дисертації були отримані такі результати досліджень: розроблено інформаційну модель персонального і групового соціальних портретів, на основі OSINT-технології, що дозволило знизити ризики порушення приватності та законів при збереженні повноти та якості вхідних даних; удосконалено метод управління соціальними процесами, що відрізняється від відомих синтезом загальної системної моделі управління в єдину тріадну ієрархічну систему, що дозволило знизити невизначеність у вирішенні завдань вибору різних варіантів управлінських рішень; отримала подальший розвиток модель процесу визначення значущості параметрів соціального профілю, що відрізняється від відомих використанням удосконаленого коефіцієнту значущості, що дозволяє знизити невизначеність вхідних даних та підвищити оперативність прийняття рішень; розроблені у роботі методи та моделі є науково-практичною основою для диференціації консолідованої інформації та розроблені на їх базі алгоритми та програми дозволили зменшити часові витрати на аналіз великих обсягів даних при збереженні рівня об'єктивності оцінювання порівняно з залученням експертів.

Ключові слова: системи підтримки прийняття рішень, диференціація даних, консолідована інформація, нечіткі критерії, OSINT-технології, структури великих даних.

ANNOTATION

Buslov P.V. Methods and models of differentiation of consolidated information data for supporting solutions. – Manuscript. – Qualifying scientific work as the manuscript.

Dissertation on competition of Candidate of Technical Sciences by specialty 05.13.06 – Information Technologies. – Cherkasy State Technological University, Cherkasy, 2021.

The dissertation is devoted to solving the scientific and practical problem of increasing management decision-making efficiency by using models and methods of differentiation of consolidated information data.

The purpose of the dissertation determined the need to set and solve several specific research problems: to analyze existing models and methods of consolidated information data differentiation on the example of the construction of the social portrait, to justify the choice of research objectives; to develop fuzzy criteria to determine the significance of the social profile information components; to develop information models of personal and group social portraits formation based on OSINT-technology of legal obtaining and use of information from open sources; to improve the method of modeling the social profile using the optimization of the transformation of Big Data structures; to investigate the developed models and methods of forming social portraits and develop practical recommendations for their application.

The object of the paper is the processes of consolidated information differentiation. The subject of the paper is models and methods of consolidated information data differentiation.

Currently, decision support systems are widely used in various spheres of human activity. They need to process large amounts of consolidated information to

function. Such processing often has to be done under time constraints and using insufficient computing resources of decision-making information systems. Decision-making systems mainly operate in real-time, which imposes stringent requirements on these systems' timeliness.

The management decision-making speed increase may be obtained by improving the methods of data processing management. Based on the consolidated information characteristics, data differentiation methods, which allow the processing of large amounts of data, are used to reduce data processing time. However, the currently existing models and methods of data differentiation do not allow to fully obtain the proper values of the operability indicators. Therefore, there is a contradiction between the increasing volume of processed information and requirements for speed of decision-making and existing models and methods of processing consolidated information, which determines the relevance of the purpose of this dissertation research

The dissertation research was conducted using the mathematical apparatus of graph theory, database theory, and the concept of non-relational data warehouses, Big Data technology, text analytics technology, parallel data processing methods, methods of research and construction of neural networks, multimedia data analysis methods.

To achieve the aim of the dissertation, the following research results were obtained: an information model of personal and group social portraits was developed, based on OSINT-technology, which allowed reducing the risks of confidentiality violations while maintaining the completeness and quality of the input data; the method of management of social processes has been improved, which differs from the known by the synthesis of the general system model of management into a single triadic hierarchical system, which allowed reducing the uncertainty in solving the problems of selecting different options for management decisions; the model of the process of determining the significance of the social profile parameters has been further developed, which differs from the known ones by using an improved coefficient of significance, which reduces the uncertainty of the initial data and increases the speed of decision-making; the methods and

models developed in the work are a scientific and practical basis for the differentiation of consolidated information, and the algorithms and programs developed on their basis allowed reducing the time spent on the analysis of large amounts of data while maintaining the level of objectivity of the evaluation compared to the involvement of experts.

The practical significance of the results is that the methods and models developed in the work are a scientific and practical basis for the differentiation of consolidated information. The algorithms and programs obtained on their basis made it possible to reduce the time loss for the analysis of large amounts of data while maintaining the level of objectivity of the evaluation by 7% compared to inviting experts.

The developed Big Data processing models allowed carrying out the transformation of structures of large data sets by 5% faster compared to classical models.

The model developed on the basis of OSINT-technology has reduced the time to find the necessary information by 8%.

The dissertation research results were used in the work of Horizont LLC Kharkiv, Radiant LLC Kharkiv, Teploenergosistemy LLC Kharkiv and in the educational process of the EC department of Kharkiv National University of Radio Electronics.

Keywords: decision support systems, data differentiation, consolidated information, fuzzy criteria, OSINT-technologies, Big Data structures.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Васильченков О. Г., Евсина Н. А., Сальников Д. В., Буслов П. В. Реализация фильтра с постфильтрационным принятием решения на микропроцессорных архитектурах с векторным расширением для обеспечения показателей эффективности судебной экспертизы // Вісник Черкаського державного технологічного університету. №4(17). 2020. С.93-102.

Особистий внесок: розробка векторного розширення для показників ефективності судової експертизи.

2. 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.

Особистий внесок: визначення специфіки використання моделей міжособистісної комунікації у соціальних групах.

3. 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.

Особистий внесок: проектування інформаційного забезпечення системи соціального профілювання.

4. 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.

Особистий внесок: вибір апарату формалізації об'єктів дослідження і визначення критеріїв значущості складових інформації соціального профілю.

5. Mozhaiev M., Buslov P. Method of modeling of a social profile using big data structure transformation optimization // Сучасні інформаційні системи 2021. Т.5 №1. С. 12-17.

Особистий внесок: розробка принципів і архітектури систем обробки інформації Big Data.

6. 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

Особистий внесок: побудова класифікації об'єктів систем консолідації інформації.

Наукові праці апробаційного характеру:

7. Pavlo Buslov, Viktoria Shvedun, Volodymyr Streltsov Modern tendencies of data protection in the corporate systems of information consolidation// Problems of Infocommunications. Science and Technology: 5th International Scientific-Practical Conference. Kharkiv, Ukraine, 2018 P. 65-67

8. Можаяєв М. О., Буслов П. В., Мелашенко О. П. Діагностика функціонування розподіленої інформаційної системи судової експертизи./ Застосування інформаційних технологій у діяльності правоохоронних органів : зб. матеріалів круглого столу (09 грудня 2020 р., м. Харків) / МВС України, Харк. нац. ун-т внутр. справ. – Харків : ХНУВС, 2020. – С.80-82.

9. Можаяєв О.О., Буслов П. В. Сучасний стан рекрутингу та його використання для потреб інформаційної системи Національної поліції України /Шлях успіху і перспективи розвитку (до 26 річниці заснування Харківського національного університету внутрішніх справ) : матеріали міжнар. наук.-практ. конф. (м. Харків, 20 листоп. 2020 р.) / редкол.: Д. В. Швець (голова), О. М. Бандурка, С. М. Гусаров та ін.; МВС України, Харків. нац. ун-т внутр. справ. Харків : ХНУВС, 2020. С. 345-347.

10. Можаяєв 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. 18.

11. Буслов П.В. Розробка класифікації систем консолідації інформації в комерційній діяльності підприємств/Актуальні питання забезпечення кібербезпеки та захисту інформації: тези доповідей учасників IV міжнародної

науково-практичної конференції (21-24 лютого 2018 р.). Київ. Європейський Університет.2018.- С.29.

12. Буслов П.В. Дослідження сучасного стану та перспектив розвитку систем консолідації інформації/ Modern methods, innovations, and experience of practical application in the field of technical sciences: conference proceeding (27-28 december 2017) Radom, Republic of Poland: Radom Academy of economics. 2017. P. 12-15.

ЗМІСТ

| | |
|--|-----|
| ВСТУП..... | 11 |
| РОЗДІЛ 1. СУЧАСНИЙ СТАН ТА СПЕЦИФІКА РОЗВИТКУ СИСТЕМ КОНСОЛІДАЦІЇ ІНФОРМАЦІЇ..... | 16 |
| 1.1. Аналіз сучасних тенденцій в сфері консолідації інформації в групах користувачів в мережі Інтернет..... | 16 |
| 1.2. Аналіз сучасних моделей систем консолідації інформації..... | 29 |
| 1.3. Аналіз завдань систем консолідації інформації..... | 34 |
| Висновки до першого розділу..... | 46 |
| РОЗДІЛ 2. УДОСКОНАЛЕННЯ МЕТОДІВ КОНСОЛІДАЦІЇ ІНФОРМАЦІЇ У ПРОЦЕСАХ СОЦІАЛЬНОГО ПРОФІЛЮВАННЯ..... | 48 |
| 2.1. Метод моделювання соціального профілю з використанням оптимізації перетворення структур Big Data..... | 48 |
| 2.2. Метод управління соціальними процесами на основі з'єднання тріад..... | 67 |
| 2.3. Застосування методу системологічного класифікаційного аналізу для розробки фрагменту концептуальної моделі функцій системи консолідації інформації..... | 73 |
| Висновки до другого розділу..... | 79 |
| РОЗДІЛ 3. РОЗРОБКА МОДЕЛЕЙ ПРЕДСТАВЛЕННЯ ДИФЕРЕНЦІАЛІЗОВАНОЇ ІНФОРМАЦІЇ В ІНТЕЛЕКТУАЛЬНИХ СИСТЕМАХ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ..... | 80 |
| 3.1. Розробка нечітких критеріїв для визначення значущості складових інформації соціального профілю..... | 80 |
| 3.2. Розробка моделі представлення знань про соціальний профіль... | 106 |
| 3.3. Розробка моделі управління знаннями у соціальній групі..... | 116 |
| Висновки до третього розділу..... | 126 |

| | |
|---|-----|
| РОЗДІЛ 4. ПРАКТИЧНА РЕАЛІЗАЦІЯ РОЗРОБЛЕНИХ МЕТОДІВ ТА МОДЕЛЕЙ ПІДТРИМКИ РІШЕНЬ ДЛЯ ДИФЕРЕНЦІАЦІЇ ДАНИХ КОНСОЛІДОВАНОЇ ІНФОРМАЦІЇ..... | 127 |
| 4.1. Розробка принципів і архітектури систем обробки інформації Big Data..... | 127 |
| 4.2. Алгоритм обробки диференціалізованих даних для соціального профілювання..... | 134 |
| 4.3. Побудова інформаційної моделі формування соціального портрета особистості за допомогою технології OSINT..... | 144 |
| Висновки до четвертого розділу..... | 162 |
| ВИСНОВКИ..... | 164 |
| СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ..... | 166 |
| ДОДАТКИ..... | 176 |

ВСТУП

Актуальність теми. На даний час системи підтримки прийняття рішень широко використовуються в різних сферах життєдіяльності людини. Основним завданням цих систем є вибір із заданої множини альтернатив деякої частини, яка на думку особи, що приймає рішення, і результатом незалежних експертних оцінок, є кращими, ніж інші. При цьому необхідно зробити вибір критеріїв за якими проводиться оцінка та визначити вагомість кожного критерію. Дуже часто таких критеріїв буває значна кількість, що істотно ускладнює роботу системи.

Найчастіше інформація, яка повинна бути оброблена і на підставі якої необхідно створити управлінське рішення є консолідованою. Необхідність обробки значних обсягів консолідованої інформації, в умовах обмежених ресурсів сучасного інформаційного простору тягне за собою різного роду завдання, пов'язані з недоліком програмних та апаратних обчислювальних ресурсів інформаційних систем. Прийняття рішень необхідно проводити у режимі реального часу, що потребує зменшення часу необхідного на обробку консолідованої інформації. Забезпечити ці показники можна за рахунок підвищення оперативності прийняття управлінських рішень. Підвищення оперативності прийняття управлінських рішень можна отримати за рахунок удосконалення методів управління обробкою даних. Виходячи з особливостей консолідованої інформації, для зменшення часу на її обробку, використовуються методи диференціації даних, що дозволяють обробити значні масиви даних. Однак, існуючі на даний час, моделі та методи диференціації даних не дозволяють у повному обсязі отримати належні значення показників оперативності. Тому виникає протиріччя між підвищенням обсягу оброблюваної інформації і вимогами до оперативності прийняття рішення та існуючими моделями і методами обробки консолідованої інформації. Враховуючі вище наведене виникає актуальна

науково-практична задача розробки моделей і методів диференціації даних консолідованої інформації.

Зв'язок роботи з науковими програмами, планами, темами.

Дисертаційну роботу виконано в Харківському національному університеті радіоелектроніки в рамках плану науково-технічної діяльності університету, в рамках наукових напрямків Харківського національного університету Повітряних Сил імені Івана Кожедуба, планів наукової і науково-технічної діяльності Повітряних Сил Збройних Сил України та у рамках науково-дослідних робіт (НДР): "Ореанда-Пошук" ДР № 0118U000135д, "Моніторинг" ДР № 0101U001015, "Візуалізація" ДР № 0101U001291.

Мета і задачі дослідження. Метою дисертаційної роботи є підвищення оперативності прийняття управлінських рішень шляхом використання моделей та методів диференціації даних консолідованої інформації.

Зазначена мета зумовила необхідність постановки та вирішення таких завдань:

- провести аналіз існуючих моделей та методів диференціації даних консолідованої інформації на прикладі побудови соціальних портретів, обґрунтувати вибір напрямку дослідження;
- розробити нечіткі критерії для визначення значущості складових інформації соціального профілю;
- розробити інформаційні моделі формування персонального і групового соціальних портретів на основі OSINT-технології легального отримання і використання інформації з відкритих джерел;
- удосконалити метод моделювання соціального профілю з використанням оптимізації перетворення структур Big Data;
- дослідити розроблені моделі та методи формування соціальних портретів та розробити практичні рекомендації щодо їх застосування.

Об'єктом дослідження є процеси диференціації консолідованої інформації.

Предметом дослідження є моделі та методи диференціації даних консолідованої інформації.

Методи дослідження. В рамках даного дисертаційного дослідження виявляється необхідність вивчення і використання ряду технологічних і теоретичних засобів: математичного апарату теорії графів, теорії баз даних і концепції нереляційних сховищ даних, технології Big Data, технологій текстової аналітики, методів паралельної обробки даних, методів застосування нейронних мереж, методів аналізу мультимедіа даних.

Наукова новизна отриманих результатів полягає в тому, що:

- вперше розроблено інформаційну модель персонального і групового соціальних портретів, на основі OSINT-технології, що дозволило знизити ризики порушення приватності та законів при збереженні повноти та якості вхідних даних;;

- удосконалено метод управління соціальними процесами, що відрізняється від відомих синтезом загальної системної моделі управління в єдину тріадну ієрархічну систему, що дозволило знизити невизначеність у вирішенні завдань вибору різних варіантів управлінських рішень;

- отримала подальший розвиток модель процесу визначення значущості параметрів соціального профілю, що відрізняється від відомих використанням удосконаленого коефіцієнту значущості, що дозволяє знизити невизначеність вхідних даних ті підвищити оперативність прийняття рішень при збереженні рівня об'єктивності оцінювання.

Практичне значення отриманих результатів полягає в тому, що розроблені у роботі методи та моделі є науково-практичною основою для диференціації консолідованої інформації. Розроблені на їх базі алгоритми та програми дозволили зменшити часові витрати на аналіз великих обсягів даних при збереженні рівня об'єктивності оцінювання порівняно з залученням експертів на 7%.

Розроблені моделі обробки Big Data дозволили здійснити перетворення структур великих масивів даних на 5% швидше порівняно з класичними моделями.

Розроблена модель на основі OSINT-технології дозволила зменшити час на пошук необхідної інформації на 8%.

Результати дисертаційних досліджень використано в роботі ТОВ «Горизонт» м. Харків, ТОВ «Радіант» м. Харків, ТОВ «Теплоенергосистеми» м. Харків та в учбовому процесі кафедри ЕОМ Харківського національного університету радіоелектроніки

Особистий внесок здобувача. Дисертаційна робота є завершеним самостійним науковим дослідженням автора, що містить теоретичні положення, практичні розробки, висновки та пропозиції, які отримано й сформульовано особисто автором, та які в комплексі уможливають вирішення важливого наукового завдання щодо вдосконалення процесів підтримки рішень для диференціації даних консолідованої інформації. У наукових працях, написаних у співавторстві, здобувачу належать такі результати: розроблено векторне розширення для показників ефективності судової експертизи; визначено специфіку використання моделей міжособистісної комунікації у соціальних групах; спроектовано інформаційне забезпечення системи соціального профілювання; запропоновано апарат формалізації об'єктів дослідження і визначено критерії значущості складових інформації соціального профілю; розроблено принципи і архітектуру систем обробки інформації Big Data; побудовано класифікацію об'єктів систем консолідації інформації.

Апробація результатів дисертації. Ключові положення дисертаційної роботи було оприлюднено та обговорено на: IV міжнародній науково-практичній конференції Європейського Університету «Актуальні питання забезпечення кібербезпеки та захисту інформації», 21-24 лютого 2018 р., м. Київ; міжнародній науково-практичній конференції «Шлях успіху і перспективи розвитку (до 26 річниці заснування Харківського національного

університету внутрішніх справ)», 20 листопада 2020 р., м. Харків; круглому столі Харківського національного університету внутрішніх справ, «Застосування інформаційних технологій у діяльності правоохоронних органів», 09 грудня 2020 р., м. Харків; VIII міжнародній науково-технічній конференції Черкаського державного технологічного університету «Проблеми інформатизації», 26–27 листопада 2020 р, м. Черкаси; international research and practice conference Radom academy of economics «Modern methods, innovations, and experience of practical application in the field of technical sciences», 27-28 december 2017, Radom, Republic of Poland; 5th International Scientific-Practical Conference «Problems of Infocommunications. Science and Technology», 2018, Kharkiv, Ukraine.

Публікації. Основні наукові положення та здобутки дисертаційного дослідження викладено в 12 друкованих працях, зокрема в 6 статтях у наукових фахових виданнях України та інших держав (з них 1 стаття входить до міжнародної наукометричної бази Web of Science, 1 в закордонному журналі (Болгарія), 4 статі у наукових фахових журналах України) та в 6 публікаціях у матеріалах міжнародних наукових конференцій, з яких 1 індексується у міжнародній наукометричній базі SCOPUS.

Структура дисертації. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел і додатків. Дисертацію викладено на 180 сторінках, з них 165 сторінок основного тексту. Робота включає 10 рисунків, 12 таблиць та 3 додатки (на чотирьох сторінках). Список використаних джерел налічує 117 найменувань (на 11 сторінках).

РОЗДІЛ 1. СУЧАСНИЙ СТАН ТА СПЕЦИФІКА РОЗВИТКУ СИСТЕМ КОНСОЛІДАЦІЇ ІНФОРМАЦІЇ

1.1 Аналіз сучасних тенденцій в сфері консолідації інформації в групах користувачів в мережі Інтернет

У сучасних умовах питанням формування та функціонування соціальних груп присвячено наукові праці таких авторів, як: Р. Мертон, Л. С. Виготський, Р. Л. Кричевський, Е. М. Дубовська та ін.

Так, Р. Мертон пропонує визначати соціальну групу в якості сукупності людей, «які певним чином взаємодіють один з одним на основі встановлених зв'язків, усвідомлюють свою приналежність до групи і вважаються її членами» [20].

Вчений Л. С. Виготський досліджує співвідношення соціальної й колективної психології та вважає, що тільки частина психології входить до психології групи [7].

Автори Р. Л. Кричевський та Е. М. Дубовська присвячують власне дослідження виключно малим соціальним групам, не розглядаючи їх інші різновиди [16].

При цьому слід зазначити, що недостатньо дослідженими залишаються проблеми застосування комунікативних моделей в межах функціонування соціальних груп, зокрема, віртуальних співтовариств, в залежності від їх особливостей.

Виходячи з вищезазначеного, необхідним є дослідження особливостей застосування моделей комунікації у соціальних групах та віртуальних співтовариствах.

Первинне поняття групи було пов'язано з її розумінням як відомої кількості людей, об'єднаних в межах спільних інтересів чи справ. У сучасних умовах група є ключовим компонентом суспільної соціальної структури, відповідно традиційно вона розуміється в якості сукупності людей, яких об'єднує визначена істотна ознака (наприклад, спільна діяльність чи спільні характеристики економічного, демографічного, етнографічного, психологічного типу). Зазначене поняття є традиційним для більшості наук, проте в соціології традиційно застосовується поняття «соціальна група».

Соціальною групою вважається сукупність людей, якій притаманна відносна стійкість. При цьому необхідно зазначити, що люди, які входять до складу соціальної групи, мають спільні інтереси, цінності та поведінкові норми, що формуються в межах історичного розвитку суспільства. У межах кожної з соціальних груп втілено визначені взаємозв'язки індивідів як між собою, так і з суспільством в цілому. При цьому необхідно зазначити, що ці взаємозв'язки можуть підлягати регулюючому впливу як формальних, так і неформальних соціальних інститутів. У залежності від особливостей взаємозв'язків у соціальних групах формуються внутрігрупові норми поведінки.

Соціальним групам притаманні такі особливості: стабільність існування завдяки стійкій взаємодії їх членів; достатній ступінь згуртованості й єдності; однорідність складу; можливість бути структурною одиницею масштабніших соціальних спільнот.

Виділяють наступні категорії соціальних груп (табл. 1.1) [26].

Зокрема, у межах первинної групи взаємозв'язки між її членами характеризуються безпосередністю та міжособистістю з наявністю високого ступеня емоційності. Відповідно, враховуючи функцію соціалізації індивіда, первинну групу можна вважати ланкою, що сполучає особистість із суспільством.

Таблиця 1.1

Категорії соціальних груп

| Найменування ознаки | Різновиди | Приклади |
|--|---------------------|---|
| Відповідно до характеру взаємодії | Первинні | Сім'я, група ровесників |
| | Вторинні | Організації виробничого, політичного, релігійного спрямування |
| Відповідно до способу здійснення взаємодії | Формальні | Підприємства, організації |
| | Неформальні | Клуби за інтересами, дружні компанії |
| Відповідно до кількісного складу | Малі | «Діада» |
| | | «Тріада» |
| | Великі | Трудові колективи, підприємства |
| | | Організації соціально-класового, професійного, політичного типу |
| Відповідно до соціально значимих ознак | Реальні | Розподіл за статтю, віком, доходом, національністю, сімейним станом, професією, місцем проживання |
| | Номінальні (умовні) | Студенти, пасажери-пільговики, пенсіонери |

Джерело: складено на підставі [26]

Вторинна група містить більше членів та характеризується їх підпорядкованістю для досягнення визначеної мети. Вторинним групам притаманні безособовість і формальність. Ключова увага у зазначених групах приділяється умінню членів групи виконувати визначені функції замість їх унікальних якостей.

Що стосується формальних груп, то у них присутній юридичний статус, відповідно, взаємодія їх членів регулюється за допомогою системи встановлених норм, правил і законів. У формальних групах присутня свідомо поставлена мета та нормативно закріплена ієрархічна структура. Вони функціонують відповідно до адміністративно встановленого регламенту. Неформальні групи створюються у стихійному порядку, як результат об'єднання інтересів, поглядів, і, відповідно, в їх межах устанавлюються міжособистісні взаємодії. У неформальних груп відсутні юридичний статус та офіціальна регламентація [13].

Відносно малих груп слід відзначити, що вони дозволяють налагоджувати контакти невеликій кількості людей, об'єднаних для здійснення спільної діяльності. Частіше за все мала група може виглядати як «діада» (дві особи) чи «тріада» (три особи). Зокрема, «тріада» вважається більш стабільною. Що стосується великих груп, то вони є чисельними за складом і формуються для досягнення визначеної мети. Взаємодія індивідів у межах подібних груп переважно здійснюється опосередковано. До великих груп також належать численні сукупності людей, які характеризуються наявністю загальних інтересів та однаковим положенням у суспільній соціальній структурі.

У реальних групах виділяються соціально значущі критерії. У номінальних групах традиційно проводяться соціологічні дослідження чи статистичний облік. Відповідно, дані групи вважаються умовними соціальними категоріями [15].

Поряд із соціальними групами існує можливість виділення квазигрупи, яка являє собою нестійку соціальну спільність, що виникає спонтанно та не є

стійкою (їм притаманні сторонність і короткочасність). Найбільш близьким до соціальних груп різновидом квазігруп є соціальне коло. Соціальні кола формуються з метою інформаційного обміну та можуть бути представлені у наступних формах (табл. 1.2) [30].

Таблиця 1.2

Форми соціальних кіл

| Форма соціальних кіл | Характеристика | Приклади |
|----------------------|--|---|
| Контактні | Постійні зустрічі на основі визначених умов | Спортивні змагання |
| Професійні | Виключно професійний обмін інформацією | Взаємодія в межах визначеної сфери професійної діяльності |
| Статусні | Обмін інформацією між людьми в межах визначеного соціального статусу | Аристократичні кола |
| Дружні | Сумісне проведення заходів | Групи друзів, компанії |

Джерело: складено на підставі [30]

При цьому необхідно зазначити, що квазігрупи можуть бути перехідними стадіями до соціальних груп, якщо вони у процесі свого функціонування стануть організованими, стійкими, структурованими.

Соціальні групи фактично є підґрунтям для створення віртуальних співтовариств, які існують на базі соціальних мереж, і яким притаманні такі характеристики, як наявність визначених меж та присутність особливої системи взаємовідносин, які, в свою чергу, ґрунтуються на особливостях формування та функціонування соціальних груп у соціокультурній практиці.

Так, визначення меж соціальної мережі, яка, у свою чергу, базується на соціальних групах, здійснюється за допомогою технічних засобів, наприклад, обчислення коефіцієнту кореляції, який дозволяє зробити висновок щодо взаємодії між користувачами соціальної мережі.

У цьому контексті необхідно зазначити, що індивіди в межах як соціальних груп, так і соціальних мереж мають можливість взаємодіяти за допомогою комунікації.

Якщо розглядати комунікацію в соціальних мережах у глобальному інформаційному просторі, то вона являє собою практично безпосереднє, контактне спілкування користувачів між собою.

Подібна ситуація обумовлена тим, на нинішньому етапі існує можливість трансляції звуку та відео в режимі он-лайн у глобальній мережі на відміну від початкових стадій розвитку технологій комунікації в мережі Інтернет, коли здійснювалося лише дистанційне спілкування через використання визначених програмних продуктів. Відповідно, доцільно проаналізувати особливості комунікативних процесів у соціальних групах.

Так, комунікація у загальному випадку може розглядатися як форма людської діяльності, що виражається в інформаційному обміні для взаєморозуміння партнерів. У межах комунікації можуть вирішуватися різні завдання: інформаційний обмін, вираження взаємних відносин людей, взаємний вплив, порозуміння та співчуття. Враховуючи описану багатофункціональність комунікації виникає можливість виділення її ключових аспектів (рис. 1.1) [3].



Рис. 1.1. Аспекти комунікації

Джерело: складено на підставі [3]

Зокрема, у межах інтерактивного аспекту комунікація розуміється як взаємодія індивідів у процесі їх кооперації. Гносеологічний аспект передбачає, що людина може виступати в якості суб'єкта й об'єкта соціокультурного пізнання. Аксиологічний аспект орієнтований на дослідження комунікації як процесу ціннісного обміну. Семіотичний аспект формує специфічну знакову систему комунікації.

Процеси комунікації протягом періоду становлення групи можуть проходити такі стадії (табл. 1.3) [3].

Таблиця 1.3

Стадії становлення процесів комунікації у соціальних групах

| Найменування стадії комунікативного процесу в соціальній групі | Характеристика |
|--|--|
| Формування | Виділення лідера, визначення правил комунікації, постановка завдань |
| Протест | Конфлікт між індивідами та підгрупами, неприйняття лідера й установок |
| Визначення норм | Формування стабільної структури групи й її норм, адаптація членів групи |
| Співробітництво | Подолання міжособистісних суперечностей, вирішення питань загального характеру |

Джерело: складено на підставі [3]

Існує можливість розподілу комунікації відповідно до контексту, в межах якого вона здійснюється. Відповідно до зазначеного критерію комунікація може бути міжособистісною, груповою та масовою.

Якщо більш докладно розглядати міжособистісну комунікацію, то доцільно відмітити, що вона являє собою безпосередній контакт людей. Міжособистісна комунікація містить три взаємопов'язані компоненти, перцептивний та інтерактивний. Зокрема, комунікативний компонент стосується обміну інформацією між партнерами по спілкуванню враховуючи передачу й прийом думок і почуттів. Інтерактивний компонент забезпечує обмін діями між сторонами. Перцептивний компонент стосується особливостей взаєморозуміння людей і побудови міжособистісних відносин

Міжособистісна комунікація, в свою чергу, поділяється на особистісну і рольову. Першій притаманний більш експресивний характер, пов'язаний з самовираженням «Я». Що стосується другої, то вона є основою організаційної системи комунікацій.

Групова комунікація традиційно має місце, коли люди взаємодіють у межах малих соціальних груп. Групова комунікація характеризується закономірностями, присутніми у міжособистісній комунікації, проте, їй притаманні й специфічні властивості (наявність лідерства та керівництва в групі; стимулюючі складові; орієнтація на реалізацію визначених завдань; вироблення групового рішення; аналіз проблеми; комунікативна структура.

У межах масової комунікації інформація поширюється у суспільстві через соціальні засоби: друковані видання, телебачення, радіо, кіно тощо, в результаті чого повідомлення одразу отримують великі групи людей. Виходячи з цього, першочерговим завданням масової комунікації у нинішньому суспільстві є формування «глобальної зв'язку» між людьми [8].

Доцільно зазначити, що комунікація може здійснюватися за допомогою вербальних і невербальних засобів. Зокрема, провідним і універсальним засобом спілкування для людини є саме вербальна комунікація (у вигляді усної й письмової мови). Що стосується невербальної комунікації, то вона

уможливорює спілкування без словесної допомоги (жести, міміка, знакові системи тощо). Зазначені способи спілкування мають назви первинної та вторинної мови або природних і штучних.

Якщо повернутися до міжособистісної комунікації, то необхідно врахувати, що її визначають основні різні цілі, завдання та мотиви учасників. В залежності від цих факторів існує можливість використання різнохарактерних моделей міжособистісної комунікації.

Найбільш широке поширення отримала лінійна (класична) модель, яка розглядає комунікацію як дію, у межах якої відправник кодує ідеї та почуття у певний вид повідомлення і потім відправляє його одержувачу, використовуючи будь-який канал. Якщо повідомлення досягло одержувача, подолавши різного роду перешкоди, то комунікація вважається успішною.

Перевага даної моделі пов'язана з тим, що з її появою виникло уявлення про швидкість і кількість переданої інформації. При цьому лінійна модель має низку обмежень:

- вона механістична і відображає переважно технічні способи комунікації;
- людина включається в неї лише в якості «джерела» або «приймача» інформації;
- вона абстрагується від змісту і сенсу переданої інформації, приділяючи увагу лише її кількості;
- комунікативний процес в даній моделі носить односпрямований характер, зворотній зв'язок відсутній.

Виходячи з вищезазначених особливостей, лінійна модель придатна для впливу засобів масової інформації, де одержувач повідомлення розглядається як об'єкт впливу [33].

Іншою моделлю міжособистісної комунікації є трансакційна модель. Вона представляє комунікацію як процес одночасного відправлення й одержання повідомлень комунікаторами. Дана модель набагато краще описує процеси комунікації, ніж лінійна.

Принцип дії трансакційної моделі заснований на припущенні, що кожен член групи, вступаючи в контакти з іншими індивідами, володіє обмеженим «набором» станів власного «Я». Вибір певного стану надає взаємодії конкретного змісту. При цьому процес взаємного обміну інформацією може бути:

- паралельним, коли відсилання інформації тягне за собою доречну, очікувану природну реакцію, а процес комунікації протікає гладко, без напруження, на основі взаєморозуміння;

- дотичним, коли сторони обмінюються інформацією, але мають різні позиції, неадекватно сприймають внутрішній стан партнера, що зумовлює напруженість у відносинах і далі – розрив угод, відмову в спілкуванні і соціально-психологічні конфлікти.

Трансакційна модель є корисною при виявленні можливих перешкод в комунікаційних процесах, а також пошуку можливостей для зміни поведінки партнера в ході обміну інформацією. Однак аналіз трансакцій стосується лише міжособистісних комунікацій і не може служити основою для аналізу комунікаційних мереж [24].

Існує також інтерактивна (секторна) модель міжособистісної комунікації. Вона являє собою не просто процес передачі повідомлення від відправника до одержувача, в ході якого перший кодує, а другий декодує інформацію. Важливим елементом цієї моделі є наявність зворотного зв'язку. Інтерактивна модель, подібно до лінійної, зображує комунікацію як низку дискретних актів, що мають початок і кінець, причому ключовим компонентом у них є відправник повідомлення, тому що саме від нього залежить реакція одержувача інформації.

Традиційно лінійна й інтерактивні моделі вважаються застарілими порівняно з трансакційною. Однак для опису процесів міжкультурної комунікації і розуміння її специфіки більше підходить саме інтерактивна модель. Вона складається з наступних елементів: відправник (джерело) — кодування — повідомлення — канал — декодування — отримувач —

зворотний зв'язок. У цьому процесі створюються різні «шуми», що перешкоджають ефективній комунікації. При цьому ефективність комунікації характеризується тим, що передана інформація повинна бути зрозуміла у відповідності з її початковим значенням [13].

Останнім часом усе більшої актуальності набуває використання діалогічної моделі соціальної комунікації, розробленої в межах парадигми, де ключове значення має ефект діалогу як смислового контакту, заснованого на здатності і прагненні суб'єктів до адекватного тлумачення комунікативних намірів партнерів по спілкуванню.

На вибір моделі міжособистісної комунікації суттєво впливають характеристики процесу спілкування психологічного та соціального характеру. Зокрема, серед психологічних характеристик комунікативного акту перш за все необхідно виділити ті, що являють собою мотиваційний бік спілкування: мета, намір та задум. Ці характеристики стосуються когнітивної складової комунікативного акту.

Що стосується соціальних характеристик комунікативного акту, то вони включають стильові прийоми учасників спілкування, а також їх ролі та норми статусного та ситуативного характеру.

Зокрема, стильові прийоми учасників акту спілкування передбачають використання визначеного мовного стилю, стратегій і тактик комунікації [10].

Під ролями традиційно розуміються спосіб і сценарій поведінки, який слід уважати придатним для конкретної ситуації чи їх групи. Так, ролі та норми статусного характеру визначають особливості поведінки індивіда, які відповідають його соціальному становищу (вік, стать, клас, посада тощо) або статусу. Відповідно, на початку кожного акту комунікації його учасники повинні цілком усвідомлювати власну та партнерську соціальні ролі [10]. Ролі та норми ситуативного характеру формуються безпосередньо протягом комунікативного процесу та чинять на нього суттєвий вплив. Зокрема, індивід може бути лідером, і, відповідно, прагне контролювати весь

комунікативний процес. Або людина може бути посередником, який відстежує хід комунікативного процесу. У деяких випадках можлива присутність нестандартних суджень тощо.

При цьому, якщо повернутися до комунікації в межах віртуальних співтовариств у соціальних групах, то доцільно виділити такі її категорії:

- діалогова комунікація;
- полілогова комунікація.

Зокрема, стосовно діалогової комунікації, то вона традиційно здійснюється через використання електронної пошти або програм, що містять мережеві протоколи миттєвого обміну повідомленнями (ICQ, Viber тощо).

Відносно полілогової комунікації необхідно відмітити, що вона забезпечується через участь користувачів у різнохарактерних форумах, веб-конференціях, чатах тощо.

У свою чергу, форуми та веб-конференції можуть бути універсальними, тематичними чи довідково-інформаційними.

У цілому, дослідження особливостей усіх зазначених видів та моделей комунікації у соціальних групах, зокрема, у віртуальних співтовариствах, є корисним для вдосконалення процесів організації електронної комерції. Зокрема, консолідація інформації щодо психологічних характеристик, стратегій поведінки, ролі та статусу користувачів соціальних мереж надає можливість суттєво вдосконалити комерційну діяльність компанії. Зокрема, доцільним є створення системи управління взаємовідносинами з клієнтами CRM (англ. – Customer Relationship Management). Використання подібної системи уможливить отримання інформації про клієнтів з різнохарактерних джерел з паралельним усуненням дублювання даних; уніфікацію структури інформації щодо клієнтів до єдиного виду та формування вітрин даних про клієнтів.

У загальному випадку про соціальні мережі можна говорити в різних аспектах: як про соціальне явище (установлення соціальних зв'язків між

людьми), як про універсальний інструмент соціологічного аналізу, і нарешті, як про Інтернет-послугу або Інтернет-сервіс стосовно побудови соціальної мережі у Всесвітній павутині для одержання соціального капіталу [31].

