

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
БУДІВНИЦТВА І АРХІТЕКТУРИ**

автоматизації і інформаційних технологій

(факультет)

інформаційних технологій проєктування та прикладної математики

(кафедра)

**ПОЯСНЮВАЛЬНА ЗАПИСКА
ДО КВАЛІФІКАЦІЙНОЇ РОБОТИ
НА ЗДОБУТТЯ ОСВІТНЬОГО РІВНЯ «МАГІСТР»**

на тему: «Розробка інформаційної системи прогнозування аварійних ситуацій
на водопровідних мережах міста»

ФРАНЧУК ВЛАДИСЛАВ ОЛЕГОВИЧ

(прізвище, ім'я та по батькові студента повністю)

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
БУДІВНИЦТВА І АРХІТЕКТУРИ**

автоматизації і інформаційних технологій

(факультет)

інформаційних технологій проєктування та прикладної математики

(кафедра)

ЗАТВЕРДЖУЮ

Завідувач кафедри ІТППМ
д.т.н., професор Бородавка Є.В.

„___” _____ 2025 року

**ПОЯСНЮВАЛЬНА ЗАПИСКА
ДО КВАЛІФІКАЦІЙНОЇ РОБОТИ
НА ЗДОБУТТЯ ОСВІТНЬОГО РІВНЯ «МАГІСТР»**

на тему: "Розробка інформаційної системи прогнозування аварійних ситуацій
на водопровідних мережах міста"

Виконав: студент II-го курсу, групи ІСТм-24

Спеціальності: 126 «Інформаційні системи та
технології»

(шифр і назва спеціальності)

Франчук В.О.

(прізвище та ініціали)

Керівник к.т.н., проф. Теренчук С.А.

(прізвище та ініціали)

Рецензент к.т.н., доц. Шабала Є.Є.

(прізвище та ініціали)

Київ, 2025 р.

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
БУДІВНИЦТВА І АРХІТЕКТУРИ**

Факультет: автоматизації і інформаційних технологій

Кафедра: інформаційних технологій проектування та ПМ

Освітній рівень: «магістр за ОПІ»

Спеціальність: 126. «Інформаційні системи та технології»

ЗАТВЕРДЖУЮ

Завідувач кафедри ІТПМ
д.т.н., професор Бородавка Є.В.

„___” _____ 2025 року

**З А В Д А Н Н Я
ДО ВИКОНАННЯ КВАЛІФІКАЦІЙНОЇ РОБОТИ
НА ЗДОБУТТЯ ОСВІТНЬОГО РІВНЯ «МАГІСТР»**

Франчук Владислав Олегович

1. Тема роботи: Розробка інформаційної системи прогнозування аварійних ситуацій на водопровідних мережах міста
затверджена наказом ректора КНУБА № 1905/23/25 від «12» листопада 2025 р.
2. Керівник роботи: Теренчук С.А., к.т.н, професор
кафедри інформаційних технологій проектування і прикладної математики
3. Строк подання студентом роботи до захисту: грудень 202 рік
4. Зміст пояснювальної записки за розділами:
 - Р.1. Аналіз предметної області та постановка задачі
 - Р.2. Концептуальне та функціональне призначення
 - Р.3. Математичне забезпечення інформаційної технології
 - Р.4. Розробка інформаційного забезпечення
 - Р.5. Тестовий приклад програми
5. Інформаційні слайди:
 - С.1 Дерево цілей системи
 - С.2. Декомпозиція моделі системи
 - С.3. Концептуальна модель бази даних
 - С.3. Даталогічна модель бази даних
 - С.4. Програмне забезпечення системи. Тестовий приклад програми

6. Календарний план виконання кваліфікаційної роботи

Види робіт та їх зміст	Дата виконання
Р. 1. Аналіз предметної області та постановка задачі	Листопад 2025 р.
Р. 2. Концептуальне та функціональне призначення	Листопад 2025 р.
Р. 3. Математичне забезпечення інформаційної технології	Листопад 2025 р.
Р. 4. Розробка інформаційного забезпечення системи	Листопад 2025 р.
Р. 5. Тестовий приклад програми	Листопад 2025 р.
Остаточне оформлення роботи	Грудень 2025 р.
Направлення роботи на рецензування, перевірку на плагіат	Грудень 2025 р.
Попередній захист роботи на кафедрі	Грудень 2025 р.

7. Консультанти розділів атестаційної випускної роботи

Розділ	Прізвище, ініціали та посада консультанта	Перевірів	
		дата	підпис
Розділ 1.	к.т.н., доц. Баліна О.І.	05.12.2025 р.	
Розділ 2.			
Розділ 3.			
Розділ 4.			
Розділ 5.			

8. Дата видачі завдання: 22 вересня 2025 року

Керівник

(підпис)

Теренчук С.А.

(прізвище та ініціали)

Магістрант

(підпис)

Франчук В.О.

(прізвище та ініціали)

РЕЗЮМЕ

Київський національний університет будівництва і архітектури

Франчук Владислав Олегович

факультет автоматизації і інформаційних технологій,

група ІСТм-II

Тема кваліфікаційної роботи:

Розробка інформаційної системи прогнозування аварійних ситуацій на

водопровідних мережах міста

освітній рівень: магістр,

спеціальність: 126 «Інформаційні системи і технології»,

Науковий керівник: Теренчук Світлана Анатоліївна ,

кандидат технічних наук, професор кафедри інформаційних технологій

проектування та прикладної математики

Обсяг роботи. Кваліфікаційна робота магістра складається: розділів 5, стор. 102, таблиць 11, рис. 43, завдання, анотація, вступу, висновків, списку використаних джерел.

Актуальність даної теми. Робота присвячена вирішенню задачі кластеризації даних, та підготовки даних до прогнозування аварійних ситуацій. В роботі наведені методи кластеризації даних, за допомогою яких проводиться підготовка даних до побудови прогнозу, а також створена програмна реалізація методу «Гірської кластеризації».

В першому розділі проводиться аналіз предметної області та постановка задачі, аналіз діючої моделі функцій системи мереж водопостачання, класифікація нестандартних ситуацій на мережах.

В другому розділі формуються цільовий аналіз призначення ІТ, функціональні моделі реалізації процесів ІТ, постановка основних функціональних задач, що вирішуються у складі ІТ.

У третьому розділі спроектовано ціль проведення кластерного аналізу даних, об'єкти кластеризації та їх атрибути, нормалізація атрибутів об'єктів кластеризації, алгоритм кластеризації об'єктів.

В четвертому розділі приводяться аналіз технології збору, передачі та обробки інформації, опис вхідної інформації, концептуальна модель бази даних, фізична модель бази даних, опис основних алгоритмів.

В п'ятому розділі наводиться введення початкових даних для контрольного прикладу, розрахунок контрольного прикладу.

Ключові слова: аварійна ситуація, кластеризація, алгоритми.

Keywords: emergency situation, clustering, algorithms.

Якість оформлення проекту. Кваліфікаційна робота магістра оформлена у відповідності до діючих нормативних документів та методичних вказівок до виконання дипломної роботи для студентів спеціальності 126. «Інформаційні системи і технології».

Загальний висновок стосовно роботи та присвоєння авторіві освітнього рівня «магістр». Робота виконана на високому рівні, студент продемонстрував високий рівень теоретичної підготовки та сформованих практичних навичок в області сучасних інформаційних технологій. Заслугує оцінки «добре».

Науковий керівник _____ / проф. Теренчук С.А. /
(підпис)

Посада, місце роботи. КНУБА, пр-т. Повітряних Сил, 31, професор кафедри інформаційних технологій проектування та прикладної математики

«8» грудня 2025 р.

АНОТАЦІЯ

«Розробка інформаційної системи прогнозування аварійних ситуацій на водопровідних мережах міста».

Кваліфікаційна робота магістра за спеціальністю: 126. «Інформаційні системи і технології» – Київський національний університет будівництва та архітектури. – Київ, 2025.

Робота присвячена вирішенню задачі кластеризації даних, та підготовки даних до прогнозування аварійних ситуацій. В роботі наведені методи кластеризації даних, за допомогою яких проводиться підготовка даних до побудови прогнозу, а також створена програмна реалізація методу «Гірської кластеризації».

Ключові слова: аварійна ситуація, кластеризація, алгоритми.

SUMMARY

"Development of information technology for forecasting emergencies in water supply networks of the city.

Master's attestation master's degree in specialty: 126. "Information systems and technologies". - Kyiv National University of Civil Engineering and Architecture - Kiev, 2025.

The project is dedicated to solving the problem of clustering data and prepares the data for forecasting situations. The paper presents the methods of data clustering with which the data preparing for the construction of the forecast, as well as set up the software implementation of the method "Mountain Clustering."

Keywords: emergency situation, clustering, algorithms.

ЗМІСТ

Вступ

1. Аналіз предметної області та постановка задачі

1.1. Загальна характеристика функціонування систем водопостачання міста

1.1.1 Системи водопостачання населених пунктів

1.1.2 Класифікація систем водопостачання

1.1.3. Основні вимоги до матеріалів конструкцій та обладнання водопровідних мереж

1.2 Аналіз діючої моделі функцій системи мереж водопостачання

1.3 Класифікація нестандартних ситуацій на мережах водопостачання міста

1.3.1. Розладнання стиків

1.3.2. Витоки з колодязів

1.3.3. Хімічна та електрохімічна корозія труб

1.3.4. Механічні пошкодження труб

1.3.5. Гідравлічні удари

1.3.6. Брак при спорудженні

2. Концептуальне та функціональне проектування системи

2.1. Цільовий аналіз призначення ІТ

2.2. Функціональні моделі реалізації процесів ІТ

2.3. Постановка основних функціональних задач, що вирішуються у складі ІТ

2.3.1 Методи і форми збору інформації про випадки ушкодження на водопровідних мережах

2.3.2 Задачі що вирішуються у складі ІТ

2.3.2.1 Кластеризація даних

2.3.2.2 Метод FCM для кластеризації даних

3. Математичне забезпечення інформаційної технології

3.1 Ціль проведення кластерного аналізу даних

3.2 Об'єкти кластеризації та їх атрибути

3.3 Нормалізація атрибутів об'єктів кластеризації

3.4 Алгоритм кластеризації об'єктів

3.4.1 Алгоритм нечітких с-середніх

3.4.2 Основні ідеї методу гірської кластеризації

3.4 Приклад роботи алгоритму «Гірської кластеризації»
об'єктів

4. Розробка інформаційного забезпечення інформаційної технології

4.1 Аналіз технології збору, передачі та обробки інформації

4.2 Опис вхідної інформації

4.3 Концептуальна модель бази даних

4.4 Фізична модель бази даних

4.5 Опис основних алгоритмів

4.5.1 Алгоритм ініціалізації діалогового вікна

4.5.2 Алгоритм доступу до редагування бази даних

4.5.3 Алгоритм кластеризації

5. Приклад використання системи

5.1 Введення початкових даних для контрольного прикладу

5.2 Розрахунок контрольного прикладу

Список літератури

ВСТУП

В даний час аварійність трубопроводів в Україні щорік зростає. Аварії приносять збитки, пов'язані з простоєм трубопроводів, втратами води, затопленнями, витратами на виконання ремонтних робіт. Вчасно вжиті заходи для їх запобігання істотно підвищують надійність водопостачання. Саме тому, прогнозування аварійних ситуацій є актуальним. Але розв'язання цієї задачі вимагає розробку універсальних інформаційних систем (ИС) та технологій, які дозволяють проводити прогноз аварійних ситуацій. Точний прогноз у свою чергу передбачає використання ретельно відібраної інформації.

Метою даної роботи є створення спеціальних програмних засобів (ПЗ), які дозволяють підготувати дані до розрахунку прогнозу аварійних ситуацій.

Для досягнення цієї мети необхідно вирішити наступні задачі:

- Дослідження особливостей предметної області;
- Дослідження існуючих методів та алгоритмів кластеризації даних, вибір оптимальних.
- Розробка адаптованих алгоритмів кластеризації до вимог предметної області;
- Дослідження ефективності розроблених алгоритмів;

В першому розділі даної роботи проводиться дослідження предметної області. Проводиться класифікація та аналіз основних причин виникнення аварійних ситуацій на системах водопостачання. Визначаються особливості проектування в перспективі особливостей предметної області.

У другому розділі проводиться цільовий аналіз інформаційної технології. Розроблюється дерево цілей розроблюваної системи. Демонструється функціональні моделі реалізації процесів ІТ, у вигляді чорної скриньки та її багаторівневої декомпозиції. Розглядається FCM метод кластеризації даних.

У третьому розділі проводиться розробка математичного забезпечення інформаційної системи. Розглядається ціль кластеризації для системи прогнозування аварійних ситуацій. Проводиться розробка принципів нормалізації різнотипових даних. Аналізуються два основних методи нечіткої

кластеризації, а саме алгоритм «Нечітких с-середніх» та «Гірської кластеризації». Представляється приклад розрахунку за типовими даними для обраного методу кластеризації.

У четвертому розділі проводиться розробка інформаційного забезпечення інформаційної технології. Аналіз технології збору, передачі та обробки інформації у вигляді DFD – діаграми. Проводиться опис вхідної інформації, та усесторонній огляд бази даних. Розглядаються основні алгоритми реалізовані у процесі створення інформаційної технології.

У п'ятому розділі розглянуті основні принципи функціонування системи. Представляються скріншоти які відображають основний функціонал інформаційної системи, а також етапи роботи, на прикладі типових даних.

1.1 Загальна характеристика функціонування систем водопостачання міста

1.1.1 Системи водопостачання населених пунктів

Система водопостачання – це комплекс інженерних споруд, які призначені для забору води від джерела, її очистки (у випадку необхідності), зберігання запасів води та її постачання до місця споживання.