Багато сервісів Інтернет, що дозволяють людям установлювати зв'язки, автоматично формують соціальні мережі. Відповідно, на певному етапі було створено сервіс, головною метою якого було нагромадження соціального капіталу, тобто особистих ділових зв'язків у вигляді соціальної мережі. У результаті з'явився Інтернет-сервіс стосовно побудови соціальних мереж.

Соціальні групи в мережі Інтернет відкривають перед дослідниками якісно нові можливості щодо аналізу бізнесу, політики, різнохарактерних соціологічних досліджень.

Аналіз соціальних груп в мережі Інтернет дозволяє одержати інформацію, що згодом може бути корисною для комерційної діяльності підприємств із урахуванням особливостей членів соціальної групи. Однак на сучасному етапі відсутній універсальний метод систематизації подібної інформації [6].

Зв'язок членів соціальної групи в мережі Інтернет можна математично моделювати за допомогою графу, в якому вершини є учасниками соціальної групи, а ребра – взаєминами між ними. Математичний апарат аналізу графів дозволяє розрахувати цілу низку параметрів і дати кількісні відповіді на багато питань.

В аналізі соціальних груп в мережі Інтернет на базі теорії графів виділяють:

- розрахунок індексів для соціальної групи в цілому й для окремих членів соціальної групи;
- виділення підструктур у соціальній групі [1].

Проте, аналіз соціальних груп у мережі Інтернет за допомогою графів можна здійснювати винятково з математичної точки зору без врахування якісних параметрів, що характеризують учасників соціальної групи. Для

всебічного аналізу соціальних груп у мережі Інтернет доцільно використовувати системи консолідації інформації.

Консолідація є комплексом методів і процедур, що спрямовані на отримання даних з різних джерел, забезпечення необхідного рівня їхньої інформативності і якості, доведення їх до єдиного формату, в якому існує можливість їх завантаження до сховища даних або аналітичної системи [2].

Системи консолідації даних у соціальних групах у мережі Інтернет дозволяють:

- підвищити швидкість доступу до даних соціальних груп;
- забезпечити компактність зберігання інформації щодо учасників соціальних груп;
- автоматично підтримувати цілісність структури даних про членів соціальних груп;
- здійснювати контроль несуперечності даних щодо соціальних груп.

Таким чином, використання систем консолідації даних для аналізу соціальних груп у мережі Інтернет є оптимальним з погляду їхньої обробки на конкретній аналітичній платформі та прийняття відповідних управлінських рішень, орієнтованих на позитивний економічний ефект від комерційної діяльності. Усе вищезазначене підкреслює актуальність поточного дослідження.

1.2. Аналіз сучасних моделей систем консолідації інформації

Консолідація даних полягає в об'єднанні безлічі комп'ютерів і серверів в єдину середу, призначену для вирішення певного роду завдань, наприклад, наукових проблем або складних розрахунків. Згодом в такій структурі накопичується безліч даних, розподілених по обчислювальним вузлам і сховищам. Зазвичай додатки, що виконуються в розподіленому обчислювальному середовищі, звертаються тільки до одного з джерел даних. Одна, при виникненні необхідності в одночасному доступі до декількох

джерел, виникають труднощі, так як ці джерела можуть містити різноманітні дані і методи доступу до них, а також розташовуватися на відстані одне від одного. Крім того, користувачам, які здійснюють аналіз накопичених даних, зручно звертатися до єдиного джерела інформації, формуючи запити і отримуючи результати у однаковому форматі. Таким чином, головною проблемою підходу до зберігання інформації в розподілених обчислювальних системах є різноманітність і віддаленість джерел даних. Рішенням даної проблеми є створення точки централізованого доступу, що забезпечує єдиний інтерфейс звернення до всіх джерел даних обчислювального хмари в режимі реального часу. У цьому контексті необхідно вибрати найбільш прийнятний підхід і відповідну платформу, що забезпечує таку консолідацію.

Розподілені системи і, зокрема, бази даних, як правило, пропонують певну модель узгодженості. Ці моделі гарантують, що, якщо певні умови будуть виконані, то можна очікувати що система буде мати певні властивості (тобто узгодженість або доступність або стійкість до розривів мережі). У цьому контексті розглядається популярний набір ACID-властивостей, який гарантує, що транзакції бази даних обробляються надійно, в порівнянні з протилежною моделлю BASE, яка є похідною з теореми CAP, але прагне забезпечити набір властивостей, що відрізняється від ACID

1. Концепція ACID. Традиційно реляційні бази даних в змозі забезпечувати і дотримуватися властивості ACID, яка є набором властивостей, які гарантують, що транзакції бази даних обробляються надійно і, що база даних залишається узгодженою в умовах паралельного доступу і відмови системи. ACID є аббревіатурою, яка означає наступне: атомарність, цілісність, ізоляція і довговічність.

Атомарність. Атомарність означає, що транзакції бази даних повинні дотримуватися принципу "все або нічого".

З огляду на, що транзакція є серією інструкцій для виконання, для транзакції, щоб бути "атомарною", всі інструкції повинні бути виконані, або

якщо одна (або більше) інструкція не може бути виконана, вся транзакція виявляється невдалою, і база даних повинна залишатися незмінною.

Цілісність. Цілісністю забезпечується те, що тільки допустимі дані записуються в базу даних, гарантуючи, що якщо транзакція пройшла успішно, база даних переходить з одного узгодженого стану в інший узгоджений стан. Якщо з будь-якої причини виникає помилка під час транзакції, то будь-які зміни, вже внесені, буде автоматично скасовано (відкат змін), тому забезпечується те, що база даних залишається в узгодженому стані.

Ізоляція. Ізоляція вимагає, щоб транзакції, що знаходяться в процесі, і ще не зафіксували свої зміни, повинні залишатися ізольованими від будь-якої іншої транзакції. Тому виконання транзакції не повинно вплинути на виконання інших паралельних транзакцій. Ізоляція важлива, тому що, поки транзакції в процесі виконання, стан системи не може бути узгодженим, так як транзакції тільки забезпечують що система буде перебувати в узгодженому стані після завершення транзакції. Якщо транзакція не була запущена в ізоляції, вона зможе отримати доступ до даних з системи, які в цей момент не є узгодженими.

Довговічність. Довговічність гарантує, що будь-яка транзакція в базі даних є постійною і не буде втрачена. База даних повинна мати можливість відновити підтвержені поновлення, зроблені транзакцією при будь-якому вигляді збою системи (апаратному або програмному). Багато баз даних реалізують довговічність, записуючи транзакції в журналі транзакцій так, що ці транзакції можуть бути виконані знову, щоб відтворити стан системи, який був безпосередньо перед відмовою. Транзакція вважається завершеною тільки після того, як вона буде записана в журнал.

Модель ACID означає перевагу узгодженості розподілених баз даних, вона песимістична і забезпечує узгодженість в кінці кожної транзакції. Ця суворая узгодженість не завжди може бути необхідною в залежності від вимог предметної області. Крім того, цю модель, дуже важко реалізувати в

розподіленій системі, не створюючи вузьких місць. Наприклад, протокол двофазної фіксації, який забезпечує атомарність розподілених транзакцій, вимагає блокування всього вузла, поки вузол очікує повідомлення. Ця модель також створює багато накладних витрат, так як передбачає передачу великої кількості повідомлень. Крім того, якщо ми візьмемо до уваги теорему CAP, забезпечення суворої узгодженості може забезпечуватися тільки на шкоду до доступності або до стійкості до розривів мережі. Якщо ми використовуємо двофазний протокол фіксації, ми в змозі забезпечити узгодженість вузлів, що призводить до втрати в доступності. Так як різні додатки можуть мати різні вимоги до узгодженості, бази даних NoSQL, як правило, забезпечують альтернативні моделі узгодженості.

2. Концепція BASE. BASE, як припускає її назва, є логічною протилежністю ACID (гра англійських слів base і acid, які позначають хімічні терміни). У той час як ACID зосереджується більше на узгодженості, і в цьому сенсі є песимістичною, BASE зосереджується на доступності, забезпечуючи узгодженість в кінцевому рахунку (англ. – eventual consistency), і в цьому сенсі є оптимістичною. BASE – це аббревіатура, що означає набір властивостей: Basically Available, Soft-state, Eventual consistency. Це нерозривно пов'язано з теоремою CAP і було також запропоновано Еріком Брюєром у 2000 році на симпозіумі про принципи розподілених обчислень.

Якщо здійснити аналіз існуючих нині моделей систем консолідації інформації, то є можливість отримати декілька їх градацій:

- системи консолідації фінансової інформації;
- в залежності від архітектури збереження інформації, що використовується в системі;
- в залежності від моделі об'єднання інформації;
- в залежності від способу обробки запитів користувачів;
- системи консолідації інформації про клієнтів;
- системи консолідації інформації про зовнішнє середовище компанії;

- в залежності від методу аналізу інформації.

Зокрема, якщо більш докладно розглянути системи консолідації фінансової інформації, то доцільно відмітити, що сьогодні на ринку вони представлені у вигляді пакетних програм, систем планування ресурсів компанії та стандартних інструментів ETL (англ. – Extraction, Transformation, Loading – витяг, перетворення, завантаження).

В свою чергу, що стосується пакетних програм, то вони займають переважну частку ринку та можуть бути представлені в якості аналітичних систем класу CPM (англ. – Corporate Performance Management – системи управління ефективністю компанії) та спеціалізованих систем фінансової консолідації [4].

Зокрема, аналітичні системи класу CPM є досить поширеними. Це – системи типу ORACLE Hyperion, SAP BPC, IBM Cognos TM1. Вони мають порівнянну функціональність, і їм притаманні подібні процеси впровадження. Вибір між ними визначається необхідністю інтеграції з вже наявними в компанії системами консолідації інформації, комерційною доцільністю і особистими уподобаннями. Спеціалізовані системи фінансової консолідації менш поширені. Серед найбільш відомих на вітчизняному ринку специфічних систем фінансової консолідації слід виділити такі, як: Oracle Hyperion Financial Management (HFM) і IBM Cognos Controller. Так, Oracle Hyperion має порівняно більшу гнучкість налаштувань (наприклад, у частині алгоритмів консолідації й основних консолідаційних поправок), що робить її більш адаптивною щодо конкретного замовника. Однак це може призвести до більш тривалого і витратного впровадження. Друга система, IBM Cognos Controller, – менш гнучка і адаптивна відносно заданого алгоритму консолідації та параметричних налаштувань, але має і таку перевагу, як швидкість налаштування виходячи із заданих системою меж. Менша поширеність спеціалізованих систем фінансової консолідації багато в чому пояснюється необхідністю залучення для їх створення фахівців визначеного профілю. Відповідно, такі системи є вузько спеціалізованими і мають більш

тривалий термін окупності. Однак перевагами спеціалізованих систем фінансової консолідації для користувача є більш швидке налаштування і, відповідно, скорочення витрат на впровадження.

Стосовно стандартних інструментів ETL, то вони являють собою різновид ключових процесів управління сховищами даних. Стандартні інструменти ETL надають наступні можливості:

- отримання інформації з джерел зовнішнього походження;
- трансформація інформації та її адаптація до потреб бізнес-моделі;
- завантаження інформації до сховища даних.

Що стосується систем планування ресурсів підприємства (англ. – Enterprise Resource Planning System – ERP), то вони являють собою корпоративні інформаційні системи, орієнтовані на автоматизацію процесів обліку й управління. Традиційно ERP-системи будуються відповідно до модульного принципу і дозволяють охопити всі ключові процеси діяльності компанії [9].

1.3. Аналіз завдань систем консолідації інформації

Серед загальних завдань будь-якої системи консолідації інформації слід виділити такі:

- вибір джерел інформації;
- розробка стратегії консолідації інформації;
- оцінка якості інформації;
- збагачення інформації;
- очищення інформації;
- перенесення інформації до сховища даних.

При цьому необхідно враховувати, що робота будь-якої системи консолідації інформації у відповідності зі стадіями формування та використання консолідованої звітності ґрунтується на таких завданнях:

- збір і структурування вихідної інформації;

- процес консолідації інформації;
- використання консолідованої звітності.

Однак, в залежності від призначення систем консолідації інформації, їх завдання можуть суттєво відрізнятися.

Так, якщо зупинитися на завданнях найбільш актуальних на сучасному етапі систем консолідації інформації про клієнтів, то серед їх специфічних завдань виділяються такі:

- збір інформації з безлічі джерел;
- реструктуризація інформації про клієнтів і приведення структури до єдиного вигляду;
- пошук і консолідація схожих записів;
- підготовка вітрин даних та розрахунок показників.

Зокрема, системи консолідації інформації, побудовані відповідно до підходу CRM, дозволяють вирішити такі завдання.

1. Збереження потенційних клієнтів. У малому та середньому бізнесі конкуренція є дуже високою. Компанії докладають значні зусилля для того, щоб залучити клієнтів. Порівняно з іншими витратами на залучення клієнтів виділяється значний бюджет. І дуже важливо, щоб всі ці фінансові ресурси і зусилля привели до очікуваного результату.

2. Контроль роботи співробітників і стандартизація роботи з клієнтами. Без загальної стандартизованої CRM-системи контакти співробітників відбуваються досить хаотично. CRM-система дозволяє розміщувати інформацію про всі вхідні та вихідні контакти в одному сховищі, звідки її можна в будь-який момент отримати.

3. Накопичення статистичної бази, що також дуже важливо для успішного розвитку будь-якого бізнесу. Завдяки використанню CRM-системи вся робоча інформація збирається в одній загальній базі в стандартизованому вигляді. В з'являється можливість аналізувати статистику роботи, складати різнохарактерні звіти (багато з яких вже в готовому вигляді

присутні в CRM-системах), тобто аналізувати роботу і планувати подальшу роботу більш усвідомлено.

4. Отримання остаточних рішень, які можуть слугувати підґрунтям для побудови власної системи роботи. У кожній CRM-системі закладено безліч готових інструментів, які дозволяють перевести роботу на якісно новий рівень. Наприклад, інтеграція CRM-системи з телефонією дозволяє фіксувати всі дзвінки, зберігати нові контакти та аналізувати якість роботи відділу продажів з клієнтськими запитами. У малому та середньому бізнесі частіше за все роботу з клієнтами організує безпосередньо власник бізнесу. Здебільшого у нього немає експертів і напрацювань щодо організації роботи з клієнтами. Впровадження CRM-системи дозволяє отримати відповідний інструмент прийняття доцільних рішень щодо роботи відділу продажів.

Таким чином, завдяки вирішенню вищезазначених завдань, CRM-системи уможливають:

- отримання загальної для компанії стандартизованої бази контактів (клієнтів, контрагентів);
- здійснення ефективного контролю якості роботи відділу продажів в будь-який момент часу;
- отримання статистики й аналітики ефективності роботи з клієнтськими запитами;
- планування підвищення якості роботи та розробку стратегії розвитку бізнесу.

Що стосується завдань більш масштабних систем консолідації інформації про зовнішнє середовище компанії, то вони орієнтовані на вирішення таких завдань:

- аналіз поведінки споживачів;
- оцінка конкурентних дій;
- дослідження ринкових тенденцій;
- підтримка прийняття рішень щодо обрання раціональної стратегії й тактики діяльності компанії;

- розробка та впровадження комплексу маркетингу компанії;
- моніторинг і контроль ефективності функціонування компанії.

У контексті аналізу моделей систем консолідації інформації в залежності від архітектури збереження інформації, що використовується в системі, є можливість виділити такі способи їх побудови.

1. Перенесення структур – це переміщення існуючих аналітичних структур на одну платформу. Даний підхід привабливий для компаній, яким потрібно мінімізувати витрати. Подібна операція дозволяє скоротити кількість серверів, що використовуються, а також розвантажити фахівців, які зайняті їх обслуговуванням. Необхідно відзначити, що перенесення структур не передбачає зміну додатків або інтеграцію даних. Модель даних, метадані, логіка перетворень і технологія підготовки звітності залишаються без змін, коригуються лише клієнтські інтерфейси.

2. Проект «з нуля» – створення системи консолідації інформації в компанії без використання існуючих раніше систем чи інших розробок.

3. Визначення корпоративного стандарту і подальший перехід до нього – встановлення існуючих сховищ або вітрини даних в якості корпоративного стандарту та негайне або поступове переведення інших структур до цього стандарту. Даний підхід можна рекомендувати в тих випадках, коли спостерігається поглинання однією компанією іншої. Відповідно, сховище даних головної компанії стає корпоративним стандартом, і в нього вбудовується сховище даних поглиненої компанії.

4. Синхронізація. Даний підхід є більш прийнятним, якщо компанії необхідно стандартизувати довідкові дані для об'єктів основних даних, таких, як клієнти, продукти, постачальники тощо серед великої кількості операційних програм. Досить часто дані для цих об'єктів збираються та зберігаються в різних операційних програмах у різних форматах. В результаті з'являються численні копії ідентичних записів, що порушує узгодженість і стандартизованість даних.

5. Узгоджені вітрини даних – це ще один спосіб консолідувати вітрини даних, не вдаючись до їх фізичного об'єднання. Це внесення вимірювань в структуру кожної вітрини, щоб вони були узгоджені одна з одною. Даний підхід спирається на методологію Ralph Kimball, відповідно до якої компанії створюють одну тимчасову проміжну область, дані з якої використовуються для заповнення узгоджених вітрин даних. Завдяки введенню цієї нової проміжної області консолідується надлишкові дані, тобто скорочуються витрати.

6. Вітрина з вітрин даних. Якщо компанія є децентралізованою, і тільки частині корпоративних користувачів необхідна консолідована інформація, доцільним є створення вітрини даних з усіх існуючих аналітичних структур. У відповідності з даним підходом необхідна корпоративна інформація отримується з існуючих структур за допомогою стандартних інструментів ETL. Перевага цього підходу полягає в тому, що не потрібно змінювати існуючі середовища. Крім того, не існує необхідності збільшувати штат фахівців і придбати програмні інструменти, що дозволяє заощадити значні кошти.

7. Розподілені запити. При застосуванні цього методу існуючі неінтегровані аналітичні структури також залишаються на місці. Однак, замість інструментів ETL і пакетної обробки, що використовуються для створення консолідованого подання, в даному випадку таке подання будується за допомогою засобів SQL (англ. – Structured Query language – мова структурованих запитів) [12].

При цьому необхідно додатково проаналізувати варіанти переміщення існуючих аналітичних структур на одну платформу. У цьому контексті доцільно здійснити огляд систем консолідації інформації в залежності від моделі об'єднання інформації. Відповідно, можливість виділити декілька варіантів побудови таких систем:

- об'єднання систем зберігання, розміщених на декількох серверах, в межах єдиного сервера;

- безпосереднє підключення декількох гетерогенних серверів до єдиної системи зберігання даних;

- об'єднання системи зберігання на базі архітектури мереж зберігання даних SAN (англ. – Storage Area Network – мережа зберігання даних) [11].

Зокрема, що стосується об'єднання систем зберігання, розміщених на декількох серверах, в межах єдиного сервера, то даний варіант побудови систем консолідації інформації знижує гостроту проблем адміністрування й економить простір, який займає обладнання, за рахунок централізації системи зберігання на одному потужному сервері.

Безпосереднє підключення декількох гетерогенних серверів до єдиної системи зберігання даних стало можливим завдяки наявності гетерогенних систем зберігання високого класу, спеціально розроблених для інформаційних центрів. Вони значно знижують вартість зберігання даних за рахунок створення єдиної високопродуктивної системи зберігання для декількох серверів, усуваючи необхідність придбання окремих пристроїв для кожного типу серверів. Цей підхід також спрощує адміністрування й економить площу, яку займає обладнання.

При використанні об'єднання системи зберігання на базі архітектури мереж зберігання даних SAN створюється загальний пул ресурсів зберігання, що спільно використовуються, включаючи різні типи накопичувачів. Ця модель покращує ефективність і знижує складність управління, підвищуючи масштабованість, доступність і надійність зберігання даних. При подібній консолідації інформації SAN виступає як засіб, що дозволяє великому числу серверів мати доступ до єдиного ресурсу зберігання даних.

Якщо звернутися до аналізу систем консолідації інформації в залежності від способу моделювання обробки запитів користувачів, то можна виділити наступні відповідні моделі.

1. Модель нечітких множин. Дана модель базується на теорії нечітких множин, яка допускає часткову належність елемента множини на відміну від традиційної теорії множин. У даній моделі весь масив документів описується

як набір нечітких множин термів. Кожен терм визначає монотонну функцію приналежності до документів даного масиву. У даній моделі здійснюється перевизначення логічних операторів для врахування можливості часткової приналежності множини. Запити виражаються так само, як і в булевій моделі, і обробляються з використанням перевизначених логічних операторів.

2. Розширена булева модель, яка є модифікацією булевої моделі. В цій моделі зроблено спробу зняти обмеження, властиві булевій моделі. Фактично вона є гібридом булевої моделі та розглянутої нижче векторної моделі. У ній передбачається, що терм описує зміст документа з деякою точністю, яка виражається у вигляді ваги терма. Для визначення ваги терма використовується статистика зустрічальності терма з відповідною процедурою нормалізації. Причому визначати ваги можна як для термів документа, так і для термів запиту.

3. Векторна модель, запропонована Дж. Селтоном. У даній моделі документи та запити користувачів представляються у вигляді n -мірних векторів у n -мірному векторному просторі. Розмірність векторного простору n являє собою загальну кількість різних термів у всіх документах. Все розмаїття словоформ кожного терма приводиться до деякої основи. У розгляді не беруть участь службові, малоінформативні і високочастотні словоформи.

4. Модель, заснована на латентно-семантичному аналізі. В теорії інформаційного пошуку дану модель прийнято називати латентносемантичним індексуванням. Латентно-семантичний аналіз (англ. – Latent Semantic Analysis – LSA) – це метод отримання та представлення контекстозалежних значень слів за допомогою статистичної обробки великих масивів текстових документів. У цій моделі сукупність всіх контекстів, в яких зустрічається і не зустрічається дане слово, задає безліч взаємних обмежень, які дозволяють визначити подібність смислових значень слів і множин слів між собою.

5. Ймовірнісні моделі. Дані моделі пошуку засновані на застосуванні методів теорії ймовірності. В них використовуються статистичні показники, що визначають ймовірність релевантності документа пошуковому запиту користувача. В основі цих моделей лежить принцип імовірнісного ранжування (англ. – Probabilistic Ranking Principle – PRP). Суть цього принципу полягає в тому, що документи в колекції повинні ранжуватися за зниженням ймовірності їх релевантності запиту користувача. Обчислення цієї вірогідності є ключовою частиною моделі, і цим більшість імовірнісних моделей відрізняється одна від одної. Особливості ймовірнісної моделі полягають в наступному. В моделі враховуються взаємозалежності та зв'язки термів і визначаються такі основні параметри, як ваги термів запиту та форма подібності «запит-документ».

Що стосується моделей об'єднання інформації в системах консолідації інформації, то вони можуть бути наступними:

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

Системи консолідації інформації про клієнтів на сучасному етапі відіграють важливу роль для забезпечення конкурентоспроможності компанії і можуть бути складовою систем консолідації інформації про зовнішнє середовище компанії.

В нинішніх умовах інформація про клієнтів є основою сучасного бізнесу, переважно стосовно сектору B2C. Аналіз даної інформації і прийняті на його основі рішення є досить часто визначальними як у межах оперативного, так і стратегічного планування. Збір інформації про клієнтів вимагає великого обсягу часу, зусиль та фінансових ресурсів. З іншого боку, відсутність централізації і множинність первинних оперативних даних породжують проблему їх якості. Один клієнт може бути присутнім одразу в декількох системах. Основним завданням, яке дозволяють вирішити системи

консолідації інформації про клієнтів, є усунення подібних дублікатів. Відповідно, впровадження систем консолідації інформації про клієнтів приносить наступні вигоди: підвищує якість вихідних даних, звітності й аналітики; централізує всі дані про клієнтів, дозволяє отримувати агреговану інформацію з клієнтської бази; підвищує якість маркетингових акцій; підвищує рівень довіри і лояльність клієнтів; знижує ризики впровадження і виконання бізнес-орієнтованих проектів. Серед завдань, що вирішуються за допомогою використання систем консолідації інформації про клієнтів, слід виділити такі: збір даних з різнохарактерних джерел; реструктуризація інформації про клієнтів і приведення структури до єдиного виду; пошук і консолідація схожих записів; підготовка вітрин даних і розрахунок показників [5].

Системи консолідації інформації про клієнтів нині представлені на ринку переважно такими категоріями:

- ETL-засоби, які вже були описані вище та являють собою рішення для де-дублікації даних;
- вбудовані механізми у системах класу CRM (англ. – Customer Relationship Management – система управління взаємовідносинами з клієнтами).

Зокрема, CRM-система – це будь-яке програмне забезпечення, яке допомагає успішно контролювати роботу з клієнтами, а також здійснювати їх упорядкування та планування. Більшість CRM систем, наприклад, Siebel CRM мають API (англ. – Application Programming Interface – інтерфейс прикладного програмування) для підключення зовнішніх рішень по де-дублікації даних.

Зокрема, доцільно виділити три CRM-підходи:

- оперативний – автоматизація бізнес-процесів, пов'язаних з роботою зі споживачами, для персоналу, який взаємодіє з клієнтами;
- співробітницький – програма забезпечує взаємодію зі споживачами без необхідності участі у цьому процесі персоналу;

- аналітичний – здійснення аналізу інформації про споживачів для досягнення різнохарактерних цілей.

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 йому цінності (моніторинг реакції споживача на наданий йому сервіс);
- збір інформації про діяльність контрагентів та конкурентів компанії;
- засоби інтеграції корпоративної інформації.

Зокрема, стосовно консолідації інформації про сприйняття споживачем запропонованої йому цінності традиційно необхідним є впровадження CALS-технологій (англ. – Continuous Acquisition and Life cycle Support – технологія інтегрованої інформаційної підтримки виробів). Консолідація інформації у даному випадку передбачає не тільки створення заснованих на єдиних (для споживача і виробника) стандартах баз даних, а й формалізацію досвіду спільної взаємодії шляхом спільного коригування стратегічних напрямків розвитку. Відповідно, в системі консолідованої інформації збираються знання про всі бізнес-процеси, які формують життєвий цикл цінності, що надається споживачу. Саме знання про структуру життєвого циклу дозволяють більш точно ідентифікувати сфери, у яких можна вдосконалити стратегічну відповідність компанії.

Другим напрямком консолідації інформації про зовнішнє середовище компанії є збір інформації про діяльність контрагентів та конкурентів компанії. При цьому методологія ВІ (англ. – Business Intelligence – бізнес аналіз) дозволяє не просто створювати системи моніторингу ринкового оточення компанії, а розробляти бази даних і банки реакцій для

орієнтованого на принципи ситуаційного менеджменту механізму управління. Початковий збір даних для роботи інструментарію ВІ здійснюється через відповідний моніторинг. Головною відмінністю цього підходу є трансформація складу елементів системи моніторингу та перевизначення принципів їх функціонування. Подібний моніторинг дозволяє відслідковувати відповідність траєкторії розвитку планових орієнтирів компанії, а також виступає засобом превентивної трансформації даної траєкторії. Так, у відповідь на виникнення системного події, за допомогою інструментарію ВІ проводиться оцінка стратегічної відповідності компанії з подальшим переглядом системи атракторів і стратегічних орієнтирів.

Відносно засобів інтеграції корпоративної інформації (англ. – Enterprise Information Integration – ЕІІ) слід відмітити, що вони відрізняються більш розвиненими засобами доступу та покращеною технологією побудови розподілених запитів. Крім того, вони допомагають вирішити деякі питання стосовно формування звітів.

Поряд з аналізом систем консолідації інформації про зовнішнє середовище компанії потребують дослідження моделі систем консолідації інформації в залежності від методу аналізу інформації.

Так, в залежності від методу аналізу інформації у системах консолідації інформації існує можливість застосування наступних підходів.

1. Метод аналізу можливостей (англ. – Opportunity Analysis) який уможлиблює висвітлення ризиків і можливостей. Використання даного методу дозволяє зосередити основну увагу саме на тих заходах, які є для компанії необхідними, з метою використання нових ефективних методик управління та вдосконалення процесів стратегічного розвитку.

2. Метод альтернативних випадків (англ. – Alternative Outcomes). Даний підхід уможлиблює вироблення декількох дієвих варіантів розвитку подій, а також оцінити довгострокові перспективи розвитку компанії.

3. Метод аналізу подій (англ. – Event Analysis). Цей підхід дозволяє здійснити аналіз останніх подій з точки зору визначення шляхів досягнення переваг над конкурентами в контексті окреслення їх загальних й особливих дій в межах стратегії їх поведінки.

4. Метод аналізу від протилежного (англ. – Linchpin Analysis) традиційно застосовується для зміни базових припущень або повної відмови від них стосовно конкурентів та переосмислення власної стратегії. Даний метод уможливорює змінити устояні стереотипи стосовно конкурентів і об'єктивно оцінити їх дії і, як результат, сформувані ефективні стратегії поведінки.

5. Метод аналізу конкуруючих гіпотез (англ. – Analysis of Competing Hypotheses) – уможливорює зіставлення різнохарактерних аналітичних висновків і пояснень відносно стратегічних дій конкурентів з відповідним виробленням низки гіпотез. Незважаючи на суттєві витрати часу, який потрібен для аналізу непотрібних гіпотез, даний метод забезпечує досягнення вагомих результатів.

6. Метод протидії конкурентові його «копіюванням» застосовується в межах стратегії «діяти на випередження» в довгостроковій перспективі [4].

У цілому, враховуючи вищезазначене, актуальними і необхідними є розробка та впровадження системи консолідації інформації про клієнтів в межах системи консолідації інформації про зовнішнє середовище компанії, адаптованої до умов вітчизняного бізнесу, з метою підвищення конкурентоспроможності компанії. Подібна система забезпечить такі технологічні переваги:

- можливість підтримки різних платформ;
- висока продуктивність за рахунок розподілених обчислень;
- практично лінійна масштабованість [17].

Висновки до першого розділу

1. Показано, що консолідація є комплексом методів і процедур, що спрямовані на отримання даних з різних джерел, забезпечення необхідного рівня їхньої інформативності і якості, доведення їх до єдиного формату, в якому існує можливість їх завантаження до сховища даних або аналітичної системи.

Зазначено, що системи консолідації даних у соціальних групах у мережі Інтернет дозволять:

- підвищити швидкість доступу до даних соціальних груп;
- забезпечити компактність зберігання інформації щодо учасників соціальних груп;
- автоматично підтримувати цілісність структури даних про членів соціальних груп;
- здійснювати контроль несуперечності даних щодо соціальних груп.

2. Здійснено аналіз існуючих нині моделей систем консолідації інформації та отримано наступні їх градації: системи консолідації фінансової інформації; в залежності від архітектури збереження інформації, що використовується в системі; в залежності від моделі об'єднання інформації; в залежності від способу обробки запитів користувачів; системи консолідації інформації про клієнтів; системи консолідації інформації про зовнішнє середовище компанії; в залежності від методу аналізу інформації.

3. Визначено, що завдяки вирішенню вищезазначених завдань, CRM-системи уможливають:

- отримання загальної для компанії стандартизованої бази контактів (клієнтів, контрагентів);
- здійснення ефективного контролю якості роботи відділу продажів в будь-який момент часу;

- отримання статистики й аналітики ефективності роботи з клієнтськими запитамми;
- планування підвищення якості роботи та розробку стратегії розвитку бізнесу.

Що стосується завдань більш масштабних систем консолідації інформації про зовнішнє середовище компанії, то вони орієнтовані на вирішення таких завдань: аналіз поведінки споживачів; оцінка конкурентних дій; дослідження ринкових тенденцій; підтримка прийняття рішень щодо обрання раціональної стратегії й тактики діяльності компанії; розробка та впровадження комплексу маркетингу компанії; моніторинг і контроль ефективності функціонування компанії.

РОЗДІЛ 2. УДОСКОНАЛЕННЯ МЕТОДІВ КОНСОЛІДАЦІЇ ІНФОРМАЦІЇ У ПРОЦЕСАХ СОЦІАЛЬНОГО ПРОФІЛЮВАННЯ

2.1. Метод моделювання соціального профілю з використанням оптимізації перетворення структур Big Data

В сучасних умовах відкриваються можливості для вирішення дуже складних і неоднозначних завдань, пов'язаних з оцінкою соціальної напруженості в розрізі окремих фокус-груп, прогнозуванням наслідків впровадження певних нововведень в життя суспільства, ідентифікації і верифікації цифрового образу особистості (для запобігання компрометації людей зловмисниками за допомогою несанкціонованого доступу до особистих облікових записів, підробки акаунтів з використанням технічних засобів і методами штучного інтелекту), розробкою сервіс-орієнтованих інтелектуальних Інтернет-технологій і так далі.

Особлива перспективність полягає в інтеграції новітніх інформаційних технологій, пов'язаних зі збором і обробкою великих неструктурованих даних (англ. Big Data), інтелектуальним аналізом даних (англ. Data Mining), розподіленими базами даних.

Великі дані (англ. Big Data) [1] – сучасний напрямок в сфері інформаційних технологій, що включає в себе ряд методів, інструментів і технологій для збору та аналізу великих обсягів різноманітних структурованих і неструктурованих даних з метою отримання ефективних результатів в умовах розподілу інформації за численними вузлів обчислювальної мережі і її постійного оновлення. Також допустимо застосування згаданого терміну і до самих даних, оброблюваних за допомогою цієї технології. Спочатку термін Big Data використовувався в

академічному середовищі для вирішення завдань, пов'язаних з проблемою зростання і різноманіття даних. Перші рішення на основі Big Data з'явилися в другій половині нульових років і розглядалися в якості альтернативи класичним реляційним СУБД в питаннях бізнес-аналітики (Business Intelligence) [2].

Вже згадана технологія може бути застосована у випадках, коли даних занадто багато, щоб їх можна було обробляти традиційними засобами, зокрема сервером з реляційної СУБД. Однак точні кількісні значення інформації, при яких дані стають «великими» не визначені. Область застосування технології залежить від обчислювальних апаратних потужностей і кількості записів в базі даних, тобто в одних випадках гігабайти даних вже можна розглядати як Великі дані, тому що обчислювальна система не справляється з їх своєчасної обробкою, а в інших - петабайт інформації прийнятно обробляються за допомогою класичних методів і, отже, не є «великими». Big Data часто асоціюють з фреймворком Hadoop [3], хоча в основі цієї технології також лежить концепція NoSQL-сховищ даних [4].

Під аналізом Big Data розуміється як аналіз масивів даних в рамках можливостей персонального комп'ютера, так і аналіз в рамках можливостей систем управління базами даних. При цьому як в першому, так і в другому випадку при формуванні і візуалізації масивів даних виникають певні труднощі, які полягають в необхідності забезпечення скоординованої роботи комп'ютерних програм на десятках, сотнях або навіть тисячах серверів. Аналіз Big Data може бути охарактеризований за такими параметрами:

- обсяг, тобто кількість даних, що генеруються. Від цього показника залежить, чи може певний масив даних вважатися Big Data. Дані традиційно зберігаються SQL-серверах в хмарному середовищі;

- різноманіття, тобто категорія, до якої належать Big Data. Знання такої приналежності дозволяють аналітикам найбільш ефективно працювати з інформацією;

- швидкість, тобто швидкість генерування або обробки даних з метою досягнення поставлених цілей;
- мінливість, тобто нестабільність даних у часі;
- достовірність, тобто якість зібраних даних, від якої залежить точність аналізу;
- складність, тобто трудомісткість процесу кореляції та побудови взаємозв'язків між даними.

Розглянемо основні методи аналізу Big Data.

Обробка Big Data зараз передбачає, як правило, впровадження спеціальних програмних комплексів, таких як, Hadoop, що дозволяють проводити обробку великих обсягів даних на підставі концепції Map-Reduce.

Hadoop на даний момент є «де-факто» стандартом обробки Big Data. Hadoop є фреймворком, на основі якого розробляються програми для аналізу і візуалізації Big Data.

Зберігання даних в даному фреймворку здійснюється за допомогою спеціальної розподіленої файлової системи HDFS (англ. – Hadoop Distributed File System), яка лежить в основі Hadoop і дозволяє зберігати дані та надавати до них доступ відразу на декількох вузлах кластера. Таким чином, якщо один або декілька вузлів кластера виходять з ладу, то ризик втрати інформації зводиться до мінімуму, і кластер продовжує роботу в штатному режимі.

Для обробки Big Data використовується алгоритм Map-Reduce, при цьому всі стадії Map повинні завершити свою роботу до початку Reduce. Також вхідні дані вимагають попередньої обробки. Таким чином будується загальний алгоритм роботи.

Одним з найактуальніших завдань сучасних інформаційних технологій є завдання швидкої обробки великих обсягів даних. Ефективне вирішення цього завдання, дозволяє швидше приймати рішення на основі даних, отриманих в минулому.

Під Big Data розуміється широке розмаїття масивів даних, які не можуть бути належним чином оброблені традиційними додатками через свій величезний обсяг або складну побудову.