Система водопостачання населеного пункту повинна забезпечувати отримання в необхідній кількості води з природних джерел, при необхідності поліпшення її якості та передачу до місця споживання.

Одним з основних показників надійності роботи системи служить ймовірність безвідмовного функціонування протягом аналізованого періоду часу. Під відмовою системи водопостачання слід розуміти неприпустиме зниження якості її роботи в результаті наступних подій: відмова вододжерела (зниження рівня води нижче допустимого, зледеніння); аварії на водоводах або магістральних лініях мережі, пошкодження насосів, перерва в подачі електроенергії на насосну станцію; порушення нормальної роботи водоочисних споруд, що несе за собою погіршення якості води. Підвищення надійності водопроводу досягається структурним резервуванням окремих елементів системи. На території населеного пункту зазвичай існують різні категорії водоспоживачів, що пред'являють різні вимоги до кількості та якості води.

Вода різними споживачами витрачається на різні цілі, які можуть бути поділені на три основні категорії: господарсько-питні потреби, виробничі цілі на промислових підприємствах, пожежогасіння.

Відповідно до призначення об'єкта та вимогами, що пред'являються до води, а також з економічних міркувань для всіх зазначених цілей вода може подаватися одним водопроводом або для окремих категорій водоспоживачів можуть бути влаштовані самостійні водопроводи. Зазвичай в містах передбачають єдиний господарсько-протипожежний водопровід, який подає

воду для господарсько-питних потреб промислових підприємств міста, іноді для технічних потреб тих підприємств, де необхідна вода питної якості. Для окремих великих промислових підприємств міста або для групи виробництв одного району, які можуть використовувати неочищену воду, доцільно влаштовувати самостійні виробничі водопроводи.

У містах зазвичай є підприємства, які споживають відносно невелику кількість води не питної якості. Враховуючи їх розпорошеність по території міста, виявляється економічно доцільним забезпечувати їх очищеною водою з мережі міського водопроводу, ніж влаштовувати для них самостійні виробничі водопроводи.

1.1.2 Класифікація систем водопостачання

Системи водопостачання класифікуються за такими признаками(рис. 1.1.):

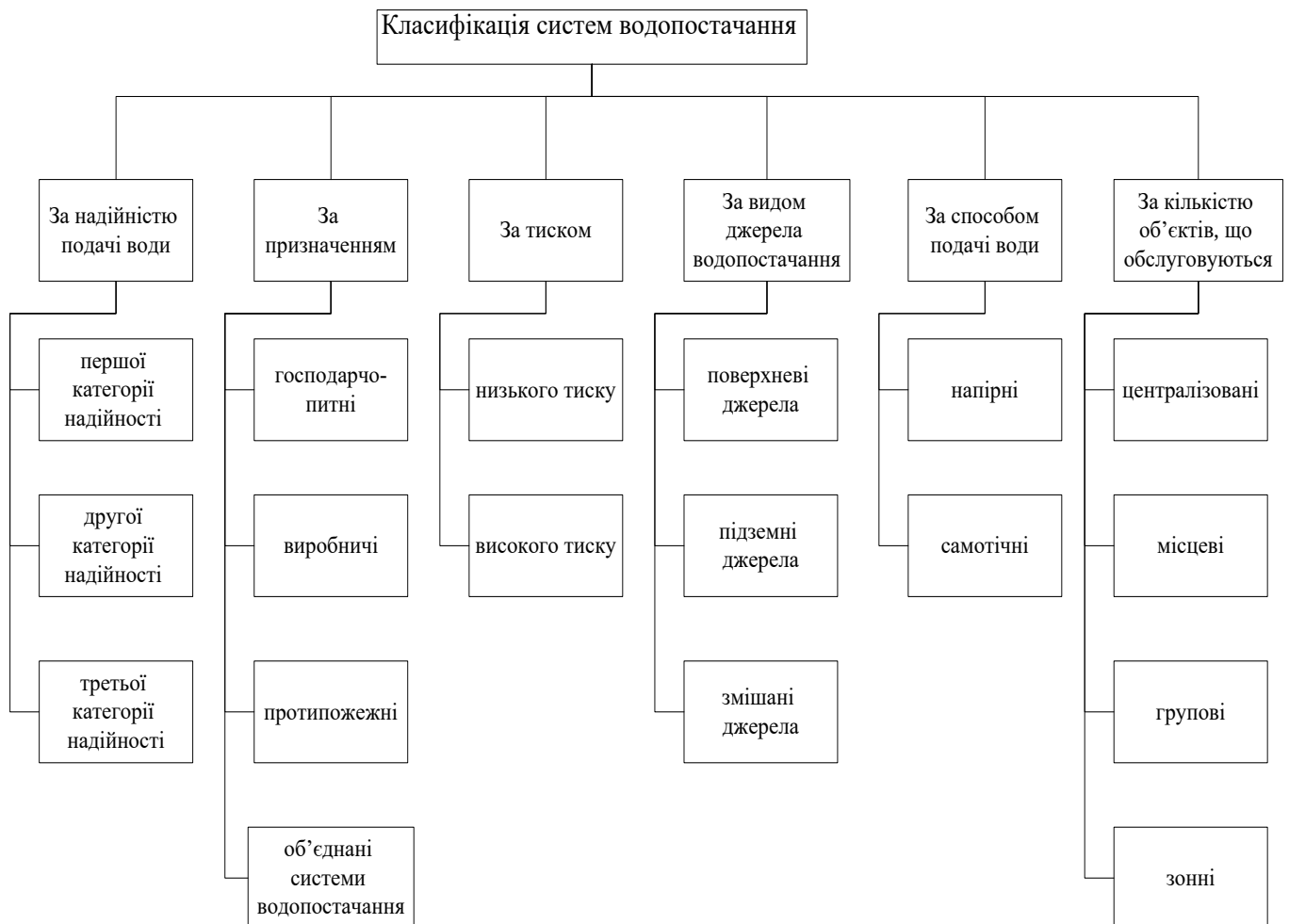


Рисунок 1.1 Класифікація систем водопостачання

За надійністю подачі води:

- *першої категорії надійності* – допускається зниження подачі води на господарчо-питні потреби не більш 30% розрахункової витрати і на виробничі потреби за аварійним графіком роботи підприємств; тривалість зниження подачі не повинна перевищувати 3 доби. Перерва в подачі води або зниження подачі нижче зазначеної межі допускаються на час вимикання ушкоджених і включення резервних елементів системи (устаткування, арматури, споруджень, трубопроводів і ін.), але не більш ніж на 10 хвилин;

- *другої категорії надійності* — величина зниження подачі води така сама, що й при I категорії; тривалість зниження подачі не повинна перевищувати 10 діб. Перерва в подачі води або зниження подачі нижче зазначеної межі допускаються на час вимикання ушкоджених і включення резервних елементів або проведення ремонту, але не більше ніж на 6 годин;

- *третьої категорії надійності* — величина зниження подачі води така сама, що й при I категорії; тривалість зниження подачі не повинна перевищувати 15 діб. Перерва в подачі води або зниження подачі нижче зазначеної межі допускається на час проведення ремонту, але не більше ніж на 24 години.

За призначенням:

- господарчо-питні, які призначені для постачання води на господарчі та питні потреби населення та працівників промислових підприємств;

- виробничі, які подають воду на технологічні потреби виробництва;

- протипожежні, які забезпечують водопостачання на пожежогасіння;

- об'єднані системи водопостачання: господарчо-протипожежні, виробничо-протипожежні або господарчо-протипожежні.

У містах та населених пунктах найчастіше влаштовують об'єднані господарчо-питні водопроводи. Цими водопроводами вода подається до промислових підприємств, якщо вони споживають у невеликих розмірах питну воду.

При великих витратах води промислові підприємства мають самостійні водопроводи, які забезпечують підприємство водою як із зовнішнього джерела (міського магістрального водопроводу), так і від місцевих джерел - поверхневих та підземних. Влаштовують системи водопостачання, які забезпечують господарчо-питні, виробничі та протипожежні потреби, тобто будують господарчо-питні та виробничі водопроводи, об'єднуючі їх з протипожежними. Перевага віддається об'єднанню протипожежного водопроводу з господарчим, а не із виробничим, тому що виробнича водопровідна мережа не охоплює усіх об'єктів підприємства. Крім того, для деяких технологічних процесів воду необхідно подавати під визначеним тиском, який змінюється при пожежогасінні, що може привести до аварії.

За тиском:

- низького тиску;
- високого тиску.

У мережі *низького* тиску вільний тиск на рівні поверхні землі мусить бути не менше 10 м, та на кожен поверх будівлі вище першого додається ще 4 м.

Системи *високого* тиску використовуються переважно для пожежогасіння. Вода до місця пожежі подається безпосередньо від гідрантів, а необхідний для пожежогасіння напір у мережі та біля стволів забезпечується стаціонарними пожежними насосами, які розміщують у насосних станціях.

За видом джерела водопостачання:

— з водопостачанням з поверхневих, підземних джерел, а також зі змішаними джерелами водопостачання.

За способом подачі води:

- напірні - із подачею води насосами;
- самотічні - при розміщенні вододжерела на висоті, що забезпечує природне водопостачання споживачів.

За кількістю об'єктів, що обслуговуються:

- централізовані;
- місцеві;
- групові;
- зонні.

Централізовані системи водопостачання зустрічаються найчастіше. Вони використовуються для подачі води до населених пунктів, промислових підприємств в них. Представляють собою централізовану систему з одним або декількома джерелами водопостачання, які забезпечують подачу води до однієї мережі.

Місцеві системи водопостачання обслуговують одну будівлю або невелику групу компактно розміщених будівель від одного поблизу розташованого джерела (наприклад, промислове підприємство, район міста). Якщо водопровід подає воду до декількох об'єктів (наприклад, групи малих населених пунктів, групи промислових підприємств), він має назву групового водопроводу. Для живлення водою під необхідним тиском різних районів населеного пункту, що мають значну різницю у геодезичних відмітках, влаштовують зонні системи.

1.1.3. Основні вимоги до матеріалів конструкцій та обладнання водопровідних мереж

Водогінна мережа є звичайно найбільш дорогою й досить відповідальною підсистемою системи водопостачання об'єкта. Дуже складний, у значній мірі випадковий процес її фактичного функціонування. Основна вимога, пропонована до водогінних мереж, - безперебійна подача

води до місць її відбору за умови забезпечення необхідних кількостей подаваної води й необхідних напорів.

Лінії водогінних мереж монтують із труб, тобто елементів, що виготовляються заводським способом. На місці будівництва виробляються лише з'єднання труб і їхнє укладання.

Відповідно до умов роботи водопровідних ліній у процесі експлуатації до них висувають наступні вимоги (рис. 1.2.):

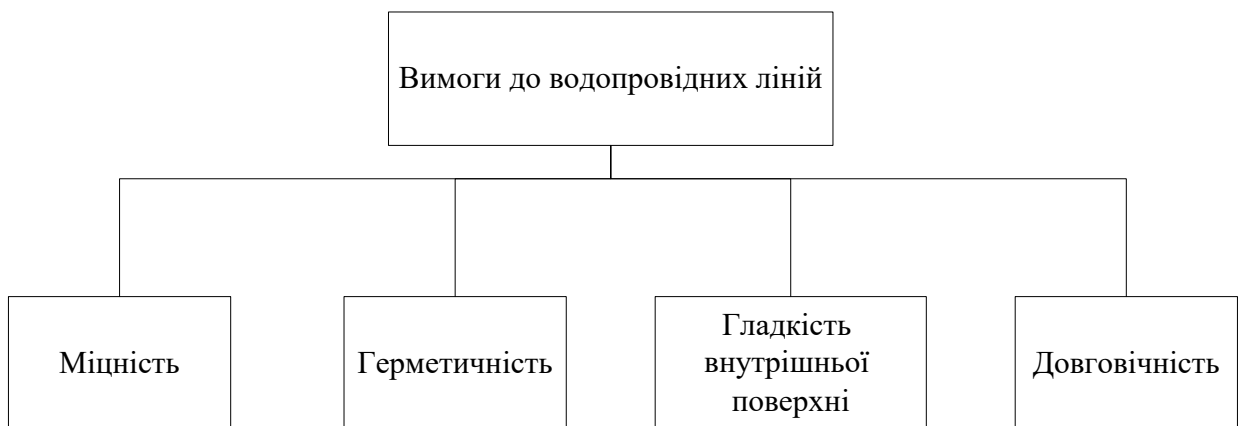


Рисунок 1.2 Основні вимоги до водопровідних ліній

- міцність, тобто високий опір всім можливим (заданим) внутрішнім і зовнішнім навантаженням;
- герметичність;
- гладкість внутрішньої поверхні їхніх стінок, що забезпечує найменші втрати напору на тертя при русі води;
- довговічність, тобто тривалий термін служби, що обумовлюється в основному досить високим опором матеріалу труб (або їхніх покриттів) зовнішньому й внутрішньому агресивному впливам води, що транспортується, ґрунтів, ґрунтових вод і т.ін.

Крім того, труби, як і всі елементи збірного будівництва, повинні забезпечувати можливість їх легкого, простого, швидкого й надійного

з'єднання (монтажу стиків). Нарешті, водопровідні лінії, як і всякі інженерні спорудження, повинні задовольняти вимогам найбільшої економічності.

Напірні водопровідні труби повинні бути розраховані на опір силам тиску води на внутрішню поверхню їхніх стінок. Розрахунковий робочий тиск визначається в результаті розрахунку мереж і водоводів, і може коливатися для різних мереж у широких межах. Відповідно до умов укладання труб вони повинні також мати відповідну міцність для опору тиску ґрунту, прогинам від власної ваги (при нещільних ґрунтах), навантаженням від транспорту та ін.

Герметичність, як самих труб, так і стикових з'єднань є найважливішою умовою успішної й економічної роботи водопроводу. Недотримання герметичності ліній викликає постійні не продуктивні витрати води й підвищує вартість експлуатації системи. Витоки води з мережі можуть призвести також до розмивання ґрунту й викликати серйозні аварії.

У системах водопостачання різних об'єктів й у різних місцевих умовах всі основні параметри, які необхідно враховувати при виборі типу труб (кількість подаваної води, внутрішній робочий тиск, характер ґрунтів та ін.), змінюється в досить широких межах.

У сучасній практиці будівництва водоводів і зовнішніх водопровідних мереж широко застосовуються чавунні, сталеві, азбестоцементні й залізобетонні труби. У цей час усе більше широке застосування у світовій практиці одержують попередньо напружені залізобетонні труби й труби із синтетичних матеріалів (пластмасові).

1.2 Аналіз діючої моделі функцій системи мереж водопостачання

Системи водопостачання являють собою комплекс елементів, що включає водоприймальні споруди, насосні станції, резервуари, очисні споруди, водоводи й водогінну мережу. Складові та функції системи

водопостачання і системи водовідведення населеного пункту зображені на рис. 1.3. Надійність функціонування зазначених споруд - одна з основних умов безперебійності забезпечення водою промислових підприємств і населених пунктів.

Під надійністю водопостачання розуміється властивість систем виконувати функції водопостачання, зберігаючи в часі встановлені технологічні показники в межах, що відповідають заданим режимам умов експлуатації.

Фізичний зміст надійності полягає в здатності систем водопостачання зберігати свої початкові технологічні характеристики в процесі експлуатації.

В населених пунктах існує велика кількість споживачів води. Умови споживання ними води можуть відрізнятись, також як їх вимоги до якості води. Для визначення витрат води, які необхідно подати в населений пункт із врахуванням особливостей, усіх водоспоживачів можна згрупувати так: комунальний сектор, підприємства, поливання, тваринницькі ферми, пожежогасіння, спеціальні споживачі.

Протягом року водоспоживання коливається залежно від сезону (літній чи зимовий, в деяких випадках може бути ще весняно-літній та осінньо-зимовий), режиму роботи підприємств, агрокліматичних умов, зміни сільськогосподарських робіт тощо. Коливання водоспоживання є і протягом місяців та тижнів. Істотніші зміни водоспоживання бувають протягом доби. Ці коливання залежать від нічного й денного часу, режиму роботи підприємств, проведення різноманітних сільськогосподарських робіт, погодних умов, особливостей побуту місцевого населення, кількості робочих і вихідних днів, свят, проведення культурних, спортивних та інших заходів.

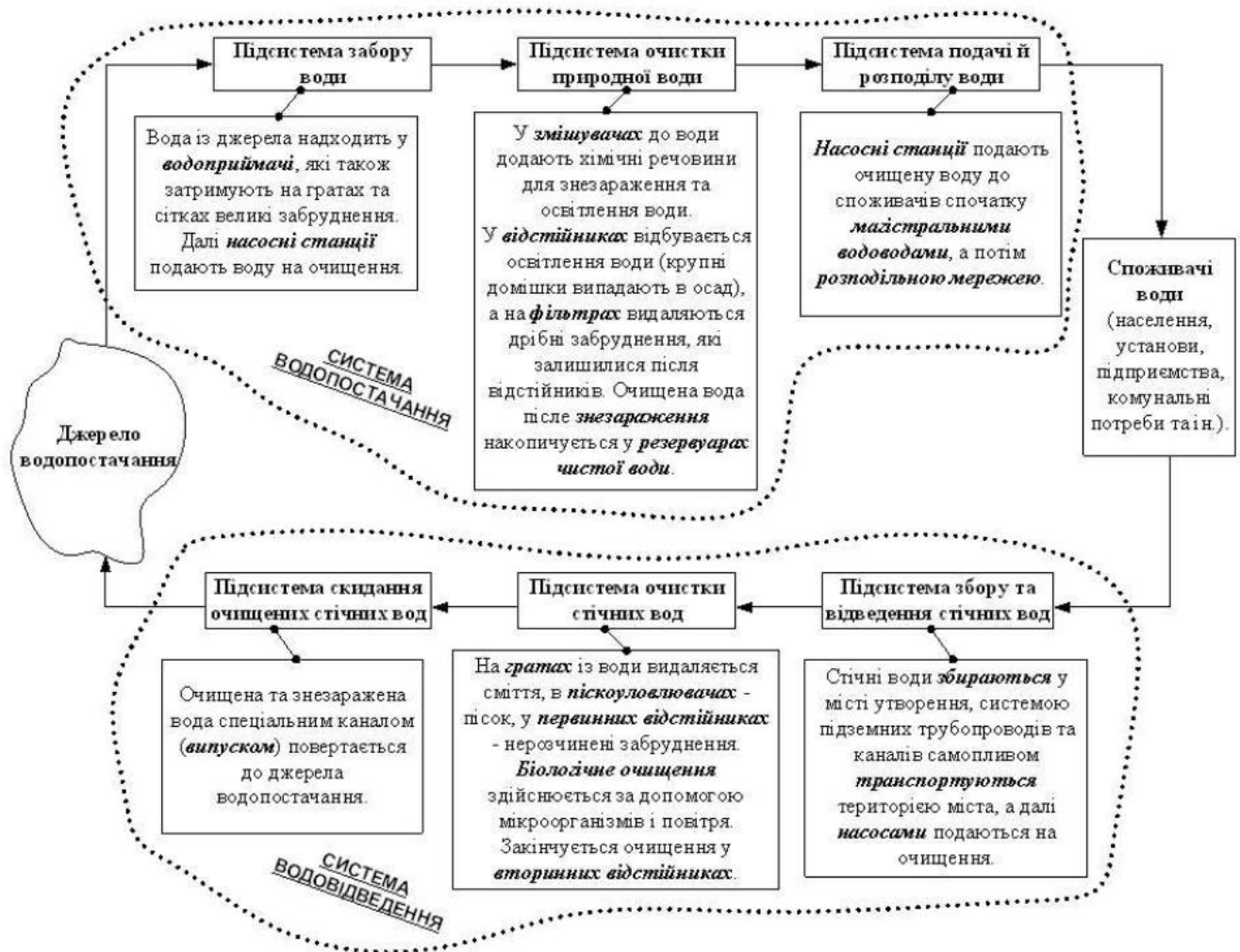


Рисунок 1.3 Склад і функції системи водопостачання і системи водовідведення населеного пункту

Стійкість системи господарчо-питного водопостачання визначається можливістю зберегти свою працездатність і здібність подавати необхідну кількість води при руйнуванні окремих її елементів.

Схема водопостачання для міст з використанням поверхневих вододжерел зображена на (рис. 1.4). Вода поступає у водоприймач (оголовок) 1 та по самотічним трубам 2 подається до берегового колодязя 3, а із нього насосною станцією першого підйому (НС-I) до очисних споруд – відстійників 5 та фільтрів 6. Після очищення та знезаражування хлором або іншим засобом, вода подається до резервуарів чистої води (РЧВ) 7, від яких забирається насосною станцією другого підйому НС – II, і по водоводам 9 до

баку водонапірної башти 10, після чого подається магістральною 11 та розподільчою 12 мережами до споживачів 13.

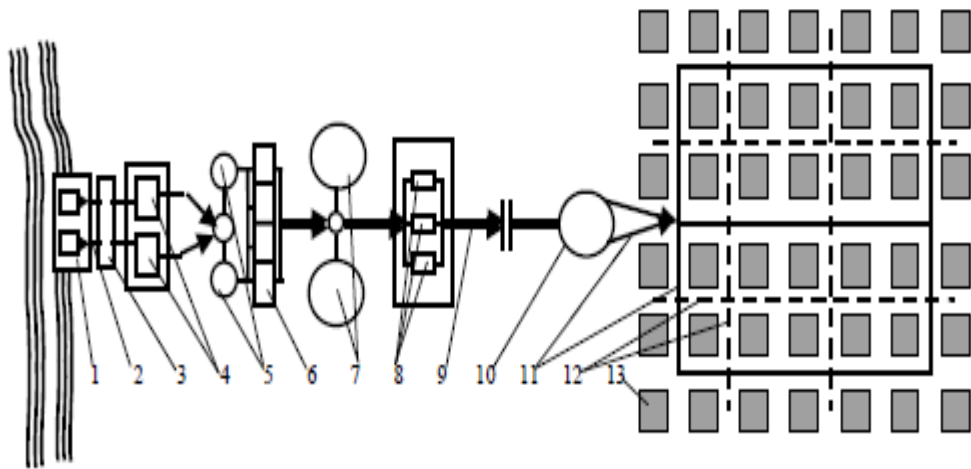


Рисунок. 1.4 Схема водопостачання населеного пункту

1- оголовок, 2 – самотічні труби, 3 – береговий колодязь, 4 – насосна станція першого підйому, 5 – відстійники, 6 – фільтри, 7 – резервуари чистої води, 8 – насосна станція другого підйому, 9 – водоводи, 10 – водонапірна башта, 11 – магістральна мережа, 12 – розподільча мережа, 13 – водоспоживачі

Сучасні системи водопостачання у містах, на промислових і інших об'єктах складні і розгалужені. Система водопостачання являє собою великий і складний комплекс різних будинків і споруд, трубопроводів, енергетичних пристроїв і ліній передач. Споруди ці часто розкидані на великій території і віддалені одна від іншої на значні відстані. Типовий генеральний план водопровідних очисних споруд зображений на рис. 1.5.

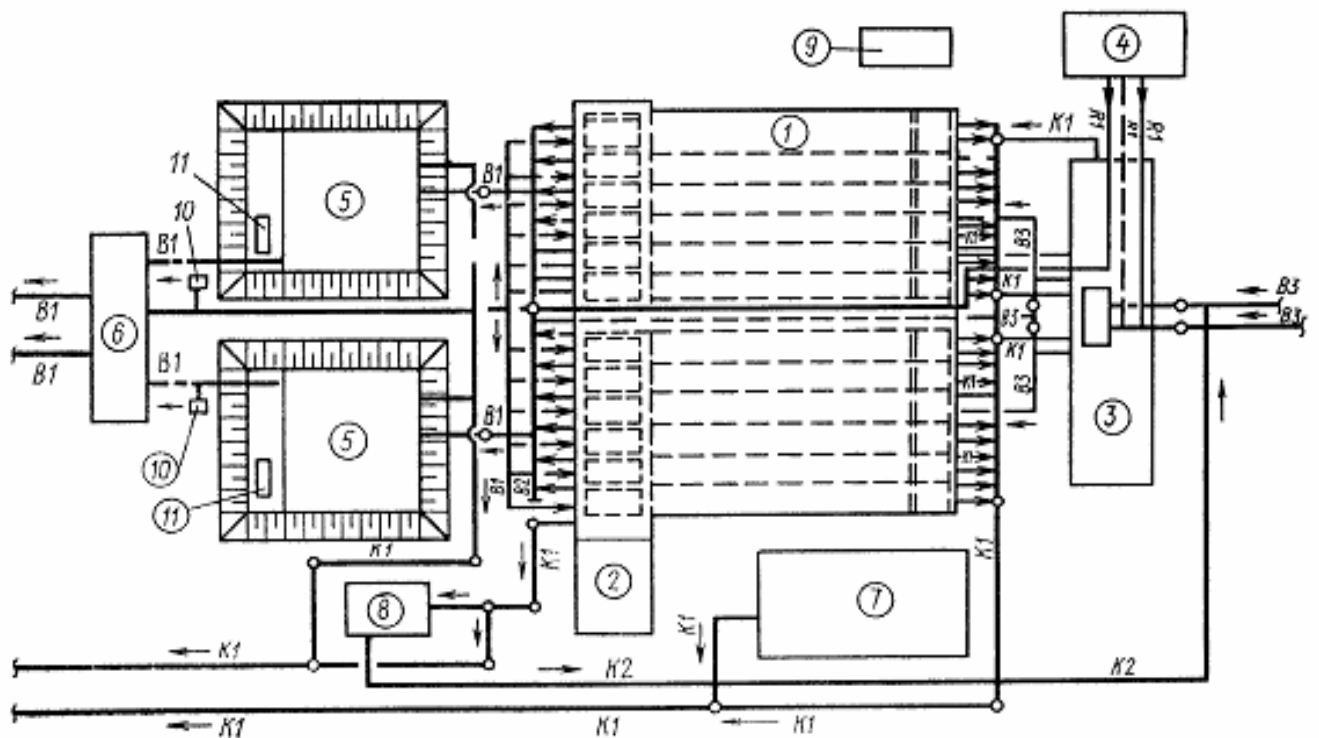


Рисунок 1.5 Генеральний план водопровідних очисних споруд:

1 - блок входних пристроїв та контактних освітлювачів з мікрофільтрами; 2 - блок службових приміщень; 3 - реагентне господарство; 4 - хлораторна з витратним складом хлору; 5 - резервуари питної води; 6 - насосна станція 2-го підйому; 7 - піскове господарство; 8 - споруда обробки промивної води; 9 - споруда обробки осадка; 10 - протирадіаційне укриття; 11 - камери для забору води; 12 - камери фільтрів - поглотителів; 13 - затворний бак коагулянту або вапна; 14 - бак мокрого зберігання коагулянту; 15 - витратний бак коагулянту; 16 - витратний бак вапна з механічною мішалкою; 17 - дозуючий агрегат коагулянту або вапна; 18 - гідроциклон; 19 - мішалка слабого розчину вапняного молока; 20 - мішалка міцного розчину вапняного молока; 21 - ємність для розмиву вапняного тіста; 22 - бак - сховище вапняного тіста; 23 - насос-дозатор вапняного молока; 24 - контактна камера; 25 - мікрофільтри або барабанні сітки; 26 - перегородчатий змішувач; 27 - контактний освітлювач; 28 - насос подачі промивної води; 29 - безгравійна трубчата розподільча система; 30 - відвід промивної води; 31 - насос 2-го підйому

Особливу небезпеку для систем водопостачання несуть надзвичайні ситуації природного характеру пов'язані зі змінами поверхні (землетруси, зсуви, просідання земної поверхні, суфозії та інші), вражаючі фактори надзвичайних ситуацій техногенного характеру – вибухи (ударна хвиля), зруйнування будинків (падіння важких уламків, пожежі, що призводять до зруйнування водопровідних мереж).