Складність аналізу Big Data полягає в специфіці їх збору, курирування, поділу, зберігання, передачі, візуалізації і збереження конфіденційності інформації. Під аналізом Великих даних часто розуміється застосування прогнозової аналітики або інших передових методів з метою вилучення з безлічі даних певної корисної інформації. Точність при аналізі Big Data допомагає приймати більш раціональні рішення. У свою чергу, прийняття найкращих рішень дозволяє збільшити виробничу ефективність, скоротити витрати і знизити ризики.

Аналіз Big Data може застосовуватися в таких областях, як відстеження кон'юнктури ринку, запобігання поширенню епідемій і боротьба зі злочинністю. Вчені, керівники найбільших компаній, працівники сфери мас-медіа і реклами, а також урядові органи часто стикаються з труднощами при аналізі масивів даних величезних обсягів в таких областях, як пошук в мережі Інтернет, інформаційні технології в сфері бізнесу і фінансів тощо. Робота вчених, особливо метеорологів, медиків, які вивчають геноми, дослідників, що працюють в області вивчення засобів комунікації, фізиків, які створюють складні симулятори, а також біологів і екологів часто обмежується можливостями обробки величезних масивів даних.

Обсяги масивів даних постійно збільшуються, оскільки розширюються можливості збору інформації за допомогою недорогих мобільних пристроїв, засобів цифрової аерофотозйомки, камер, мікрофонів, зчитувачів радіочастотних міток і бездротових сенсорних мереж.

Бази даних є незамінним інструментом в розробці програмного забезпечення, орієнтованого на оброблення Big Data. Будь-яка діяльність в інформаційному суспільстві спирається на зберігання даних і їх обробку. З появою і розвитком нових завдань для зберігання Big Data зростає актуальність пошуку рішень для роботи з ними. До недавнього часу

реляційна модель була ідеальним варіантом, але зараз для деяких специфічних проектів більше підходять нові моделі зберігання даних, які отримали назву нереляційні або NoSQL. На даний момент існують вже сотні різних СУБД, кожна з яких має свої переваги і недоліки і підходить для вирішення певних завдань. Щоб мати можливість підібрати найкращий інструмент для конкретно поставленого завдання необхідно мати інструменти тестування і порівняльного аналізу баз даних. Для зберігання даних необхідно не тільки мати інформацію про кожен вид баз даних, а й мати можливість випробувати їх для свого завдання, провести тестування. В результаті існує можливість:

- систематизувати інформацію на одному стенді;
- надати інструменти для тестування і звіти про виконані тестові завдання.

Для дослідження і порівняльного аналізу використовуються найбільш популярні в своєму типі бази. Для відповідного вибору наводиться рейтинг видання DB-Engines, який будується на основі запитів в пошукових системах і числа результатів по ним, так само враховуючи обсяг обговорень в соціальних мережах і число вакансій, пов'язаних з цією технологією. Результати згруповані за обраними нами моделям і представлені в табл. 2.1.

Виходячи з рейтингу доцільно обрати для дослідження найбільш затребувані бази даних для кожної моделі:

- Oracle – реляційна СУБД;
- Redis – сховище ключ-значення;
- Cassandra – розподілене сховище;
- MongoDB – документно-орієнтована СУБД;
- Neo4j – БД на основі графів.

Слід зазначити, що найбільш розповсюдженою є реляційна СУБД Oracle (рис. 2.1).

Таблиця 2.1

Загальний рейтинг баз даних

| Модель | Реляційна | Ключ-значення | Розподілене сховище | Документо-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 |

Джерело: складено на підставі [3; 4]

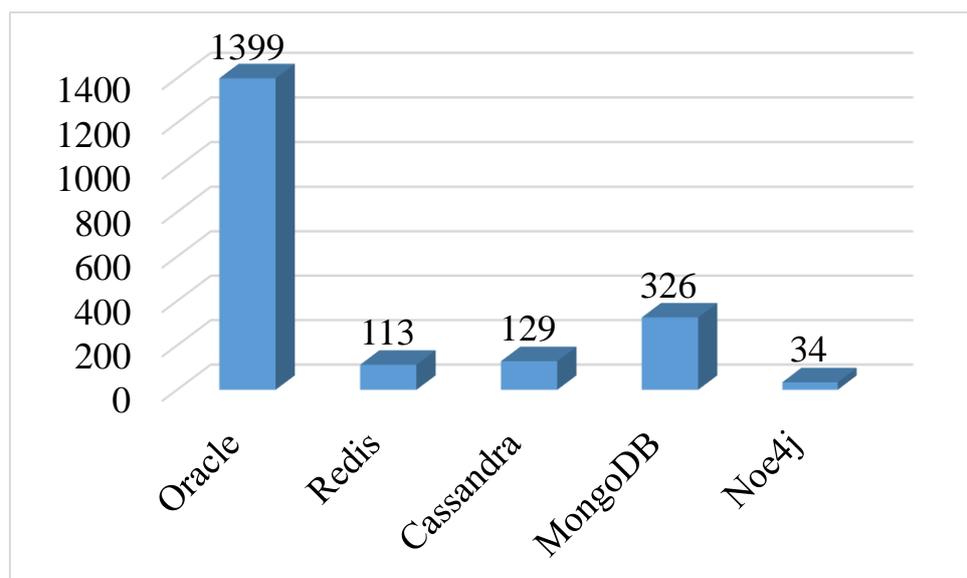


Рис. 2.1. Розповсюдженість баз даних

Джерело: складено на підставі [3; 4]

Далі для обраних для дослідження баз даних наводиться детальна характеристика, специфіка і область застосування. Для цього необхідно з'ясувати, як в кожній з них подається інформація.

Зокрема, у реляційній моделі дані групуються в колекції або таблиці, які зберігаються в структурованому вигляді і зв'язуються за допомогою атрибутів. База даних будується на створенні відносин між ними, які дозволяють пов'язувати різні таблиці між собою, тим самим перетворюючи їх у єдине сховище.

Сховище ключ-значення представляє дані у вигляді асоціативного масиву. У такій системі не потрібна побудова схеми даних і суворої типізації даних, в ній відсутні зв'язки між значеннями. Така простота забезпечує масштабованість і швидкодію.

Документно-орієнтована БД являє собою систему зберігання ієрархічних структур документів, які можна представити у вигляді дерева. Структура дерева починається з кореневого вузла і може мати кілька внутрішніх і листових вузлів. Дані не мають строгих зв'язків між собою, але ця модель дозволяє створювати складні структури.

Для графової моделі дані представляються у вигляді вузлів графа, а зв'язки між ними – у вигляді ребер. Вона дозволяє створювати щільні зв'язки між об'єктами, не створюючи жорсткої структури.

Розподілені сховища зберігають дані, групуючи їх за стовпцями. Рядки і стовпці такої моделі утворюють розріджену матрицю, зв'язки між даними не такі суворі, як в реляційної моделі. В якості ключів можна використовувати як рядки, так і стовпці. Далі необхідно позначити параметри, які є важливими для тестування. Для цього потрібно зрозуміти, які основні вимоги пред'являються до баз даних і яким чином їх можна досліджувати.

Однією з найбільш важливих характеристик є висота швидкодії системи. Швидкість обробки даних багато в чому визначає швидкодію всього програмного комплексу. Для тестування часу відгуку необхідно провести

запити кожного типу (читання, зміна, додавання, видалення даних) і провести порівняння часу їх виконання.

Масштабованість баз даних є наступним важливим параметром. Для будь-якої системи масштабованість, тобто можливість справлятися зі збільшенням навантаження при додаванні ресурсів, є обов'язковим критерієм. У разі СУБД розглядаються два види масштабування: горизонтальна і вертикальна. При горизонтальному масштабуванні збільшується число серверів СУБД, що дозволяє розподілити навантаження між ними. Воно може ускладнити роботу в сильно пов'язаних моделях баз даних. У свою чергу, вертикальне масштабування має на увазі збільшення потужності одного сервера. В цьому випадку не доводиться розподіляти дані, але доводиться замінювати апаратне забезпечення сервера або додавати нові вузли.

СУБД повинна забезпечувати цілісність даних до руйнування і знищення, пов'язаних з несправністю технічних засобів, системними помилками і помилковими діями користувачів. Вона включає в себе:

- відновлення даних;
- захист від помилок при видаленні даних;
- коректну роботу в розрахованому на багато користувачів режимі.

База даних не здатна забезпечити цілісність інформації, що в ній зберігається, і не може вважатися надійним.

Попередні параметри визначали надійність і ефективність СУБД, але навіть найшвидша і безпечна база даних виявиться марною, якщо дозволені запити не дозволяють виконувати всі необхідні для даного завдання маніпуляції з даними. Тому останнім параметром варто вказати можливість мови запитів, яка використовується для звернення до бази даних. Нижче наводиться порівняльна табл. 2.2, в якій показані можливості масштабування системи і мови запитів для обраних для дослідження баз даних.

Таблиця 2.2

Загальний рейтинг баз даних

| База даних | Масштабування | Запити |
|------------|---------------|---------------|
| Oracle | Вертикальне | SQL |
| Regis | Вертикальне | Команди Redis |
| Cassandra | Горизонтальне | CQL |
| MongoDB | Горизонтальне | Java Script |
| Neo4j | Вертикальне | Cypher |

Джерело: складено на підставі [3; 4]

У цілому, питання оптимального вибору рішення для зберігання даних стає все більш актуальним з розвитком нових моделей баз даних. Представлені аналіз і порівняльна характеристика дозволяють створити уніфікований підхід до тестування і дослідження СУБД, структури представлення даних і мови запитів, що в свою чергу дозволить ефективно використовувати ресурси системи.

В даний час використовується значна кількість платформ і систем Big Data. Системи обробки Big Data є фреймворками, для використання яких необхідно зістикувати їх з іншими фреймворками. В аналітичному звіті Big Data Analytics Market Study, наводиться така діаграма інфраструктур Big Data, впроваджених на підприємствах, представлена в розрізі розмірів підприємств (рис. 2.2).

Нижче описано специфіку методів і технологій, що використовуються для Big Data в обчислювальних задачах.

1. Технології Big Data. Розроблені програмні платформи і технології розподілених обчислень і Big Data адаптуються до вирішення масштабних завдань в області економіки, бізнесу, соціології, державного управління тощо.

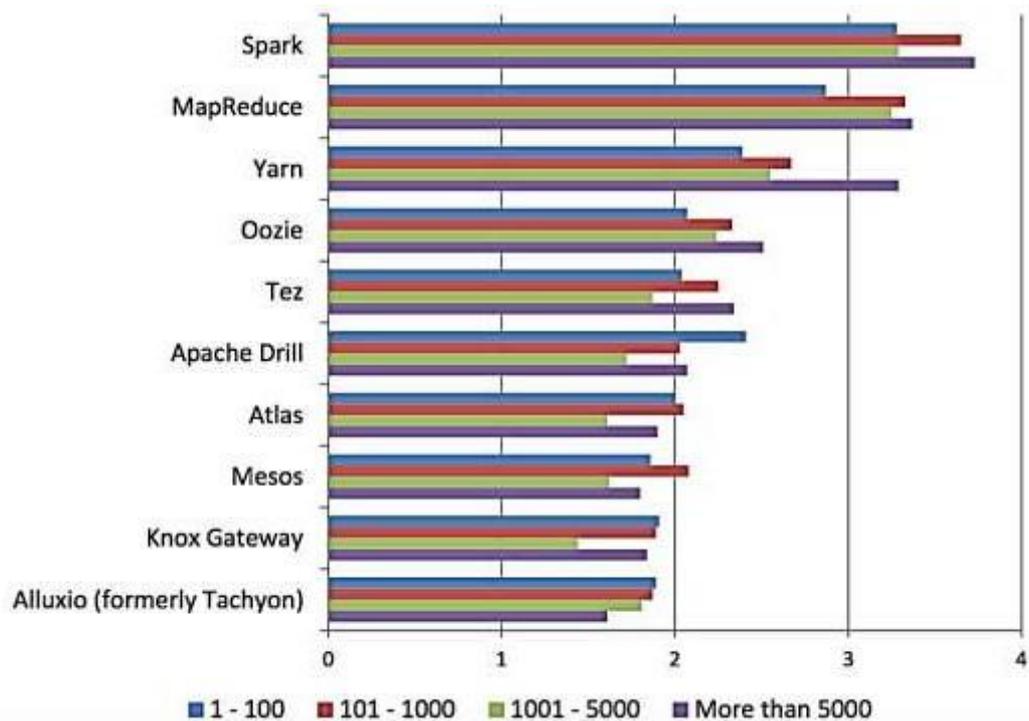


Рис. 2.2. Інфраструктури Великих даних, впроваджених на підприємствах

Джерело: складено на підставі [3; 4]

При цьому критично важливими є зміни в управлінні даними, ІТ-інфраструктури та компетенціях персоналу.

У світі Великих даних сучасні технології роблять можливим обробку і аналіз величезної кількості даних, в деяких випадках – ВСІХ даних, що стосуються того чи іншого явища (не покладаючись на випадкові вибірки) в їх первозданному вигляді – структуровані, неструктуровані, потокові.

2. Управління Big Data. Стратегічним завданням Лабораторії хмарних технологій є аналітика Big Data. Наявний досвід щодо застосування цих рішень дозволяє створювати програмні платформи, які в сукупності зі спеціалізованими прикладними пакетами, адаптованими для таких платформ, дозволяють досягти якісно нового рівня обробки й аналізу даних, принципово неможливого без застосування даних технологій.

3. Аналітика Big Data. Нові аналітичні додатки висувають наступні вимоги до платформи для роботи з Big Data:

- об'єднувати і управляти усім розмаїттям, швидкістю та обсягом, достовірністю і обґрунтованістю даних;
- мати можливість застосовувати передову аналітику до інформації в її вихідній формі;
- візуалізувати всі доступні дані для спеціального аналізу;
- наявність середовища проектування для створення нових аналітичних додатків;
- можливість оптимізації робочого навантаження і планування;
- управління і захист.

4. Персонал Big Data. Однією з найважливіших компонент є підготовка фахівців, здатних розвивати парадигму Big Data:

- розв'язувати задачі;
- розвивати платформу і побудовані на ній рішення;
- супроводжувати й адмініструвати систему;
- пропонувати нові підходи.

5. Завдання, що стоять перед Лабораторією хмарних технологій і Big Data:

- розгортання платформи управління Big Data;
- створення кластерних, хмарних і / або грид-інфраструктур для зберігання, передачі, обробки й аналізу Великих даних;
- навчання хмарним і грид-технологіями (користувачів, адміністраторів і розробників);
- адаптація пакетів прикладних програм (тобто програм) для роботи в цих інфраструктурах;
- надання хмарних ресурсів і інфраструктури для користувачів;
- вибір рішень і методів для організації зберігання і управління даними (SQL, NoSQL (сховища типу «ключ-значення», масштабовані розподілені сховища, документоорієнтовані СУБД, графові СУБД);

- управління життєвим циклом даних (створення, обробка, аналіз, систематизація, візуалізація, формування звітів, видалення);
- дослідження в області аналітики Big Data – методи аналізу і передбачувальна модель (математична статистика, аналіз часових рядів, кластерний аналіз, кореляційний аналіз, регресійний аналіз, нейронні мережі, генетичні алгоритми, нечіткі алгоритми, розпізнавання образів й ін.).

6. Вимоги до завдань замовника. Big Data припускають щось більше, ніж просто аналіз величезних обсягів інформації. Проблема не в тому, що організації створюють величезні обсяги даних, а в тому, що більша їх частина представлена в форматі, що погано відповідає традиційному структурованому формату БД, – це веб-журнали, відеозаписи, текстові документи, машинний код або, наприклад, геопросторові дані. Все це зберігається в безлічі різноманітних сховищ, іноді навіть за межами організації. В результаті організації можуть мати доступ до величезного обсягу своїх даних і не мати необхідних інструментів, щоб встановити взаємозв'язок між цими даними і зробити на їх основі значущі висновки. Поняття великих даних має на увазі роботу з інформацією величезного обсягу і різноманітного складу, вельми часто оновлюваної та такої, що знаходиться в різних джерелах з метою збільшення ефективності роботи, створення нових продуктів і підвищення конкурентоспроможності.

У зв'язку з цим, потенційний замовник, перш за все, повинен визначити:

1) потребу в зберіганні, передачі, обробці й аналізі досить великих обсягів різноманітних даних, що підпадають під визначення «Big Data». Накопичений світовий досвід показує, що дані в банківській сфері (в задачах залучення клієнтів, оцінки позичальників, протидії шахрайству й ін.), в сфері телекомунікацій (в задачах підвищення якості зв'язку, виявлення шахрайства та ін.), в торгівлі (прогнозування трендів купівельного попиту, оптимізація цін і проведених акцій тощо), в медицині (визначення найбільш ефективних методів лікування, контроль ходу лікування тощо) і ряді інших областей

(маркетинг, енергетика, страхування, житлово-комунальне господарство) відносяться до категорії «Big Data». До вимог «обсягу» можна додати вимоги високої швидкості приросту даних і потреби у високій швидкості їх обробки, різноманітні даних, тобто потреби одночасної обробки різних типів структурованих і неструктурованих даних);

2) обчислювальна ресурсомісткість процедур обробки й аналізу цих даних.

Найбільш часто в якості апаратної платформи для систем обробки Big Data використовуються багатопроцесорні обчислювальні системи без спільного використання ресурсів (англ. – Shared Nothing Architecture), які можуть забезпечити масово-паралельну обробку даних, масштабовану без деградації на сотні і тисячі вузлів. Існує також ряд апаратно-програмних рішень, орієнтованих на обробку Big Data. Дані рішення є готовими до установки в ЦОД, телекомунікаційні шафи, містять кластер серверів і програмне забезпечення для масово-паралельної обробки. Найбільш відомими апаратно-програмними комплексами для обробки Big Data є такі:

- Aster MapReduce Appliance корпорації Teradata;
- Big Data Appliance корпорації Oracle;
- Greenplum Appliance корпорації EMC.

Існують також комплекси для аналітичної обробки в оперативній пам'яті: система Hana компанії SAP і система Exalytics корпорації Oracle, побудована на основі реляційної СУБД TimesTen і багатовимірної СУБД Essbase. У список найбільш популярних технологій обробки Big Data в даний час, мабуть, входять NoSQL, MapReduce і Hadoop, а також паралельні СУБД. Термін NoSQL (Not Only SQL – не тільки SQL) позначає набір підходів до реалізації сховищ даних, які засновані на моделі даних, відмінній від реляційної.

Інтерес до технологій NoSQL блискавично виник після того, як корпорація Google на початку 2000-х опублікувала документацію про розподілені файлові системи.

BigTable, яка здатна обробити 20 Петабайт інформації в день, є базисом пошукової системи Google і популярних сервісів Gmail, Google Maps, Google Earth й ін. Особливістю NoSQL-технологій є ідея необмеженого масштабування і відмова від узгодженості даних на догоду продуктивності їх обробки. Системи класу NoSQL проектуються з розрахунком на необмежене горизонтальне масштабування: додавання або видалення вузлів в кластері не повинно позначитися на працездатності системи. На відміну від реляційних СУБД, NoSQL системи не підтримують транзакції з ACID-властивостями (англ.: Atomicity – атомарність, Consistency – узгодженість, Isolation – ізолюваність, Durability – довготривалість).

В силу відсутності підтримки транзакцій NoSQL-рішення не можуть використовуватися в додатках, де зазначені властивості є необхідністю, наприклад, в системах обслуговування бірж і банків. Однак в системах обслуговування запитів багатомільйонної web-аудиторії користувачів властивості ACID, незважаючи на всю їх привабливість, забезпечити практично неможливо. Тому в NoSQL-системах, що обробляють Big Data, узгодженість даних жертвують заради досягнення двох інших властивостей: доступність даних і стійкість до розбиття даних по вузлах обчислювальної системи.

Подібна особливість розподілених обчислень – можливість забезпечити не більше двох властивостей з трьох, зазначених вище, – відома як теорема CAP (аббревіатура від англійських назв властивостей: Consistency, Availability, Partition tolerance), хоча, строго кажучи, не є такою в силу відсутності чіткої формальної постановки задачі. На сьогодні існує велика кількість різноманітних NoSQL-систем. Відповідно до класифікації порталу nosql-database.org, розрізняють сховища «ключ-значення», документ-орієнтовані системи, стовпчасті сховища і системи обробки графів.

Зокрема, сховища «ключ-значення» будуються на основі асоціативних масивів, що дозволяють працювати з даними по ключу; будь-яка інформація

про структуру значень не зберігається. Такі сховища використовуються в основному для зберігання зображень та створення спеціалізованих файлових систем, як кешей для об'єктів. Приклади систем «ключ-значення»: Redis, Scalaris, Riak, DynamoDB й ін.

Документ-орієнтовані системи призначені для зберігання ієрархічних структур даних (документів). Документи можуть бути згруповані в колекції, причому колекції можуть містити інші колекції. Системи даної групи застосовуються в системах управління вмістом, видавничій справі, документальному пошуку й ін. Приклади подібних систем: MongoDB, Berkeley DB XML, SimpleDB, CouchDB й ін.

Стовпчасті сховища засновані на ідеї зберігання даних на диску не по рядках, як це роблять реляційні СУБД, а по колонках. З точки зору SQL-клієнта, дані представлені в вигляді таблиць, проте фізично ці таблиці є набором колонок, кожна з яких представляє собою таблицю з одного поля. При цьому фізично, на диску, значення одного поля зберігаються послідовно. При виконанні вибірки даних така організація зберігання забезпечує скорочення часу обробки за рахунок виконання читання тільки тих полів, які потрібні в запиті. Модель даних включає в себе поняття вершин, ребер і властивостей. Обробка даних виконується у вигляді послідовності з двох кроків: Map і Reduce. На кроці Map узел-майстер отримує вхідні дані і розподіляє їх по вузлам-робочим. На кроці Reduce вузол-майстер виконує згортання даних, попередньо оброблених робочими вузлами, і відправку кінцевого результату користувачеві. Широко відомий експеримент корпорації Google, в ході якого проводилося сортування 1 петабайт даних за допомогою фреймворка MapReduce. Дані були представлені у вигляді 10 трлн. записів розміром 100 байт кожна. Кластер з 4000 комп'ютерів виконав сортування безпрецедентного для такого типу задач обсягу даних за 6 годин 2 хвилини.

Hadoop – проект фонду Apache Software Foundation, який являє собою фреймворк для розробки і виконання розподілених програм, що працюють на

кластерах з сотень і тисяч вузлів. Використовується для реалізації високонавантажених веб-сайтів (наприклад, Yahoo! і Facebook), розроблений на Java в рамках обчислювальної парадигми MapReduce. Hadoop складається з наступних частин:

- набір інфраструктурних програмних бібліотек і утиліт Hadoop Common;
- розподілена файлова система HDFS;
- система для планування завдань і управління кластером YARN;
- платформа програмування і виконання розподілених MapReduce-обчислень Hadoop MapReduce. З Hadoop асоційований ряд проектів і технологій, багато з яких спочатку розвивалися в рамках цього проекту, але згодом стали самостійними.

Розподілене сховище даних HIVE забезпечує управління даними в HDFS і надає мову запитів HiveQL, засновану на SQL. Запити HiveQL транслуються в MapReduce-завдання Pig, середовище виконання і високорівневу мову для опису обчислень в Hadoop. Програми Pig також транслуються в MapReduce-завдання. Паралельні СУБД, на противагу NoSQL-рішень, не відмовляються від реляційної моделі даних. Базовою концепцією паралельних СУБД є фрагментний паралелізм, який має на увазі горизонтальну фрагментацію кожної таблиці бази даних по дисках кластерної системи. Спосіб фрагментації визначається функцією фрагментації, яка для кожного запису таблиці обчислює номер обчислювального вузла кластера, де повинен зберігатися даний запис. На кожному вузлі кластерної системи запускається паралельний агент (ядро СУБД), що обробляє запити користувачів. Один і той же запит паралельно виконується кожним агентом над «своїми» фрагментами таблиць бази даних, і потім отримані часткові результати зливаються в результуючу таблицю. Незважаючи на незалежну обробку агентами «своїх» фрагментів бази даних, для отримання коректного результату необхідно виконувати пересилання

записів під час виконання запиту (при виконанні операції з'єднання двох таблиць по загальній колонці).

Для організації таких пересилань виконується розпаралелювання послідовного плану запиту: на додаток до звичайних реляційних операцій (з'єднання, вибірка, проекція і ін.). В потрібні місця плану запиту вставляється спеціальна операція *exchange*, яка забезпечує пересилку «чужих» записів відповідним паралельним агентам і прийом «своїх» записів від них, не порушуючи ходу виконання запиту. В силу того, що різні фрагменти таблиці можуть мати істотно різні розміри (така ситуація носить назву «Перекіс даних»), в паралельних СУБД для балансування завантаження паралельних агентів використовується техніка часткової реплікації бази даних. До класу паралельних СУБД можна віднести системи *Greenplum*, *Netezza*, *EXASOL* й ін.

Слід ще сказати про взаємопроникнення технологій паралельних СУБД і *MapReduce*. *Hadoop* забезпечує комунікаційну інфраструктуру, що об'єднує вузли кластера, на яких запускаються екземпляри СУБД *PostgreSQL*. Запити користувача на мові *SQL* транслуються в завдання для середовища *MapReduce*, які далі передаються в екземпляри СУБД. У системах *Greenplum* і *nCluster* модель *MapReduce* реалізується всередині СУБД, і можливостями цих реалізацій можуть користуватися розробники аналітичних додатків: в *Greenplum Database* – поряд з *SQL*, а в *nCluster* – з *SQL*. Згадаємо також бібліотеки і фреймворки, що використовуються для інтелектуального аналізу *Big Data*. *Apache Mahout* є вільною бібліотекою алгоритмів машинного навчання з відкритим кодом. Реалізація бібліотеки виконана на мові *Java*, масштабованість досягається за рахунок використання фреймворка *Hadoop*. Інструментарій, що надається бібліотекою, зараз дозволяє реалізувати рекомендаційні системи, кластеризації та класифікацію *Big Data*.

Вищезгаданий апаратно-програмний комплекс *Big Data Appliance* корпорації *Oracle* містить інтегровані програмні засоби на основі мови *R* і фреймворка *Hadoop*; підтримка мови *R* вбудована також в СУБД *Oracle* і

Netezza. Вільні фреймворки RHIPE і RHadoop дозволяють інтегрувати можливості мови R і фреймворка Hadoop для аналітичної обробки Big Data.

Головними джерелами отримання Big Data є датчики, вимірювальні пристрої, обладнання мереж стільникового зв'язку, пристрої аудіо- відео реєстрації, потоки повідомлень всередині соціальних мереж, блоги і новинні стрічки, супутники. За даними компанії Cisco [5], обсяг мобільного трафіку в 2014 році склав 2,5 екзабайт на місяць, а до 2019 року він збільшиться більш ніж в 9,5 разів. Розвиток і повсюдне поширення електронних джерел інформації стало поштовхом до впровадження технологій BD в різних галузях діяльності. Це пов'язано в тому числі зі збільшенням накопичення інформації в мережі Internet, що стало особливо помітно зі збільшенням популярності соціальних мереж. На початку 2010-их років багато розробників ринку IT-продуктів почали використовувати розглянуту технологію безпосередньо в контексті обробки Big Data і випускати власні аналітичні рішення.

Незважаючи на удавану перевагу перед класичними засобами, аналітика соціальних медіа мережі Інтернет має ряд проблем. Основними недоліками наявних методів соціальної аналітики є [6-8]:

- низька поширеність технологій для роботи з великими обсягами даних, що призводить до неефективного використання обчислювальних ресурсів і додатковим економічним витратам;
- спрямованість існуючих рішень на вузький спектр завдань, пов'язаних в великій мірі зі статистичними соціальними дослідженнями і маркетингом;
- труднощі роботи з інформаційними структурами, що володіють високим ступенем неструктурованості і гетерогенності;
- наявність у соціальних даних таких властивостей, як суперечливість, неповнота, лінгвістична некоректність, емоційна забарвленість, що заважають їх об'єктивній оцінці і інтерпретується;
- питання правомірності використання видобутої інформації третіми особами;

- відсутність інтеграції різних інформаційних технологій аналізу текстових і мультимедіа даних в рамках єдиної системи;
- складності зберігання і захисту інформації.

Для усунення вищевказаних недоліків необхідно створення нових і доопрацювання існуючих методів обробки персоніфікованих соціальних даних, складання алгоритмів аналізу різнорідної текстової та мультимедійної інформації, моделей зберігання і представлення вихідних, проміжних аналізованих і результуючих даних.

В якості інструменту покращення ефективності методів обробки персоніфікованих соціальних даних було обрано теорію графів та алгоритми перерахування розрізів.

Результати соціального профілювання, що ґрунтуються на аналізі загальнодоступних джерел інформації, повинні володіти рядом властивостей, такими як, явна структурованість, масштабованість, представлення як у вигляді таблиць, так і зображень (графи, діаграми) [9]. Це необхідно для забезпечення можливості їх використання в прикладних задачах, пов'язаних з дослідженнями суспільства. В даний час найбільш актуальними областями застосування соціальних профілів є маркетинг і правоохоронна діяльність [7,8]. До першої групи належать завдання колаборативної фільтрації, рішення яких дозволяє підвищити якість рекламних кампаній за рахунок врахування інтересів цільової аудиторії і передбачення її реакції. Як наслідок, це дозволяє оптимізувати витрати на маркетингові дослідження (проведення опитувань, формування цільових груп тощо), таргінг просувати товари і послуги, а також підвищити лояльність клієнтів. В області правоохоронної діяльності завдання побудови соціальних профілів перетинається з розвідкою на базі відкритих джерел (OSINT). Органи забезпечення правопорядку використовують персональні і групові профілі для оцінки рівня соціальної напруженості і ймовірних реакцій суспільства на провокаційні події, формування «груп ризику», збору досьє для слідчо-судових заходів.

2.2. Метод управління соціальними процесами на основі з'єднання тріад

Відомо, що моделювання є одним з ефективних методів дослідження складних систем, в тому числі і соціальних. Побудова моделей, які адекватно відображають особливості будови, функціонування, розвитку і поведінки соціальних систем в різних складних ситуаціях, є важко вирішуваною проблемою. При цьому дослідники вказують на різницю між моделями і системами, вважаючи, що система – це математична абстракція, яка служить моделлю динамічного явища. Вивчивши різні наукові підходи до моделювання соціальних процесів і об'єктів (далі – СПО), автор дотримується такої методології моделювання як послідовності побудови моделей для вивчення будови і властивостей СПО, в яких, очевидно, відбивається непереборний суб'єктивізм автора, викликаний його цілями і завданнями дослідження.

При цьому автор дотримується думки, що процедура моделювання як абстрактна система складається з множини (Φ) моделей різного типу, взаємопов'язаних (H) одна з одною і організованих в певну структуру (Str), яка володіє при даних характеристиках (P) моделей і відносин між ними такими властивостями, які дозволяють досягти поставленої мети (Z) найкращим чином.

Далі розглянемо побудову моделей за окремими етапами.

На першому етапі будується абстрактна модель соціального процесу або об'єкта. В основу побудови абстрактної моделі покладена гіпотеза про те, що СПО підкоряються системним законам і принципам. Це дозволяє представити абстрактні моделі в формі системних структурних (організаційних) моделей.

Системна модель заснована на понятті системи і являє собою множину концептів (об'єктів, понять) і множини зв'язків між ними, що утворюють

єдине ціле. Ця єдність може бути представлена як в ієрархічній, так і багатопов'язаній, а також тріадній формі.

Структурна модель є найпершою сходинкою до конкретизації будови системи. Структурна модель – це системна модель, у якій конкретизовано цілісність. Структура (будова) складається з множини концептів і зв'язків між ними. При цьому структура може бути також лінійною, ієрархічною, багатопов'язаною, мережевою, петлевою тощо.

Відзначимо, що абстрактні моделі відображають лише склад і будову реальної системи і не відображають стан системи і режими її функціонування.

На другому етапі будуються знакові моделі як продовження абстрактних моделей, але вже з конкретизацією логічних і знакових відносин між концептами (об'єктами, поняттями), що відображає якісну відповідність реальним системам. Це знакові когнітивні моделі (далі – ЗКМ).

На третьому етапі відбувається кількісне розширення когнітивних моделей із зазначенням не тільки знакових, але і кількісних відносин між концептами, а також кількісних характеристик останніх. Це числові когнітивні моделі (далі – ЧКМ). У підсумку виходить алгебраїчна модель оцінки статичних станів (статична модель) системи.

На четвертому етапі відбувається врахування динамічних властивостей самих концептів, що характеризують індивідуальну інерційність сприйняття і переробки ними в часі вхідної інформації.

Результатом цього етапу є побудова лінійної когнітивної динамічної моделі (далі – ЛДМ), що відбиває особливості руху лінеаризованої системи при переході з одного стану в інший.

На п'ятому етапі будуються нелінійні когнітивні динамічні моделі (далі – НДМ) у формі нелінійних диференціальних рівнянь, що відображають вплив зовнішніх і внутрішніх факторів на траєкторію руху системи.

Відзначимо, що на останніх трьох етапах робиться математизація соціальних процесів і систем в формі алгебраїчних (статичних моделей – СМ), а також динамічних, лінійних і нелінійних моделей.

На заключному, шостому етапі в результаті дослідження систем і процесів на основі даних класів моделей повинні бути з'ясовані закономірності:

- в будові системи, а також у формуванні її синергетичних властивостей;
- в поведінці системи при дії різних збурень;
- в появі точок біфуркації в процесі функціонування системи.

Тут досліджуються функціональні моделі контролю (*K*), управління (регулювання) (*УР*), планування (*Пл*) і розвитку (*P*).

Виявлені закономірності і властивості можуть використовуватися для передбачення і пояснення поведінки системи.

На основі системного аналізу всі розглянуті вище моделі можуть бути об'єднані в єдину тріадну систему моделей, розташованих за наступною ієрархією:

- рівень 1 абстрактних (системних, структурних) моделей (АМ), що відображають знання про будову системи;
- рівень 2 пізнання якісних і кількісних характеристик системи (знакові (ЗКМ) і числові (ЧКМ) когнітивні моделі);
- рівень 3 пізнання форм руху системи (статичні (СМ), лінійні динамічні (ЛДМ) і нелінійні динамічні моделі (НДМ));
- рівень 4 пізнання функціональних властивостей системи (функцій контролю, управління (регулювання), планування та розвитку).

Системне мислення у формі тріад сприяє породженню нових пов'язаних одна з одною тріад, що підтверджує логічну послідовність і безперервність процесу системного мислення. Покажемо, що кожен процес діяльності людини, спрямований на досягнення мети, можна представити у

вигляді з'єднання тріад, кожна з яких представляє етап життєвого циклу процесу руху до мети.

В основі всіх тріад руху відповідно до принципу переходу лежить деяка базова тріада: «Початковий (вихідний) стан (1) – Спосіб (технологія, алгоритм) переходу (2) – Новий (кінцевий) стан (3)».

Як приклад розглянемо задачу задоволення деякої множини P_m соціальних потреб людського суспільства, моделлю якого виступає деяка множина Z цілей. Процес виникнення множини P_m соціальних потреб в різних ситуаціях відноситься до класу неформалізованих задач, які вирішуються за допомогою інтелекту людини. Тому будемо вважати, що досягнення множини Z цілей повністю забезпечує задоволення цих соціальних потреб.

Отже, наприклад, в якості соціальних потреб розглянемо: створення деякої множини N_b інноваційних продуктів, їх реалізація на ринку і отримання прибутку. При цьому множина Z цілей будемо вважати початковою множиною, яка сформована на основі аналізу множини потреб.

Процес досягнення множини Z цілей розіб'ємо на t етапів, кожному з яких буде відповідати своя тріада (рис. 1).

Етап 1. На даному етапі множини Z цілей як множини бажаних кінцевих результатів тим чи іншим способом декомпозиється на множину M_z підцілей. Подальша конкретизація підцелей як в просторі й в часі, так і в кількісному й якісному відношенні дозволяє сформулювати множину 3 завдань. Таким чином на першому етапі формується тріада «Мета – Підцілі – Завдання».

Етап 2. На другому етапі множину 3 завдань відображають на множину F функцій, пов'язаних з рішенням цих завдань. Найбільш поширеними функціями при інноваційній діяльності стосовно дослідження соціальних профілів є такі: проектування, планування, управління, організація, контроль, діагностика, прогнозування, конструювання, виробництво, збір, зберігання й обробка інформації, адаптація, навчання, розвиток тощо. Для кожного завдання визначається свій алгоритм A рішення в залежності від

того, в якому режимі виконується функція (ручному, автоматичному, автоматизованому тощо). Бажано, щоб алгоритм був оптимальним в сенсі обраного критерію (максимальної точності, швидкодії, ресурсозбереження тощо). Таким чином, на другому етапі формується тріада: «Завдання – Алгоритм – Функція».

Етап 3. На цьому етапі множину F виконуваних функцій відображають на множину S функціональних систем, які повинні: забезпечити надійне, точне, своєчасне й якісне виконання функцій при заданих умовах функціонування. Для кожної системи впливає з деякого різноманіття вибрати ту структуру (Stz), яка забезпечила б ефективне функціонування системи. Параметризація і призводить до конкретної моделі реалізації системи. Отже, сформована тріада: «Функція – Структура – Система».

Етап 4. Далі множину S систем відображають на множину K конструкцій цих систем. Тут конструкція системи розглядається як образ її практичної реалізації. Різноманіття способів конструювання, технічних рішень, структур системи породжує множини B_k варіантів конструкцій. Варіанти відрізняються один від одного комбонуванням елементів, використанням матеріалом, фізичними принципами дії. З них для кожної системи вибирається такий варіант, який задовольняв би заданим критеріям. Наприклад, для проектованої системи управління технічними об'єктами це різноманіття конструкцій виражається в тому, що система може бути реалізована в одному з наступних класів: механічних, електромеханічних, електронних, гідромеханічних, пневмомеханічних тощо систем. Таким чином, сформована тріада: «Система – Варіанти конструкцій – Конструкція».

Етап 5. На цьому етапі множини K конструкцій відображають на множини P_n виробничих процесів. При цьому множина K конструкцій повинна бути погоджена з множиною T_x технологій їх виготовлення в даному виробництві. Етап закінчується формуванням тріади: «Конструкція – Технологія – Виробництво».

Етап 6. Далі множина Π_n виробничих процесів породжує множину N_e видів продукції, що випускається. Однак виробничі процеси не можуть бути реалізовані без множини R ресурсів: трудових, паливно-енергетичних, матеріальних, фінансових, технічних, технологічних, інформаційних тощо. В результаті формується тріада: «Виробництво – Ресурси – Продукція».

Етап 7. На сьомому етапі множина N_e інноваційних видів продукції, що випускаються, реалізується на ринку за деякою ринковою ціною і за вирахуванням витрат і податків виходить множина P_r прибутків, в досягненні яких і була сформульована мета (потреба). Так як процес реалізації множини інноваційних видів продукцій на ринку відбувається в умовах протидії зовнішнього середовища у вигляді конкуренції, то його неможливо ефективно здійснити без множини C_p способів і засобів підтримки конкурентоспроможності продукцій на ринку. В результаті формується тріада: «Продукція – Підтримка конкурентоспроможності – Прибуток».

Етап 8. На заключному етапі здійснюється множина J оцінок відповідності множині P_r отриманих прибутків від реалізації інноваційних видів продукції множини Z початкових цілей, тобто формується тріада: «Прибуток – Оцінка – Мета».

Весь процес досягнення множини Z цілей представимо у вигляді з'єднання тріад (рис. 2.3).

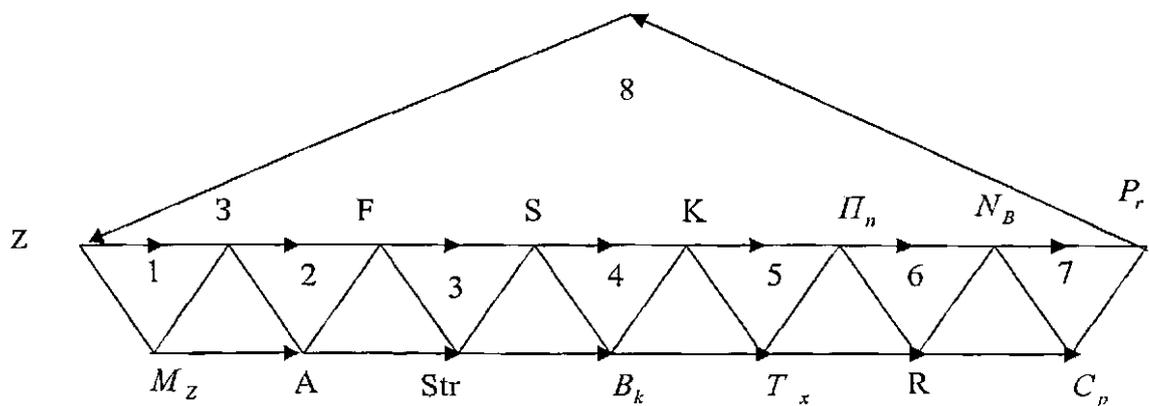


Рис. 2.3. Системна модель досягнення соціальних цілей за допомогою тріад

Джерело: авторська розробка

Відзначимо, що вільні кінці триад також можна з'єднати послідовно один з одним, що призведе до утворення нових триад. При цьому нижчий рівень відображає перехід від однієї технології до іншої.

У кожній триаді доводиться вирішувати завдання вибору (прийняття рішення) в силу різноманіття можливих варіантів рішення. При цьому критерії вибору повинні бути адекватними оцінюваним характеристиками об'єктів.

Таким чином, дана системна концепція досягнення мети через побудову послідовно з'єднаних триад як елементарних систем охоплює всі найважливіші етапи організаційних дій, об'єднуючи їх в єдину систему зі зворотним зв'язком. Даний підхід знижує невизначеність у вирішенні завдань вибору різних варіантів при побудові проміжних об'єктів. З точки зору даного підходу оптимізація шляху до досягнення мети лежить в розвитку і вдосконаленні кожної триади.

2.3. Застосування методу системологічного класифікаційного аналізу для розробки фрагменту концептуальної моделі функцій системи консолідації інформації

У сучасних умовах існує об'єктивна необхідність дослідження систем консолідації інформації, враховуючи те, що вони дозволяють забезпечити оптимальний формат даних для їх подальшої обробки в межах вирішення визначеного аналітичного завдання стосовно обраної сфери дослідження. Особливо це стосується комерційної діяльності, коли принципово важливим є здійснення реструктуризації інформації, усунення повторень ідентичної інформації, що, у кінцевому рахунку підвищує загальну якість корпоративної інформації і відповідно – рівень довіри споживачів до підприємства.

На нинішньому етапі не існує повноцінної класифікації систем консолідації інформації. Присутні лише окремі спроби здійснити їх типологію відповідно на основі різнохарактерних технічних критеріїв,

зокрема, відносно архітектури збереження інформації, моделі об'єднання інформації, способу обробки запитів користувачів. Крім того, існують вузькоспеціалізовані типології систем консолідації інформації, зокрема, такі, що орієнтовані на накопичення й обробку знань, наприклад, виключно про фінанси, клієнтів чи зовнішнє середовище компанії [9; 12].

Проте в межах усіх наведених типологій присутня складність встановлення першочерговості загальних ознак об'єктів при поділі їх множини на підмножини.

При цьому недоліками здійснення подібних типологій слід уважати наступні:

- вибір із підмножини ознак об'єктів однакового ступеня спільності найбільш суттєвого (при цьому ступінь спільності дослідженого об'єкту передбачає, в якій мірі припускається його вихід за межі заданого критерію еквівалентності);

- складність визначення доцільної кількості ознак і ступенів, а також глибини класифікації.

Враховуючи вищезазначене, у роботі пропонується здійснити дослідження та побудову класифікації систем консолідації інформації та властивостей об'єктів класифікації.

Так, на основі системологічного класифікаційного аналізу в роботі було досліджено та розроблено класифікацію систем консолідації інформації за функціональним призначенням, яка містить класифікацію об'єктів (табл. 2.3) та класифікацію властивостей (табл. 2.4).

При цьому метарівень системного опису видів систем консолідації інформації відповідає дуже високому ступеню абстракції та являє собою інформаційні системи. Відповідно, метаскладовою інформаційних систем у контексті даного дослідження є системи консолідації інформації.

Таблиця 2.3

Класифікація інформаційних систем

| | | | | | | |
|----------------------|---------------------------------|--|--|---|--|--|
| Інформаційні системи | Системи консолідації інформації | Деталізовані системи консолідації інформації | Системи консолідації інформації про зовнішнє середовище | Системи консолідації інформації про політико-правовий стан | | |
| | | | | Системи консолідації інформації про економічний стан | | |
| | | | | Системи консолідації інформації про природно-географічні умови | | |
| | | | | Системи консолідації інформації про техніко-технологічний рівень розвитку | | |
| | | | | Системи консолідації інформації про соціально-культурні особливості | | |
| | | | Системи консолідації інформації про внутрішнє середовище | Системи консолідації інформації про клієнтів | | |
| | | | | Системи консолідації інформації про посередників | | |
| | | | | Системи консолідації інформації про постачальників | | |
| | | | | Системи консолідації інформації про конкурентів | | |
| | | | | Системи консолідації інформації про ринкові умови | | |
| | | | Системи консолідації інформації про підприємство | Системи консолідації інформації про виробництво | | |
| | | | | Системи консолідації інформації про фінанси | | |
| | | | | Системи консолідації інформації про менеджмент | | |
| | | | | Системи консолідації інформації про маркетинг | | |
| | | | | Системи консолідації інформації про збут | | |
| | | | Агреговані системи консолідації інформації | | | |

Джерело: авторська розробка

На першому рівні системи консолідації інформації як складова інформаційних систем на метарівні поділяються на деталізовані та агреговані системи консолідації інформації. У свою чергу, на другому рівні деталізовані системи консолідації інформації, як складова систем консолідації інформації

на макрорівні, поділяються на системи консолідації інформації на макрорівні – про зовнішнє середовище, системи консолідації інформації про внутрішнє середовище та системи консолідації інформації про підприємство.

У свою чергу, на третьому рівні системи консолідації інформації про зовнішнє середовище, які відносяться до систем консолідації інформації на макрорівні, поділяються на системи консолідації інформації на мікрорівні: системи консолідації інформації про політичний стан, системи консолідації інформації про економічний стан, системи консолідації інформації про природно-географічні умови, системи консолідації інформації про техніко-технологічний рівень розвитку та системи консолідації інформації про соціально-культурні особливості.

Відповідно, на третьому рівні системи консолідації інформації про внутрішнє середовище, які також відносяться до систем консолідації інформації на макрорівні, поділяються на наступні системи консолідації інформації на мікрорівні: системи консолідації інформації про клієнтів, системи консолідації інформації про посередників, системи консолідації інформації про постачальників, системи консолідації інформації про конкурентів і системи консолідації інформації про ринкові умови.

До систем консолідації інформації про підприємство на третьому рівні (який також є макрорівнем класифікації, що підлягає дослідженню та розробці) відносяться такі системи консолідації інформації на макрорівні: системи консолідації інформації про виробництво, системи консолідації інформації про фінанси, системи консолідації інформації про менеджмент, системи консолідації інформації про маркетинг та системи консолідації інформації про збут.

Класифікацію властивостей представлено у табл. 2.4.

Таблиця 2.4

Класифікація властивостей інформаційних систем

| | | | | |
|---|---|--|---|---|
| Здійснюють пошук й обробку та зберігання інформації | Здійснюють накопичення знань та надають інформацію для прийняття рішень | Здійснюють накопичення знань та надають деталізовану інформацію для прийняття рішень | Здійснюють накопичення знань та надають зведену інформацію про зовнішнє середовище | Здійснюють накопичення знань та надають інформацію щодо законодавчого регулювання функціонування ринку |
| | | | | Здійснюють накопичення знань та надають зведену інформацію щодо тенденцій зміни валового національного продукту, рівня інфляції, оподаткування, регулювання грошової маси |
| | | | | Здійснюють накопичення знань та надають зведену інформацію щодо особливостей клімату й інфраструктури території |
| | | | | Здійснюють накопичення знань та надають зведену інформацію щодо рівня розвитку науки й техніки |
| | | | | Здійснюють накопичення знань та надають зведену інформацію щодо соціально-культурних умов і демографічної ситуації |
| | | | Здійснюють накопичення знань та надають зведену інформацію про внутрішнє середовище | Здійснюють накопичення знань та надають зведену інформацію щодо мотивації, прихильності, намірів клієнтів та їх відношення до підприємства |
| | | | | Здійснюють накопичення знань та надають зведену інформацію щодо умов функціонування посередників та співробітництва з ними |
| | | | | Здійснюють накопичення знань та надають зведену інформацію щодо характеристик продукції та умов співробітництва з постачальниками |
| | | | | Здійснюють накопичення знань та надають зведену інформацію щодо ринкової позиції конкурентів |
| | | | | |

| | | | | |
|--|--|---|--|---|
| | | | | Здійснюють накопичення знань та надають зведену інформацію щодо тенденцій зміни ринкової кон'юнктури |
| | | | Здійснюють накопичення знань та надають зведену інформацію про внутрішнє середовище підприємства | Здійснюють накопичення знань та надають зведену інформацію щодо виробничих процесів на підприємстві |
| | | | | Здійснюють накопичення знань та надають зведену інформацію щодо фінансового стану підприємства |
| | | | | Здійснюють накопичення знань та надають зведену інформацію щодо стратегії управління підприємством |
| | | | | Здійснюють накопичення знань та надають зведену інформацію щодо комплексу маркетингу підприємства |
| | | | | Здійснюють накопичення знань та надають зведену інформацію щодо організації збутової й логістичної діяльності на підприємстві |
| | | Здійснюють накопичення знань та надають узагальнену інформацію для прийняття рішень | | |

Джерело: авторська розробка

Побудована класифікація відповідає правилам формальної логіки:

- використовується одна основа розподілу;
- члени класифікації взаємно виключають один одного (відсутній перетин понять) [4; 27].

У цілому, в процесі виконання роботи була самостійно досліджена та розроблена класифікація систем консолідації інформації, яка має наступні переваги:

- формування системного уявлення щодо наявних систем консолідації інформації, відповідно до їх функціонального призначення;
- зручність вибору для використання необхідних систем консолідації інформації, як для фахівців підприємств, організацій, установ, так і для звичайного користувача;

- можливість чіткого відображення функціональних властивостей об'єктів класифікації на підставі сформульованих родовидових визначень;
- чітка відповідність об'єктів і властивостей систем консолідації інформації.

Висновки до другого розділу

1. Запропоновано тріадну систему моделей, розташованих за наступною ієрархією:

- рівень 1 абстрактних (системних, структурних) моделей (АМ), що відображають знання про будову системи;
- рівень 2 пізнання якісних і кількісних характеристик системи (знакові (ЗКМ) і числові (ЧКМ) когнітивні моделі);
- рівень 3 пізнання форм руху системи (статичні (СМ), лінійні динамічні (ЛДМ) і нелінійні динамічні моделі (НДМ));
- рівень 4 пізнання функціональних властивостей системи (функцій контролю, управління (регулювання), планування та розвитку).

2. Розроблено системну концепцію досягнення мети соціального профілювання через побудову послідовно з'єднаних тріад як елементарних систем, що охоплює всі найважливіші етапи організаційних дій, об'єднуючи їх в єдину систему зі зворотним зв'язком. Даний підхід знижує невизначеність у вирішенні завдань вибору різних варіантів соціальних портретів при побудові проміжних об'єктів. З точки зору даного підходу оптимізація шляху до досягнення мети лежить в розвитку і вдосконаленні кожної тріади.

3. На основі системологічного класифікаційного аналізу досліджено та побудовано класифікацію систем консолідації інформації за функціональним призначенням, яка містить класифікацію об'єктів та класифікацію властивостей.

РОЗДІЛ 3. РОЗРОБКА МОДЕЛЕЙ ПРЕДСТАВЛЕННЯ ДИФЕРЕНЦІАЛІЗОВАНОЇ ІНФОРМАЦІЇ В ІНТЕЛЕКТУАЛЬНИХ СИСТЕМАХ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ

3.1. Розробка нечітких критеріїв для визначення значущості складових інформації соціального профілю

Процес прийняття рішень – це певна послідовність дій, орієнтованих на вирішення проблем системи, який містить аналіз ситуацій, генерацію альтернатив, прийняття рішень.

Прийняття рішень, які ліквідують розрив між сьогоденням і майбутнім станом системи являє собою обґрунтований вибір з наявних альтернатив напрямів діяльності. Загальна схема процесу прийняття рішень зображена на рис. 3.1.

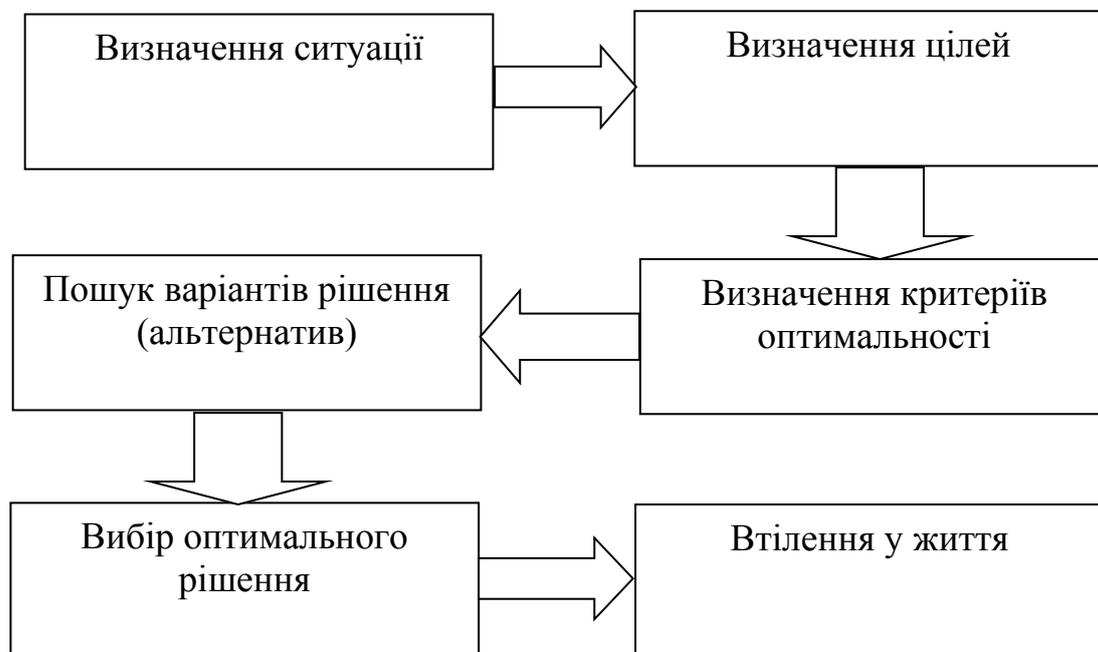


Рис. 3.1. Етапи процесу прийняття рішення

Джерело: складено на підставі [73]

При більш детальному розгляді, основні фази прийняття рішень в соціально-економічних системах полягають у наступному:

- збір інформації про можливі проблеми;
- виявлення та визначення джерел виникнення проблем;
- постановка цілей вирішення проблем;
- обґрунтування стратегії вирішення проблем;
- пропозиція альтернатив рішень;
- вибір оптимального варіанта;
- внесення корективів і узгодження рішення;
- реалізація рішення.

Ефективність прийнятого рішення при здійсненні управління соціально-економічною системою базується на наступних принципах:

- обґрунтованість – управлінські рішення повинні прийматися на підставі максимально вичерпних і достовірних даних;
- своєчасність – рішення не повинні ні відставати, ні випереджати потреби і завдання соціально-економічної системи;
- повнота змісту – рішення повинні охоплювати всю соціально-економічну систему;
- повноважність – суб'єкт управління повинен строго дотримуватися тих правил і повноважень, які йому присвоєно;
- узгодженість з раніше прийнятими рішеннями – забезпечення чіткого причинно-наслідкового зв'язку розвитку системи;
- гнучкість – використання численних можливостей, тобто наявність точних критеріїв, ясних цілей і вичерпної інформації.

Підтримка прийняття рішень є можливою через застосування наступних методів.

1. Методи теорії корисності. Теорія корисності носить аксіометричний характер і дозволяє одночасно вимірювати корисність і суб'єктивну ймовірність, яка дана як об'єктивно відома величина. Це знайшло

відображення в моделі суб'єктивної очікуваної корисності, де ймовірність вже визначається як ступінь впевненості в здійсненні тієї чи іншої події.

2. Метод дерев рішень – полягає в розбитті задачі на ряд підзадач, а їх, в свою чергу, – на інші підзадачі, і так далі. У частині вершин дерева рішень вибір здійснюється безпосередньо особою, що приймає рішення, в іншій частині – на основі суб'єктивної ймовірності звершення подій. Дерево рішень завершується наслідками, кожному з яких приписується певна корисність. Ймовірність кожного результату підраховується як добуток суб'єктивних ймовірностей на шляху, що йде від вершини дерева рішень, а від кінця до початку дерева рішень вибирається результат з найбільшою суб'єктивною очікуваною корисністю. Метод дерев рішень дозволяє визначити оптимальну послідовність дій.

3. Багатокритерійна теорія корисності передбачає, що варіанти рішень мають оцінки за багатьма критеріями. В якості додаткових до загальних аксіомам виступають умови незалежності, на підставі яких доводяться теореми про вид функції корисності. При виконанні умови суворої умовної незалежності по корисності, функція корисності має або адитивний, або мультиплікативний вигляд,

Багатокритерійна теорія корисності, як і попередні методи, будується аксіоматичним способом. Перевагою багатокритерійної теорії корисності відзначається детальна опрацьованість процедур.

4. Метод аналізу ієрархій – методологічна основа для вирішення завдань вибору альтернатив за допомогою їх багатокритеріального рейтингування. Основне застосування методу – підтримка прийняття рішень за допомогою ієрархічної композиції завдання і рейтингування альтернативних рішень. Перерахуємо можливості цього методу:

1) метод дозволяє провести аналіз проблеми, яка подається у вигляді ієрархічно впорядкованих: головної мети рейтингування можливих рішень;

груп однотипних факторів; групи можливих рішень; системи зв'язків, що вказують на взаємний вплив чинників і рішень;

2) метод дозволяє провести збір даних з проблеми. У відповідності з результатами ієрархічної декомпозиції модель ситуації прийняття рішення має кластерну структуру, тобто можливі рішення і фактори, що впливають на пріоритети рішень, розбиваються на кластери;

3) метод дозволяє оцінити суперечливість даних і мінімізувати її.

4) метод дозволяє провести синтез проблеми прийняття рішення, тобто проводиться аналіз проблеми та збір даних по всім кластерам і за допомогою спеціального алгоритму розраховується підсумковий рейтинг – набір пріоритетів альтернативних рішень. Властивості цього рейтингу дозволяють здійснювати підтримку прийняття рішень. Крім того, метод дозволяє побудувати рейтинги для груп факторів, що дозволяє оцінювати важливість кожного фактора;

5) метод дозволяє організувати обговорення проблеми, сприяє досягненню загального рішення;

6) метод дозволяє оцінити важливість врахування кожного рішення і важливість врахування кожного чинника, що впливає на пріоритети рішень;

7) метод дозволяє оцінити стійкість прийнятого рішення.

Структура моделі прийняття рішення в методі аналізу ієрархій є наступною:

- набір альтернативних рішень;
- головний критерій рейтингування рішень;
- набір груп однотипних факторів, що впливають на рейтинг;
- безліч спрямованих зв'язків, що вказують на вплив рішень, критеріїв та факторів один на одного.

5. Наукові методи прийняття рішень. Науковий метод складається з трьох етапів:

1) спостереження – об'єктивний збір та аналіз даних з проблеми і ситуації;

2) формулювання гіпотези – виявлення наявних альтернатив і їх наслідків, а також складання прогнозу, заснованого на цих спостереженнях;

3) верифікація – перевірка гіпотези шляхом спостережень результатів рішення. Так як система складається з взаємопов'язаних частин, то до наукових методів прийняття рішень відносяться ще й системна орієнтація і використання моделей, що дозволяють приймати об'єктивні рішення в складних ситуаціях оцінки альтернатив.

6. Методи прогнозування прийняття рішень. Прогнозування – це метод, який використовує накопичені в минулому знання і поточні припущення майбутнього з метою його визначення, який можна використовувати як основу для планування.

Кількісні методи прогнозування – використовуються, коли діяльність в минулому мала певну тенденцію, яку можна продовжити в майбутньому, і є достатньо інформації для виявлення достовірних тенденцій і залежностей. Два типових методу кількісного прогнозування – це аналіз часових рядів і причинно-наслідкове моделювання.

Зокрема, аналіз часових рядів заснований на припущенні, що те, що трапилося в минулому дає достатнє наближення в оцінці майбутнього. Причинно-наслідкове моделювання використовується в ситуаціях з більш, ніж однією змінною і розглядає залежність між чинником і іншими змінними.

Якісні методи прогнозування використовуються експертами, що володіють інформацією, достатньою для виявлення тенденції або достовірної залежності між змінними. До таких методів належать:

- думка журі – цей метод полягає в поєднанні і усередненні думок експертів в релевантних сферах;
- модель очікування потреб в майбутньому;
- метод експертних оцінок, який являє собою процедуру, що дозволяє групі експертів приходити до згоди.

7. Моделі прийняття рішень. Існує три типи моделей прийняття рішень:

1) класична модель прийняття рішень – представляє собою альтернативний вибір, який повинен бути найбільш економічно ефективним і обґрунтованим логікою.

2) адміністративна модель – ця модель прийняття рішень використовується, коли необхідно прийняти незапрограмовані рішення, які не можуть бути кількісно визначені. Вона носить описовий характер і вчить швидко приймати рішення в невизначених умовах;

3) політична модель прийняття рішень – поєднує в собі деякі характерні риси як адміністративної, так і класичної моделі.

Незапрограмовані в зв'язку з браком інформації рішення необхідно прийняти, але все альтернативні варіанти повинні бути розглянуті, і прийняте рішення повинно мати властивість максимізації.

Розглянемо три популярних типу архітектур систем підтримки прийняття рішень:

- функціональна система підтримки прийняття рішень;
- система підтримки прийняття рішень на основі дворівневого сховища даних;
- система підтримки прийняття рішень на основі трирівневого сховища даних.

Функціональна система підтримки прийняття рішень є найбільш простою з архітектурної точки зору. Характерною рисою функціональної системи підтримки прийняття рішень є те, що аналіз здійснюється з використанням даних з оперативних систем. Серед переваг функціональної системи підтримки прийняття рішень слід виділити такі:

- швидке впровадження за рахунок відсутності етапу перевантаження даних в спеціалізовану систему;
- мінімальні витрати за рахунок використання однієї платформи.

Серед недоліків необхідно виокремити наступні:

- єдине джерело даних потенційно звужує коло питань, на які може відповісти система;

- в силу відсутності етапу очищення даних оперативні системи характеризуються дуже низькою якістю даних з точки зору їх ролі в підтримці прийняття стратегічних рішень;

- велике навантаження на оперативну систему, тобто складні запити можуть привести до зупинки роботи оперативної системи.

Дворівневе сховище даних будується централізовано для надання інформації. Для підтримки такої архітектури необхідна виділена команда професіоналів в області сховищ даних.

Серед переваг дворівневого сховища даних виділяються такі:

- дані зберігаються в єдиному екземплярі;
- мінімальні витрати на зберігання даних;
- відсутні проблеми, пов'язані з синхронізацією кількох копій даних;
- дані об'єднуються на рівні об'єкта, що дозволяє мати єдину картину системи.

Однак недоліками такої архітектури є наступні:

- дані не структуруються для підтримки потреб окремих користувачів або груп користувачів;
- можливі проблеми з продуктивністю системи;
- можливі труднощі з розмежуванням прав користувачів на доступ до даних.

Трирівневе сховище даних являє собою єдине централізоване джерело інформації. Кінцеві користувачі мають можливість доступу до детальним даним сховища для отримання більш повної картини стану системи.

Його перевагами є такі:

- створення і наповнення сховища даних відбувається з єдиного стандартного джерела очищених даних;
- сховища даних синхронізовані і сумісні;
- гарантована продуктивність.

Недоліки: існує надмірність даних, що веде до зростання вимог на їх зберігання.

Особливості дослідження соціально-економічних об'єктів вимагають застосування методів прийняття рішення, які враховують невизначеність умов оцінки та функціонування об'єктів. Особливостями змісту методів прийняття рішень при управлінні соціально-економічними об'єктами є такі:

1) елементи методу: ситуації, цілі, обмеження, рішення, переваги мають, перш за все, змістовний характер і тільки частково визначаються кількісними характеристиками;

2) знаходження найкращого рішення не може бути повністю формалізоване;

3) елементи методу описуються характеристиками, частина з яких може бути виміряна об'єктивно, а для іншої частини, можливим є тільки суб'єктивний вимір;

4) в ряді випадків доводиться застосовувати метод прийняття рішення в умовах невизначеності, обумовленої неповним описом проблемної ситуації і неможливістю достатньо точної оцінки очікуваних наслідків.

Для аналізу пропозицій в системі підтримки прийняття рішень використовуються різні методи. Це можуть бути:

- інформаційний пошук;
- інтелектуальний аналіз даних;
- пошук знань в базах даних;
- міркування на основі прецедентів;
- імітаційне моделювання;
- еволюційні обчислення і генетичні алгоритми;
- нейронні мережі;
- ситуаційний аналіз;
- когнітивне моделювання та ін.

Деякі з цих методів були розроблені в рамках штучного інтелекту.

Процес розпізнавання включає в себе ряд етапів, одним з яких є процес класифікації. У міру розширення областей застосування для систем розпізнавання образів розширюються області використання і алгоритмів

класифікації. Вони вбудовуються в системи діагностики, побудови моделей, адаптивного і оптимального управління і ін.

Система розпізнавання образів складається з декількох підсистем: датчики вихідної інформації про об'єкт, виокремлювачі інформативних ознак, класифікатор. Датчики сприймають вихідну інформацію про об'єкт і перетворюють її до вигляду, зручного для подальшої обробки на цифровій обчислювальній техніці.

Виділення інформативних ознак і класифікація – близькі по духу завдання, і провести грань між ними буває важко. Чим краще справляється зі своїми завданнями виокремлювачі інформативних ознак, тим більше спрощується робота класифікатора, і навпаки, дієвий класифікатор не потребує виокремлювачів інформативних ознак. На практиці існує золота середина, коли помірна складність обох пристроїв дає прекрасні результати класифікації.

Інформативність ознак – поняття відносне. Одна і та ж система ознак може бути інформативною для вирішення однієї задачі розпізнавання і не інформативною – для іншої. Оцінка інформативності ознак залежить від того, що від чого потрібно відрізнити, тобто від списку образів, що розпізнаються. Залежить вона і від типу вирішальних функцій (B). Так що, вказати «типові», «часто використовувані» ознаки неможливо. Для кожного завдання (для кожної навчальної вибірки A) потрібно знаходити свою інформативну множину описових ознак X_n .

До першого складу ознак (система x_g) відносяться неформалізовані шляхом, що визначаються на основі досвіду та інтуїції фахівця. Формальні методи застосовуються для перевірки цієї вихідної системи на достатність і необхідність. Серед усіх W можливих систем ознак достатньою будемо вважати систему, яка при заданих S , D і A забезпечує витрати N , що не перевищують певний поріг N_0 . Під витратами N тут розуміється вартість вимірювання ознак (N_x) і вартість втрат, що викликаються помилками розпізнавання (N_r): $N = N_x + N_r$. Необхідною є достатня система мінімальної

складності (вартості). Таким чином, фактично вирішується завдання наступного типу $W = \arg \min N(X) / S, D, A, N_0$ серед усіх W систем.

Вирішальним критерієм інформативності ознак в задачі розпізнавання образів є, звичайно, величина втрат від помилок R . Навіть якщо розподіли генеральних сукупностей відомі, обчислення втрат R пов'язано з дуже великими витратами машинного часу. Тому застосовуються критерії, що обчислюються простіше, і, разом з тим, жорстко, якщо вони не є однозначно, корельовані з оцінкою втрат R .

В основі більшості алгоритмів не тільки розпізнавання, але і багатьох інших завдань аналізу даних лежать так звані гіпотези, що приймаються для спрощення постановки задачі класифікації.

Роль гіпотези особливо велика у вирішенні проблеми забезпечення адекватності математичного методу, що впливає на всі етапи рішення соціологічної задачі, починаючи з введення і операціоналізації основних понять і закінчуючи інтерпретацією результатів, отриманих за допомогою математичного формалізму.

На етапі підготовки вихідних даних виникнення питань, пов'язаних з вибором формальної моделі, обумовлено необхідністю:

- забезпечення певної однорідності досліджуваної сукупності об'єктів (що потрібно для коректності застосування методу);
- аналізу взаємозв'язків між змінними і тими латентними факторами, які ці змінні віддзеркалюють;
- аналізу впливу на досліджувані закономірності статистичних залежностей між змінними, що підлягають спостереженню.

Гіпотеза компактності полягає в тому, що реалізації одного і того ж добре організованого способу зазвичай відображаються в просторі ознак у вигляді геометрично близьких точки, утворюючи «компактні» згустки.

Звичайно, вона підтверджується не завжди. Якщо, наприклад, серед ознак є багато випадкових, неінформативних, то такий випадок відповідає поганій організації образів, і точки одного і того ж образу можуть виявитися

далекими одна від одної. Але додатково передбачається, що робота по організації образів вже проведена, і в багатовимірному просторі ознак знайдено такий «інформативний» підпростір, в якому точки одного класу дійсно утворюють компактні згустки, що явно виділяються. Гіпотеза компактності оперує значеннями евклідових відстаней між векторами в просторі описових характеристик. Однак на деяких прикладах можна показати, що важливу роль в задачах аналізу даних відіграють не стільки самі відстані, скільки відносини між ними.

Формулювання гіпотези компактно спирається на поняття «відстань». Якщо визначити відстані між усіма парами точок множини A , то можна побудувати повний граф, який з'єднує всі точки між собою. Якщо на цьому графі виділити дві вершини O і P , позначити довжину, що зв'язує їх ребра через L , а довжину мінімального з суміжних йому ребер через K , то відношення довжин цих ребер $A = L / K$ буде характеризувати міру локальної неоднорідності відстаней або відстань між точками пар. Користуючись цим методом, можна обчислити відстані між усіма парами з n точок, і побудувати граф без петель, який з'єднує між собою всі точки ребрами, сумарна довжина J яких мінімальна. Такий граф називається «найкоротшим незамкнутим шляхом». Тепер відстанню між двома будь-якими точками будемо вважати суму J – довжину тих ребер, за якими проходить шлях між ними по найкоротшому незамкнутому шляху. Локальна «неоднаковість» щільності точок, сусідство «неоднорідних» по довжині ділянок найкоротшого незамкнутого шляху легко визначається.

Виходячи з цього, гіпотезу l -компактності ($АН$) можна сформулювати наступним чином: реалізації одного і того ж добре організованого способу відображаються в просторі ознак в A -близькі точки, утворюючи « X -компактні» згустки.

За аналогією з гіпотезою компактності можна розрізнити гіпотези унімодальної ($ГУ$), полімодальної ($ГП$) і локальної ($Л$) l -компактності, а також всі їх проектні версії.

Гіпотеза компактності також дає основу для оцінки інформативності простору ознак через прояв характеристик компактності. З неї випливає, що для хорошого розпізнавання образів бажано, щоб відстані між точками кожного образу були малими, а відстані між точками різних образів були по можливості великими. Компактність (щільність) B образу, представленого в навчальній вибірці HB ; точками $1, 2, \dots, n$, можна характеризувати середньою довжиною ребер, що з'єднує їх повний граф.

Рознесеність образів в просторі характеристик можна оцінювати через середнє відстань між усіма парами точок з різних образів B_y .

На підставі вищесказаного інформативність простору ознак буде тим більшою, чим більше ця величина

Таким чином, в результаті ототожнення об'єктів з точками на осі значень деякої міри об'єктів отримуємо матрицю розмірністю $n * k$, де n – кількість об'єктів, що класифікуються; k – кількість інформативних ознак, або в нашому випадку це будуть фактори, найбільш повно відображають картину соціального профілю.

В даний час володіння інформацією про соціальні профілі як особистості, так і соціальної групи в цілому становить значний інтерес в більшості сфер діяльності людини і суспільства: економіці, освіті, політиці, безпеці, забезпеченні комфортних умов життєдіяльності тощо.

Тому побудова соціального портрета індивідуума, групи, а також всього суспільства в цілому є важливою науковою проблемою, на вирішення якої спрямовані численні дослідження.

Вихідною точкою пошуку інформації про особу, групу і спільноту особистостей є одержання відповіді на питання «Для чого?». Тоді стануть зрозумілими область пошуку, тип уявлення і місце зберігання цієї інформації. Це означає, що нами наповнюється інформаційна мозаїка соціального портрета, де кожен елемент являє собою самостійний сегмент, і в той же час тісно пов'язаний з безліччю інших елементів, пов'язаних між собою. В якості інструментарію, що дозволяє виконувати такі дослідження,

можна використовувати OSINT – концепцію, методологію і технологію легального отримання і використання інформації з відкритих джерел.

Одним із найважливіших аспектів дослідження є точна оцінка важливості соціальних об'єктів, яка ускладнюється проблемами верифікації, суб'єктивністю експертних оцінок і чутливістю даних до впливу людського фактора. Формалізованими критеріями можна вважати тільки рейтинги або оцінки, представлені в числовому або процентному вигляді.

Таким чином, виникає актуальне наукове завдання вибору апарату формалізації об'єктів дослідження і визначення критеріїв значущості складових інформації соціального профілю.

При виборі засобів і методів вирішення поставленої перед пропонованими дослідженнями мети автори зупинили свій погляд на апараті теорії нечіткої логіки.