Найбільш слабкими ланками системи водопостачання є всі наземні будинки і споруди, у тому числі трансформаторні підстанції, насосні станції, обладнані складною системою зв'язку й автоматики, будинок очисних споруд і інші наземні об'єкти. Основні засоби забезпечення надійності систем водопостачання: при проектуванні, при будівництві, при експлуатації, при виконанні ремонтних робіт, резервування.

Системи водопостачання, що живлять окремі категоризовані міста або декілька міст, у числі яких є категорійні міста й об'єкти особливої важливості, повинні базуватися не менше, ніж на двох незалежних джерелах води, один з яких слід передбачати підземним. Для підвищення надійності системи водопостачання і можливості виконання ремонтних робіт комунікацій насосних станцій, головних, очисних і інших споруд проектують таким чином, щоб при виході з ладу яких-небудь споруд, чи елементів системи їх можна було відключити, не порушуючи ритму роботи всієї системи. Для цих цілей передбачають велику кількість перемичок, що дозволяють подавати воду в будь-який напірний трубопровід і відключати ушкоджені споруди і лінії. Передбачаються обвідні лінії (байпаси), що дозволяють подавати воду, минаючи ці ушкоджені споруди, наприклад повз відстійники на фільтри, повз фільтри в резервуари чистої води.

Необхідно відзначити, що для забезпечення надійності роботи систем водопостачання можуть вирішуватись питання, зв'язані з посиленням будівельних конструкцій і вузлів різних технологічних будинків, споруд і мереж, створенням запасів будівельних матеріалів і устаткування для

швидкої їхньої заміни та ін. Заражена вода може подаватися споживачу тільки для технічних цілей і тільки після того, як ступінь зараження й отруєння знизяться до безпечних меж, чи буде зроблене її відповідне очищення. Для цього найчастіше передбачають введення у воду хімічних реагентів (звичайно хлору) за допомогою хлораторних установок на насосних станціях з наступною обробкою її на очисних спорудах. Порядок обробки й очищення води, зараженої отруйними речовинами і бактеріальними засобами, залежить від характеру і ступеня зараження, наявних реагентів.

Очищена вода надходить у резервуари чистої води. Якщо передбачити захист від зараження, резервуари чистої води забезпечать досить надійний аварійний запас на випадок виходу з ладу водозабірних споруд системи.

Особливо важливою ланкою в системі водопостачання є насосні станції, що забезпечують подачу води безпосередньо до споживачів. З огляду на це, насосні станції першого і другого підйомів повинні будуватися з урахуванням їх захисту від впливу засобів ураження (ударної хвилі й інших вражаючих факторів), стихійного лиха, дії вторинних факторів при аваріях та катастрофах.

Водогінна мережа, як правило, будується за кільцюваною, тупикові лінії роблять тільки як виключення для забезпечення водою споживачів, що допускають перерву подачі води для виробничих потреб на час ліквідації аварії. Закільцьована водогінна мережа забезпечує можливість маневру водою шляхом обходу ушкоджених чи зруйнованих ділянок, якщо збереглися насосні станції і резервуари чистої води.

Водопровідні колодязі, у тому числі і ті, у яких встановлені гідранти і перемикаючі засувки, можуть виявитися під завалами. Тому їх виносять за зону можливого завалу від будинків і споруд, біля яких вони розміщуються.

Для взаємного резервування автономні системи водопостачання окремих підприємств з'єднують одну з іншою і з міськими системами, з дотриманням санітарно-гігієнічних вимог. На сполучних лініях ставлять дві

засувки, закриті й опечатані. Це необхідно в тих випадках, коли якість води в системах різна. Для з'єднання мереж господарсько-питного водопроводу з мережами, що подають воду, непридатну для пиття, необхідна перемичка, що забезпечує повітряний розрив між ними.

Водопостачання, що базується на артезіанських свердловинах, є більш надійним. Стійкість роботи артезіанських свердловин може бути підвищена проведенням заходів, щодо захисту від зараження води на поверхні землі і створенням надійного енергоживлення.

1.3 Класифікація нестандартних ситуацій на мережах водопостачання міста

Під нестандартною ситуацією слід розуміти відхилення одного або декількох параметрів на водопровідній мережі, насосній станції, РЧВ, артезіанській свердловині від розрахункових. Нестандартна ситуація характеризується відмовою, порушенням діяльності системи, об'єкта.

- Відмова – порушення діяльності об'єкта – один з основних досліджуваних параметрів при аналізі будь-якої технічної системи. Відповідно розрізняють наступні відмови:
- Раптова відмова – відмова, яка характеризується раптовою зміною одного або декількох заданих параметрів об'єкта.
- Поступова відмова – відмова, яка характеризується поступовою зміною одного або декількох заданих параметрів об'єкта.
- Збій – відмова, яка призводить до короточасного порушення працездатності і яка самоліквідується.
- Збій який чергується – багаторазово виникаючий збій одного і того характеру.
- Конструкційна відмова – відмова, яка виникає внаслідок порушень встановлених правил та (або) норм конструювання.
- Промислова відмова – відмова, яка виникає внаслідок порушення встановленого процесу виготовлення або ремонту об'єкта.
- Експлуатаційна відмова – відмова, яка виникає внаслідок порушення встановлених правил та (або) умов експлуатації об'єкта. Експлуатаційна відмова – явище випадкове по моменту виникнення, але має досить з'ясовані причини, які або закладені в самих вузлах

системи водопостачання внаслідок багатьох неконтрольованих параметрів матеріалів і деталей або викликані впливом зовнішнього середовища. Відмова на кінцевій стадії експлуатації (*відмови, обмежуючі строк служби*) – явища закономірні, викликані природними процесами, які протікають у вузлах і обумовлені принципами їх дії, конструкцією, використаними матеріалами, та технологією виготовлення.

Належність відмови до тої чи іншої стадії “життєвого циклу” системи водопостачання з’ясовує природу відмови, задачі та методи її дослідження: - на першій стадії – виявлення неякісних вузлів; на другій стадії – виявлення причин відмови та використання отриманої інформації для пошуку способів збільшення ресурсу, розвитку надійних методів оцінки надійності.

Протягом багатьох років підприємством реєструються такі види аварій (рис. 1.6.):

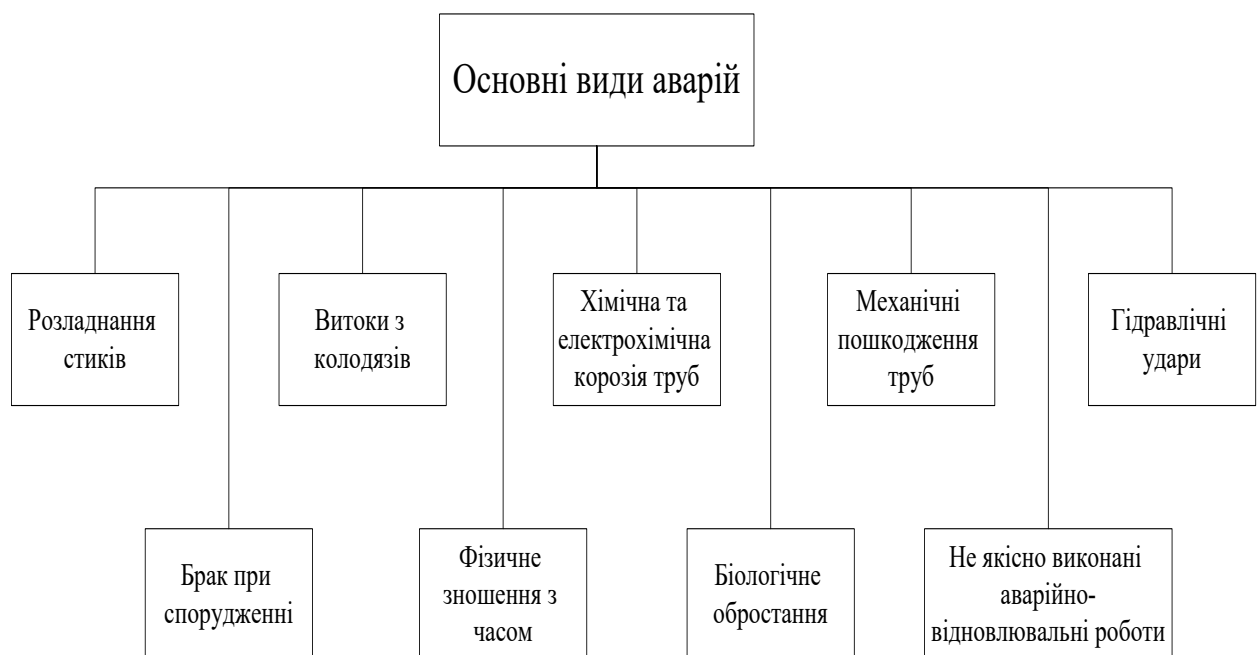


Рисунок 1.6 Основні види аварій

Тепер розглянемо більш детально кожну аварійну ситуацію, частоту її виникнення та причини, що зумовлюють її появлення.

1.3.1. Розладнання стиків

Найчастіше розладнання стиків виникає на чавунних водоводах та на водоводах, які складаються з різних матеріалів. Так у 1986 році з 1822 аварій в системі ПРВ 420 шт. прийшлося на розладнання стиків (табл. 1.1). А у 1998 році на 2614 аварій за рік - 419 прийшлося на розладнання стиків. Протягом 12 років кількість розладнання стиків коливалась у межах 419 - 546 шт. на рік.

Таблиця 1.1 - Кількість аварій на чавунних водоводах в залежності від терміну їх експлуатації

Кількість аварій на чавунних водоводах в залежності від терміну їх експлуатації.			
5 років	5-10 років	10-15 років	Понад 15 років
58	82	103	257

Головною причиною розладнання стиків у більшості випадків є фізичне зношення з часом матеріалу труби, набивки, просади основи під трубопроводом. Як тільки починається незначний витік з трубопроводу у той же час вода, що витікає починає розмивати основу під трубопроводом. Пісок насичується водою виникає пульпа, яка більш не підтримує водовод і як наслідок виникає перелом труби. Якщо на початковій стадії аварії (розладнання стику) витіки з трубопроводу не помітні, то після розлому труби вода витікає з водоводу повним перетином і аварія стає помітною. На поверхні землі утворюються провали. Схематично місце аварії (розладнання стику) зображено на рис. 1.7 :

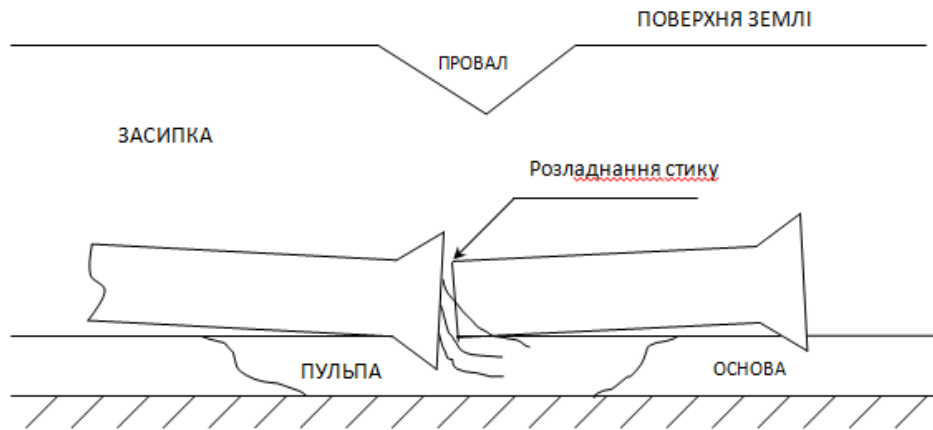


Рисунок 1.7 Схематичний вигляд розладнання стику

Були випадки коли до зони промоїни потрапляли дороги, тротуари, будівлі. Тому для запобігання такого роду аварій існують спеціальні підрозділи підприємства, які намагаються виявити такого роду аварії на початковій стадії їх утворення. Сигналом для пошуку є різка, нічим не зумовлена зміна тиску та витрати води в на ділянці мережі.