Емпірично були підібрані параметри, що впливають на значимість окремих характеристик соціального профілю, і визначені наступні їх характеристичні функції.

Актуальність (ні, та, частково) – «свіжість» інформації, що залежить від часу появи інформації в досліджуваній вибірці. Період оновлення даних соціального профілю становить одну добу. Це обґрунтовується активністю середньостатистичного сучасного користувача мережі Інтернет: зміни, що відбуваються за менший інтервал часу будуть незначними, тоді як при врахуванні більшого інтервалу буде губитися актуальна інформація. У перспективі, з прискоренням темпів цифровізації суспільства, період оновлення даних буде знижуватися. З огляду на прив'язку діяльності людей до календарних циклів, можна обґрунтувати втрату актуальності здебільшого динамічної соціальної інформації через місяць після своєї публікації з огляду на появу в цей період оновлених відомостей. Далі зниження актуальності висловлювання слід характеризувати степеневим законом. Характеристична функція показника актуальності в даному випадку приймає вид:

$$\mu_{vol} = \begin{cases} 0, & vol = 1 \wedge time = " \text{майбутнє} "; \\ 1, & x_{vol} \leq 1 \wedge time = " \text{минуле} "; \\ \frac{1}{1+a^2}, & x_{vol} > 1, \end{cases} \quad (3.1)$$

де vol – показує факт наявності більш свіжої інформації;

$time$ – час висловлювання;

x_{vol} – час, що минув з моменту публікації;

$a = \frac{x_{vol}-1}{30}$ – коефіцієнт старіння інформації.

Авторитетність (відсутня, низька, висока) джерела – ключовий параметр, що впливає на оцінку висловлювання експертами при ускладненнях визначення його достовірності наявними у експертів доказовими засобами. Дана характеристика визначається емпіричними доказами:

$$\mu_{auth} = \begin{cases} 0, & x_{auth} = 0; \\ 2x_{auth}^2, & 0 < x_{auth} \leq 0.5; \\ 1 - 2(x_{auth} - 1)^2, & 0.5 < x_{auth} \leq 1, \end{cases} \quad (3.2)$$

де x_{auth} – показує частку експертів, які вважають джерело надійним.

Важливість (не має відношення, другорядна, важлива, критична) інформації – показує ступінь пов'язаності з іншими характеристиками соціального профілю. Параметр залежить від їх вагових коефіцієнтів, тому множина елементів СП з високим показником важливості може не збігатися з множиною ключових елементів, отриманих в ході оцінки метрики центральності. Ступінь приналежності нечіткої величини до важливості інформації описується формулою:

$$\mu_{val} = \begin{cases} 0, x_{val} = 0; \\ \frac{2x_{auth}^2}{nlink^2}, 0 < x_{val} \leq \frac{nlink}{2}; \\ 1 - \frac{2(x_{val}-nlink)^2}{nlink^2}, \frac{nlink}{2} < x_{val} < nlink; \\ 1, x_{val} = nlink, \end{cases} \quad (3.3)$$

де $nlink$ – загальна кількість унікальних зв'язків з досліджуваним об'єктом;
 x_{val} – к кількість зв'язків з соціальними об'єктами, які мають вагу більше деякого порогового значення.

Аргументованість (недостатньо, досить) – враховує наявність доказів і посилянь з метою виявлення інформаційного шуму і бездоказових заяв, та залежить від авторитетності інформаційних джерел:

$$\mu_{arg} = \begin{cases} 0, x_{arg} = 0; \\ \frac{1}{e^{2na}}, x_{arg} > 0, \end{cases} \quad (3.4)$$

де x_{arg} – кількість можливих аргументів;

na – величина, що показує недолік авторитетності джерел в середньому і максимальному значенні, позначимо її як

$$na = 2 - \max_{x_{arg}}(\mu_{auth}) - \frac{\sum^{x_{arg}} \mu_{auth}}{x_{arg}}. \quad (3.5)$$

Унікальність (раніше невідомо, відомо, широко відомо) інформації – дозволяє відокремити факти від бездоказових висловів, та визначає ступінь, зворотний поширеності висловлювання в різних джерелах інформації на поточний момент часу:

$$\mu_{unq} = \begin{cases} 1, x_{unq} = 0; \\ 1 - \frac{2b^2}{ns^2}, 0 < x_{unq} \leq \frac{ns}{2}; \\ \left(\frac{ns-1-b}{ns}\right)^2, \frac{ns}{2} < x_{unq} < ns; \end{cases} \quad (3.6)$$

де x_{unq} – кількість джерел, що містять досліджуване висловлювання, з датою публікації раніше, ніж в оригіналі;

ns – загальна кількість джерел в соціальному профілі, що містять досліджуване висловлювання;

$b = \sum^{x_{unq}} \mu_{auth}$ – сумарна авторитетність джерел, що містять досліджуване висловлювання з датою публікації раніше, ніж в оригіналі.

Достовірність (брехня, правда) – висловлювання при наявності протилежних думок залежить від найбільш авторитетної з них, яка підкріплена безліччю фактів або є новаторською. При відсутності суперечливих точок зору характеристика достовірності висловлювання вважається рівною 1. Показник залежить від ступеня аргументованості та унікальності:

$$\mu_{ver} = \begin{cases} 1 - \max_{i=1}^{x_{ver}} (ver_i), x_{ver} \geq 1; \\ 1, x_{ver} < 1, \end{cases} \quad (3.7)$$

де x_{ver} – кількість альтернативних точок зору;

ver – показник верифікації, позначимо його як

$$ver = \max[\mu_{arg}(\mu_{auth}), \mu_{unq}(\mu_{auth}, ns_{min})].$$

Таким чином, підсумковий ваговий коефіцієнт описується наступною формулою:

$$v = \mu_{vol} \cdot \mu_{val}(u_{min}) \cdot \mu_{ver} \cdot ver. \quad (3.8)$$

Розроблений коефіцієнт значущості характеристик соціального профілю дозволяє враховувати поділ неструктурованого різноманіття вихідної інформації за рівнем її потенційної користі, ґрунтуючись на величинах авторитетності джерел, актуальності, аргументованості, достовірності, унікальності та пов'язаності об'єктів цифрового соціального середовища. Використання критерію значущості в рамках розробленої математичної моделі цифрового соціального середовища покликане вплинути на забезпечення несуперечності елементів соціальних профілів, що створюються. Також він може застосовуватися при проведенні аналізу настроїв та сприяти визначенню семантичного змісту висловлювань.

В ході виконання завдання соціального профілювання проводиться обробка інформації різного ступеня структурованості. При цьому можливо використовувати наступні моделі даних:

- мережева – для зберігання даних кінцевого соціального профілю та подання їх у вигляді графів;
- реляційна – для зберігання інформаційної карти соціального профілю;
- постреляційна і NoSQL – для зберігання неструктурованих вихідних даних і динамічного контенту, в тому числі мультимедійного.

В першу чергу проводиться збір даних до загального нереляційного сховища, де дані зберігаються «як є», але із зазначенням сервісної інформації (номер запису, дата пошуку, джерело, коментарі) в окремих стовпцях для полегшення подальшої обробки. При наявності вже відомої ключової інформації або її знаходженні в зібраних даних потрібно додати її в таблицю інформаційної карти, стовпці якої вказують на списки знайдених атрибутів, ідентифікатор соціального профілю і посилання на співвіднесений з ним динамічний контент. Рядки таблиці інформаційної карти співвідносяться з різними соціальними профілями. Дані динамічного контенту з нереляційних сховищ в залежності від свого типу і ступеня структурованості можуть

переноситься до інших сховищ, що пропонують специфічний функціонал, такий як можливість ведення обліку версій контенту, внесення змін і зауважень в режимі реального часу, аналіз нетекстових даних тощо. Інформація кінцевого соціального профілю відтворюється у вигляді вершин і ребер графа. Кожна характеристика повинна співвідноситися з одним із шаблонів (суб'єкт, інтерес), які мають власні набори атрибутів. Також для кожного об'єкта СП вводяться поняття ваги, що позначає значимість, і настроїв, на який вказує емоційне забарвлення висловлювання. Як приклади основних категорій об'єктів у задачі побудови соціального профілю можуть виступати такі:

- вже згадана персона – інформація (ПІБ, вік ...) про людину, яка є центральним об'єктом соціального профілю, що будується, та має найбільше число зв'язків з іншими соціальними об'єктами;
- персона, що згадується – коротка інформація про людину, яка не розглядається при побудові даного соціального профілю, але має прямий або опосередкований зв'язок з даною персоною;
- адреса – відомості про географічне місцезнаходження, співвідносне з деякими соціальними явищами й об'єктами;
- діяльність – дані про соціальну активність людини чи спільноти, такі як програмування, професійний спорт або благодійність;
- захід – відомості про соціальне явище, що має точне місце і час проведення, а також дозволяє об'єднати безліч учасників до спільноти і визначити їх загальні характеристики;
- праці – описує різні матеріальні і нематеріальні досягнення людини або групи, отримані в результаті діяльності;
- організація – інформація про різні установи і спільноти, що асоціюються з людьми, перерахованими в соціальному профілі;
- спеціалізація – спеціальність людини або конкретний напрям діяльності;

- телефон – номер телефону, код і супутня інформація;
- хобі – інформація про захоплення та інтереси людей (важливо не плутати поняття «діяльність» і «хобі»);
- аккаунт – ідентифікатори облікових записів web-ресурсів та адреси електронної пошти, що належать конкретній особі або спільноті.

Потрібно враховувати можливі типи зв'язків між елементами соціального середовища. Вони дозволяють розмежувати вплив соціальних об'єктів між собою. Можливі відносини між об'єктами і типами зв'язків соціального профілю представлені в табл. 3.1.

Знак «±» в матриці означає можливість об'єднання однорідних об'єктів обраним зв'язком. Важлива умова при плануванні зв'язків полягає в тому, що в графі СП повинна бути тільки одна компонента зв'язності, а вершина типу «Вже згадана персона» повинна бути досяжна з інших вершин. Зв'язки є повноцінними елементами моделі представлення соціального профілю і можуть мати власні набори атрибутів. Обмеження на розташування зв'язків залежать тільки від зручності подальшого аналізу і представлення даних.

Концептуальну модель представлення даних соціальних профілів пропонується розширити і привести у відповідність комплексній і математичній моделей цифрового соціального середовища.

Далі потрібно розкрити елементи і взаємозв'язки чотирьох базових категорій інфологічної моделі представлення результатів соціального профілювання.

Таблиця 3.1

Матриця інцидентності графа уявлення соціального профілю

| | Адресація | Зв'язок з організацією | Спеціалізація організації | Участь у заходах | Зв'язок з персоною | Хобі персони | Діяльність персони | Колектив організації | Контакти персони | Спеціалізація персони | Асоційовані праці |
|---------------------------|-----------|------------------------|---------------------------|------------------|--------------------|--------------|--------------------|----------------------|------------------|-----------------------|-------------------|
| Персона, що розглядається | 1 | 0 | 0 | -1 | -1 | 1 | 1 | 1 | ±1 | 1 | 1 |
| Персона, що згадується | 1 | 0 | 0 | -1 | 0 | 0 | 1 | 1 | ±1 | 1 | 1 |
| Адреса | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Діяльність | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 0 |
| Заходи | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Праці | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 |
| Організація | 1 | -1 | 1 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 1 |
| Спеціалізація | 0 | 0 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 |
| Телефон | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Хобі | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 0 | 0 |
| Аккаунт | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |

Джерело: авторська розробка

Категорія об'єктів соціального профілю людини, яка враховує поділ даних на статичну і динамічну частини, зображується наступним чином:

$$PSP = \left\{ \begin{array}{l} ppid, PNAME, PDATE, PLOC, PLINK, PGLINK \\ PTLINK, PDLINK, PSOURCE, TESLINK \end{array} \right\}, \quad (3.9)$$

де $ppid$ – унікальний числовий ідентифікатор соціального профілю особистості;

$PNAME = \{FNAME, LNAME, PATR, ALIAS\}$ – множина імен даної персони (П.І.Б., псевдоніми);

$PDATE$ – множина ключових дат життя персони;

$PLOC$ – множина пов'язаних з персоною геолокацій, що включає координати розташування $pcord$, їх назви $PNAMELOC$, важливі дати $PLOCDATE$ та коментарі $PLOCNOTE$, причому

$$PLOC = \left\{ \begin{array}{l} pcord, PNAMELOC \\ PLOCDATE, PLOCNOTE \end{array} \middle| PLOCDATE \subset PDATE \right\}; \quad (3.10)$$

$PLINK$ – множина посилань на інші СП, з якими встановлені зв'язки, причому

$$PLINK = \left\{ ppid \middle| \begin{array}{l} \exists x, y \in PSP: x \cap y \neq 0, \\ ppid(x) \neq ppid(y) \end{array} \right\}; \quad (3.11)$$

$PGLINK$ – посилання на безліч пов'язаних спільнот, що враховують соціальні ролі персони $role$ та сутність відношень з групою $matter$, причому

$$PGLINK = \left\{ \begin{array}{l} ppid, pgid, \\ role, matter \end{array} \middle| \begin{array}{l} \exists x \in PSP, : \exists y \in GSP: \\ ppid \in x, pgid \in y, x \cap y \neq 0 \end{array} \right\}; \quad (3.12)$$

$PTLINK$ – множина посилань на найбільш значущі для персони соціальні явища і характеристики, причому

$$PTLINK = \left\{ ppid, idtopic \middle| \begin{array}{l} \exists x \in PSP, : \exists y \in FACT: \\ ppid \in x, idtopic \in y, x \cap y \neq 0 \end{array} \right\}; \quad (3.13)$$

$PDLINK$ – множина посилань на об'єкти динамічної частини з найбільш стійкими зв'язками, причому

$$PDLINK = \left\{ ppid, dynid \middle| \begin{array}{l} \exists x \in PSP, : \exists y \in PDYN: \\ ppid \in x, dynid \in y, x \cap y \neq 0 \end{array} \right\}; \quad (3.14)$$

$PSOURCE$ – відомості про джерело вихідної інформації, із зазначенням web-посилання $sourcelink$, назви $sourcenamе$ та дати звернення до джерела $sourceget$, причому

$$PSOURCE = \left\{ \begin{array}{l} ppid, sourcelink, x \\ sourcename, sourceget \end{array} \middle| x \in PSP \cap SOURCE \right\}; \quad (3.15)$$

де $SOURCE$ – множина джерел інформації в мережі Інтернет;

$TESLINK$ – посилання на словники.

Динамічний контент розглядається як самостійна множина:

$$PDYN = \left\{ \begin{array}{l} dynid, DATATYPE, DSERVICE, DNOTE, DTAG, \\ PDLINK, DGLINK, DTLINK, DSOURCE, TESLINK \end{array} \right\}; \quad (3.16)$$

де $dynid$ – унікальний числовий ідентифікатор динамічного соціального об'єкта; $DATATYPE = \{ \text{текстовий, аудіо, відео, графіка, геодані, бінарний} \}$ – тип даних;

$DSERVICE = \{ servicetype \ servicevalue \}$ – сервісна інформація у вигляді параметр-значення про динамічний об'єкт;

$DNOTE$ – коментарі до динамічного об'єкту;

$DTAG$ – множина тематичних тегів:

$$DTAG = \left\{ \begin{array}{l} \text{Істота, Організація, Діяльність, Адреса, Телефон} \\ \text{Електронна пошта, Досягнення, Спеціалізація, ...} \end{array} \right\} \quad (3.17)$$

$PDLINK$ – множина посилань на інформаційні карти ПСП);

$DGLINK$ – множина посилань на інформаційні карти ГСП, причому

$$DGLINK = \left\{ \begin{array}{l} dynid, pgid \end{array} \middle| \begin{array}{l} \exists x \in PDYN, : \exists y \in GSP: \\ dynid \in x, pgid \in y, x \cap y \neq \emptyset \end{array} \right\}; \quad (3.18)$$

DTLINK – множина посилань на соціальні явища, причому

$$DTLINK = \left\{ dynid, idtopic \left| \begin{array}{l} \exists x \in PDYN, : \exists y \in FACT: \\ dynid \in x, idtopic \in y, x \cap y \neq 0 \end{array} \right. \right\}; \quad (3.19)$$

DSOURCE – відомості про джерело вихідної інформації, причому

$$DSOURCE = \left\{ \begin{array}{l} dynid, sourcelink, x, \\ sourcename, sourceget \end{array} \left| x \in PDYN \cap SOURCE \right. \right\}; \quad (3.20)$$

TESLINK – посилання на словники.

Соціальні явища (далі – СЯ) (теми, факти) – складають структуровану множину даних, відокремлену від інформаційних карт СП персон і груп, що має власний набір унікальних атрибутів:

$$FACT = \left\{ \begin{array}{l} idtopic, TNAME, TTYPE, TFEATURE, TNONE, TLINK, \\ PTLINK, GTLINK, DTLINK, TSOURCE, TESLINK \end{array} \right\}, \quad (3.21)$$

де *idtopic* – унікальний ідентифікатор СЯ;

TNAME – множина найменувань або ключових слів соціального явища, теми або факту;

TTYPE = {подія, соціальний об'єкт, соціальний зв'язок} – види СЯ;

TFEATURE – множина атрибутів, характеристик соціального явища;

TNOTE – коментарі до СЯ;

TLINK – множина посилань на інші СЯ, причому

$$TLINK = \left\{ idtopic \left| \begin{array}{l} \exists x \in FACT: x \cap y \neq 0, \\ idtopic\langle x \rangle \neq idtopic\langle y \rangle \end{array} \right. \right\}; \quad (3.22)$$

PTLINK – множина посилань на пов'язані ПСП;

GTLINK – множина посилань на пов'язані ГСП, причому

$$GTLINK = \left\{ pgid, idtopic \left| \begin{array}{l} \exists x GSP, \exists y \in FACT: \\ pgid \in x, idtopic \in y, x \cap y \neq \emptyset \end{array} \right. \right\}; \quad (3.23)$$

DTLINK – множина посилань на динамічний контент;

TSOURCE – відомості про джерело вихідної інформації про соціальне явище:

$$TSOURCE = \left\{ idtopic, sourcelink, x \left| \begin{array}{l} sourcename, sourceget \\ x \in FACT \cap SOURCE \end{array} \right. \right\}; \quad (3.24)$$

TESLINK – посилання на словники.

Характеристики СП груп не статичні в часі, тому вони не включають цілком персональні соціальні профілі своїх членів. Категорія даних, що відносяться до технічного ГСП, являє собою:

$$GSP = \left\{ pgid, PGNAME, GTAG, PGDATE, PGLOC, GLINK, \right. \\ \left. PGLINK, DGLINK, GTLINK, GSOURCE, TESLINK \right\}; \quad (3.25)$$

де *pgid* – унікальний ідентифікатор групи;

PGNAME = {*RNLAM*, *ABBR*, *TRNM*} – множина назв спільноти (офіційне, аббревіатура, переклад) за наявності;

GTAG – множина тегів для опису соціального статусу групи;

PGDATE – множина ключових дат групи;

PGLOC – множина пов'язаних з спільнотою розташувань, що включає координати *pgcoord*, назви локацій *PGNAMELOC*, дати *PGLOCDATE* та коментарі *PGNOTE*:

$$PGLOC = \left\{ pgcoord, PGNAMELOC, \left. \begin{array}{l} PGLOCDATE \subset PDATE \cup PGDATE \\ PGLOCDATE, PGNOTE \end{array} \right| \right\}, \quad (3.26)$$

GLINK – множина посилань на інші профілі груп, з якими встановлені зв'язки, причому

$$GLINK = \left\{ pgid \left| \begin{array}{l} \exists x \in FACT: x \cap y \neq 0, \\ pgid\langle x \rangle \neq pgid\langle y \rangle \end{array} \right. \right\}; \quad (3.27)$$

PGLINK – множина посилань на учасників групи);

DGLINK – множина посилань на атрибути групи з динамічного контенту;

GTLINK – посилання на пов'язані СЯ;

GSOURCE – відомості про джерело вихідних даних групи, причому

$$GSOURCE = \left\{ \begin{array}{l} pgid, sourcelink, x, \\ sourcename, sourceget \end{array} \left| x \in GSP \cap SOURCE \right. \right\}; \quad (3.28)$$

TESLINK – посилання на словники.

У моделі соціального профілю для інформації, що одержується з різних словників і довідників, виділено окрему множину:

$$TESAURUS = \{definition, SEMANTICPROPERTIES, VALUE\}, \quad (3.29)$$

де *definition* – поняття, визначення або термін зі словника;

SEMANTICPROPERTIES – множина семантичних властивостей терміну;

VALUE – значення поняття, визначення або терміну.

Розглянуті властивості розроблених моделей забезпечують можливість подальшої алгоритмізації аналітичних функцій для підтримки осіб, які приймають рішення, в межах нової методики соціального профілювання. Надана можливість врахування природи і основних характеристик зібраних відомостей спрощує виявлення явних і неявних зв'язків між елементами соціального середовища.

Таким чином, в результаті проведених досліджень розроблені нечіткі критерії для визначення значущості складових інформації соціального профілю, що дозволить істотно підвищити адекватність моделі соціального профілю.

Зокрема, визначено вимоги до типу і змісту даних для побудови соціального профілю людини. Вихідна інформація є загальнодоступною, тобто для неї повинна бути позначена можливість обробки третіми особами. Вона збирається з соціальних медіа мереж Інтернет і може відноситися до наступних типів даних: текст, мультимедіа тощо. Зміст даних має відповідати об'єктам соціального середовища: персонам, співтовариствам і соціальним явищам. Інформація соціального профілю може бути розподілена в просторі і часі.

До якості інформації ставляться такі вимоги: дані про соціальне середовище повинні бути актуальними, повними і достовірними. Відповідність цим вимогам уможливило створення соціальних профілів, які використовуються в якості інформаційної основи для задач формалізації людського фактора в соціо-економічній діяльності.

Запропоновано нову модель нечіткого коефіцієнта значущості параметрів соціального профілю, відмінністю якого є можливість врахування як формалізованих об'єктивних, так і суб'єктивних показників, що важко інтерпретуються, для оцінки вихідної інформації СП. Перевагою використання критерію значущості перед залученням експертів є менші часові витрати на аналіз великих обсягів даних при збереженні рівня об'єктивності оцінювання.

Запропоновано інфологічну модель для представлення статичної та динамічної частин соціального профілю, що дозволяє спроектувати схему інформаційного забезпечення системи соціального профілювання в рамках розробленої методики збору і аналізу гетерогенних соціальних даних з відкритих джерел мережі Інтернет. Відмінність моделі полягає у використанні принципів теорії метаграфії і реалізації заплутаної структури соціального середовища у вигляді чотирьох великих категорій соціальної інформації, що відноситься до персон, співтовариств і соціальних явищ, що дозволяє подолати проблему гетерогенності і нерівноцінності вихідних даних.

3.2. Розробка математичної моделі оцінки ступеня впливу учасників соціальної групи

Виходячи з аналізу попереднього матеріалу, можна визначити сукупність основних вимог, яким повинна задовольняти система підтримки прийняття рішень, а саме вона повинна мати можливість:

- змінювати базову основу прийнятих рішень про соціальний профіль груп в залежності від складу інформації, необхідної якості рішень і рівня стратифікаційного аналізу всіх соціальних груп;
- враховувати ієрархічний характер послідовної оцінки окремих компонентів соціального профілю на різних рівнях його аналізу;
- враховувати в загальному випадку багатоцільовий характер оцінок, виходячи з умов соціально-економічного розвитку;
- оперувати з різномірними даними, що визначають стан економіки і соціальної сфери.

Проектування систем підтримки прийняття рішень, їх реалізація та функціонування повинні базуватися на послідовному застосуванні принципів системного підходу (рис. 3.2).

Зазначені принципи проектування систем підтримки прийняття рішень послідовно реалізуються на різних етапах формування баз знань діагностичного комплексу соціальних профілів. Послідовність і зміст етапів формування баз знань такої системи наведено в табл. 3.2.

В цілому база даних системи підтримки прийняття рішень відповідно до загальних вимог до процедури моніторингу розглядається як відкрита система, що взаємодіє з користувачем і володіє високою зв'язністю, тобто не є простою сумою окремих індикаторів соціального профілю.



Рис. 3.2. Ключові принципи проектування системи підтримки прийняття рішень щодо діагностики соціальних профілів

Джерело: складено на підставі [92]

Зазначені принципи проектування систем підтримки прийняття рішень послідовно реалізуються на різних етапах формування баз знань

діагностичного комплексу соціальних профілів. Послідовність і зміст етапів формування баз знань такої системи наведено в табл. 3.2.

Таблиця 3.2

Зміст етапів формування баз знань підсистеми оцінки соціальних профілів

| Етапи | Зміст |
|--|---|
| Ідентифікація проблеми | Проводиться аналіз завдань оцінки соціальних профілів, виявляється структура оцінок, визначаються найбільш важливі форми їх інтегрального виразу та взаємозв'язки з урахуванням соціально-економічного розвитку, оцінюються необхідні ресурси, конкретизуються цілі і завдання. |
| Концептуалізація опису діагностичного завдання | Визначається структура діагностичних завдань за етапами їх стратифікаційного уявлення в ієрархії причинно-наслідкових зв'язків соціально-економічних характеристик, визначається метрика уявлень інтегральних характеристик соціального профілю |
| Нормалізація | Здійснюється представлення основних понять і відносин соціального профілю в метриці його інтегральної оцінки |

Джерело: складено на підставі [76]

В цілому база даних системи підтримки прийняття рішень відповідно до загальних вимог до процедури моніторингу розглядається як відкрита система, що взаємодіє з користувачем і володіє високою зв'язністю, тобто не є простою сумою окремих індикаторів соціального профілю.

Важливою властивістю, що відображає значимість бази даних для роботи системи підтримки прийняття рішень, є її багатоцільовий характер:

прийняті рішення про оцінку соціального профілю можуть відповідно до його соціально-економічної структури враховувати кілька інтегральних показників оцінки його рівня.

Це, в свою чергу, обумовлює різноманіття і ієрархічність структури бази даних й її високу адаптивність (структурну стійкість).

Таким чином, база даних містить велику кількість складних різнорідних понять, які характеризуються певним набором атрибутів і їх значень, багатозначністю і багатоаспектністю, високою зв'язністю понять, об'єднаних в ієрархічну (мережеву – на нижніх рівнях деталізації) багаторівневу систему, в якій порядок (пріоритет) рівнів визначається контекстом використання знань.

Типи даних, які використовуються для оцінки соціального профілю, характеризуються великою різноманітністю послідовності цілих і дійсних чисел списку (вектору) елементів, на якому визначено порядок відносин, пропозиції природної мови, розподілені (вірогіднісно або нечітко) числові або буквено-числові характеристики, чіткі або нечіткі логічні конструкції тощо. Різні типи даних взаємообумовлені і мають високу зв'язність на різних рівнях стратифікованої оцінки різних груп населення.

На етапі формалізації понять і відносин між ними важливий вибір адекватного математичного апарату. У цьому контексті найбільш адекватним є апарат теорії нечітких множин. На користь такого вибору свідчить, по-перше, те, що завдання синтезу бази знань і системи прийняття рішень в системі підтримки прийняття рішень діагностичного комплексу соціальних профілів є недостатньо формалізованою, а, по-друге, – існує можливість забезпечення достатнього рівня спільності опису (при цьому детермінований і статистичний описи виходять з нечіткого як окремі випадки).

Формалізований опис виділених понять, які характеризують різні аспекти соціального профілю, є певною системою відносин між індикаторами соціального профілю і простором ознак, які відповідають

певному образу соціального профілю. Даний опис реалізується за схемою: поняття (властивість) – змінна (набір змінних) – стан змінної.

Змінною в подальшому будемо називати формалізоване уявлення станів економіко-фінансової сфери і соціально-політичної стабільності, що відображає один з аспектів опису соціального профілю. Таке уявлення завжди співвідноситься з певною процедурою вимірювання агрегованих індикаторів соціального профілю. Це, в свою чергу, вимагає визначення базової (параметричної) множини, що використовується для розрізнення цих станів в світлі аналізу соціального профілю. Стан кожної змінної визначається завданням таких її значень, які можуть оцінюватися в одній з шкал: найменувань, порядку, інтервалів або відносин з використанням як чітких, так і нечітких точкових чи інтервальних оцінок.

Залежно від складності поняття для його опису може використовуватися скалярне, векторне або тензорне уявлення.

Іншим важливим моментом при формалізації поняття є математичні властивості простору соціального профілю, які визначаються властивостями вимірювальної шкали: простір станів може бути, наприклад, повністю або частково впорядкованим.

Наступним кроком є формалізація відносин між соціальними профілями. У теорії систем використовуються наступні типи моделей, що описують відносини між елементами: ланцюжкова, ієрархічна, реляційна, мережева, матрична. Результати узагальненого формалізованого опису моделей наведені в табл. 3.3.

Таблиця 3.3

Основні типи моделей діагностики соціальних профілів у базах знань системи підтримки прийняття рішень

| Тип моделі | Зміст |
|------------|-------|
|------------|-------|

| | |
|------------|---|
| Ланцюжкова | Застосовується для подання взаємозв'язку елементів, на множині яких існує відношення лінійного упорядкування |
| Ієрархічна | Застосовується для опису відносин на множині підмножин характеристик соціального профілю. При цьому виділяється покриття, що включає кілька супідрядних рівнів. |
| Реляційна | Є окремим випадком ієрархічних моделей, коли на множині елементів відсутнє ранжування за важливістю. Використовується для опису залежностей між поняттями, наприклад, загального поняття з приватними, коли приватні поняття мають однакову важливість |
| Мережева | Використовується для опису класів понять, структура яких має високу зв'язаність. На відміну від ієрархічних, в даних моделях кожен елемент одного рівня може замикатися на кілька різних елементів більш високого рівня |
| Матрична | Застосовується для опису відносин в структурі «елементи – функції», що служить для досягнення спільної мети системи. |

Джерело: складено на підставі [82; 83]

Зокрема, ланцюжкова модель застосовується для подання взаємозв'язків елементів, на множині яких існує відношення лінійного впорядкування. Наприклад, X може бути безліччю соціально-економічних характеристик соціального профілю, В цьому випадку в ланцюжковій моделі

V – це множина соціально-економічних феноменів, що визначають оцінки інтегрального опису складових соціального профілю (наприклад, стратифікаційний і орієнтаційний аналіз, аналіз рівня фінансової забезпеченості тощо.), Z – множина станів зазначених вище феноменів, що визначає метрику їх вимірювання.

Ієрархічна модель застосовується для опису відносин на множині підмножин деякої множини елементів, при цьому можна виділити покриття (або розбиття), у тому числі кілька супідрядних рівнів, причому кожен елемент даного рівня замикається тільки на один елемент більш високого рівня.

Дана модель відображає стратифікаційний характер послідовного аналізу соціального профілю, при якому досліджуваний соціально-економічний феномен поступово уточнюється в міру його інформаційного наповнення на різних стадіях соціально-економічного аналізу.

Реляційна модель є окремим випадком ієрархічної, коли відсутнє ранжування множини елементів X за важливістю.

Мережева модель використовується для опису класів понять, структура яких має високу зв'язність. У ній кожен елемент одного рівня може замикатися на кілька елементів більш високого рівня. Це характерно для представлення структури даних і взаємозв'язку складних соціально-економічних феноменів соціального профілю, таких, як стан кредитно-фінансова сфери, соціальна стабільність тощо.

Матричні моделі використовуються для опису відносин в структурі «елементи – функції». Вони успішно можуть бути застосовані при визначенні складу діагностичних засобів соціально-економічного аналізу соціального профілю і виробленні рекомендацій щодо порядку їх застосування і відповідно до мети.

Крім зазначених, для формування баз знань можуть бути використані моделі, що є різновидом ієрархічних.

Це, в першу чергу, фреймові моделі і міжфреймової мережі, семантичні мережі і семіотичні моделі.

Фреймові моделі представляють собою типову структуру даних для опису певних ситуацій, побудовану на основі відносин «ціле – частина» та «абстрактне – конкретне».

Для подання класифікаційних структур та їх пошуку використовують міжфреймової мережі, в яких об'єднані фрейми, пов'язані певним відношенням подібності.

Більш загальним є уявлення знань у вигляді семантичної мережі, тобто мережевої моделі, що використовує фрейми, вузлами яких є концепції або об'єкти (поняття), пов'язані дугами, що описують відносини між вузлами. При цьому кожен вузол описується набором ознак і їх значень.

Основна перевага фреймових і семантичних моделей уявлення знань – це можливість об'єднання декларативних і процедурних знань. При цьому процедурні знання представляються у вигляді так званих приєднаних процедур, що використовуються в якості значень характеристик об'єкта (слотів).

Розглянуті моделі є різновидом більш загальних семантичних моделей, що включають три рівні знань;

- синтаксичний (символьний), що відображає відносини між знаками уявлення характеристик групи;
- семантичний, що відображає смисловий аспект знаків;
- прагматичний, що включає відносини знаків до понять, що визначають їх застосування.

Подання знань у вигляді правил (продукцій) в розглянутих вище фреймових, семантичних і семіотичних моделях володіє такими перевагами, як гнучкість, природність і простота. При цьому база знань формується за допомогою набору правил, що складаються з пар «атрибут (властивість) – значення» або трійок «атрибут – об'єкт – значення». Така система ефективна

для представлення об'єктів у вигляді дворівневих (бінарних) зв'язків. Це істотно обмежує сферу застосування подібних систем, так як побібний підхід ускладнює опис об'єктів, що мають складну структуру взаємопов'язаних елементів.

Ще один спосіб представлення знань в системах підтримки прийняття рішень реалізується за допомогою логіки предикатів. При цьому основними символами мови логіки предикатів є константи і змінні, що визначають соціально-економічні індикатори і функціональні символи, що визначають функціональні залежності індикаторів показниками соціально-економічної стабільності. Таким чином можна представляти складні ієрархічні структури оцінки соціального профілю.

Таким чином, попередній аналіз дозволяє зробити висновок про те, що з позицій ефективності роботи системи підтримки прийняття рішень з оцінки соціального профілю найбільш зручними є такі методи представлення знань, як фрейми і семантичні мережі. Обумовлено це рядом обставин. По-перше, наявністю великого числа нечітко визначених різнорідних і взаємозалежних елементів соціально-економічних характеристик, що визначають стан соціально-економічної безпеки. Це змушує на кожному етапі аналізу соціального профілю розглядати складові його елементи як єдине ціле. Подання такої мережі у вигляді продукційних правил, а тим більше у вигляді нечітких продукцій, надзвичайно складно, і, крім того, час релаксації такої мережі велике. Застосування ж системи класифікацій (в загальному випадку, як видно, нечіткої) дозволяє уникнути зазначених недоліків. По-друге, ієрархічним характером уявлення інтегральних характеристик соціального профілю, обумовлених змістом і внутрішньою структурою послідовної реалізації стратометричної концепції аналізу соціального профілю. По-третє, – нечіткістю самих класифікаційних схем, що описують основні поняття соціально-економічних показників і їх взаємозв'язки.

Слід при цьому зазначити, що алгоритми та функціонал соціальних мереж постійно удосконалюються і змінюються, тому ті критерії, які

підходили для оцінки рівня впливу учасників груп у соціальних мережах на конкурентоспроможність товарів або послуг раніше, вже є нерелевантними.

Відповідно, актуальним є створення такого набору критеріїв, який дозволяє обрати метод підрахунку, що задовольняє конкретним вимогам підприємства [14; 25].

Критерії оцінки рівня впливу учасників груп у соціальних мережах на конкурентоспроможність товарів і послуг є вкрай різноманітними оскільки сьогодні соціальні мережі дають можливість отримувати великий обсяг статистичних показників. Відповідно, важливо знайти підходи правильної інтерпретації отриманої інформації. Якщо кілька років тому головним способом оцінки рівня впливу учасників груп у соціальних мережах на конкурентоспроможність товарів і послуг був підрахунок кількості учасників груп соціальної мережі, то в сучасних умовах така модель вже не є досконалою. На зміну цій моделі прийшла модель залучення користувачів.

Зокрема, серед раціональних критеріїв оцінки рівня впливу учасників груп у соціальних мережах на конкурентоспроможність товарів і послуг доцільно виділити наступні.

1. Чистий приріст – характеризує активність зростання соціальної мережі в заданий період часу. Цей показник розраховується як різниця між кількістю нових учасників групи у соціальній мережі та кількістю користувачів, які залишили цю групу.

2. Відсоток неактивних акаунтів.

3. Відсоток цільової аудиторії у групі соціальної мережі – концентрація цільової аудиторії у групі не повинна бути нижче 80%.

4. Активне ядро – критерій, який вказує на кількість унікальних користувачів, які вчинили від одного і більше дій в обраній групі за певний період часу. У свою чергу, ядра мають власні індекси – від одного до десяти.

Як показує практика, до поняття лояльної аудиторії включаються ті користувачі, які зробили в обраній групі від трьох і більше цільових дій.

6. Охоплення користувачів – при підрахунку враховуються тільки перегляди у безпосередній групі та в новинній стрічці.