1.3.2. Витоки з колодязів

Як явище виникає рідше ніж пориви та розладнання стиків. Локалізація аварій такого типу легша внаслідок конкретного місця їх розташування. Витоки в колодязях виникають в наслідок пошкоджень на запірній арматурі, фланцевих з'єднаннях, пожежних гідрантах та інших елементах водопровідної мережі, які розташовані в водопровідних колодязях. Схематично місце аварії (виток з колодзя) зображено на рис. 1.8:

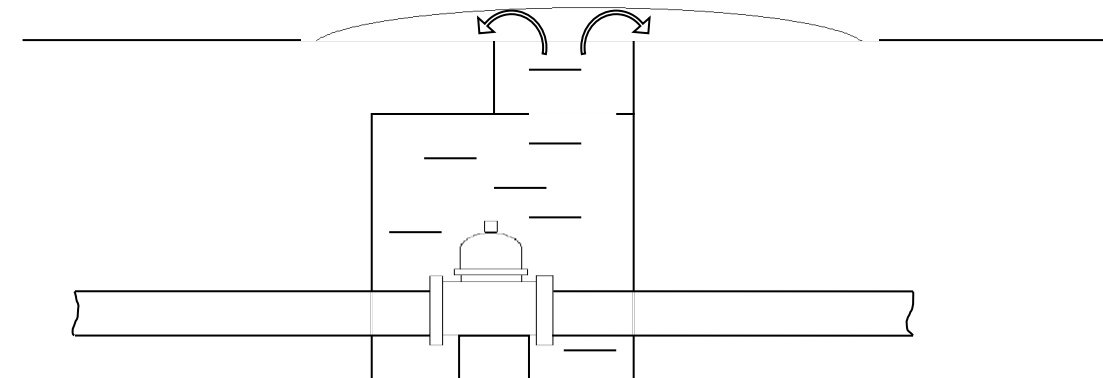


Рисунок 1.8 Схематичний вигляд витоку з колодзя

1.3.3. Хімічна та електрохімічна корозія труб

Корозія трубопроводів — процес руйнування трубопроводів під дією зовнішнього навколишнього (породи — пісок, глина, суглинки) і внутрішнього (пластова вода, емульсія, нафта, що вміщає сірководень) середовищ.

За характером взаємодії металу труб із середовищем розрізняють хімічну і електрохімічну корозію трубопроводів.

Процес хімічної корозії відбувається в середовищах, що не є провідниками електричного струму (сухі гази, органічні рідини - нафтопродукти, спирти і ін.), причому інтенсивність корозії зростає з підвищенням температури - в результаті на поверхні металів утворюється оксидна плівка.

До хімічної корозії схильні абсолютно всі метали - і чорні, і кольорові. Активні кольорові метали (наприклад - алюміній) під впливом корозії покриваються оксидною плівкою, що перешкоджає глибокому окисленню і що захищає метал. А такий мало активний метал, як мідь, під впливом вологи повітря набуває зеленуватого нальоту - патини. Причому оксидна плівка захищає метал від корозії не у всіх випадках - лише якщо кристалохімічна структура плівки, що утворилася, подібна будові металу, інакше - плівка нічим не допоможе.

Сплави схильні до іншого типу корозії: деякі елементи сплавів не окислюються, а відновлюються (наприклад, в поєднанні високої температури і тиску в сталях відбувається відновлення воднем карбідів), при цьому сплави повністю втрачають необхідні характеристики.

Відбувається наступне: метали (сплави) втрачають частину атомів (вони переходять в електролітичний розчин у вигляді іонів), електрони, що заміщають втрачені атоми, заряджають метал негативним зарядом, тоді як електроліт має позитивний заряд. Утворюється гальванічна пара: метал руйнується, поступово всі його частки стають частиною розчину. Електрохімічну корозію можуть викликати блукаючі струми, що виникають

при витоку з електричного ланцюга частини струму у водні розчини або в ґрунт і звідти, - в конструкції з металу. У тих місцях, де блукаючі струми виходять з металоконструкцій назад у воду або ґрунт, відбувається руйнування металів. Особливо часто блукаючі струми виникають в місцях руху наземного електротранспорту (наприклад, трамваїв і ж/д локомотивів на електричній тязі).

1.3.4. Механічні пошкодження труб

Механічна міцність – здатність матеріалу сприймати зусилля робочих навантажень, не руйнуючись і не утворюючи пластичних деформацій понад гранично встановлені величини. Міцність технологічного обладнання забезпечується правильним підбором матеріалу з врахуванням характеру і величини зовнішніх навантажень, що діють на апарат.

В результаті механічних впливів матеріал трубопроводу зазнає високих внутрішніх напружень, які здатні зумовити не тільки утворення нещільностей у швах і з'єднаннях, але й спричинити повне їх руйнування.

Високі внутрішні напруження виникають при підвищеному тиску в апаратах та в результаті навантажень динамічного характеру.

Виявляють такі основні причини механічного пошкодження трубопроводу (рис. 1.9):

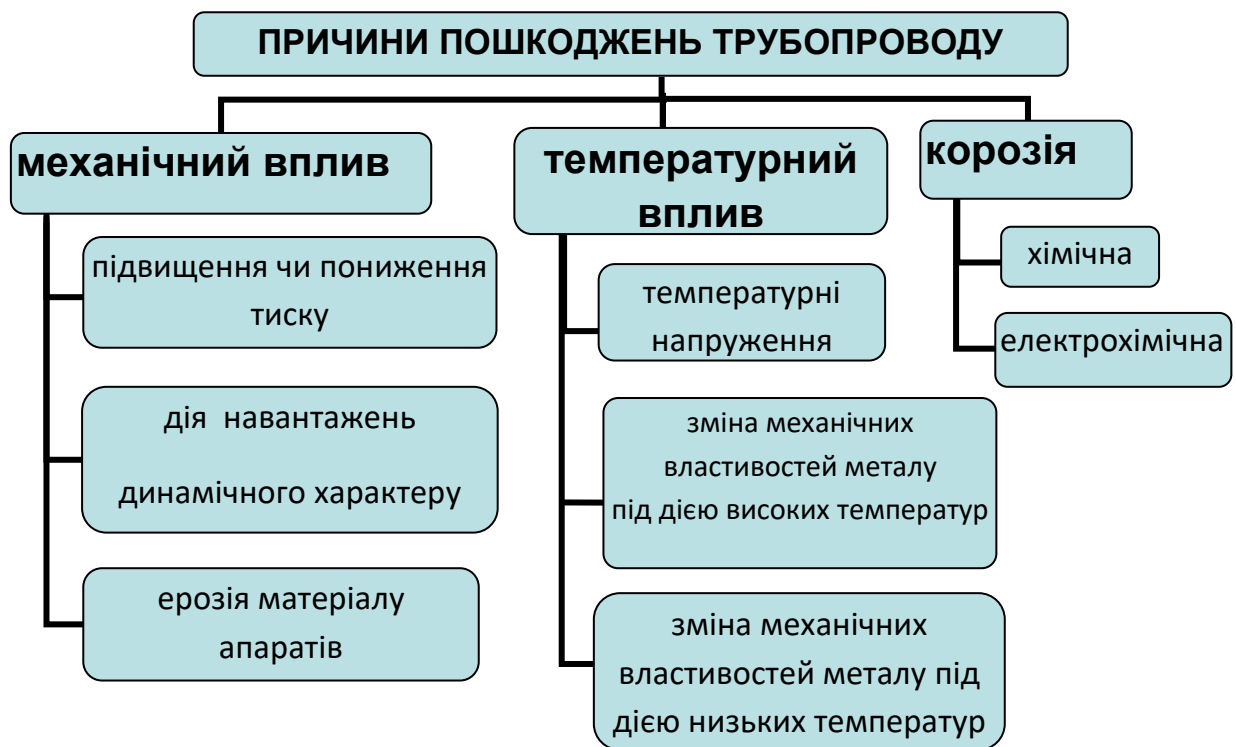


Рисунок 1.9 Схема класифікації причин механічних пошкоджень трубопроводу

1.3.5. Гідравлічні удари

Гідравлічним ударом є короткочасне, але різке і сильне підвищення тиску в трубопроводі при раптовій зупинці потоку рідини, що рухався по ньому. Як правило, це явище виникає при заповненні трубопроводів, коли повітря встигає вийти через спеціально відкритий кран, але перетину цього крану не вистачає, щоб пропустити весь потік такою, що раптово досягла його нестискуваної рідини. Такий же ефект виникає і при швидкому закритті вентиля, що різко перекриває потік. Останнє особливо актуально в наші дні, коли старі гвинтові кран-букси, що мимоволі закривалися плавно (адже крутити маховичок треба багато зворотів, і тому шток перекриває просвіт вентиля досить повільний), замінюються сучасними кульовими кранами, що «перерізають» потік всього за чверть звороту одним рухом руки.

Особливості явища гідроудару

Гідроудар через свою природу має декілька істотних особливостей, про які не можна забувати.

1. Висока швидкість процесу

Перш за все, слід врахувати високу швидкість процесу. Оскільки швидкість переміщення кордонів зон з різним тиском при високій жорсткості труби і заглушки визначається швидкістю поширення пружних деформацій в рідині, тобто швидкістю звуку, все відбувається за дуже короткий час.

2. Умови відриву рідини. Сильні і слабкі гідроудари

У фазі розрідження відривши рідині від заглушки відбувається не завжди. Для цього швидкість потоку має бути досить високою, а стінки труби — досить жорсткими, щоб удар вийшов різким. Якщо удар виявиться дуже слабким (або дуже плавним), то порожньої області в заглушки не утворюється, хоча у будь-якому випадку у фазі розрідження тиск усередині труби, у тому числі безпосередньо в заглушки, буде менше, ніж тиск навколишньої рідини зовні.

1.3.6. Брак при спорудженні

Браком у роботі водопровідних підприємств є погіршення якості очищення води проти стандарту або різке зниження напорів у водопровідній мережі.

Браком у роботі водопровідних підприємств вважається:

- Зниження якості води проти встановленого стандарту, послідувало в результаті виходу з ладу обладнання або приладів для очищення та дезінфекції води, а також внаслідок недбалої роботи обслуговуючого персоналу.
- Засмічення оголовка, самопливних труб і берегового колодязя. Засмічення решіток водоприймача плаваючими предметами: донним льодом, шугою та ін. Засмічення споруд з

очищення води, хоча і не викликало зменшення подачі води, але призвело до погіршення її якості.

- Засмічення піску на фільтрах, що викликає різке зниження їх продуктивності.
- Зниження напору у всій мережі або в окремих її ділянках, викликане неправильним регулюванням роботи споруд і режиму водопровідної мережі.
- Погіршення якісних показників роботи споруд і обладнання, викликане порушенням правил технічної експлуатації з боку обслуговуючого персоналу.

Аварії і брак в роботі можуть мати місце за наступними причин:

- а) з вини персоналу: оперативного, ремонтного, лабораторного і керівного;
- б) з вини заводів-постачальників і будівельно-монтажних організацій;
- в) по стихійним обставинам (повінь, шугообразованіє, ураган та ін.);
- г) за іншими: з вини абонентів, енергопостачальної організації та ін.

До аварій і браку в роботі по прямій вини оперативного персоналу та персоналу лабораторії відносяться помилкові відключення і включення обладнання, невжиття заходів, що запобігають падіння тиску в мережі, погіршення якості очищення води, неправильна дозування реагентів і т.д.

До аварій та шлюбу з вини ремонтного персоналу ставляться випадки недоброякісної ремонту обладнання та водопровідної мережі.

До аварій і браку в роботі з вини керівного персоналу можуть ставитися випадки, що сталися внаслідок невиконання протиаварійних заходів, несвоєчасного усунення аварійних вогнищ, допуску до роботи

ненавченого і неперевіреного персоналу, несвоєчасного проведення профілактичних випробувань та ін.

До аварій і браку в роботі з вини заводів-постачальників, монтажних і будівельних організацій можуть ставитися ті випадки, за якими встановлена вина заводу або цих організацій (дефекти виготовлення і монтажу конструкції, недоброякісний матеріал і ін.). У цих випадках вина заводу або будівельно-монтажної організації повинна бути встановлена документально, тобто зафіксована відповідними актами комісії, фотознімками, лабораторними випробуваннями та ін., якщо при цьому дефекти не могли бути виявлені і усунені в період будівництва, монтажу та пробної роботи.

Аваріями через стихійних явищ вважаються тільки ті порушення режиму роботи устаткування, які не могли бути передбачені при проектуванні і будівництві водопроводу, а також не могли бути своєчасно попереджені персоналом підприємства.

До числа аварій і браку в роботі по іншим причинам відносяться: випадки механічних пошкоджень обладнання сторонніми особами, випадки, що сталися з вини абонентів, внаслідок неправильних дій персоналу або дефектів обладнання у абонентів; випадки, обумовлені виною енергопостачальної організації (міської електричної станції, ТЕЦ та ін.).

2.1. Цільовий аналіз призначення ІТ

Основним призначенням розроблюваної ІТ є підготовка статистичних даних до прогнозування аварійних ситуацій на водопровідних мережах міста. Дані отримані від користувачів (житлово-експлуатаційні контори) організовуватимуться та зберігатимуться у базу даних. Підготовка даних відбуватиметься за допомогою алгоритму кластеризації. Програмний комплекс розділятиме певний масив даних про аварії на деякі підгрупи, це збільшить точність прогнозу. Створюватимуться файли з даними, які надалі використовуватимуться у системі прогнозування.

Багаторівнева декомпозиція цілей створення та функціонування ІТ надана у графічному вигляді «Дерева цілей» та зображена на Рис 2.1.

Дерево цілей - це структурована, побудована за ієрархічним принципом (розподілена по рівнях, ранжируваних) сукупність цілей системи, в якій виділені генеральна мета («вершина дерева»); підпорядковані їй підцілі першого, другого і подальшого рівнів ("гілки дерева"). Назва «дерево цілей» пов'язане з тим, що схематично представлена сукупність розподілених по рівнях цілей нагадує по виду перевернуте дерево.

Метод дерева цілей орієнтований на одержання повної і відносно стійкої структури цілей, проблем, функцій, напрямків, тобто такої структури, яка мало змінюватиметься протягом певного періоду часу при неминучих змінах, що відбуваються в будь-якій системі, яка розвивається.