7. Залучення за охопленням – замість формування новинної стрічки за часом використовується більш складний алгоритм видачі новин, що враховує інтереси конкретного користувача.

Показник за охопленням дає уявлення про реальну залученість в обраній групі соціальних мереж.

8. Негативний відгук.

9. Частота відгуку – соціальні мережі враховують швидкість зворотного зв'язку компанії на запити своїх користувачів. Також враховується частка питань користувачів, які були відпрацьовані. Сьогодні велике значення має час швидкості відповіді і рішення проблем користувачів.

10. Кількість обговорень.

У цілому, розробка системи критеріїв оцінки рівня впливу учасників груп у соціальних мережах на конкурентоспроможність товарів або послуг є продуманим менеджментом, в основі якого лежить аналіз великих обсягів інформації з використанням статистичних методів її обробки [23; 28; 32].

3.3. Розробка моделі управління знаннями у групі

Серед моделей управління знаннями у групах традиційно найбільш відомими є наступні.

1. Модель Нонака й Такеучі. Вона має назву чотирьохфазної моделі SEKI (соціалізація, екстерналізація, комбінація, інтерналізація).

Зокрема, соціалізація є методом обміну неявних знань за допомогою спостереження, наслідування, практики й участі у формальних і неформальних співтовариствах і груп. Цей процес переважно стосується створення фізичного або віртуального простору, де дане співтовариство може взаємодіяти на соціальному рівні.

Екстерналізація є методом вираження неявного знання у вигляді явних концепцій. Оскільки неявні знання високо засвоєні, цей процес є ключем до обміну знаннями й створення системи знань.

Комбінація є методикою інтеграції понять до системи знань. Деякі приклади або випадки будуть являти собою синтез у формі звіту про огляд, аналіз тенденцій, коротке резюме або нову базу даних.

Інтерналізація є технікою, що втілює явне знання у неявне.

Її елементами є наступні:

- дві форми знання – неформалізоване (неявне) і формалізоване (явне);
- динаміка взаємодії – передача знань від фази до фази, від циклу до циклу;
- три рівні соціальної агрегації – індивід, група, контекст;
- чотири фази створення знання;
- умови створення знання;
- структура компанії, орієнтована на знання.

2. Модель Хедлунда. Головним атрибутом моделі є аналіз двох наборів концепцій:

- тип неявного знання й тип явного знання (кожний тип має три форми знання: пізнавальне, навичка, втілене), а також чотири рівні соціальної агрегації або чотири рівні носія (індивідууми, малі групи, організації, міжорганізаційна область).

Хедлунд здійснює опис рушійних сил перенесення знання й трансформації, які виражені процесами:

- оформленням й інтернаціоналізацією, взаємодією яких є рефлексія;
- поширенням і засвоєнням, взаємодією яких є діалог;
- асиміляцією й розсіюванням, які відносяться до витягу знання із середовища й уведенню знання в середовище.

3. Модель Ерла. Дана модель відображає розходження між даними, інформацією й знанням. Знання поділяються на три категорії (три рівні):

- прийняте знання – «наука» (дані);

- здійсненне знання – «судження» (інформація);
- потенційне знання – «досвід» (знання).

Зокрема, потенційне знання має найбільшу потенційну віддачу. Модель припускає, що компанія може створювати й захищати знання, використовуючи активи знання й чотири функції:

- інвентаризація – картографування індивідуального й організаційного знання;
- аудит – оцінка природи й величини запланованого незнання, розвиток знання через пізнавальні дії;
- соціалізація – створення подій, які уможливають для людей обмін неявним знанням;
- випробування – проблема невідомого незнання розглядається шляхом навчання на базі досвіду, дії й дослідження незвичайних ситуацій.

У цьому контексті доцільно розглянути школи управління знаннями (табл. 3.4).

Грунтуючись на цих положеннях, модель управління знаннями Ерла містить технічні та соціальні компоненти: система знання; мережі; працівники, що набувають знань у компанії.

4. Модель Караянуса. Дана модель об'єднує в межах управління знаннями інформаційні технології з управлінським й організаційним пізнанням. В оформленому вигляді модель позначена як «Мережа організаційного знання» (англ. – Organizational Knowledge Network), або ОК-Net. Ця модель побудована на трьох ключових елементах: мета пізнання, метанавчання й метазнання.

Таблиця 3.4

Школи управління знаннями

| Школи | Напрямки |
|-----------------------|--|
| Технократичний підхід | |
| 1. Системна | Заснована на традиціях систем знань і експертних систем. Центральна ідея – одержати знання й зробити |

| | |
|-------------------------|---|
| | його доступним для тих, хто може його використовувати |
| 2. Картографічна | Пов'язана з відображенням організаційного знання на базі встановлення зв'язку між знанням і людьми. Використовується підтримка з боку інформаційних технологій. |
| 3. Процесна | Є результатом реінжинірингу бізнес-процесів. Базується на двох основних ідеях: - бізнес-процеси підсилюються, коли оперативному персоналу буде надане потрібне для даного завдання знання; - передбачається, що процеси управління традиційно є більш ємними відносно знань, ніж бізнес-процеси, тобто контекстуальне знання й знання «зразкової практики» є важливими. |
| Економічний підхід | |
| 4. Комерційна | Базується на комерційному захисті й використанні активів знань (інтелектуальної власності): патенти, авторські права й торговельні марки. Знання трактується як актив або товар. Даний підхід вимагає створення технології управління інтелектуальною власністю як рутинного процесу. |
| Бихевіористський підхід | |
| 5. Організаційна | Інтенсивно використовуються співтовариства для активізації обміну й створення знання, які можуть бути інтерорганізаційними або інтраорганізаційними та часто бувають міждисциплінарними. |
| 6. Просторова | Базується на використанні простору для того, щоб сприяти обміну знаннями. Соціалізація використовується як засіб обміну знаннями і їхнього створення. |
| 7. Стратегічна | Управління знанням розглядається як розмірність конкурентної стратегії. Метою є створення й використання активів знання на базі різноманітних можливих засобів. |

Джерело: складено на підставі [64]

У моделі використовується матриця, що складається з послідовних циклів знання, у яких індивідуум або компанія перебуває на чотирьох стадіях розуміння й незнання: незнання про незнання, незнання про розуміння, розуміння незнання, розуміння.

Позначається ситуація в одній із цих осередків (квадрантів). Дія спрямована на управління переходом з одного стану в інший. Переходи здійснюються двома шляхами:

- здатністю до взаємодії;
- погодженістю дій.

Взаємодія здійснюється за допомогою інформаційної технології ефективним способом. Правильно скерована, вона породжує не тільки спіральне знання або метазнання, але також і знання методів навчання.

5. Модель Уіга. Основним принципом даної моделі є те, що знання повинні бути організованими й синхронізованими, щоб бути корисними й цінними.

Серед ключових компонентів даної моделі виділяють наступні:

- завершеність;
- пов'язаність;
- конгруентність;
- перспективи й цілі.

Зокрема, завершеність має на увазі, наскільки відповідні знання можна одержати з даного джерела. При цьому джерела варіюються від людського розуму до баз знань.

Пов'язаність – підтверджує, що відносин між різними об'єктами знань є зрозумілими й чітко визначеними.

Конгруентність – база знань вважається конгруентною, коли всі факти, концепції, перспективи, цінності, судження, реляційні посилання й зв'язки між об'єктами є послідовними.

Перспективи й цілі - це техніка, через визначається точка зору для досягнення конкретної мети

Узагальнена схема цієї моделі базується на чотирьох положеннях, на яких повинна сфокусуватися ініціатива по управлінню знанням:

- фундамент управління знаннями складається з того, як створюється знання, як воно використовується при рішенні проблем і прийнятті рішень, як воно проявляється з пізнавальної точки зору в технологіях і процедурах;
- необхідні постійні дослідження знання;

- необхідна оцінка знання для кожного бізнесу до початку, у процесі й після закінчення;

- необхідне активне управління знаннями протягом усього життєвого циклі компанії, на всіх стадіях життєвого циклу товару й інновацій.

Фактично ця модель намагається визначити різні рівні інтернаціоналізації знань і, отже, може розглядатися як подальше уточнення моделі Нонака у квадранті інтерналізації.

6. Модель Едвісона – модель інтелектуального капіталу.

Центральним компонентом моделі є схема організації активів компанії, чотири компоненти інтелектуального капіталу й їхня взаємодія для створення вартості:

- людський капітал, включаючи знання й навички, які можуть бути конвертовані у вартість;

- структурний капітал як допоміжна інфраструктура компанії;

- бізнес-активи, які визначаються як структурний капітал, що використовується для створення вартості в комерційному процесі;

- інтелектуальна власність – інтелектуальні юридично захищені активи.

Динамічний характер моделі полягає в створенні вартості за двома напрямками:

- інновації;

- продукти й послуги, що виникають у результаті комерціалізації інновацій.

7. Модель Сноудена – модель екології управління знаннями. Дана модель заснована на когнітивній науці, семіотиці й епістемологічній прагматиці.

У моделі розвивається орієнтована на дію система знань, що складається із чотирьох елементів:

- явне й неявне знання;

- активи знання;

- віра;

- визначеність і невизначеність рішень відносно цілей і причинних відносин.

Відповідно, складається матриця рішень, що буде управляти процесом із чотирьох типів перехідних дій:

- обміном явним знанням через системи й структури;
- обміном неявним знанням через психосоціальні механізми;
- перетворенням неявного знання в явне на основі;
- вивільненням неявного знання через віру й її рушійні сили. Все це веде до екології управління знаннями у групі.

8. Модель Інкпена та Динура – процеси управління знаннями. Зазначена модель є емпіричною, створеною для експлікації й передачі знання між партнерами.

9. Модель Бурена – управління інтелектуальним капіталом. Модель розроблена робочою групою по ефективному управлінню знанням у межах віртуальної організації, складеної із практиків управління знаннями у різних областях.

Модель включає два набори характеристик:

- які мають відношення до запасів інтелектуального капіталу, включаючи: людський капітал, інноваційний капітал, процесний капітал, і клієнтський капітал;
- які мають відношення до фінансового функціонування й ефективності бізнесу.

Зокрема, у моделі вводиться поняття «критична точка посилення потенціалу управління знаннями» і «критичні процеси управління знаннями».

10. Модель Деспре й Шаувеля – таксономія управління знаннями. У даній моделі виділено чотири аспекти, що зустрічаються в більшості досліджень:

- час – у зв'язку з лінійним і спрощеним поданням когнітивного процесу;
- тип – стосовно неявного і явного знання;
- рівень – стосовно до різних рівнів соціальної агрегації;
- контекст – ніякий елемент знання не має змісту поза даним контекстом.

Крім того, у зазначеній моделі виділено основних кластерів діяльності:

- бізнес-інтелект;
- бенчмаркинг;
- побудова сховищ даних;
- програмне забезпечення для групової роботи;
- співтовариства практики;
- інновація й синергії, творчість;
- навчання, компетенції, розвиток службовців.

Крім представлених вище традиційних моделей управління знаннями, в групах останнім часом застосовуються й їх інноваційні різновиди.

1. Модель фон Крога та Руса – відноситься до організаційної теорії пізнання й є першою моделлю, що точно розрізняє індивідуальні й соціальні знання.

Ця модель аналізує наступні аспекти:

- чому й яким чином знання приходять до компанії;
- що знання означає для працівників, а також компанії в цілому;
- які існують перешкоди організаційного управління знаннями.

Зазначена модель досліджує природу управління знаннями з погляду: штатних співробітників компанії, комунікації й зв'язку, організаційної структури й розташуванням компанії, мережею між членами компанії, управління людськими ресурсами [29].

2. The Choo Sense-Making KM Model – фокусується на:

- змісті рішень;
- створенні знань;
- навичках прийняття рішень.

Модель містить три тісно взаємозалежних процеси, що відіграють важливу роль у розкритті бачення знань компанії, це – потенціал для створення знань.

Зокрема, Sense Making гарантує, що компанії будуть адаптуватися й продовжувати процвітати в динамічному й складному середовищі за рахунок діяльності щодо конкурентної розвідки й інтерпретації відповідної інформації, що, у свою чергу, дозволяє їм зрозуміти зміни, тенденції й

сценарії відносно поведінки клієнтів, постачальників, конкурентів й інших зовнішніх суб'єктів зовнішнього та внутрішнього середовища.

Knowledge Creation – це процес, що дозволяє компанії створювати або здобувати, організовувати й обробляти інформацію з метою одержання нових знань за допомогою організаційного навчання. Нові отримані знання дозволяють компанії відкривати нові здатності й можливості, створювати нові продукти й нові послуги, поліпшувати й перепроектувати свої існуючі організаційні процеси.

Decision Making передбачає, що компанія повинна вибрати найкращий варіант серед тих, які є правдоподібними та прагни до його реалізації на основі організаційної стратегії [18].

3. Модель Voisot I-Space КМ – заснована на ключовій концепції, що дозволяє відрізнити інформацію від даних, підкресливши, що інформація є тим, що спостерігач буде витягати з даних залежно від його очікування або попередніх знань.

Модель можна розглядати як тривимірний куб з наступними параметрами:

- від некодифікованого до кодифікованого;
- від конкретного до абстрактного;
- від інфузії до дифузії.

Модель пропонує цикл соціального навчання, що дозволяє моделювати динамічний потік знань через низку етапів:

- scanning – мудрість досягається даних, що є загальнодоступними або розповсюджуваними;
- problem-solving – проблеми вирішуються у межах структури розміщення знань, коли знання стають кодифікованими;
- abstraction – знову кодифікована мудрість розповсюджується на широке коло ситуацій, коли знання стають більш абстрактними;
- siffusion – нова мудрість, що отримується з цільової групи населення в кодифікованій й абстрактній формі, коли знання стають розсіяними;

- absorption – знову кодифіковані ідеї застосовуються до різних ситуацій, що породжують новий досвід навчання, коли знання поглинається й викликав вивчену поведінку та, відповідно, стає не кодифікованим;

- ampracting – абстрактне знання фіксується в конкретних практиках, наприклад, у правилах або моделях поведінки, коли знання стають конкретними.

Модель Voisot I-Space КМ навіть порівнюється з живим організмом. Їхній процес росту й розвитку активів знань усередині компанії постійно змінюється. Це означає, що компаніям необхідно використовувати динамічну стратегію, що враховує динамічний характер організаційного циклу навчання [19].

4. Комплекс моделей адаптивних систем (англ. – Intelligent Complex Adaptive Systems – ICAS) – передбачає розгляд компанії як складної адаптивної системи. Складні адаптивні системи містять у собі безліч незалежних агентів, які взаємодіють один з одним на місцевому рівні, і їхня спільна поведінка приводить до виникнення складних адаптивних явищ.

Системи ICAS засновані на принципах кібернетики, що використовує комунікації й механізми контролю для того, щоб зрозуміти, описати й спрогнозувати, що життєздатна компанія повинна робити [21].

Адаптивні системи включають безліч незалежних агентів, що взаємодіють між собою. Їхня поведінка уможливорює появу деяких складних обставин адаптації. Загальна модель складного поведіння є результатом всіх взаємодій. Усередині адаптивної моделі інтелектуальні елементи складаються з людей, які самоорганізуються, але дехто може залишитися в складі загальних ієрархій компаній [22].

Фактично завдання полягає в тому, щоб використовувати переваги, що виникають як результат взаємодії людей у групі, зберігаючи глобальне відчуття єдності. Таким чином компанія може вирішити проблеми шляхом створення можливостей, використання ресурсів, як внутрішніх, так і зовнішніх, з метою отримання конкурентних переваг, в межах груп соціальних мереж.

Висновки до третього розділу

1. Запропоновано раціональні критерії оцінки рівня впливу учасників груп у соціальних мережах на конкурентоспроможність товарів і послуг: чистий приріст – характеризує активність зростання соціальної мережі в заданий період часу; відсоток неактивних акаунтів; відсоток цільової аудиторії у групі соціальної мережі; активне ядро; охоплення користувачів; негативний відгук; частота відгуку; кількість обговорень.

Зазначено, що розробка системи критеріїв оцінки рівня впливу учасників груп у соціальних мережах на конкурентоспроможність товарів або послуг є продуманим менеджментом, в основі якого лежить аналіз великих обсягів інформації з використанням статистичних методів її обробки.

2. Запропоновано нову модель нечіткого коефіцієнта значущості параметрів соціального профілю, відмінністю якого є можливість врахування як формалізованих об'єктивних, так і суб'єктивних показників, що важко інтерпретуються, для оцінки вихідної інформації СП. Перевагою використання критерію значущості перед залученням експертів є менші часові витрати на аналіз великих обсягів даних при збереженні рівня об'єктивності оцінювання.

3. Запропоновано інфологічну модель для представлення статичної та динамічної частин соціального профілю, що дозволяє спроектувати схему інформаційного забезпечення системи соціального профілювання в рамках розробленої методики збору і аналізу гетерогенних соціальних даних з відкритих джерел мережі Інтернет. Відмінність моделі полягає у використанні принципів теорії метаграфії і реалізації запутаної структури соціального середовища у вигляді чотирьох великих категорій соціальної інформації, що відноситься до персон, співтовариств і соціальних явищ, що дозволяє подолати проблему гетерогенності і нерівноцінності вихідних даних.

РОЗДІЛ 4. ПРАКТИЧНА РЕАЛІЗАЦІЯ РОЗРОБЛЕНИХ МЕТОДІВ І МОДЕЛЕЙ ПІДТРИМКИ РІШЕНЬ ДЛЯ ДИФЕРЕНЦІАЦІЇ ДАНИХ КОНСОЛІДОВАНОЇ ІНФОРМАЦІЇ

4.1. Розробка принципів і архітектури систем обробки інформації Big Data

Розроблені в ході дисертаційного дослідження моделі, методика збору та обробки неструктурованих даних цифрового соціального середовища і супутні алгоритми побудови соціального профілю спрямовані, в тому числі, на рішення практичної задачі інструментального переоснащення з метою більш ефективного виконання комплексу управлінських завдань. Відповідно, було проведено огляд сфер діяльності, пов'язаних з обробкою соціальних даних, результатом якого став список напрямків, потенційно зацікавлених в прикладному використанні системного вирішення соціального профілювання:

- сфера торгівлі;
- фінансові організації;
- страхові компанії;
- кадрові служби підприємств;
- правоохоронні органи;
- органи охорони здоров'я;
- органи місцевого самоврядування;
- науково-дослідні й освітні установи.

Далі доцільно перейти до розгляду проблем, що вирішуються через використання соціальних профілів, в інших областях.

Фінансові та страхові організації на постійній основі мають справу з ситуаціями, схильними до ризиків людського фактора. Їх оцінка дуже значима і впливає на репутацію і доходи організації. Соціальні профілі пропонується використовувати для складання документа-характеристики

клієнтів. Грунтуючись на них, відповідальна особа, яка приймає рішення, може дозволити або відмовити клієнту в послугах компанії, підібрати йому найбільш прийнятний пакет послуг, а також удосконалювати скрипт діалогу для підвищення лояльності клієнтів і оптимізації ведення переговорів з діловими партнерами.

Використання соціальних профілів в кадрових службах підприємств може спростити процедуру працевлаштування і підвищити ефективність роботи колективу. Система соціального профілювання може застосовуватися для автоматизованої генерації і перевірки резюме на відповідність списку вимог, що пред'являються вакантною посадою, а також для розсилки пропозицій про працевлаштування потенційним співробітникам на підставі інформації про їхні інтереси і досягнення. Крім того, соціальні профілі можуть використовуватися керівниками для вирішення управлінсько-організаційних завдань з метою оптимізації діяльності робочого колективу і підвищення продуктивності праці.

Сфера охорони здоров'я зацікавлена в розробці системи автоматизованої діагностики пацієнтів та систематизації результатів новітніх медичних досліджень з метою вдосконалення методів лікування захворювань. Система побудови соціальних профілів з модулем знеособлення персональних даних уможливує спільне використання спеціалізованої медичної інформації та даних про життя пацієнтів із загальнодоступних цифрових соціальних медіа.

Органи державної влади та місцевого самоврядування зацікавлені в реалізації економічно і соціально значущих національних і регіональних проектів. Впровадження програмно-інструментального комплексу соціального профілювання сприяє вирішенню ряду ключових завдань, серед яких:

- формування системи підготовки кадрів, спрямованої на навчання основам підвищення продуктивності праці, в тому числі за допомогою використання цифрових технологій і платформних рішень;

- забезпечення оптимальної доступності для населення (в тому числі для жителів населених пунктів, розташованих у віддалених місцевостях) медичних організацій, що надають первинну медико-санітарну допомогу;

- створення механізмів розвитку комфортного міського середовища, комплексного розвитку міст та інших населених пунктів з урахуванням індексу якості міського середовища.

Складання групових соціальних профілів дозволить підвищити ефективність освітньо-виховної роботи. На їх основі педагог може вносити корективи в навчальний план з метою впровадження індивідуального підходу до кожного учневі. Також соціальні профілі допоможуть більш ефективно запобігати і вирішувати конфліктні ситуації в колективі учнів. Сфера науки надає безліч варіантів впровадження моделей соціальних профілів, починаючи з розробки методів людинно-машинної взаємодії і закінчуючи гуманітарними дослідженнями: наприклад, для оцінювання рівня життя людей певного стану в різні історичні періоди, реакції сучасників на ключові події і так далі.

Архітектура системи обробки Big Data використовує складні системи, в яких можна виділити кілька компонентів або шарів (Layers). Зазвичай виділяють чотири рівні компонентів таких систем: прийом, збір, аналіз даних і представлення результатів. Цей поділ є значною мірою умовним, так як, з одного боку, кожен компонент, в свою чергу, може бути розділений на підкомпоненти, а з іншого боку деякі функції компонентів можуть перерозподілятися в залежності від розв'язуваної задачі і використовуваного програмного забезпечення, наприклад, виділяють зберігання даних в окремий шар.

Для роботи з Big Data розробниками систем створюються моделі даних, змістовно пов'язані з реальним світом. Розробка адекватних моделей даних є складним аналітичним завданням, що виконується системними архітекторами і аналітиками. Модель даних дозволяє створити математичну модель взаємодій об'єктів реального світу і включає в себе опис структури

даних, методи маніпуляції даними і аспекти збереження цілісності даних. Опис розробки моделей даних не є завданням цієї роботи. Для зберігання даних використовуються розподілені системи різних типів. Це можуть бути файлові системи, бази даних, журнали, механізми доступу до загальної віртуальної пам'яті. Більшість систем зберігання орієнтовані виключно на роботу з Big Data та мають вкрай обмежене число функцій, що пояснюється внутрішньою складністю створення високоефективних розподілених систем.

Для того, щоб робота з даними відбувалася швидше, системи зберігання і обробки даних розпаралелюються в кластері (cluster – це група комп'ютерів, об'єднаних мережею для виконання єдиного завдання). Однак, відповідно до гіпотези Брюера, неможливо забезпечити одночасну узгодженість (несуперечність) даних, доступність даних і стійкість системи до відокремлення окремих вузлів. Гіпотеза доведена для транзакцій типу ACID (англ. – Atomic, Consistent, Isolated, Durable) і відома під назвою CAP-теореми (англ. – Consistency, Availability, Partition tolerance).

1. Прийом даних (англ. – Data Ingestion). Джерела даних мають різні параметри, такі як:

- частота надходження даних з джерела;
- обсяг порції даних;
- швидкість передачі даних;
- тип даних, що надходять, та їх достовірність.

Для ефективного збору даних необхідно встановити джерела даних. Це можуть бути:

- сховища даних;
- постачальники агрегованих даних;
- API будь-яких датчиків;
- системні журнали;
- згенерований людиною контент в соціальних мережах, в корпоративних інформаційних системах;

- геофізична інформація;
- наукова інформація;
- успадковані дані з інших систем. Джерела даних визначають вихідний формат даних. Наприклад, ми можемо самостійно проводити погодні дослідження на території аеропорту, використовувати дані, що надходять від літаків, що злітають і сідають, закупити дані з супутників, що пролітають над аеропортами, у місцевої метеослужби, а також знайти їх десь в мережі в іншому місці.

У загальному випадку для кожного джерела необхідно створювати власний складальник (Data Crawler для збору інформації в мережі і Data Acquisition для проведення вимірювань). Прийом даних полягає в початковій підготовці даних від джерел з метою приведення даних до загального формату представлення даних. Цей єдиний формат вибирається відповідно до прийнятої моделі даних. Виконуються перетворення систем вимірювання, типів (типізація), верифікація. Обробка даних змістовно не впливає на наявну в даних інформацію, але може змінювати її уявлення (наприклад, приводити координати до єдиної системи координат, а значення – до єдиної розмірності).

2. Збір даних (англ. – Data Staging). Етап збору даних характеризується безпосередньою взаємодією з системами зберігання даних. Встановлюється точка збору, в якій зібрані дані забезпечуються локальними метаданими і поміщаються в сховище або передаються для подальшої обробки. Дані, що з яких-небудь причин не пройшли точку збору, ігноруються.

Для структурованих даних проводиться перетворення з вихідного формату за заздалегідь заданими алгоритмами. Це найбільш ефективна процедура в разі, якщо структура даних відома. Однак якщо дані представлені в двійковому вигляді, структура і зв'язки між даними загублені, то розробка алгоритмів і заснованого на них програмного забезпечення для обробки даних може виявитися вкрай скрутною.

Для напівструктурованих даних потрібна інтерпретація даних, що надходять, і використання програмного забезпечення, що вміє працювати з мовою опису даних. Істотним плюсом напівструктурованих даних є те, що в них часто містяться не тільки самі дані, але метадані у вигляді інформації про зв'язки між даними і способи їх отримання [117].

Розробка програмного забезпечення для обробки напівструктурованих даних являє собою досить складну задачу. Однак є значна кількість готових конвертерів, які можуть, наприклад, витягти дані з формату XML в сформоване табличне представлення.

Найбільшого обсягу робіт вимагає обробка неструктурованих даних. Для їх переведення до заданого формату може знадобитися створення спеціального ПЗ, складна ручна обробка, розпізнавання і вибіркового ручний контроль.

На етапі збору проводиться контроль типів даних і може виконуватися базовий контроль достовірності даних. Наприклад, координати молекул газу, що містяться в будь-якій області, не можуть лежати за межами цієї області, а швидкість не може істотно перевищувати швидкість звуку. Для того, щоб уникнути помилок типізації, необхідно перевіряти правильність налаштувань одиниці виміру. Наприклад, в одному наборі даних висота може вимірюватися в кілометрах, а в іншому – в футах. У цьому випадку необхідно зробити перетворення висоти в ті одиниці виміру, які прийняті у моделі, що використовується.

При зборі дані систематизуються і забезпечуються метадані, збережені в пов'язаних метаданих. При наявності великої кількості джерел даних може знадобитися управління збором даних для того, щоб збалансувати обсяги інформації, що надходять з різних джерел.

3. Аналіз даних (англ. – Analysis Layer). Аналіз даних, на відміну від збору даних, використовує інформацію, що міститься в самих даних. Аналіз може проводитися як в реальному часі, так і в пакетному режимі. Аналіз даних становить основне по трудомісткості завдання при роботі з Big Data.

Існує безліч методик обробки даних: предикативний аналіз, запити та звітність, реконструкція з математичної моделі, трансляція, аналітична обробка та інші. Методики використовують специфічні алгоритми в залежності від поставлених цілей. Наприклад, аналітична обробка може бути:

- аналізом зображень, соціальних мереж, географічного розташування;
- розпізнаванням за ознаками, текстовим аналізом, статистичною обробкою;
- аналізом голосу;
- транскрибуванням.

Алгоритми аналізу даних також, як і алгоритми обробки даних, спираються на модель даних. При цьому під час аналізу може бути використано кілька моделей, які задають загальний формат даних, але по-різному моделюють змістовні процеси, дані про яких ми обробляємо. При аналізі методів штучного інтелекту, зокрема нейронних мереж, проводиться динамічне навчання моделей на різних наборах даних.

При аналізі даних проводиться ідентифікація сутностей, що описуються даними, на підставі наявної в даних інформації і використовуваних моделей. Сутністю аналізу є аналітичний механізм, який використовує аналітичні алгоритми, управління моделями та ідентифікацію сутностей для отримання нової змістовної інформації, що є результатом аналізу.

Для аналізу даних також використовуються методи штучного інтелекту на нейронних мережах, що не розглядаються в даній роботі.

4. Представлення результатів (англ. – Consumption Layer). Результати аналізу даних надаються на рівні споживання. Є кілька механізмів, що дозволяють використовувати результати аналізу Big Data.

1). Моніторинг метаінформації – це підсистема відображення в реальному часі істотних параметрів роботи системи, завантаженості обчислювачів, розподіл завдань в кластері, розподіл інформації в сховищах,

наявність вільного місця в сховищах, надходження даних від джерел, активності користувачів, відмов обладнання тощо.

2). Моніторинг даних – це підсистема відображення в реальному часі процесів прийому, збору і аналізу даних, навігація за даними.

3). Генерація звітів, запити до даних, подання даних у вигляді візуалізації на дашбордах (англ. – Dashboard), в форматі PDF, інфографіки, зведених таблицях і коротких довідках.

4). Перетворення даних і експорт в інші системи, інтерфейс з ВІ-системами.

4.2. Побудова інформаційної моделі формування соціального портрета особистості за допомогою технології OSINT

На сьогоднішній день з урахуванням бурхливого розвитку інформаційних технологій і соціальних процесів гостро стоїть проблема ефективного виконання управлінських завдань і прийняття рішень при роботі з великими неструктурованими гетерогенними масивами даних, на основі соціального портрета особистості. Це пов'язано зі значним ускладненням як первинних даних, так і структури керуючих інформаційних систем.

Для моделювання та дослідження таких систем в даний час широке застосування отримала OSINT (англ. – Open Source INTelligence) – концепція, методологія і технологія легального отримання і використання інформації з відкритих джерел. На даний час найбільш докладні описи змістовного наповнення OSINT викладені в збірниках «NATO Open Source Intelligence Handbook» (2006–2017 p.p.), «NATO Open Source Intelligence Reader» (2006–2017 p.p.) і «NATO Intelligence Exploitation of the Internet» (2002 p.).

Пошук за відкритими джерелами (OSINT) – процес, в ході якого проводиться виявлення, вибір, збір і аналіз інформації, що знаходиться у

вільному доступі, та може дозволити істотно підвищити ефективність систем прийняття управлінських рішень.

Методика роботи за принципами OSINT вже тривалий час активно використовується в бізнес-колах провідних країн світу для пошуку і отримання законними шляхами інформації про партнерів або конкурентів. Одним з базових «золотих правил» такої концепції є те, що близько 90% інформації, необхідної для аналізу і прийняття відповідних рішень, розміщені у відкритих джерелах. На рис. 4.1 наведені можливі різні джерела, які використовуються для консолідації інформації в технології OSINT.



Рис. 4.1. Різноманітні джерела інформації OSINT

Джерело: складено на підставі [117]

На тлі стрімкого розвитку сучасних інформаційних технологій даному виду пошукової діяльності приділяється все більше уваги. Різниця між новачком, який шукає в Інтернеті інформацію і OSINT продажем досить помітна: там, де початківець побачить фото, репости, групи і сторінки, на які підписані людина або організація в соціальній мережі – фахівець побачить активність, дати публікацій, фон на фотографіях, можливі причини підписки на певні групи, кола спілкування. Найчастіше людина використовує один або кілька псевдонімів в мережі, а, значить, за знайденим нікнеймом через

запит можна знайти й іншу її активність в Інтернеті, наприклад, в соцмережі або на форумах, які відвідує людина. І це лише кілька прикладів. При цьому зібрана інформація служить основою для подальшої обробки, очищення (шляхом оцінки надійності джерел отримання та достовірності відомостей), аналітичного узагальнення та інтерпретації кінцевих результатів.

Таким чином, базова ідея OSINT – це цілеспрямований збір інформації (Harvesting) про об'єкт зацікавленості з метою подальшої обробки та різновекторного контент-аналізу отриманих даних (створення «портрета» особистості, виявлення неочевидних фактів або зв'язків, прогноз її поведінки тощо).

OSINT зручний тим, що:

- передбачає набагато менше ризиків: не порушується чиясь приватність і закони;
- ця технологія дешевше – не потрібно якимось додаткове обладнання і дорогий софт;
- до такої інформації легко отримати доступ (зайти в мережі Інтернет), і частіше за все вона завжди актуальною.

Існує два основні методи збору інформації.

1. Пасивний. В цьому випадку пошукачу інформації неможливо видати себе і те, що він шукає. Пошук обмежується контентом на сайті об'єкта дослідження, архівною або кеш інформацією, незахищеними файлами.

2. Активний. Цей метод використовується для Інтернет-розвідки набагато рідше. Для отримання інформації досліджується ІТ-інфраструктуру компанії, здійснюється активна взаємодія з комп'ютерами і машинами. Використовуються просунуті техніки для отримання доступу до відкритих портів, сканування вразливостей і серверних веб-додатків. У цьому випадку інформаційна розвідка може бути легко розпізнана. Соціальна інженерія теж відноситься сюди.

Вибір методу збору інформації залежить від того, на яких вона збирається, а також від того, які дані потрібні. Важливо розуміти, що не завжди те, до чого можна легко отримати доступ, є легальним.

Будь-яка інформація в публічному доступі може виявитися небезпечною: соціальні мережі, фотографії, дані з профілів і сайтів третіх сторін, публічна документація тощо. Адже в сукупності з іншими даними вона може розповісти хакерам те, що вони шукають.

Зупинимося на основних напрямках, які фахівці з інформаційної безпеки досліджують на регулярній основі.

1. Метадані файлів. У них можна знайти:

- дату створення документа;
- імена користувачів;
- моделі принтерів;
- ППЗ, встановлене на комп'ютерах;
- іноді геолокацію.

Інформація про встановлені програми й їх версії, наприклад, дасть можливість вибрати найбільш вразливі і підібрати експлойти. Імена користувачів, в свою чергу, стануть потенційними логінами в особистих або корпоративних системах.

2. Конфіденційна документація. Навіть в самих просунутих компаніях і серйозних державних структурах якийсь засекречений документ може випадково опинитися у відкритому доступі. Конфіденційна інформація може включати в себе політику створення паролів, а також використовуване ППЗ і сервіси.

3. Дані про домен. Існує дуже багато інструментів, які допомагають зібрати всі дані з сайту (в тому числі, які не видно звичайним користувачам), наприклад:

- е-мейли,
- телефони,
- факси,

- технології, на яких побудований сайт,
- криптографічні сертифікати, які використовуються на конкретному домені.

Після дослідження основного домену необхідно дослідити, як компанія організовує свої Інтернет-ресурси. Серед субдоменів зазвичай є погано захищені сайти для тестування нових технологій. Такі субдомени можуть містити якісь важливі документи, залишені на сервері.

4. Серверні веб-додатки, інтернет речей. Індексуватися можуть сервера, роутери, камери відеоспостереження, вебкамери, онлайн накопичувачі тощо. Крім того, що до деяких можна отримати доступ просто пройшовши за посиланням, ці девайси містять в собі технічну інформацію, геолокацію, відкриті порти, запущені сервіси, доменне ім'я, пов'язане з пристроєм, інтернет провайдерів, веб-технології.

Розвиток в OSINT здійснюється наступним чином

1. Освоєння базових технік, наприклад Google dorks.
 2. Пошук цікавих підходів до використання інструментів і технік та написання невеликих звітів з візуалізацією результатів.

3. Максимальна анонімність. При вивченні OSINT значний час приділяється забезпеченню своєї безпеки при пошуку. Це потрібно, щоб компанія або людина не змогли розпізнати, що здійснюється збір якоїсь інформації.

Відповідно, нижче показано кілька подібних практик:

- створення фейкових профілів;
- використання андроїд-емуляторів;
- VPN;
- браузер Tor;

Нижче наведено декілька ключових інструментів OSINT.

1. Shodan - пошуковик по пристроях, підключених до мережі (у т.ч. інтернет речей і веб-додатки). Розділ «Exploze» допоможе почати пошуки, так як туди збираються запити користувачів. Щоб отримати доступ до

розширеного пошуку, треба зареєструватися. У платних версіях надається доступ до більшої кількості пристроїв, а також необмежена кількість пошукових запитів в добу.

2. MaltegoMaltego – ПЗ, яке збирає всі дані разом, допомагає побачити взаємозв'язки і зробити висновки. Результат візуалізується у вигляді дерева, що збирає в єдину систему IP-адреси, е-мейли, телефони, домени тощо.

3. Google Dorks – це запити в Google з використанням спеціальних операторів.

4. Foca - програма, яка допомагає з вивантаженням, класифікацією і аналізом файлів на віддаленому веб-сервері. Для цього вона сканує певний домен за допомогою пошукових систем Google, Bing, DuckDuckGo.

5. Spuse – пошуковик з технічної інформації веб-сайтів. З його допомогою існує можливість знайти різноманітні дані, на предмет кшталт вразливостей, IP-адрес, субдоменів і SSL / TLS.

Соціальний портрет являє собою інформаційну структуру, що описує соціальні властивості окремої людини чи спільноти, причому ця інформація має властивість ясності для людиномашинного сприйняття, що забезпечує можливість її автоматизованої обробки в різних прикладних задачах.

Виходячи з визначення, соціальний портрет представляє собою неоднорідну семантичну мережу, що складається з персоналізованих даних. На інформаційній моделі соціального портрету ґрунтуються модель соціального середовища і методика рішення задачі побудови соціальних портретів.

Описи соціальних явищ важливі для цілісного уявлення про соціальне середовище й окремі соціальні портрети: найчастіше в соціальних дослідженнях вони можуть виявитися в безлічі центральних об'єктів. Соціальні явища мають ряд властивостей:

- час і місце виникнення;
- список учасників визначення характеристик явища;

- посилання на джерела інформації, що підтверджують факт здійснення явища;

- пов'язані інші соціальні об'єкти і явища.

У завданнях управління соціальні явища представлені громадськими подіями і заходами, інноваційними впровадженнями, законодавчими ініціативами тощо. Відповідно до теорії акторномережевого аналізу, елементи соціального портрета, виражені характеристиками, поняттями, подіями, явищами, являють собою актанти, а соціальні зв'язки між елементами – предикати.

Результати побудови соціального портрета є структурованими даними, виділеними з динамічного контенту і пов'язаними з інформаційною картою в ході аналізу. Вони являють собою безмасштабну мережу, яка для подання і подальшого використання аналітиками повинна зберігатися в базах даних, здатних зберігати графи і слабоструктуровані відомості.

Відповідно інформаційна модель соціального портрета спільноти буде представлена наступним чином:

$$PSP = \{P(X, v), R(Y, v), Q\}, \quad (4.1)$$

де $P = \{P_1, \dots, P_m\}$ – множина характеристик соціальних об'єктів, пов'язаних з розглянутою персоною (теми, події, інші персони тощо);

$R = \{R_1, \dots, R_n\}$ – множина соціальних зв'язків між соціальними об'єктами даної персони (реалізованих у вигляді їх предикатів);

$Q = \{Q_1, \dots, Q_{m+n}\}$ – множина оцінок інтонацій соціальних характеристик, на основі яких будується комплексна оцінка настрою розглянутої персони Q_{avg} , що залежить від кількості соціальних характеристик, являє собою нечітку величину, причому $Q \in [-1; 1]$, m, n - кількість характеристик соціальних об'єктів і зв'язків відповідно, причому $m - 1 \leq n$; -

v – вагові коефіцієнти характеристик соціальних об'єктів і зв'язків, що являють собою нечіткі величини; X, Y – множини властивостей соціальних об'єктів і зв'язків відповідно, причому

$$\begin{cases} X = AO \times SO; \\ Y = AE \times SE \end{cases} \quad (4.2)$$

де $AO = \{AO_1, \dots, AO_f\}$ – множина атрибутів (властивостей), пов'язаних із соціальним об'єктом;

$AO_i = \{AO_{i1}, \dots, AO_{ifx}\}$ – множина значень властивості для одиничного визначення соціального об'єкта;

SO – множина синонімів і повторень визначення соціального об'єкта у вихідній інформації;

$AE_i = \{AE_{i1}, \dots, AE_{ig}\}$ – множина значень властивості для одиничного визначення соціальних відносин;

SE – множина синонімів і повторень визначення соціальних відносин у вихідній інформації;

f, g – кількість властивостей у соціальних об'єктів і зв'язків даної персони відповідно;

fx, gy – кількість входжень властивості (у вигляді синонімів в тому числі) для соціальних об'єктів і зв'язків відповідно.

Традиційно особистості не можуть розглядатися у відриві від соціуму. Люди організують формальні і офіційні спільноти різної спрямованості, причому кількість таких спільнот, що співвідносяться з конкретною персоною практично нічим не обмежена. Виникає проблема дослідження групових діяльності і властивостей, тому як особисті характеристики персони можуть змінюватися під впливом конкретної соціальної групи, що впливає на результати соціологічних досліджень. При аналізі змін характеристик персони виявляється ряд закономірностей, який необхідно формалізувати в процесах інформаційної підтримки прийняття рішень в

задачах управління. Для вирішення даної проблеми необхідно ввести поняття групового соціального портрета (далі – ГСП), який буде враховувати соціальні тенденції спільноти, об'єднані якимись загальними властивостями, а також міжособистісні взаємодії членів групи й їх поведінкові закономірності.

Соціальний портрет групи включає персональні портрети, об'єднані по ряду закономірностей в їх особистих характеристиках, а також властивості, притаманні спільноті, що безпосередньо розглядається. Характеристики діяльності групи можуть бути статичними і динамічними. Відмінність полягає в можливій кількості і типі ключових елементів: ними можуть виступати ідентифікатори персон – лідерів спільноти й ідентифікатори подій або деяких об'єктів реального світу, які виступають в ролі символів – в даному випадку людей пов'язує схоже ставлення до символу. Зважаючи на ці особливості, число внутрішніх зв'язків в ГСП більше сумарної кількості зв'язків всіх соціальних портретів членів групи. Відповідно інформаційна модель соціального портрета спільноти буде представлена наступним чином:

$$GSP = \{PSP, GSP', GP(X', v), GR(Y', v), GQ\}, \quad (4.3)$$

де $PSP = \{PSP_1, \dots, PSP_t\}$ – множина соціальних портретів членів групи GSP;

t – кількість ГСП в групі GSP;

$GSP' = \{GSP'_{11}, \dots, GSP'_{tt}\}$ – множина підгруп, що входять в соціальну групу GSP;

tt – кількість підгруп в соціальній групі GSP;

$GP = \{GP_1, \dots, GP_m\}$ – множина характеристик соціальних об'єктів, пов'язаних із групою GSP;

$GR = \{GR_1, \dots, GR_n\}$ – множина безпосередніх соціальних зв'язків групи GSP;

$$GQ = \{GQ_1, \dots, GQ_{m+n}\} \cup \{Q_{PSP_1}, \dots, Q_{PSP_t}\} \cup \{GQ_{GSP_1}, \dots, GQ_{GSP_{tt}}\} \quad -$$

множина оцінок інтонацій соціальних характеристик групи, а також підгруп і людей, що входять до неї, на основі чого будується комплексна оцінка настрою розглянутого спільноти:

$$GQ_{avg} = \frac{\sum_{l=1}^t Q_{avg1}}{t} + \frac{\sum_{ll=1}^{tt} Q_{avgll}}{tt}, \quad (4.4)$$

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

m, n – кількість соціальних характеристик і зв'язків групи відповідно, причому $m - 1 \leq n$; v – вагові коефіцієнти соціальних об'єктів і зв'язків групи GSP , що являють собою нечіткі величини;

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

$$\begin{cases} X = AO \times SO; \\ Y = AE \times SE, \end{cases} \quad (4.5)$$

де $AO = \{AO_1, \dots, AO_f\}$ – множина атрибутів (властивостей), пов'язаних із соціальним об'єктом, що належить даній групі;

$AO_i = \{O_1, \dots, O_{fx}\}$ – множина значень властивості для одиничного визначення соціального об'єкта, що належить даній групі;

SO – множина синонімів і повторень визначення соціального об'єкта, що належить даній групі, у вихідній інформації;

$AE = \{AE_1, \dots, AE_g\}$ – множина атрибутів (властивостей), пов'язаних із соціальним ставленням, що належить даній групі;

$AE_i = \{E_1, \dots, E_{gy}\}$ – множина значень властивості для одиничного визначення соціальних відносин, що належать даній групі;

SE – множина синонімів і повторень визначення соціальних відносин, що належать даній групі, у вихідній інформації;

f', g' – кількість властивостей у соціальних об'єктів і зв'язків групи GSP відповідно;

$f'x', g'y'$ – кількість входжень властивостей (у вигляді синонімів в тому числі) для соціальних об'єктів і зв'язків групи GSP відповідно.

При аналізі групових СП спрощується визначення неявних властивостей персон. Наприклад, «якщо у однієї людини з групи не визначена будь-яка властивість, а у інших вона приблизно однакова, то з високою ймовірністю можна зробити висновок, що у першої людини присутня та ж властивість». Розроблені інформаційні моделі персонального і групового соціальних портретів дозволяють враховувати всі основні властивості об'єктів, що досліджуються, їх тональність і значимість, а також проводити аналіз по визначенню неявних залежностей. Наступним кроком є перехід до розгляду різноманіття елементів цифрової соціального середовища.

Таким чином, аналіз соціального портрета, дозволяє більш повно охарактеризувати поведінку як соціального індивідуума, так і соціального середовища в цілому.

В результаті проведених досліджень були створені інформаційні моделі персонального і групового соціальних портретів, які дозволяють проводити аналіз по визначенню неявних залежностей за допомогою технології OSINT.

4.3. Алгоритм обробки диференціалізованих даних для соціального профілювання

Процес побудови соціального профілю людини розділяється на етапи, в ході яких вирішуються важливі підзадачі:

- пошук і збір попередньої інформації з публічних джерел мережі Інтернет з використанням засобів Big Data;
- первинна фільтрація зібраної інформації з метою виключити з подальшої обробки конфіденційну і явно марну, сторонню інформацію;

- аналіз вихідних даних з використанням сукупності традиційних і Big Data засобів аналітики з метою створення соціального портрета особистості або групи;

- представлення результатів аналізу в структурованому вигляді, зручному для автоматизованої обробки.

Для виконання перерахованих етапів пропонується методика збору та аналізу персоніфікованих даних, яка розглядає сукупність ознак з різних загальнодоступних джерел мережі Інтернет і виділяє з них справжні, несправжні і суперечливі відомості. Вона реалізує моделі опису взаємодій між елементами цифрового соціального середовища (рис. 4.2).



Рис. 4.2. Структура системи соціального профілювання у загальному вигляді

Джерело: авторська розробка

Однією з важливих відмінностей цієї методики є поділ безлічі інформації, що поділяється на постійну структуровану і динамічну неструктуровану частини. У методиці також пропонується використання розробленого критерію значущості для формалізації елементів соціального профілю. Узагальнена послідовність кроків методики виражається наступним чином:

- ідентифікація і пошук об'єкта соціального середовища;
- поділ вихідних даних на статичну і динамічну частини;
- фільтрація вихідної інформації від конфіденційних даних;
- виключення з розгляду даних, які помилково віднесені до соціального об'єкта, що досліджується;
- аналіз текстових і мультимедійних даних з використанням критерію значущості;
- визначення тональності соціального профілю через сукупність тональностей його складових елементів;
- візуалізація кінцевих результатів соціального профілювання.

Кожен з основних етапів алгоритму поділяється на підетапи і має низку особливостей. Перейдемо до їх розгляду.

Процес створення соціального профілю починається з виявлення та збору вихідної інформації із загальнодоступних ресурсів мережі Інтернет. Для цього потрібно провести ідентифікацію людини або спільноти, які підлягають дослідженню: встановлюються точки входження в мережевий простір – облікові записи в соціальних медіа, які виділяють конкретного користувача із загальної маси і дають йому право користуватися можливостями web-ресурсу, такими, як відправка повідомлень, підписки, голосування, завантаження і вивантаження файлів, соціальні взаємодії (наприклад, додавання користувачів до списків друзів / ігнорованих). У більшості випадків пошук людини в Інтернеті починається з введення одного або сукупності унікальних критеріїв: ім'я, e-mail, вік, ідентифікатор облікового запису. При нестачі або суперечливості наявної інформацією про

досліджуваний об'єкт може проводитися додатковий пошук за непрямими ознаками: авторському контенту, роду діяльності, відомому колу знайомих.

Людина реєструється на безлічі сайтів і зазвичай не використовує своє справжнє ім'я в якості ідентифікатора облікового запису, а вибирає псевдонім або «нікнейм». Псевдоніми можуть не відрізнятися оригінальністю і використовуватися різними людьми на різних ресурсах. Також одна людина може мати кілька нікнеймів. Це ускладнює ідентифікацію, однак, в даний час багатьма Інтернет-ресурсами реалізований механізм прив'язки користувальницьких акаунтів до відповідних облікових записів в популярних соціальних мережах. Таким чином, ідентифікація в одному з пов'язаних джерел гарантує виявлення профілів користувача в інших.

Крім П.І.Б. і псевдонімів для ідентифікації користувачів можуть використовуватися їхні контактні дані: адреса електронної пошти і номер телефону, які аналогічно можуть виступати в якості ідентифікаторів облікових записів. Також приносить користь перевірка часу і місця розташування входу в разі знаходження закономірностей (наприклад, коли повторюється одночасний доступ з однієї IP-адреси, то можна зробити висновок про належність акаунтів одній людині або близько пов'язаним людям).

З публічних профілів користувача web-ресурсу можливо отримати низку важливих відомостей про персону: ім'я, контактна інформація, часовий пояс і місце перебування, інтереси, робота, склад сім'ї, друзі, дати реєстрації та останнього відвідування сайту.

На самому ресурсі вихідна інформація добувається шляхом аналізу активності і пошуку контенту користувача. Потрібно відзначити, що збір вихідних даних може здійснюватися також із джерел, в яких досліджувана людина не зареєстрована, але згадується в будь-якому контексті.

На початку етапу збору ідентифікації інформація та перелік ключових ознак досліджуваної персони або групи заносяться аналітиком в інформаційну карту, представлену реляційною базою даних. Потім

пошуковий механізм автоматизовано проводить збір до нерелеляційного сховища знайдених неструктурованих даних в їх первинному вигляді і з зазначенням гіперпосилань. Вихідні дані можуть являти собою: сторінки, файли мультимедіа, текстові документи, двійкові дані. В рамках вирішення завдання соціального профілювання вони відносяться до динамічного контенту, фільтруються й аналізуються з метою вилучення з них структурованої персоналізованої інформації. Розподілений запис Big Data у сховищі здійснюється фреймворком. Загальне уявлення про систему збору даних наведено на рис. 4.3.

Пошуковий робот може виявити відомості про персону, представлені в упорядкованому вигляді, наприклад, у таблиці профілю користувача веб-ресурсу. Такі дані є структурованими і відразу ж заносяться в інформаційну карту. Список можливих типів характеристик показаний в табл. 4.1.

В ході фільтрації та аналізу динамічного контенту можливе повернення на етап збору вихідних даних для уточнення і виявлення нововиявлених характеристик досліджуваного об'єкта.

Отримані на попередньому етапі неструктуровані вихідні дані спочатку малопридатні для аналізу з огляду на те, що не визначена семантична повнота їх вмісту. Різні соціальні медіа можуть отримувати одну й ту ж саму інформацію, але представляти її з різним ступенем деталізації і вносити спотворення. Смысловий зміст текстової інформації може викладатися в різних формах завдяки лексичній різноманітності людської мови. Однак механізми її подання в символному вигляді на ЕОМ недосконалі: крім лінгвістичних помилок і мови на простоту сприйняття впливає вибір кодування і настройки розмітки, вбудовані в текст у вигляді тегів. Неструктуровані тексти можуть зберігатися в джерелах інформації в неповному фрагментованому вигляді. У разі мультимедійної інформації крім факту наявності фрагментованості і спотворень, важливий вплив на результати аналізу чинять її первинна якість і алгоритми стиснення.



Рис. 4.3. Структурна схема підсистеми збору даних

Джерело: складено на підставі [117]

Таблиця 4.1

Приклад упорядкованої структури, доцільної для розміщення в інформаційній карті

| Ім'я (основне) | Дата народження | Місце народження | Кореневий обліковий запис | Найбільш відомі ознаки |
|-------------------|--------------------|---------------------|------------------------------|---------------------------|
| | | | | |

Джерело: авторська розробка

Додатково існує проблема, з якою може зіткнутися будь-який аналітик, який автоматизовано збирає дані з мережі Інтернет: по ряду причин конфіденційні відомості можуть виявитися у відкритому доступі і потрапити на етапі збору в сховище вихідних даних соціального профілю. Або можлива ситуація, коли сукупність зібраних персоналізованих даних, які окремо не є конфіденційними, утворює персональні дані, чиє використання порушує право недоторканності особистого життя досліджуваної персони. Також існує певна ймовірність, що до вибірки підсистеми збору даних потраплять Інтернет-джерела з шкідливим вмістом. Ці проблеми критичні для можливості здійснення процесу побудови соціального профілю і при виникненні вимагають негайного вирішення.

Метою етапу фільтрації є усунення технічних труднощів аналітичної обробки неструктурованих соціальних даних, а також вирішення проблеми забезпечення конфіденційності і безпеки зберігання інформації, що використовується. Наведемо загальний алгоритм фільтрації вихідних даних соціального профілю, який використовується в однойменній підсистемі, структура якої зображена на рис. 4.4.

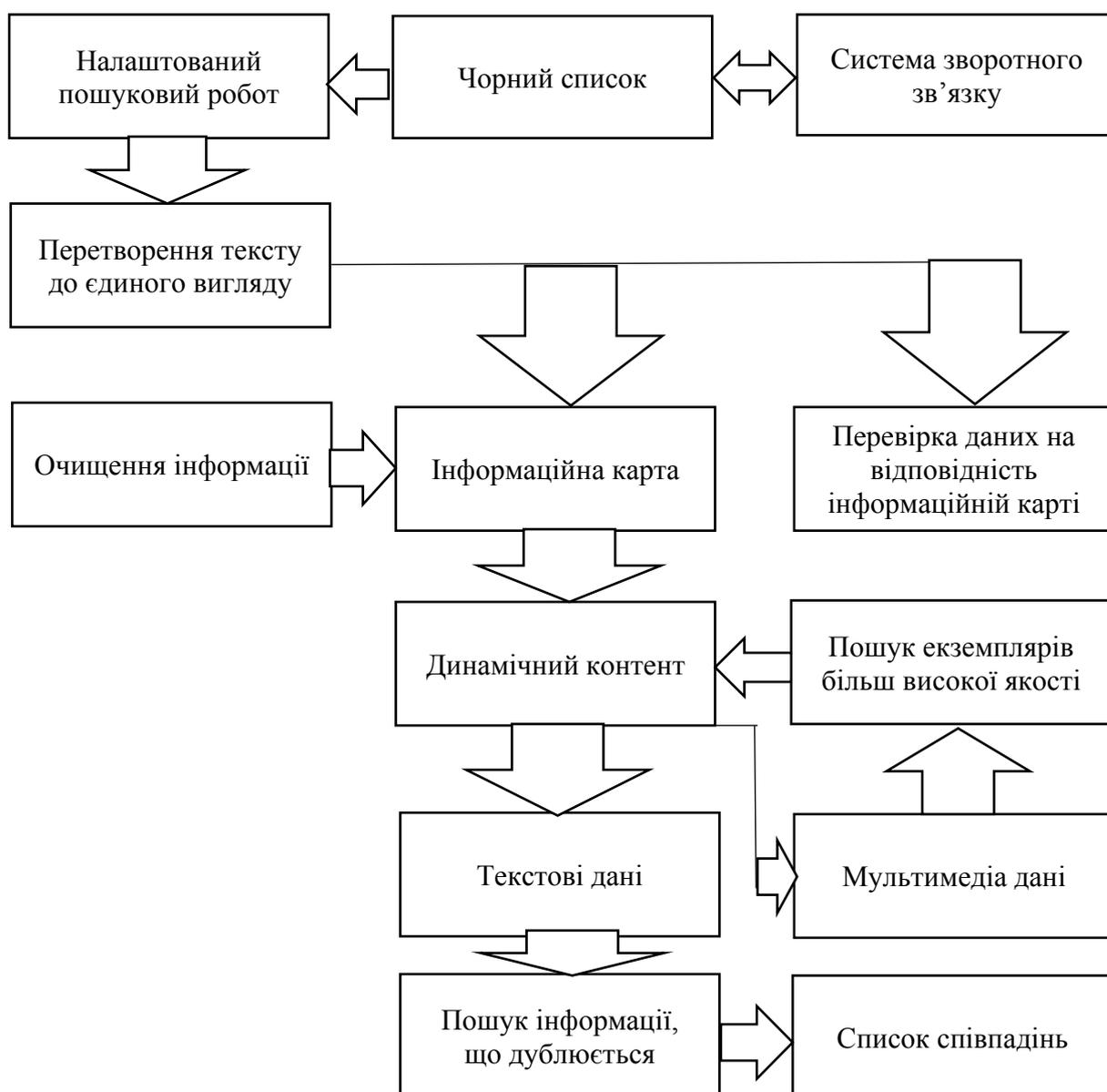


Рис. 4.4. Структурна схема підсистеми фільтрації даних

Джерело: авторська розробка

1. Вирішується проблема конфіденційності та безпеки вихідних даних. Найпростішим способом не допустити потрапляння шкідливої або секретної інформації до процесу соціального профілювання є ведення і поповнення чорного списку джерел, що виключаються з розгляду на етапі збору даних із зазначенням причин заборони збору. Також повинна існувати можливість видалення записів зі сховища вихідних даних. Для захисту персоналізованих даних від несанкціонованого доступу пропонується використовувати

шифрування. Крім цього, важливо організувати зворотний зв'язок з фахівцями, які перевіряють дані на придатність, та видають задокументоване розпорядження на використання тієї чи іншої інформації, для своєчасного редагування критеріїв фільтрації і вже зібраних даних, а також людьми, які безпосередньо мають відношення до інформації, що збирається, для отримання дозволу на обробку їх персональних даних.

2. Зібраний динамічний контент повинен перевірятися на предмет відповідності елементам інформаційної карти. Відсутність збігів у вмісті показує, що зібрана інформація не належать до теми дослідження і повинна бути виключена або інформаційна карта складена некоректно.

3. Універсальне кодування текстової інформації проводиться за допомогою програмного конвертера або в процесі збору, або над готовими записами в сховищі вихідних даних соціального профілю.

4. Очищення текстових даних від символічних конструкцій, що не несуть сенсу, зокрема, таких, як фрагменти HTML-розмітки, що повторюються, знаків пробілу і нового рядка, здійснюється за рахунок орфографічного і синтаксичного аналізу, а також порівнянням зі списком шаблонів сервісних конструкцій.

5. Фільтрація мультимедійного контенту полягає в стиранні сторонньої інформації, такої як рекламні вставки, з використанням засобів нейронних мереж і Big Data.

5. Засобами Big Data в зібраному динамічному контенті проводиться пошук співпадаючих текстів і мультимедійних об'єктів більш високої якості. Після знаходження збігів в спеціальному полі записи динамічного контенту вказується посилання на джерело і процентне співвідношення збігів. Найбільш докладні дані будуть мати пріоритет в подальшому аналізі. Також при необхідності можливий пошук аналогів у ресурсах мережі Інтернет.

Після проведення фільтрації і виявлення аналогів, динамічний контент може аналізуватися комплексними засобами Data Mining, Big Data, візуальної аналітики та машинного навчання для генерації структурованого уявлення

соціальних профілів особистостей або спільнот і цифрового соціального середовища.

Подальший крок методики полягає в аналізі всього різноманіття вихідних гетерогенних даних і побудові соціального профілю у вигляді пов'язаних структурованих текстових елементів. Окремою проблемою при цьому є виявлення неявних семантичних залежностей, неоднозначностей, спотворень й інформаційного сміття у вихідній інформації.

Ранжування вихідної інформації соціального профілю передбачає вилучення її семантики, а також структурування і розподіл за елементами моделі представлення даних соціальних профілів. Однак в даний час не існує єдиного методу вилучення семантики, що застосовується як до неструктурованих текстових, так і мультимедійних об'єктів. Тому в рамках нової методики збору й аналітичної обробки гетерогенних соціальних даних пропонується розробити два алгоритми ранжирування – для тексту та мультимедіа відповідно.

В даний час ефективна обробка неструктурованих текстових даних здійснюється методами інтелектуального аналізу текстів і обробки природної мови. На додаток до них можуть застосовуватися методи визначення емоційного забарвлення текстів для пошуку неявних залежностей і моніторингу громадської думки.

У динамічному контенті соціального профілю неструктуровані текстові дані можуть бути представлені:

- публікаціями користувача соціальних web-ресурсів;
- електронними сторінками цих ресурсів;
- стенограмами;
- мультимедійними даними, що транскрибуються;
- електронними листами і документами.

Вони є джерелом елементів інформаційної карти соціального профілю. Для досягнення цієї мети весь неструктурований текстовий контент

необхідно розбити на прості висловлювання і розподілити їх за об'єктами, зв'язками і характеристикам цифрового соціального середовища.

Алгоритм ранжирування неструктурованого початкового тексту соціального профілю складається з наступних кроків:

- визначення мови вихідного тексту на основі використання словників;
- виявлення в початковому тексті існуючих елементів інформаційної карти;
- первинний лексичний аналіз – складання лінгвістичної мережі, що є основою для подання соціального профілю у вигляді графа;
- вторинний лексичний аналіз – угруповання елементів отриманої лінгвістичної мережі:
 - а) пошук пропозицій зі словесними конструкціями з інформаційної карти (ключовими словами);
 - б) виділення зі знайдених пропозицій слів – кандидатів на ролі соціальних об'єктів і соціальних зв'язків;
 - в) рекурсивний обхід вихідного тексту на задану вручну глибину;
 - г) визначення настроїв соціальних об'єктів у вигляді сумарної тональності їх атрибутів.

Зокрема, первинний лексичний аналіз полягає в простій розбивці вихідного тексту на абстрактні логічні елементи без виділення змістовної частини. Суть цього етапу полягає в знаходженні в текстовому динамічному контенті кандидатів на роль об'єктів і зв'язків цифрового соціального середовища. Уявімо узагальнений алгоритм первинного лексичного аналізу таким чином.

1. Поділ тексту на логічні частини за критеріями: джерело, час отримання даних, переважаюча мова тощо. Поточний крок має на увазі паралельну обробку даних для скорочення часу отримання результатів. У разі неможливості застосування критеріїв, розбивка тексту проводиться за пріоритетом символів:

- знаки кінця пропозиції (точка, три крапки, знаки питання й оклику), символ нового рядка;

- знаки, що розмежовують частини пропозицій (кома, крапка з комою, двокрапка, тире), лапки, дужки, знаки табуляції;

- пробіли;

- усі інші символи.

2. Визначення граматичних основ. Відбувається подальший поділ тексту на окремі словоформи і наступне порівняння частин цих словоформ із еталонними значеннями для визначення частин мови й їх параметрів, таких як: час, особа, відмінок, відмінювання тощо. Знайдені дієслова позначаються як кандидати в елементи соціального профілю категорії «соціальне відношення» і з'єднуються з іменниками, займенниками або числівниками в називному відмінку, які є кандидатами в категорію «об'єкт» соціального профілю.

3. Визначення емоційного забарвлення тексту. Проводиться пошук знаків оклику, псевдографіки (наприклад, текстові смайли, емоджі), порівняння словоформ зі значеннями з тлумачного словника і присвоєння їм настрою (негативний, нейтральний, позитивний).

4. Визначення способу присудків. Для заповнення інформаційної карти в першу чергу розглядаються розповідні висловлювання, тому що вони можуть представляти факти на відміну від питальних і спонукальних. Решта тексту використовується для визначення характеристик об'єктів і настрою досліджуваної особи або групи.

5. Виділення в початковому тексті визначень як характеристик кандидатів в об'єкти соціального середовища, обставин – як характеристик або рідше кандидатів в зв'язки соціального середовища, доповнень – як кандидатів в об'єкти. Це досягається за рахунок встановлення приналежності словоформ, що залишилися, через аналіз морфем.

Для первинного лексичного аналізу потрібна наявність засобів обробки природної мови, заповнених тезаурусів, правил синтаксичного і лексичного

аналізу. Бажана наявність словника для оцінки тональності тексту. З огляду на дуже велику кількість даних, що обробляються, пропонується використовувати рішення на основі технології Big Data, що здійснює:

- створення пошукового індексу;
- виявлення фактів на основі аналізу контенту;
- розробку інтелектуальних фільтрів;
- синтаксичний розбір і аналіз текстового вмісту.

Тоді обробка неструктурованих текстових даних буде складатися з послідовності кроків, які описані нижче.

1. Створюється індексована структура для роботи зі зрізами даних або обмеженими сумами однорідних значень за деякою класифікаційною ознакою.

2. У створену колекцію документів поміщаються неструктуровані текстові дані з динамічного контенту соціального профілю.

3. Підключаються бази даних тезаурусів і синтаксичних правил. Словники правил обробки природної мови доповнюються характерними виразами з текстів вихідних даних.

4. Проводиться пошук статистичної інформації з використанням баз даних словників і запитів за ключовими словами.

Вторинний лексичний аналіз полягає в угрупованні елементів лінгвістичної мережі, отриманої на етапі первинного лексичного аналізу. Існує проблема автоматичної обробки, пов'язана з визначенням семантики для омонімів і словоформ, що визначають одночасно різні об'єкти (імена, аббревіатури, імена загальні, що позначають окремий об'єкт всередині однойменної категорії та інше) або один і той же об'єкт. Для їх обробки потрібна база даних синонімів. Можливим вирішенням цієї проблеми є використання інструментарію глибокого машинного навчання, налаштованого на обробку великих обсягів неструктурованих текстових даних. Остаточне заповнення сховища інформаційної карти соціального профілю здійснюється наступним чином.

1. Проводиться пошук висловлювань з елементами інформаційної карти або заданими вручну ключовими словами, які виступають в ролі підмету або доповнень, якщо вони пов'язані безпосередньо з підметом – іншим соціальним об'єктом. Присудки речень стають зв'язками соціального профілю, а доповнення – новими соціальними об'єктами.

2. Здійснюється повторний обхід тексту, що враховує знайдені на попередньому кроці об'єкти і зв'язки. На глибину обходу вихідних даних не обов'язково встановлювати обмеження, так як передбачається використання спеціалізованих засобів обробки Big Data, стійких до високих навантажень.

3. Визначається настрій соціальних об'єктів як сукупності коефіцієнтів тональностей всіх характеристик і синонімічних значень кожного об'єкта. При створенні соціальних профілів потрібно виділити в окрему характеристику настроїв досліджуваної персони або групи. Дана властивість має сильну залежність від часу і, на відміну від інших соціальних характеристик, таких як хобі, робота, контактна інформація, може змінюватися в дуже короткі проміжки часу. У той же час оцінка настрою є дуже корисною величиною в аналізі соціуму з використанням Big Data, у яких одним з ключових властивостей є швидкість появи й обробки інформації.

Кінцевим результатом етапу вторинного лексичного аналізу є повноцінний соціальний профіль. Залежно від прикладних задач особа, яка приймає рішення, вибирає потрібний їй набір тематичних тегів і відповідні цим тегами соціальні об'єкти, після чого визначає для них рівень деталізації характеристик і околиць. Соціальний профіль можна розширити й актуалізувати за рахунок аналізу мультимедійної інформації.

Обробка мультимедійних даних тим же способом, що і текстових, неможлива. Необхідно використання специфічного інструментарію для виділення семантики з нетекстових даних. В даний час існує достатня кількість готових рішень на основі інтелектуального аналізу даних і машинного навчання для обробки аудіовізуальної інформації. Розробка

нових технологічних рішень аналізу мультимедіа не є метою даного дослідження, тому будуть розглянуті тільки особливості звукового і графічного типів даних по вилученню з них персоналізованої інформації.

В рамках завдання соціального профілювання мультимедійна інформація може розглядатися як:

- дані мультимедіа, Популярні і створювані досліджуваною особою – інформація, яка може говорити про діяльність і переваги персони. Також сюди відноситься і авторський контент. Завдання аналізу полягає в порівнянні мультимедійних об'єктів з існуючими зразками в мережі Інтернет і всередині соціального профілю (наприклад, визначення музичних уподобань за кількома аудіозаписами);

- контент, який містить в собі відомості безпосередньо про розглянуту особу. Метою аналізу є виділення суттєвої інформації з самого мультимедійного об'єкта (наприклад, розпізнавання емоцій на фото, виділення смислових виразів з аудіозаписів).

Перший тип даних аналізується контентним методом – вміст мультимедійної інформації розділяється на складові частини і порівнюється між собою. Для аналізу другої категорії використовується контентно-інтерпретаційний метод, що відрізняється від контентного методу привласненням частинам контенту формалізованих визначень і дослідженням їх взаємозв'язків.

Завдання і методи обробки акустичної та візуальної інформації в ході соціального профілювання принципово різні. Обробка аудіозаписів здійснюється з метою розпізнавання мови й її інтонації. Багато з існуючих систем аналізу аудіозаписів засновані на використанні прихованих Марківських моделей і нейронних мереж. Цілі обробки зображень включають в себе розпізнавання графічних образів і тексту.

В аналізі аудіо досліджуються акустичні властивості людського голосу: спектрально-часові, кепстральні, амплітудно-частотні й інші. Незрозуміла

мова із зазначенням вихідного мультимедійного файлу й її інтонації поміщаються в динамічний контент, після чого аналізується як текст.

Визначення тональності усного мовлення має на увазі:

- запис промови людини;
- поділ аудіозапису на закінчені інтонаційні конструкції;
- визначення тону голосу в кожній з частин;
- побудова класифікатора інтонаційних конструкцій.

Технологія OCR дозволяє розпізнавати на зображенні друкований і в меншому ступені рукописний текст. Як і у випадку з аналізом аудіо, розпізнаний на зображенні текст прив'язується до вихідного об'єкту і надалі розглядається за допомогою засобів текстової аналітики.

Розпізнавання графічних образів в завданні побудови соціального профілю дозволяє виявляти людей і визначати вирази їхніх обличь, а також шукати інші об'єкти соціального середовища. Отримувані результати (імена розпізнаних осіб, їх емоції, навколишні об'єкти) в текстовому вигляді переносяться в інформаційну карту.

Існуючі рішення розпізнавання образів поділяються на методи перебору, штучні нейронні мережі і методи визначення контурів об'єкта з вивченням їх властивостей. Однак, вони не пропонують універсального підходу до обробки зображень. Наприклад, при розпізнаванні емоцій за допомогою нейронних мереж виникають проблеми, пов'язані з визначенням положення особи, відтінку шкіри через висвітлення, нечіткості і сторонніх об'єктів на зображенні. Тому не можна вибрати якийсь певний метод для аналізу великих обсягів графічних даних про соціальне середовище.

Таким чином, ранжування даних з мультимедійного контенту соціального профілю пропонується здійснювати наступним алгоритмом:

- контент розділяється за класами (аудіо, зображення);
- вміст розділяється на призначений для користувача контент і відомості про розглянуті персону або співтовариство;

- для авторського контенту проводиться порівняння мультимедійних об'єктів з існуючими зразками з мережі Інтернет і динамічного контенту соціального профілю, а також витяг сервісної інформації з метою виявлення відомостей про діяльність і переваги досліджуваної особи або соціальної групи;

- для відомостей про розглянуту особу проводиться детальний аналіз вмісту з використанням наявних інструментів аналітики мультимедіа з метою виділення суттєвої інформації про соціальне середовище з самого мультимедійного об'єкта;

- для всіх типів даних проводиться аналіз тональності на основі експертного оцінювання самого контенту і виділених з нього слів;

- результати аналізу заносяться в інформаційну карту із зазначенням посилань на оброблений контент.

Однак результати аналізу неструктурованих даних навіть з використанням спеціалізованих засобів все ще можуть бути неповними: наприклад, коли вихідна інформація містить лінгвістичні помилки, має двоякий або прихований сенс. Тому для отримання найбільш точних результатів рекомендується проведення візуального аналізу даних: особа, що приймає рішення, проводить всередині обробленої текстової інформації з підсвіченими елементами інформаційної карти автоматизований пошук на предмет невиявлених соціальних об'єктів і відносин, після чого нові правила заносяться в тезауруси і починають поширюватися на інші оброблювані дані. Крім інформації, виявленої безпосередньо під час аналізу, можливе знаходження неявних взаємозв'язків на споруджуваному графі соціального профілю.

Заключним етапом процесу соціального профілювання є візуалізація результатів, отриманих на попередніх етапах, у вигляді, придатного для подальшого використання в різних прикладних задачах моніторингу та управління соціальними процесами. Потрібно, щоб підсумкове уявлення соціального профілю було структуроване, компактне, зрозуміле для людської

і машинної обробки, повністю відображало сутність соціальних об'єктів і взаємозв'язків між ними, а також раціонально використовувало ресурси ЕОМ. В рамках розробленої системи побудови соціальних профілів за виконання перерахованих умов відповідає підсистема представлення результатів.

Через специфіку структури інформаційної карти і динамічного контенту соціального профілю, для візуалізації результатів використовується методологія теорії графів і метаграфія. Це дозволяє отримувати наочні для експертної обробки соціальні графи, які можна використовувати в різних прикладних дослідженнях. До результатів, наведених у графовому вигляді, можливе застосування відомих графових алгоритмів, що дозволяють знаходити:

- центральні об'єкти;
- найбільш схожі об'єкти;
- зв'язаність елементів соціального середовища, найближчих сусідів;
- ступінь відмінності поглядів,

а також наочно показувати взаємодії між окремими персонами і соціальними групами в різні моменти часу.