Цілі впливають з об'єктивних потреб і мають ієрархічний характер. Цілі верхнього рівня не можуть бути досягнуті, поки не досягнуті цілі найближчого нижнього рівня. В міру переміщення вниз рівнями ієрархії цілі конкретизуються. У процесі побудови та використання дерева цілей необхідно прагнути чіткого і конкретного формулювання цілей, забезпечити можливість кількісної чи порядкової оцінки ступеня їхнього досягнення.

Так зване дерево цілей тісно пов'язує між собою перспективні цілі та конкретні завдання на кожному рівні ієрархії. При цьому мета вищого порядку відповідає вершині дерева, а нижче в кілька ярусів розташовуються локальні цілі (завдання), за допомогою яких забезпечується досягнення цілей верхнього рівня.



Рисунок 2.1 Дерево цілей розроблюваної системи

1 Розробити базу даних. Створити архів статистичних даних які зберігатимуться на сервері. Такий архів значно полегшить вибірку даних для прогнозу. Оскільки база даних єдина для усіх користувачів, подібні умови аварій у різних районах можна використовувати для уточнення прогнозу за рахунок більшої вибірки.

1.1 Накопичити дані. За допомогою спеціальної форми, надаватиметься зручний та швидкий спосіб внесення даних про аварію до БД. Такі форми внесення даних прикріплені до кожної копії програми.

1.1.1 Аналіз аварії. Проводитиметься спеціально підготовленими робітниками, складатиметься протокол аварії, визначатиметься причина аварії, технічний стан об'єкту аварії та подальша придатність до експлуатації.

1.1.2 Заповнення реєстраційної картки аварії. Спеціально розроблена картка аварії забезпечуватиме просте розуміння даних аналізу аварії для подальшого їх внесення у базу даних за допомогою програмної форми спеціально підготовленими операторами.

1.2 Сорткування даних. В базі даних передбачений алгоритм для сорткування статистичних даних по декільком критеріям. Для виявлення та усунення можливих конфліктів та помилок спеціальна бригада підготовлених спеціалістів проводитиме моніторинг даних що вносяться до БД.

2 Кластеризація даних. Проводитиметься кластерний аналіз даних, та знаходження схожих груп об'єктів у вибірці. Дані розташовуватимуться у кластерах таким чином, що найбільш подібні об'єкти заходитимуться в одному кластері, тоді як об'єкти які істотно відрізняються знаходитимуться у різних кластерах.

2.1 Зведення даних до єдиної системи значень. Усі об'єкти бази даних характеризуються деяким набором якісних та кількісних показників. Для проведення кластеризації усі дані повинні бути зведені до єдиної системи значень, за допомогою деяких припущень та алгоритмів.

2.1.1 Приведення якісних даних. Якісні дані зводяться до вигляду при якому усі варіанти якісної ознаки об'єкту приймають значення 0 або 1. Якщо ознака дійсно має ідентичне значення, їй у відповідність назначається 1, іншим варіантам цієї ознаки назначається 0.

2.1.2 Приведення кількісних даних. Усі кількісні дані зводяться до єдиної системи (0..1). Обирається мінімальне й максимальне значення що приймає ознака й цим значенням у відповідність назначається 0 та 1. Ідентичним ознакам інших об'єктів назначається значення в діапазоні від 0 до 1.

2.2 Визначення відстані векторів характеристик. Дані об'єктів зводяться у вектори характеристик. За допомогою метрик визначається відстань між цими векторами.

2.3 Проведення кластеризації. За допомогою алгоритмів кластеризації та відстані між векторами характеристик кожний об'єкт приєднується до кластеру. У кожному кластері знаходяться об'єкти найбільш схожі між собою.

3 Використання програмного додатку для прогнозування аварійних ситуацій. Цей програмний додаток розраховуватиме прогноз на основі підбраного кластеру даних.

2.2. Функціональні моделі реалізації процесів ІТ

При проведенні структурного системного аналізу використовувалась поступова деталізація (декомпозиція) функціональної моделі інформаційної технології - «чорної скриньки» яка зображена на рис. 2.2.

Чорна скринька — це термін, який використовується у техніці й кібернетиці для позначення об'єкта чи системи, про принципи дії яких нічого невідомо, крім того, що певному вхідному сигналу відповідає певний вихідний сигнал.

Модель ЧС розкладається на сукупність моделей, що описують реалізацію функцій системи на різних рівнях ієрархії процесів (дерево процесів).

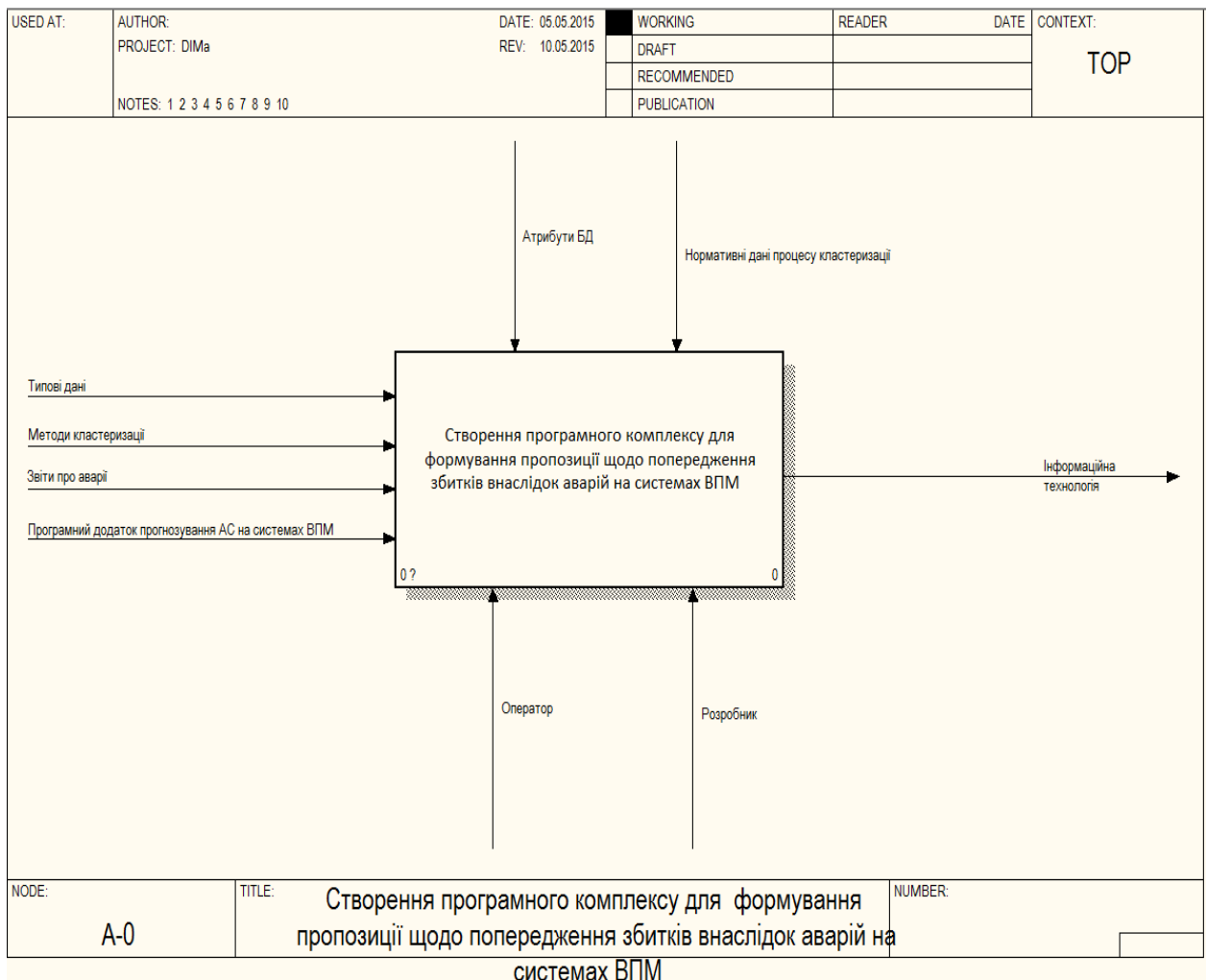


Рисунок 2.2 Чорна скринька системи

При побудові моделей декомпозиції використовувались SADT-моделі.

Основою SADT - технологій є створення і аналіз моделей систем та їх подальше використання.

SADT – моделі відповідають на запитання: як функціонує система, які процеси проходять в системі, як зробити, щоб система працювала найкращим способом і т. п.

Загальні характеристики IDEF0 моделі. Дана модель, як уже відомо з попереднього, описує функції системи, тобто як система досягає своїх цілей, які процеси в ній відбуваються, як ці процеси пов'язані між собою. Модель

являє собою деревовидну топологічну структуру і створюється на основі функціональної декомпозиції цілей та задач системи.

Декомпозиція моделі системи I рівня за допомогою функціональної моделі IDEF0 представлена на рис. 2.3, II рівня представлена на рис. 2.4 та рис. 2.5

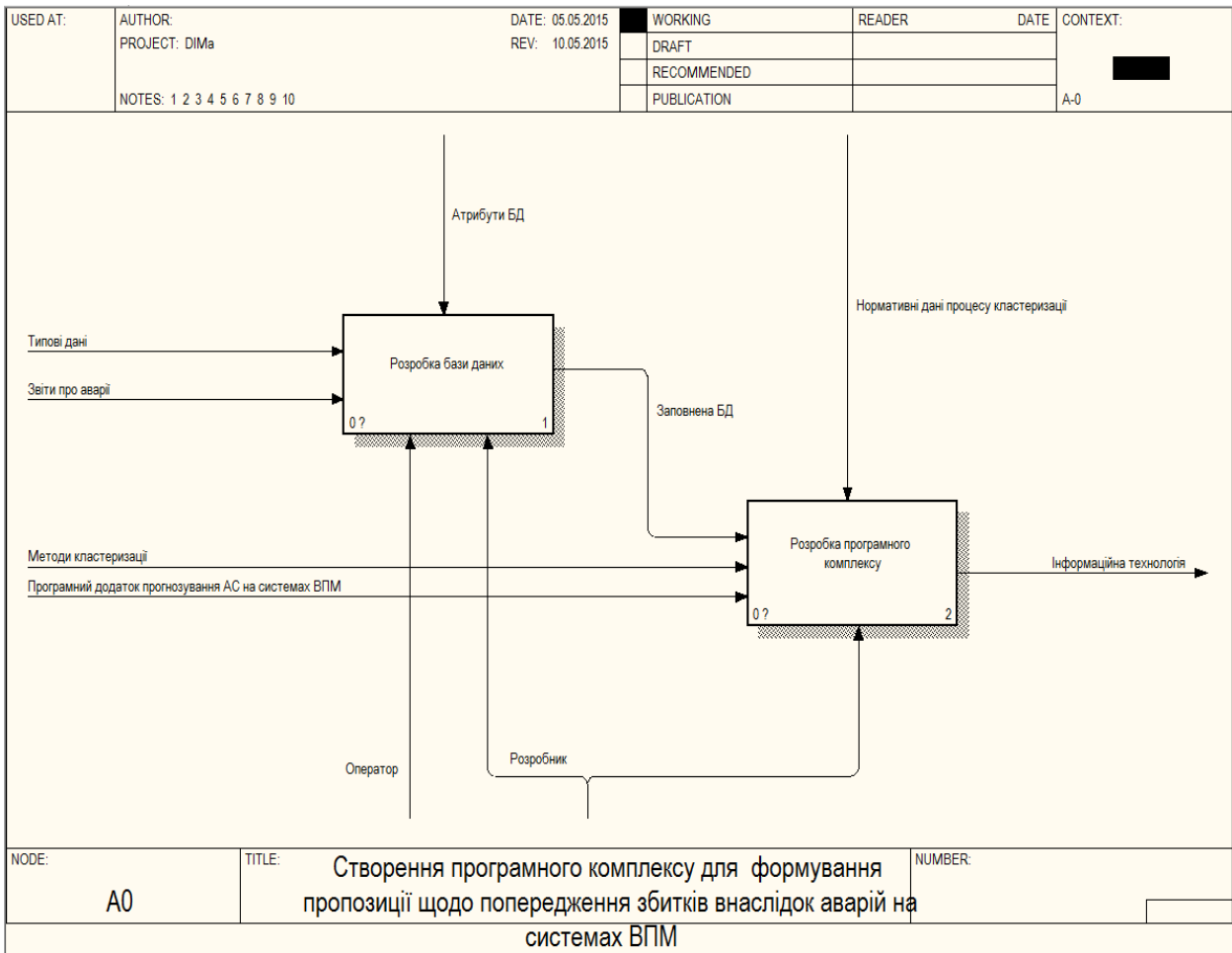


Рисунок 2.3 Декомпозиція моделі системи I рівня

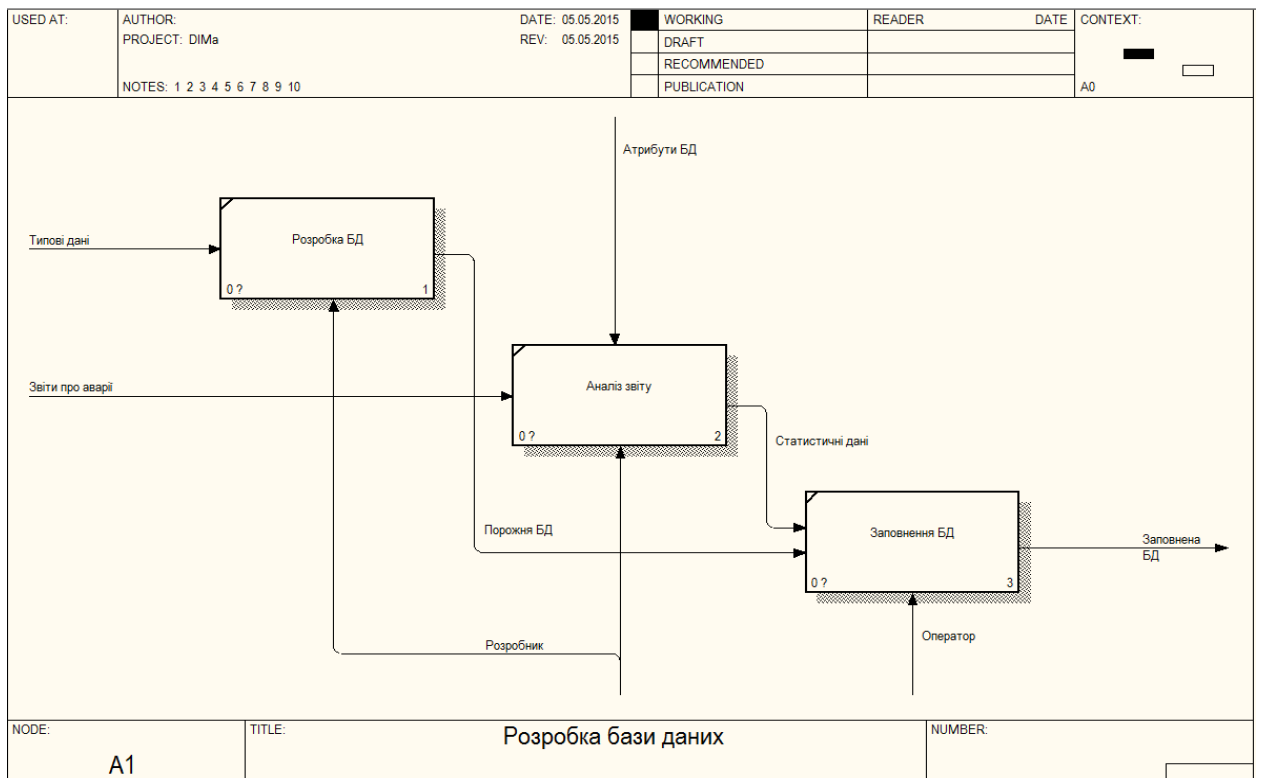


Рисунок 2.4 Декомпозиція моделі системи II рівня. Розробка БД.

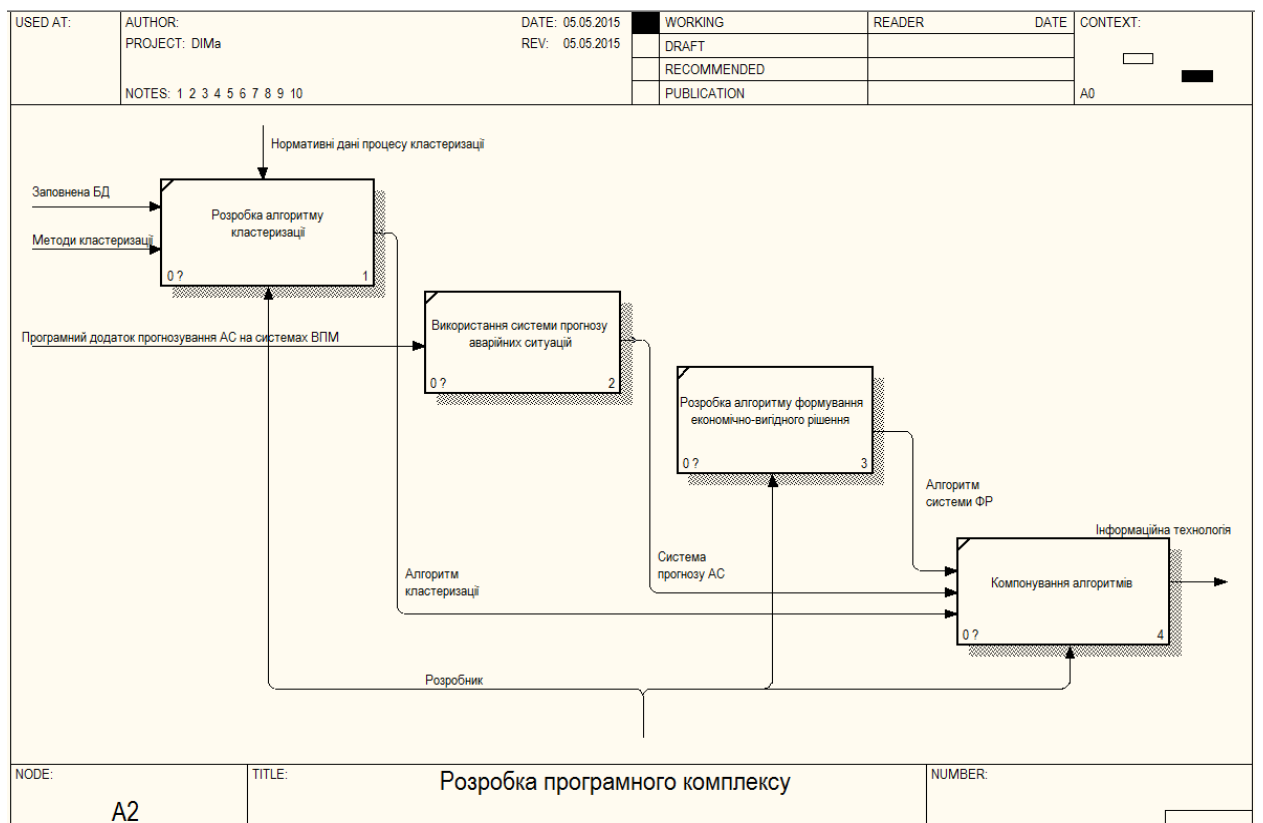


Рисунок 2.5 Декомпозиція моделі системи II рівня. Розробка програмного комплексу.

2.3. Постановка основних функціональних задач, що вирішуються у складі ІТ

2.3.1 Методи і форми збору інформації про випадки ушкодження на водопровідних мережах

Отримання вхідних даних для подальших розрахунків підсистеми є так званий «Реєстраційна картка пошкоджень на ділянці», приклад якої зображений на рис. 2.6. Він являє собою документ в якому чітко викладені такі дані:

Реєстраційна картка пошкоджень на ділянці		
1. Рік _____	2. Місяць _____	3. Число _____
4. Місце пошкодження _____		
5. Час пошкодження і ремонту:		
Час сповіщення про аварію _____	год. _____	хв. _____
Час прибуття бригади _____	год. _____	хв. _____
Час закінчення роботи _____	год. _____	хв. _____
6. Характеристика ділянки мережі (споруди) _____		

7. Причини ушкодження (очевидні і допустимі) _____		

8. Характер ремонтних робіт _____		

9. Додаткові відомості _____		

10. Карточку заповнив відповідальний за ремонт _____		

посада _____		
(підпис, ПІБ)		

Рисунок 2.6 - Реєстраційна картка пошкоджень на ділянці

Картку заповнюють на місці ушкодження працівники, що проводили ремонтні роботи. Відомості з карток потім переносять у журнал ушкоджень у хронологічному порядку інженерно-технічні працівники, у веденні яких перебуває ділянка ушкодженого трубопроводу або працівники диспетчерської служби підприємства комунального водопостачання. У формі введення даних відомості з карток можуть бути доповнені деякими даними, якими не розташовує аварійна бригада, наприклад, вказівками про строк експлуатації ушкодженої ділянки трубопроводу. Відомості про окремі ушкодження, занесені в реєстраційні картки, при досить тривалому строку спостережень служать вихідним матеріалом для визначення чисельних показників частоти й характеру ушкоджень трубопроводів, розрахунку прогнозу. Обробка отриманих статистичних даних дозволяє визначити число ушкоджень на ділянках труб різних діаметрів і матеріалу за певний строк спостережень. Будучи відповідним чином систематизовані, ці дані дозволяють обґрунтовано підійти до планування трудомісткості робіт з усунення ушкоджень, віднести їх до одиниці довжини трубопровідної мережі (наприклад, число ушкоджень на 100 км), розподілити по видах, рокам і протягом року, характеризуючи очікувану частоту появи ушкоджень. На цій основі з'являється можливість прогнозувати ймовірність виникнення ушкоджень на трубопроводах залежно від різних умов, сезонів, установлювати найбільш підданим ушкодженням ділянки труб і вчасно вживати заходів по усуненню й запобіганню причин ушкоджень. Одночасно на основі статистичних даних можуть бути оцінені нормативи тривалості проведення ремонтних робіт з усунення ушкоджень залежно від виду, характеру ушкодження й інших особливостей, а також обґрунтована потрібна чисельність персоналу аварійно-ремонтних служб і їхня оснащеність механізмами й технікою.

Дані, накопичені за період тривалої експлуатації по деяких системах комунального водопостачання, дозволяють у загальному виді характеризувати основні причини виникаючих ушкоджень і в такий спосіб

визначити головні напрямки технічної профілактики трубопровідних мереж для попередження аварії й ушкоджень.

На трубопровідній мережі систем - спостерігається кілька характерних повторюваних видів ушкоджень. На чавунних трубопроводах - це вихід закладення з розтрубів, переломи, тріщини, розриви труб і фасонних частин, свищі, розточення водою й піском стінок і розтрубів, руйнування труб від корозії й ін. На сталевих трубах розриви поздовжніх (заводського зварювання) і поперечних (монтажного зварювання) швів, а також руйнування труб від корозії. На азбоцементних труб-різного роду переломи.

Порушення герметичності стиків і вихід з них закладення (свинцю, цементу, азбоцементна, сіросплава) є часом наслідком низької якості застосовуваного каната й недотепної конопатки розтруба. Установлено, що трубопроводи, забиті недоброякісним канатом, дають найбільш високий відсоток розладу стиків. Якість свинцевого закладення знижується, якщо заливання виробляється недостатньо розігрітим свинцем, або розтруб заливається неповністю й потім свинець підливають удруге. Незалежно від застосовуваного матеріалу, який використовують при ремонті труби, ненадійними є розтрубні й муфтові стики чавунних труб зі сталевими. Пояснюється це різницею зовнішніх діаметрів сталевих і чавунних труб і муфт. Відсутність бурту на трубах і гладкій формі розтруба сталевих фасонних частин є однією із причин виходу закладення стиків трубопроводів, що мають сполучення сталевих і чавунних труб і фасонних частин.

2.3.2 Задачі що вирішуються у складі ІТ

2.3.2.1 Кластеризація даних

Кластеризація даних (англ. Data clustering) — задача розбиття заданої вибірки об'єктів (ситуацій) на підмножини, що називаються кластерами, так, щоб кожен кластер складався з схожих об'єктів, а об'єкти різних кластерів

істотно відрізнялися. Завдання кластеризації відноситься до статистичної обробки, а також до широкого класу завдань навчання без вчителя.

Визначення множини ознак, які покладаються в основу оцінки об'єктів $(x_1, x_2, x_3, \dots, x_n)$, у кластерному аналізі є одним із найважливіших завдань дослідження. Мета цього кроку повинна полягати у визначенні сукупності змінних ознак, яка найкраще відображає поняття подібності. Ці ознаки мають обиратися з урахуванням теоретичних положень, покладених в основу класифікації, а також мети дослідження.

При визначенні міри подібності об'єктів кластерного аналізу використовуються чотири види коефіцієнтів: коефіцієнти кореляції, показники віддалей, коефіцієнти асоціативності та ймовірності, коефіцієнти подібності. Найбільшого розповсюдження здобули коефіцієнти віддалей.

Існує безліч метрик обрахунку відстаней, основні з них:

Евклідова відстань

Евклідова відстань (Евклідова метрика) — формула традиційної відстані між двома точками для Евклідового простору:

$$\rho(x, x') = \sqrt{\sum_i^n (x_i - x'_i)^2} \quad (2.1)$$

Квадрат евклідової відстані

Застосовується для надання більшої ваги більш віддаленим один від одного об'єктам. Ця відстань обчислюється таким чином:

$$\rho(x, x') = \sum_i^n (x_i - x'_i)^2 \quad (2.2)$$

Відстань міських кварталів (Манхеттенська метрика)

Ця відстань є середньою різницею координат. У більшості випадків ця міра відстані призводить до таких же результатів, як і для звичайної відстані Евкліда. Однак для цього заходу вплив окремих великих різниць (викидів) зменшується (тому вони не зводяться в квадрат). Формула для розрахунку манхеттенської відстані:

$$\rho(x, x') = \sum_i^n |x_i - x'_i| \quad (2.3)$$

Відстань Чебишева

Ця відстань може виявитися корисною, коли потрібно визначити два об'єкти як «різні», якщо вони розрізняються за якоюсь однією координатою. Відстань Чебишева обчислюється за формулою:

$$\rho(x, x') = \max(|x_i - x'_i|) \quad (2.4)$$

Вибір метрики повністю є відповідальністю дослідника, оскільки результати кластеризації можуть істотно відрізнятися при використанні різних заходів.

2.3.2.2 Метод FCM для кластеризації даних

Метод FCM (Fuzzy c-means – нечітких c-середніх) для вирішення задачі нечіткої кластеризації має ітеративний характер послідовного поліпшення певного вихідного нечіткого розбиття $R(A) = \{A_v | A_v \dot{A}\}$, що задається користувачем або формується автоматично за певним евристичним правилом. На кожній з ітерацій рекурентно перераховуються значення функцій приналежності нечітких кластерів та їхніх типових представників.

Метод FCM закінчить роботу у випадку, коли відбудеться виконання заданого апріорі деякого кінцевого числа ітерацій, або коли мінімальна

абсолютна різниця між значеннями функцій приналежності на двох послідовних ітераціях не стане менше деякого апріорі заданого значення.