Соціальний граф визначає побудову соціальної мережі, тобто показує відносини конкретної особистості з іншими людьми і спільнотами. Його основною перевагою є наочність представлення пов'язаної між собою інформації про цифрове соціальне середовище і, як наслідок, простота проведення візуального аналізу. Представлення даних у метаграфовому вигляді дозволяє візуалізувати відносини елементів, що належать різним рівням соціального середовища. Так, наприклад, можливо без зміни структури соціальних профілів вивчати розподілені в часі реакції на різні явища як окремих особистостей, так і всіх спільнот, в яких складаються дані особи.

До сукупності вершин і ребер соціального графа, що виражає деяке висловлювання, повинен застосовуватися розроблений критерій значущості

параметрів соціального профілю. Отримані висновки використовуються для пріоритезації видачі результатів соціального профілювання і візуалізації соціального графа таким чином, щоб товщина ребер на зображенні графа була прямо пропорційна отриманим значенням вагових коефіцієнтів. За допомогою критерію значущості також реалізується опція приховування другорядних характеристик і доведення помилкової або неактуальною інформації.

На основі аналізу активності досліджуваної особи можливо перетворення соціального графа в граф інтересів. Його відмінність полягає в тому, що центральними об'єктами в ньому не є особи й співтовариства, а види діяльності, хобі людей. Особливе значення в інтерпретації інформації графа інтересів має тип зв'язку «інтерес – інтерес», що дозволяє знаходити неявні відносини на основі схожих характеристик декількох видів діяльності. Крім критерію значущості для зваженого графа інтересів використовується ваговий коефіцієнт зв'язку інтересів, який показує неявну кореляцію між двома видами діяльності шляхом порівняння характеристик відповідних їм груп персон.

Крім створення інтерактивних зображень соціальних графів можливе проведення вибірок із баз даних соціальних профілів, які дозволяють порівняти їх з еталонним профілем, отримувати з профілів співтовариств відомості про окремі особи і будувати статистичні діаграми.

Висновки до четвертого розділу

1. Запропоновано методику збору та аналізу персоніфікованих даних, яка розглядає сукупність ознак з різних загальнодоступних джерел мережі Інтернет, виділяє з них справжні, несправжні і суперечливі відомості та реалізує моделі опису взаємодій між елементами цифрового соціального середовища.

2. Розроблені інформаційні моделі персонального і групового соціальних портретів дозволяють враховувати всі основні властивості об'єктів, що досліджуються, їх тональність і значимість, а також проводити аналіз по визначенню неявних залежностей за допомогою технології OSINT.

3. Обґрунтовано можливість побудувати орієнтований граф, що є математичною моделлю орієнтованих інформаційних мереж, які досить часто використовуються на практиці та дозволяє вдосконалити функціонування комп'ютерних інформаційних мереж за рахунок оптимізації структур даних шляхом перетворення інформаційного списку в відповідний орієнтований граф.

ВИСНОВКИ

У дисертаційній роботі вирішена науково-практичне завдання розробки моделей та методів диференціації даних консолідованої інформації, на прикладі побудови соціальних портретів, для підвищення оперативності прийняття управлінських рішень.

В результаті чого були отримані такі результати:

1. За аналітичним оглядом наукових публікацій щодо функціонування систем підтримки прийняття управлінських рішень та методів та моделей диференціації даних встановлено, що в теперішній час актуальність придбало завдання по розробці моделей та методів диференціації даних консолідованої інформації, на прикладі побудови соціальних портретів, що дозволяє підвищити оперативність прийняття управлінських рішень.

2. Вперше розроблено інформаційну модель персонального і групового соціальних портретів, на основі OSINT-технології, що дозволило знизити ризики порушення приватності та законів при збереженні повноти та якості вхідних даних. Розроблена модель на основі OSINT-технології дозволила зменшити час на пошук необхідної інформації на 8%.

3. Удосконалено метод управління соціальними процесами, що відрізняється від відомих синтезом загальної системної моделі управління в єдину тріадну ієрархічну систему, що дозволило знизити невизначеність у вирішенні завдань вибору різних варіантів управлінських рішень до 10%.

4. Отримала подальший розвиток модель процесу визначення значущості параметрів соціального профілю, що відрізняється від відомих використанням удосконаленого коефіцієнту значущості, що дозволяє знизити невизначеність вхідних даних ті підвищити оперативність прийняття рішень до 7 % при збереженні рівня об'єктивності оцінювання.

5. Розроблені у роботі методи та моделі є науково-практичною основою для диференціації консолідованої інформації. Розроблені на їх базі алгоритми та програми дозволили зменшити часові витрати на аналіз великих обсягів

даних при збереженні рівня об'єктивності оцінювання порівняно з залученням експертів на 7%. Розроблені моделі обробки Big Data дозволили здійснити перетворення структур великих масивів даних на 5% швидше порівняно з класичними моделями.

6. Результати дисертаційної роботи впроваджені при вирішенні задач управління системою інтернет-продажів та системою дистрибуції:

· модель для подання статистичної та динамічної частин соціального профілю, яка дозволяє проектувати схему інформаційного забезпечення системи соціального профілювання в рамках розробленої методики збору і аналізу гетерогенних соціальних даних із відкритих джерел мережі Інтернет дозволило зменшити час планування та підготовки до 8% (ТОВ «Радіант» м. Харків); модель персонального та групового соціальних портретів, які дозволяють враховувати всі основні властивості досліджуємих об'єктів, їх значимість, а також проводити аналіз по виявленню неявних залежностей, з використанням технології OSINT, це дозволило зменшити витрати обчислювальних ресурсів системи до 5% у розробці плану науково-технічних досліджень в ТОВ «Радіант» м. Харків та ТОВ «Теплоенергосистеми»; метод управління соціальними процесами, що відрізняється від відомих синтезом загальної системної моделі управління в єдину тріадну ієрархічну систему, що дозволило знизити невизначеність у вирішенні завдань вибору різних варіантів управлінських рішень до 10% в ТОВ «Горизонт» та в учбовому процесі кафедри ЕОМ Харківського національного університету радіоелектроніки.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Аверченков В. И. Мониторинг и системный анализ информации в сети Интернет : монография / В. И. Аверченков. – М. : Флинта, 2011. – 160 с.
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. Агальцов, В.П. Базы данных. В 2-х т.Т. 1. Локальные базы данных : учебник / В.П. Агальцов. – М. : ИД ФОРУМ, НИЦ ИНФРА-М, 2013. – 352 с.
4. Базы данных и знаний. Лабораторный практикум для студентов специальности 1-25 01 12 – Экономическая информатика. – Минск : БГУ, 2015. – 34 с.
5. 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.
6. Берко А.Ю. Системы баз даних та знань. Книга 1. Організація баз даних та знань: підручник. – 2-е вид. / А.Ю. Берко, О.М. Верес, В.В. Пасічник. – Вид-во: «Магнолія-2006», 2015. – 440 с.
7. Берко А.Ю. Системы баз даних та знань. Книга 2. Системы управління базами даних та знань: підручник. / А.Ю. Берко, О.М. Верес, В.В. Пасічник. – Вид-во: «Магнолія-2006», 2013. – 680 с.
8. Бодров И. Сильные и слабые стороны NoSQL // корпоративный журнал компании "Инфосистемы Джет". – [Электронный ресурс] – <http://www.jetinfo.ru/stati/silnye-i-slabye-storony-nosql>.
9. Болотова А. К. Психология коммуникации / А. К. Болотова. – М. : Издательский дом Высшей школы экономики», 2015. – 297 с.
10. Бондаренко М. Ф. Информационному обществу – профессионалов в области информации (Business Intelligence и Knowledge Management на службе государству) [Электронный ресурс] / М. Ф. Бондаренко,

С. И. Маторин, Е. А. Соловьева. – Режим доступа : www.it2b.ru/it2b2.view3.page123.html

11. Бородина А.И. Лекция на тему «Модели данных» [Электронный ресурс] / А.И. Бородина. – Режим доступа: http://www.bseu.by/new/tohod/lekcii2_4.htm.

12. 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

13. Буслов П. В. Анализ особенностей застосування моделей комунікацій у соціальних групах та віртуальних співтовариствах // Ukrainian Scientific Journal of Information Security, 2017, vol. 23, issue 1, p. 39-44

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

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

16. Васильченков О. Г., Евсина Н. А., Сальников Д. В., Буслов П. В. Реализация фильтра с постфильтрационным принятием решения на микропроцессорных архитектурах с векторным расширением для обеспечения показателей эффективности судебной экспертизы // Вісник Черкаського державного технологічного університету. №4(17). 2020. С.93–102.

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

18. Влияние через социальные сети / Под общей ред. Е.Г. Алексеевой. – М. : Фонд «ФОКУС-МЕДИА», 2010. – 200 с.

19. Выготский Л. С. Психология искусства / Л. С. Выготский. – Ростов-на-Дону : Феникс, 1998. – 480 с.

20. Габриэлян А.Г. Многомерные модели данных. Презентация [Электронный ресурс] / А.Г. Габриэлян. – Режим доступа: <http://www.slideshare.net/asheg/ss-7700975>

21. Гаврилова Т.А., Хорошевский Т.А. Базы знаний интеллектуальных систем. СПб : Питер, 2000. – 384 с.
22. Гагарина Л. Г., Киселев Д. В., Федотова Е. Л. Разработка и эксплуатация автоматизированных информационных систем : учеб. пособие / Под ред. Л.Г. Гагариной. М : ИНФРА-М, 2007. – 384 с.
23. Гайна Г.А. Основи проектування баз даних. Навч. посібник. - К. : Кондор, 2007. – 208 с.
24. Гандерлой М. Автоматизация Microsoft Access с помощью VBA = Mike Gunderloy, Susan Sales Harkins; Automating Microsoft Access with VBA / Пер. с англ. С.А. Храмова. М, С.Пб., К : Вильямс, 2006. – 416 с.
25. Гарсиа-Молина Г. Системы баз данных. Полный курс / Г. Гарсиа-Молина, Дж. Ульман, Дж. Уидом. - М.: «Вильямс», 2003. – 1088 с.
26. Голицына О.Л. Базы данных : учебное пособие / О.Л. Голицына, Н.В. Максимов, И.И. Попов. – М. : Форум, 2012. – 400 с.
27. Грабер М. Введение в SQL / Мартин Грабер. – М.: Лори, 2010. – 227 с.
28. Грядовой В. И. Социальное управление. Теория, методология, практика / В. И. Грядовой, В. Я. Кикоть. – М. : Юнити-Дана, 2012. – 312 с.
29. Гудов А.М. Базы данных и системы управления базами данных. Программирование на языке PL/SQL : учебно-методическое пособие [Электронный ресурс] // А.М. Гудов, С.Ю Завозкин., Т.С. Рейн. – КемГУ. – 2009. – Режим доступа: http://unesco.kemsu.ru/study_work/method/bd/umk/book/gl_6.html.
30. Гужва В.М. Інформаційні системи і технології на підприємствах : навч. посіб. / В.М. Гужва. - К. : КНЕУ, 2001. – 400 с.
31. Дейт К. Дж. Введение в системы баз данных, 8-е издание / К. Дж. Дейт. – М.: «Вильямс», 2005. – 1328 с.
32. Додаткові матеріали та програмні засоби з розробки інтелектуальних систем [Електронний ресурс]. – Режим доступу : <http://www.pearsoneduc.com/computing>
33. Завадський І.О. Основи баз даних : навч. посіб. / І.О. Завадський. – К. : Видавець І.О. Завадський, 2011. – 192 с.

34. Зарицька О.Л. Бази даних та інформаційні системи : метод. пос. / О.Л. Зарицька. – Житомир : Вид-во ЖДУ ім. І. Франка, 2009. – 132 с.
35. Зубов В.С. Справочник программиста. Базовые методы решения графовых задач и сортировки. – М.: ИИД Дом «Филинь». – 1999. – 256 с.
36. Исаев Д. В. Автоматизированные системы формирования консолидированной финансовой отчетности : учебное пособие / Д. В. Исаев, Т. К. Кравченко. – М. : Государственный университет – Высшая школа экономики, 2006. – 336 с.
37. Исаченко А.Н. Модели данных и СУБД : уч. пос. / А.Н. Исаченко, С.П. Бондаренко. – Минск : Изд-во БГУ, 2007. – 205 с.
38. Каймакова М. В. Коммуникации в организации : текст лекций / М. В. Каймакова. – Ульяновск : Ульяновский государственный технический университет, 2008. – 73 с.
39. Калиниченко Л. А. Методы и средства интеграции неоднородных баз данных / Калиниченко Л.А. – Москва: Наука, 1983. – 420 с.
40. Калитич Г. І. Консолідація інформації, знань і мудрості як проектування і основа гармонійного поступу України / Г. І. Калити // НТІ. – 2008. – № 1. – С. 51.
41. Карпова, И.П. Базы данных : учебное пособие / И.П. Карпова. – СПб. : Питер, 2013. – 240 с.
42. Каучмен Джейсон С., Марисети Судхир Н. OCA Oracl 9x Associate DBA. Подготовка администраторов баз данных : пер. с англ. М. : Лори, 2006. – 680 с.
43. Кенинг О. Введение в групповую динамику / О. Кениг, К. Шаттенхофер. – М: ИКСР, 2014. – 176 с.
44. 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>. – Назва з екрана.
45. Кириллов, В.В. Введение в реляционные базы данных. Введение в реляционные базы данных / В.В. Кириллов, Г.Ю. Громов. – СПб. : БХВ-Петербург, 2012. – 464 с.

46. Колисниченко Д. Н. Поисковые системы и продвижение сайтов в Интернете / Д. Н. Колисниченко. – М. : Диалектика ; Вильямс, 2007. – 272 с.
47. Коломинский Я. Л. Социальная психология взаимоотношений в малых группах / Я. Л. Коломинский. – М. : АСТ, 2010. – 446 с.
48. Коннолли Т. Базы данных. Проектирование, реализация и сопровождение. Теория и практика. 3-е издание / Т. Коннолли, К. Бегг. – М. : «Вильямс», 2003. – 1440 с.
49. Кошелев В.Е. Базы данных в ACCESS 2007: Эффективное использование / В.Е. Кошелев. – М. : Бином-Пресс, 2009. – 592 с.
50. Кричевский Р. Л. Социальная психология малой группы : Учебное пособие для вузов / Р. Л. Кричевский, Е. М. Дубовская. – М. : Аспект Пресс, 2001. – 318 с.
51. Кузин А.В. Базы данных : учебное пособие для студ. высш. учеб. заведений / А.В. Кузин, С.В. Левонисова. – М.: ИЦ Академия, 2012. – 320 с.
52. Кунанець Н. Е. Консолідація інформації та інформаційна безпека / Н. Е. Кунанець, В. В. Пасічник // III Міжнародна науково-практична конференція «Інформація та економічна безпека (INFECO 2010)» : тези доповідей. – Х. : ДВНЗ УБС Харківський навчально-науковий інститут, 2010. – С. 131–133.
53. Лукьянова А. А. Концепция управления знаниями в контексте управления человеческими ресурсами / А. А. Лукьянова // Вестник Омского университета. – 2011. – № 2. – С. 23–29. – (Серия «Экономика»).
54. Люгер Джордж Ф. Искусственный интеллект: стратегии и методы решения сложных проблем : пер. с англ. / Джордж Ф. Люгер. - М. : Вильямс, 2003. – 864 с.
55. Макарова М. В., Карнаухова Г. В., Запара С. В. Информатика та комп'ютерна техніка : навч. посіб. / за ред. М. В. Макарової. 3-тє вид., переробл. і допов. Суми : Університетська книга, 2008. – 665 с
56. Маклаков С.В. ВРwin и ERwin: CASE-средства для разработки информационных систем / С.В. Маклаков. - М. : Диалог-Мифи, 2000. – 295 с.
57. Можаяев М. О., Буслов П. В., Мелашенко О. П. Діагностика функціонування розподіленої інформаційної системи судової експертизи./

Застосування інформаційних технологій у діяльності правоохоронних органів : зб. матеріалів круглого столу (09 грудня 2020 р., м. Харків) / МВС України, Харк. нац. ун-т внутр. справ. – Харків : ХНУВС, 2020. – С.80-82.

58. Можаяєв О.О., Буслів П. В. Сучасний стан рекрутингу та його використання для потреб інформаційної системи Національної поліції України /Шлях успіху і перспективи розвитку (до 26 річниці заснування Харківського національного університету внутрішніх справ) : матеріали міжнар. наук.-практ. конф. (м. Харків, 20 листоп. 2020 р.) / редкол.: Д. В. Швець (голова), О. М. Бандурка, С. М. Гусаров та ін.; МВС України, Харків. нац. ун-т внутр. справ. Харків : ХНУВС, 2020. С. 345-347.

59. Можаяєв М. О., Буслів 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. 18
Мильнер Б. З. Управление знаниями в корпорациях : учеб. пособие / Б. З. Мильнер, З. П. Румянцева, В. Г. Смирнова и др. – М. : Дело, 2006. – 304 с.

60. Нонака И. Компания – создатель знания. Зарождение и развитие инноваций в японских фирмах / И. Нонака, Х. Такеучи. – М. : Олимп – Бизнес, 2011. – 384 с.

61. Оре О. Теория графов. – М.: Наука, 1968. – 336 с.

62. Омельченко Л. Visual FoxPro 7.0.- СПб. : БХВ, 2002. – 672 с.

63. Палагин А. В., Кривой С. Л., Петренко Н. Г. Концептуальные графы и семантические сети в системах обработки естественно-языковой информации. Математические машины и системы. 2009. Т. 1. №. 3. URL : <https://cyberleninka.ru/article/n/kontseptualnye-grafy-i-semanticheskie-setiv-sistemah-obrabotki-estestvenno-yazykovoy-informatsii>

64. Парабеллум А. Социальные сети. Источники новых клиентов для бизнеса / А. Парабеллум, Н. Мрочковский, В. Калаев. – М. : Огни–Москва, 2013. – 176 с.

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

66. Пасічник В. В., Шаховська Н. Б. Сховища даних : навч. посіб. / за ред. В. В. Пасічника. Львів : Магнолія 2006, 2008. – 492 с.
67. Пирогов В. Ю. Информационные системы и базы данных: организация и проектирование : учебное пособие / В.Ю. Пирогов. – СПб. : БХВ-Петербург, 2009. – 528 с.
68. Пэддок Р., Петерсон Д. Visual FoxPro 6. Разработка корпоративных приложений. – М. : ДМК, 1999. – 552 с.
69. Роб П. Системы баз данных: проектирование, реализация и управление / П. Роб, К. Коронел. – СПб.: БХВ-Петербург, 2004. – 1040 с.
70. Романов А. А. Массовые коммуникации : Учебно-практическое пособие / А. А. Романов. – М. : Евразийский открытый институт, 2010. – 175 с.
71. Романенко В. Н. Основы использования Интернета во внешней экономической деятельности / В. Н. Романенко, Г. В. Никитина, В. В. Корец. – М. : ИВЭСЭП, 2010. – 160 с.
72. Романюк О.Н. Організація баз даних та знань / О.Н. Романюк, Т.О. Савчук. – Вінниця: ВДТУ, 2001.
73. Руденко В. Д. Базы данных в информационных системах : навч. посібник / за заг. ред. В. Ю. Бикова. К. : Фенікс, 2010. – 240 с.
74. Свами М., Тхуласираман К. Графы, сети и алгоритмы. – М. : Мир, 1984. – 455 с.
75. Семечкин Н. И. Психология социальных групп : Учебное пособие для ВУЗов / Н. И. Семечкин. – М. : Владос, 2011. – 287 с.
76. 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>.
77. Ситник Н. В. Проектування баз і сховищ даних : навч. посібн. – К. : КНЕУ, 2004. – 348 с.
78. Советов Б.Я. Базы данных: теория и практика : учебник для бакалавров / Б.Я. Советов, В.В. Цехановский, В.Д. Чертовской. - М. : Юрайт, 2013. – 463 с.

79. Соловьева Е.А. Естественная классификация: системологические основания / Е. А. Соловьева. – Харьков: ХНУРЕ, 1999. – 222 с.
80. Страхарчук А.Я. Реляційна модель даних, Нормалізація даних, Інформаційні системи і технології в банках [Електронний ресурс] / А.Я. Страхарчук, В.П. Страхарчук // Бібліотека українських підручників. – Режим доступу: http://libfree.com/134926096_bankivska_spravarelyatsiyana_model_daniv.html
81. СУБД MySQL [Електронний ресурс]. – Режим доступу : www.mysql.com 19.
82. СУБД PostgreSQL [Електронний ресурс]. – Режим доступу : www.postgresql.org
83. Теорія розробки інтелектуальних агентів [Електронний ресурс]. – Режим доступу : <http://www.williamspublishing.com>
84. Уорнер М. Виртуальные организации. Новые формы ведения бизнеса в XXI веке / М. Уорнер, М. Витцель. – М. : Добрая книга, 2005. – 296 с.
85. Ульман Л. MySQL / Лари Ульман. – СПб.: Питер, 2004. – 352 с.
86. Управление знаниями. Как превратить знания в капитал / А. Л. Гапоненко, Т. М. Орлова. – М. : Эксмо, 2008. – 400 с.
87. Федорова Д.Э. CASE-технологии / Д.Э. Федорова, Ю.Д. Семенов, К.Н. Чижик. - М. : Горячая линия Телеком, Радио и связь, 2005. – 160 с.
88. Фетискин Н. П. Социально-психологическая диагностика развития личности и малых групп / Н. П. Фетискин, В. В. Козлов, Г. М. Мануйлов. – М. : Изд-во Института Психотерапии, 2009. – 544 с.
89. Фролов С.С. Социология : Учебник для высших учебных заведений. –М. : Наука, 1994. – 256 с.
90. Хернандес М. Д. SQL-запросы для простых смертных = SQL Queries for Mere Mortals : Практическое руководство по манипулированию данными в SQL / М.Дж. Хернандес, Дж.Л Вьескас. – Москва : Лори, 2003. – 458 с.

91. Хомоненко А.Д. Базы данных: Учебник для высших учебных заведений / А.Д. Хомоненко, В.М. Цыганков, М.Г.Мальцев. – СПб.: КОРОНА принт, 2004. – 736 с.

92. Целых А. А., Дедюлина М. А. Теоретико-графовые подходы к моделированию актор-сетей в исследованиях науки и технологий. Моделирование, оптимизация и информационные технологии. 2018. Т. 6. №. 4. С. 244-259.

93. Целых Ю. А. Теоретико-графовые методы анализа нечетких социальных сетей. Программные продукты и системы. 2008. №. 2. С.48-50. URL : <https://cyberleninka.ru/article/v/teoretiko-grafovye-metodyanaliza-nechetkih-sotsialnyh-setey>

94. Чередниченко Ю. Маркетинг в Интернете. Сайт, который зарабатывает / Ю. Чередниченко. – М. : Гостехиздат, 2013. – 176 с.

95. Широков Л.А. Базы данных и знаний: Учебное пособие. Ч. 1. / Л.А.Широков. – М. : МГИУ, 2000. – 86 с.

96. Энсор Дейв. Oracle. Проектирование баз данных : пер. с англ. / Дейв Энсор, Йен Стивенсон. - К. : Издат. группа BHV, 2000. - 560 с.

97. Эрик Ландквист. Пять главных трендов в области больших данных [Электронный ресурс] / Эрик Ландквист. – Режим доступа: <http://www.pcweek.ua/themes/detail.php?ID=144474>

98. Яшин Б. Л. Культура общения: теория и практика коммуникаций : Учебное пособие. – М. : Директ-медиа, 215. – 243 с.

99. Apache HBase™ Reference Guide – [Электронный ресурс] – <http://hbase.apache.org/book.html#faq>.

100. 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.

101. Buslov P. Analysis of using the features of communication models in social groups and virtual communities // Ukrainian Scientific Journal of Information Security, 2017, vol. 23, issue 1, p. 39-44.

102. Cherneva G. Fractal Models for Approximation of Random Processes. Proceedings of the Technical University of Sofia. 2017. Vol. 67, Issue 2, P.171-176.

103. Cherneva G. Fractal Models for Approximation of Random Processes. Proceedings of the Technical University of Sofia. 2017. Vol. 67, Issue 2, P.171-176.

104. Doug Laney. 3D Data Management: Controlling Data Volume, Velocity, and Variety // Meta Group. - [Электронный ресурс] - <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.

105. Graph-based intelligence analysis. URL : <https://linkurio.us/blog/graph-based-intelligence-analysis>

106. Hows D., Membrey P., Plugge E., Hawkins T. The Definitive Guide to MongoDB: A complete guide to dealing with Big Data using MongoDB, Third Edition. – Apress, 2015. – 376 с.

107. Introduction to Views // Apache CouchDB 2.2 Documentation. – [Электронный ресурс] – <http://docs.couchdb.org/en/stable/ddocs/views/intro.html>.

108. 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. - June, 2011. pp. 27-31. - https://www.webcitation.org/6ComRQdfl?url=http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20InnovationBig%20Data/MGI_big_data_full_report.ashx.

109. 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.

110. 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.

111. 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.

112. Mozhaiev M., Buslov P. Method of modeling of a social profile using big data structure transformation optimization // *Сучасні інформаційні системи* 2021. Т.5 №1 С. 12–17.

113. Neil Biehn. The Missing V's in Big Data: Viability and Value // *Wired*. - [Электронный ресурс] - <https://www.wired.com/insights/2013/05/the-missing-vs-in-big-data-viability-and-value/>.

114. Slavov D. Programming Languages for Artificial Intelligence. *Proceedings of the Technical University of Sofia*. 2019/ Vol. 69, Issue 2, 2019, P.43-52.

115. Tiwari G. Chapter 1: NoSQL: What It Is and Why You Need it > Definition and Introduction // *Professional NoSQL*. – Packt Publishing, 2011. – 384p.

116. Robinson I., Webber J., Eifrem E. *Graph Databases*. – O'Reilly Media, 147 2013. – 178 p.

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

ДОДАТКИ

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇНаукові праці, в яких опубліковані основні наукові результати дисертації:

1. Васильченко О. Г., Евсина Н. А., Сальников Д. В., Буслов П. В. Реализация фильтра с постфильтрационным принятием решения на микропроцессорных архитектурах с векторным расширением для обеспечения показателей эффективности судебной экспертизы // Вісник Черкаського державного технологічного університету. №4(17). 2020. С.93-102.
2. 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.
3. 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.
4. 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.
5. Mozhaiev M., Buslov P. Method of modeling of a social profile using big data structure transformation optimization // Сучасні інформаційні системи 2021. Т.5 №1. С. 12-17.
6. Буслов П. В. Розробка класифікації систем консолідації комерційної інформації//Вчені записки Таврійського національного університету ім. В.І. Вернадського, серія: Технічні науки, 2018, Т.29 (68)№3. С. 105-108

Наукові праці, які засвідчують апробацію матеріалів дисертації:

7. Pavlo Buslov, Viktoria Shvedun, Volodymyr Streltsov Modern tendencies of data protection in the corporate systems of information consolidation// Problems of Infocommunications. Science and Technology: 5th International Scientific-Practical Conference. Kharkiv, Ukraine, 2018 P. 65-67 – *заочна участь*

8. Можаяєв М. О., Буслов П. В., Мелашенко О. П. Діагностика функціонування розподіленої інформаційної системи судової експертизи./ Застосування інформаційних технологій у діяльності правоохоронних органів : зб. матеріалів круглого столу (09 грудня 2020 р., м. Харків) / МВС України, Харк. нац. ун-т внутр. справ. – Харків : ХНУВС, 2020. – С.80-82. – **заочна участь**

9. Можаяєв О.О., Буслов П. В. Сучасний стан рекрутингу та його використання для потреб інформаційної системи Національної поліції України /Шлях успіху і перспективи розвитку (до 26 річниці заснування Харківського національного університету внутрішніх справ) : матеріали міжнар. наук.-практ. конф. (м. Харків, 20 листоп. 2020 р.) / редкол.: Д. В. Швець (голова), О. М. Бандурка, С. М. Гусаров та ін.; МВС України, Харків. нац. ун-т внутр. справ. Харків : ХНУВС, 2020. С. 345-347. – **заочна участь**

10. Можаяєв 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. 18. – **заочна участь**

11. Буслов П.В. Розробка класифікації систем консолідації інформації в комерційній діяльності підприємств/Актуальні питання забезпечення кібербезпеки та захисту інформації: тези доповідей учасників IV міжнародної науково-практичної конференції (21-24 лютого 2018 р.). Київ. Європейський Університет.2018.- С.29. – **заочна участь**

12. Буслов П.В. Дослідження сучасного стану та перспектив розвитку систем консолідації інформації/ Modern methods, innovations, and experience of practical application in the field of technical sciences: conference proceeding (27-28 december 2017) Radom, Republic of Poland: Radom Academy of economics. 2017. P. 12-15. – **заочна участь**

«ЗАТВЕРДЖУЮ»

Перший проректор
Харківського національного
університету радіоелектроніки

І.В. Рубан

« 26 » 02 2021 р.



АКТ

про впровадження в освітній процес результатів
дисертаційної роботи на тему:
«Моделі та методи диференціації даних консолідованої інформації для
систем підтримки рішень»
здобувача кафедри електронних обчислювальних машин
Харківського національного університету радіоелектроніки
Буслова Павла Володимировича

Комісія у складі:

Голови: начальника навчального відділу Харківського національного університету радіоелектроніки, к.т.н., доц. Міхнкової А.В.

Членів комісії: голови методичної комісії факультету Комп'ютерної інженерії та управління, к.т.н., доц. Філіппенко І.В., завідувача кафедри електронних обчислювальних машин, д.т.н., проф. Коваленка А.А., професора кафедри електронних обчислювальних машин, д.т.н., доц. Фесенко Т.Г.

встановила, що результати наукових досліджень впроваджено в освітньому процесі Харківського національного університету радіоелектроніки на кафедрі електронних обчислювальних машин (протокол засідання кафедри електронних обчислювальних машин №13 від 26.02.2021 р.).

Розглянувши матеріали роботи та організацію освітнього процесу на кафедрі, комісія відзначає, що при проведенні лекційних занять та лабораторних робіт з дисципліни «Методи та алгоритми управління ІТ-проектами», яка викладається для студентів другого рівня вищої освіти за освітньо-науковою програмою «Системне програмування» за спеціальністю 123 – «Комп'ютерна інженерія», використані наступні наукові результати дисертаційної роботи:

- модель для подання статистичної та динамічної частин соціального профілю, яка дозволяє проектувати схему інформаційного забезпечення системи соціального профілювання в рамках розробленої методики збору і аналізу гетерогенних соціальних даних із відкритих джерел мережі Інтернет;

- метод управління соціальними процесами, що відрізняється від відомих синтезом загальної системної моделі управління в єдину тріадну ієрархічну систему.

Впровадження зазначених наукових результатів дозволило підвищити якість навчального контенту з дисципліни «Методи та алгоритми управління ІТ-проектами».

Начальник навчального відділу



А.В. Міхнова

Голова методичної комісії факультету
Комп'ютерної інженерії та управління



І.В. Філіппенко

Завідувач кафедри електронних
обчислювальних машин



А.А. Коваленко

Професор кафедри електронних
обчислювальних машин



Т.Г. Фесенко

ЗАТВЕРДЖУЮ:
 Директор ТОВ «Теплоенергосистеми»
В.Ю. Дацюк
 « / / » травня 2020



АКТ
впровадження результатів наукових досліджень
Буслова Павла Володимировича,
отриманих під час виконання дисертаційної роботи

Комісія ТОВ «Теплоенергосистеми» у складі:

Голова комісії Чистякова Олена В'ячеславівна — юрисконсульт;

Члени комісії:

- Ткаченко Анна Володимирівна — інженер-проектувальник;

- Трубін Віктор Миколайович — слюсар з ремонту та обслуговуванню газового обладнання

констатують, що в управлінні системою інтернет-продажів та системою дистрибуції в цілому використано результати дисертаційних досліджень, отриманих Бусловим П.В., а саме:

- нечіткі критерії для визначення значущості складових інформації соціального профілю, які враховують природу й основні характеристики зібраних відомостей та спрощують виявлення явних і неявних зв'язків між елементами соціального середовища, що дозволило зменшити час планування та підготовки до 8%;

- модель персонального та групового соціальних портретів, які дозволяють враховувати всі основні властивості досліджуваних об'єктів, їх значимість, а також проводити аналіз по виявленню неявних залежностей, з використанням технології OSINT, що дозволило зменшити витрати обчислювальних ресурсів системи до 5%.

В результаті впровадження наукових положень та результатів дисертації Буслова П.В. отримано науково-технічний ефект, пов'язаний з підвищенням ступеня автоматизації та інтелектуалізації вирішення задач управління процесами проектування складних систем.

Застосування зазначених результатів дозволило зменшити витрати обчислювального ресурсу комп'ютерної системи до 8% та зменшити час планування та підготовки до 5%.

Даний акт впровадження не є підставою для пред'явлення претензій, пов'язаних з авторськими правами та отриманням будь-якої винагороди.

Акт складений для подання до спеціалізованої вченої ради у зв'язку із захистом дисертації Буслова П.В.



Голова комісії:

Члени комісії:

Чистякова О.В.

Ткаченко А.В.

Трубін В.М.

ЗАТВЕРДЖУЮ:
 Директор ТОВ «Горизонт»
В.М. Судаков
 «21» січня 2021 року



АКТ
впровадження результатів наукових досліджень
Буслова Павла Володимировича,
отриманих під час виконання дисертаційної роботи

Комісія ТОВ «Горизонт» у складі:

Голова комісії Чертков Юрій Миколайович — головний енергетик;

Члени комісії:

- Геращенко Галина Вікторівна — головний бухгалтер;

- Бабенко Ольга Яківна — керівник економічного відділу

констатують, що в управлінні системою інтернет-продажів та системою дистрибуції в цілому використано результати дисертаційних досліджень, отриманих Бусловим П.В., а саме:

- метод моделювання соціального профілю з використанням оптимізації перетворення структур Big Data, який, на відміну від існуючих, дозволяє оптимізувати структуру даних шляхом перетворення інформаційного списку у відповідний орієнтований граф та сприяє удосконаленню функціонування комп'ютерних інформаційних мереж, це дозволило зменшити витрати обчислювальних ресурсів системи до 7%;

- нечіткі критерії для визначення значущості складових інформації соціального профілю дозволили зменшити час планування та підготовки до 9%.

В результаті впровадження наукових положень та результатів дисертації Буслова П.В. отримано науково-технічний ефект, пов'язаний з підвищенням ступеня автоматизації та інтелектуалізації вирішення задач управління процесами проектування складних систем.

Застосування зазначених результатів дозволило зменшити витрати обчислювального ресурсу комп'ютерної системи до 7% та зменшити час планування та підготовки до 9%.

Даний акт впровадження не є підставою для пред'явлення претензій, пов'язаних з авторськими правами та отриманням будь-якої винагороди.

Акт складений для подання до спеціалізованої вченої ради у зв'язку із захистом дисертації Буслова П.В.

Голова комісії:

Члени комісії:

Чертков Ю.М.

Геращенко Г.В.

Бабенко О.Я.

ЗАТВЕРДЖУЮ:

Директор ТОВ «Радіант»

О.О. Цепковський

«16» грудня 2020



АКТ
впровадження результатів наукових досліджень
Буслова Павла Володимировича,
отриманих під час виконання дисертаційної роботи

Комісія ТОВ «Радіант» у складі:

Голова комісії Борисенко Олександр Сергійович — комерційний директор;

Члени комісії:

- Скоріков Віталій Сергійович — керівник відділу логістики;

- Андрієць Валерія Володимирівна — менеджер з дослідження товарного ринку

констатують, що в управлінні системою інтернет-продажів та системою дистрибуції в цілому використано результати дисертаційних досліджень, отриманих Бусловим П.В., а саме:

- модель для подання статистичної та динамічної частин соціального профілю, яка дозволяє проектувати схему інформаційного забезпечення системи соціального профілювання в рамках розробленої методики збору и аналізу гетерогенних соціальних даних із відкритих джерел мережі Інтернет дозволила зменшити час планування та підготовки до 8%;

- модель персонального та групового соціальних портретів, які дозволяють враховувати всі основні властивості об'єктів, що досліджуються, їх значимість, а також проводити аналіз по виявленню неявних залежностей, з використанням технології OSINT, дозволила зменшити витрати обчислювальних ресурсів системи до 5%.

В результаті впровадження наукових положень та результатів дисертації Буслова П.В. отримано науково-технічний ефект, пов'язаний з підвищенням ступеня автоматизації та інтелектуалізації вирішення задач управління процесами проектування складних систем.

Застосування зазначених результатів дозволило зменшити витрати обчислювального ресурсу комп'ютерної системи до 5% та зменшити час планування та підготовки до 8%.

Даний акт впровадження не є підставою для пред'явлення претензій, пов'язаних з авторськими правами та отриманням будь-якої винагороди.

Акт складений для подання до спеціалізованої вченої ради у зв'язку із захистом дисертації Буслова П.В.

Голова комісії
Члени комісії:



Борисенко О.С.

Скоріков В.С.

Андрієць В.В.