Крок 1. Попередньо необхідно задати такі значення: кількість шуканих нечітких кластерів V ($V > 1$), максимальну кількість ітерацій методу Epochs, параметр збіжності методу ε , а також експонентну вагу розрахунку цільової функції і центрів кластерів m (як правило, $m = 2$). Як поточне нечітке розбиття на першій ітерації методу для вибірки даних x задати деяке вихідне нечітке розбиття

$$R(A) = \{A_v | A_v \subseteq A\} \quad (2.5)$$

на V непорожніх нечітких кластерів, що описуються сукупністю функцій приналежності U_{sv} , $s = 1, 2, \dots, S$; $v = 1, 2, \dots, V$. Нечітке розбиття отримують шляхом генерації випадковим чином елементів U_v , що задовольняють умовам цільової функції.

Крок 2. Для вихідного поточного нечіткого розбиття

$$R(A) = \{A_v | A_v \subseteq A\} \quad (2.6)$$

розрахувати центри нечітких кластерів C_{vj} , $v = 1, 2, \dots, V$; $j = 1, 2, \dots, N$ та значення цільової функції $J(x, u, C)$. Кількість виконаних ітерацій покласти рівною 1.

Крок 3. Сформувати нове нечітке розбиття

$$R'(A) = \{A_v | A_v \subseteq A\} \quad (2.7)$$

вихідної множини об'єктів кластеризації A на V непорожніх нечітких кластери, що характеризуються сукупністю функцій приналежності $U_s'v$, $v = 1, 2, \dots, V$, $X_s \in A$, які визначаються за формулою:

$$u_v^s = \left(\sum_{g=1}^V \left(\frac{d(x^s, C^v)}{d(x^s, C^g)} \right)^{\frac{2}{m-1}} \right)^{-1}, v = 1, 2, \dots, V. \quad (2.8)$$

Крок 4. При цьому якщо для деякого v та деякого X_s значення $d(X_s, C_v)=0$, то для відповідного нечіткого кластера встановимо $U_s^v = 1$, а для інших кластерів $U_h^g = 0$, $g = 1, 2, \dots, V$; $h = 1, 2, \dots, S$. Якщо ж таких v для деякого X_s виявиться декілька, тобто для них значення $d(X_s, C_v)=0$, то евристично для меншого з v встановимо $U_s^v = 1$, а для інших встановимо $U_h^g = 0$, $g = 1, 2, \dots, V$; $h = 1, 2, \dots, S$.

Крок 5. Для нового нечіткого розбиття

$$R'(A) = \{A_v | A_v \subseteq A\} \quad (2.9)$$

розрахувати центри нечітких кластерів C_v та значення цільової функції $J(x, u, C)$.

Крок 6. Якщо кількість виконаних ітерацій перевищує задане число E_{procs} або ж модуль різниці $|J(x, u, C) - J^*(x, u, C)| \leq \varepsilon$, то за шуканий результат нечіткої клас-тернзації прийняти нечітке розбиття $R'(A)$ і закінчити виконання методу. У протилежному випадку вважати поточним нечітким розбиттям $R(A) = R'(A)$ і перейти до кроку 3 методу, збільшивши на 1 кількість виконаних ітерацій.

Вибір кількості кластерів V є однією з найважливіших проблем у розглянутому методі. Правильно вибрати кількість кластерів для реальних задач без будь-якої апріорної інформації про структуру даних досить складно. Існує два формальних підходи до вибору кількості кластерів. Перший підхід заснований на критерії компактності та роздільності отриманих кластерів. Логічно припустити, що при правильному виборі кількості кластерів дані будуть розбиті на компактні і добре віддільні одна від іншої групи. У іншому випадку, кластери, імовірно, не будуть

компактними і добре віддільними Існує кілька критеріїв оцінки компактності кластерів, однак питання про те, як формально і вірогідно визначити правильність вибору кількості кластерів для довільного набору даних залишається відкритим. Для методу FCM рекомендується використовувати *індекс Хіє-Бені (Xie-Beni index)*:

$$\chi = \frac{\sum_{v=1}^V \sum_{s=1}^S (u_v^s)^m d^2(x^s, C^v)}{S \min d^2(x^s, C^v)}. \quad (2.10)$$

Другий підхід заснований на *редукції кількості кластерів* і пропонує починати кластеризацію при досить великій кількості кластерів, а потім послідовно поєднувати схожі суміжні кластери.

Діагональна норма дозволяє виділяти кластери у вигляді гіпереліпсоїдів, орієнтованих уздовж координатних осей. При цьому використовуються різні формальні критерії схожості кластерів. Для діагональної норми матриця B задається у такий спосіб:

$$B_{ij} = \begin{cases} w_i, i = j; \\ 0, i \neq j, \end{cases} \quad (2.11)$$

де w_i – елементи головної діагоналі матриці, що інтерпретуються як ваги координат.

Норма Махаланобіса дозволяє виділяти кластери у вигляді гіпереліпсоїдів, вісі яких можуть бути орієнтовані в довільних напрямках. Для норми Махаланобіса матриця B розраховується через коваріаційну матрицю від x :

$$B = G^{-1}, \quad G = \frac{1}{S} \sum_{s=1}^S (x^s - \bar{x})^T (x^s - \bar{x}), \quad \bar{x} = \frac{1}{S} \sum_{s=1}^S x^s, \quad (2.12)$$

де G – коваріаційна матриця; $\bar{x}$ – вектор середніх значень даних.

Приклади ізоліній різних норм показані на рис 2.7.

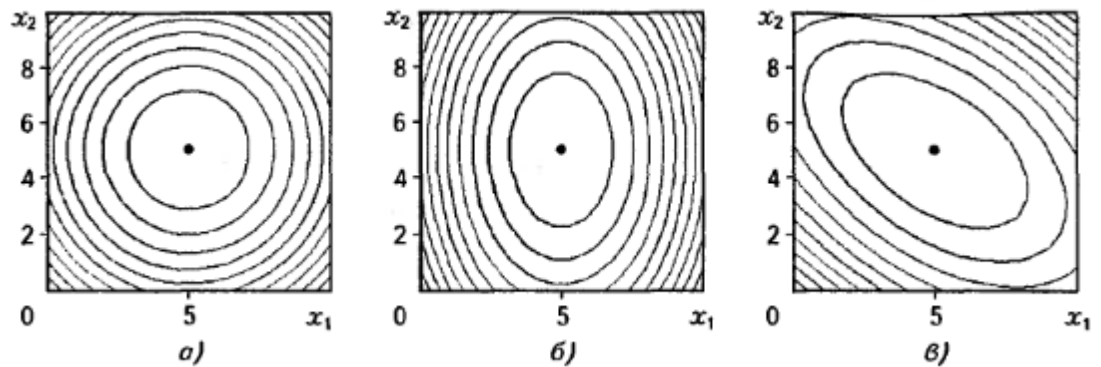


Рисунок 2.7 – Ізолінії різних норм

a – евклідова норма; b – діагональна норма; v – норма Махаланобіса

На рис. 2.8 наведений приклад кластеризації методом нечітких с-середніх при відстані Евкліда: зліва показані об'єкти кластеризації; справа показані результати нечіткої кластеризації. Центри нечітких кластерів позначені символами «+»

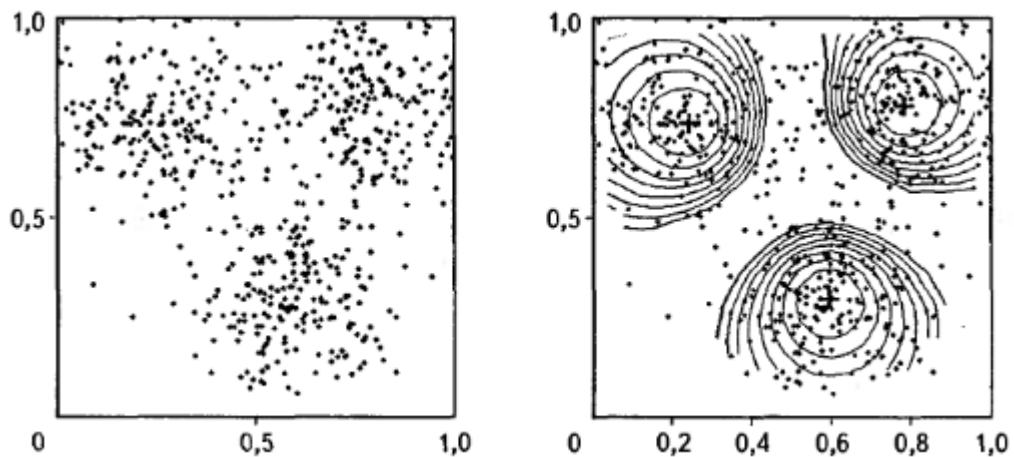


Рисунок 2.8 – Нечітка кластеризація при евклідовій нормі

Кластеризація даних включає десятки алгоритмів, кожний з яких використовується в залежності від необхідних результатів. Оскільки наша система передбачає кластеризацію записів бази даних, на невідому кількість кластерів, використовувати алгоритми в яких потребується попереднє зазначення їх кількості неможливо. У результаті аналізу методів які дозволяють алгоритмічно визначити кластери та їх кількість було обрано для подальшого аналізу методи «Гірської кластеризації» та «С - середніх».

3.1 Ціль проведення кластерного аналізу даних

Мета кластерного аналізу полягає в пошуку наявних структур, що виражається в утворенні груп схожих між собою об'єктів – кластерів. Водночас його дія полягає й у привнесенні структури в досліджувані об'єкти. Це означає, що методи кластеризації необхідні для виявлення структури в даних, яку нелегко знайти при візуальному обстеженні або за допомогою експертів.

Основні завдання кластерного аналізу зображені на рис. 3.1:

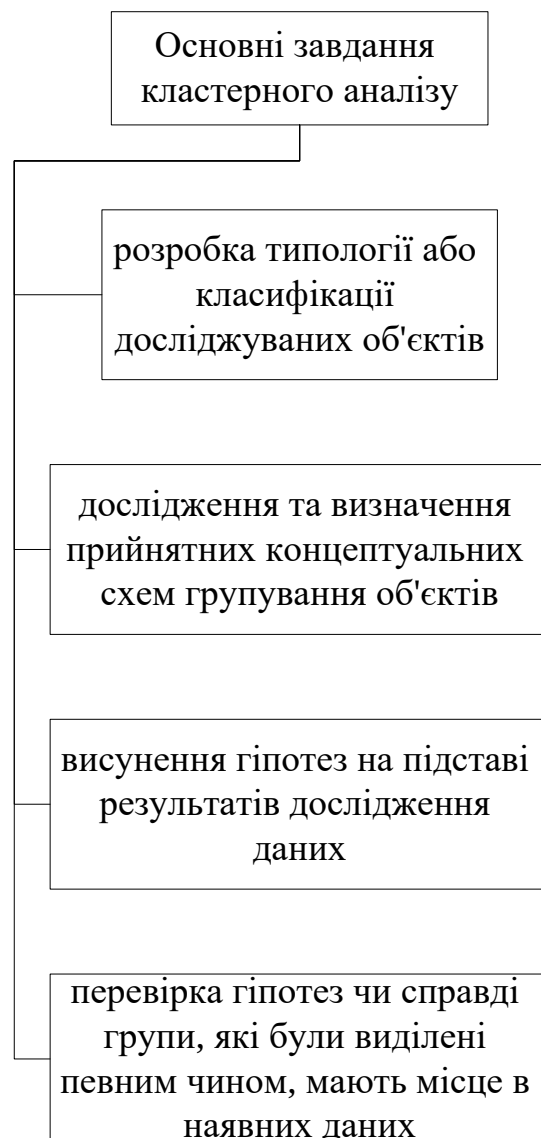


Рисунок 3.1 Основні завдання кластерного аналізу

Основна ціль проведення кластеризації у системі прогнозування аварійних ситуацій є:

- Розуміння даних шляхом виявлення кластерної структури. Розбиття вибірки на групи схожих об'єктів дозволяє спростити подальшу обробку даних і прийняття рішень, застосовуючи до кожного кластера свій метод аналізу.

Доволі об'ємна вибірка неструктурованих об'єктів може бути розподілена на кластери меншого розміру, крім того проведення розрахунку прогнозу використовуючи усю вибірку може негативно вплинути на отриманий результат.

Наприклад розраховуючи прогноз за одним типом аварій внесення у розрахунок даних що не є подібними але належать до вибірки може призвести до величезних похибок у результатах. Тобто розбиття об'єктів на групи в яких знаходитимуться схожі об'єкти дозволить більш точно провести прогноз майбутніх аварійних ситуацій.

Для аналізу кожної окремої аварії необхідно використовувати окремий, найкращий для прогнозу на основі неї алгоритм. При великих розмірах вибірки об'єктів даних підбір алгоритму для кожного об'єкту може зайняти доволі тривалий проміжок часу, й такий підбір є доволі вимогливим до ресурсів обчислювальної машини.

Оскільки в кластерах знаходяться об'єкти подібного типу справедливо припустити що для них є доцільним використовувати той самий алгоритм прогнозу. Це надзвичайно знизить час роботи алгоритму.

3.2 Об'єкти кластеризації та їх атрибути

Набір даних які використовуватимуться для кластеризації зберігатимуться у базі даних. Для подальшого розгляду інформаційного забезпечення даної системи, необхідно виділити ті інформаційні об'єкти, які потрібні для побудови таблиць на основі яких проводитиметься кластеризація. Проводитиметься вибірка необхідних атрибутів, які визначатимуть відстань(схожість) об'єктів. Для кожного атрибуту вибірки необхідно провести нормалізацію. Початкова структура вхідних даних зображена на рис 3.2.

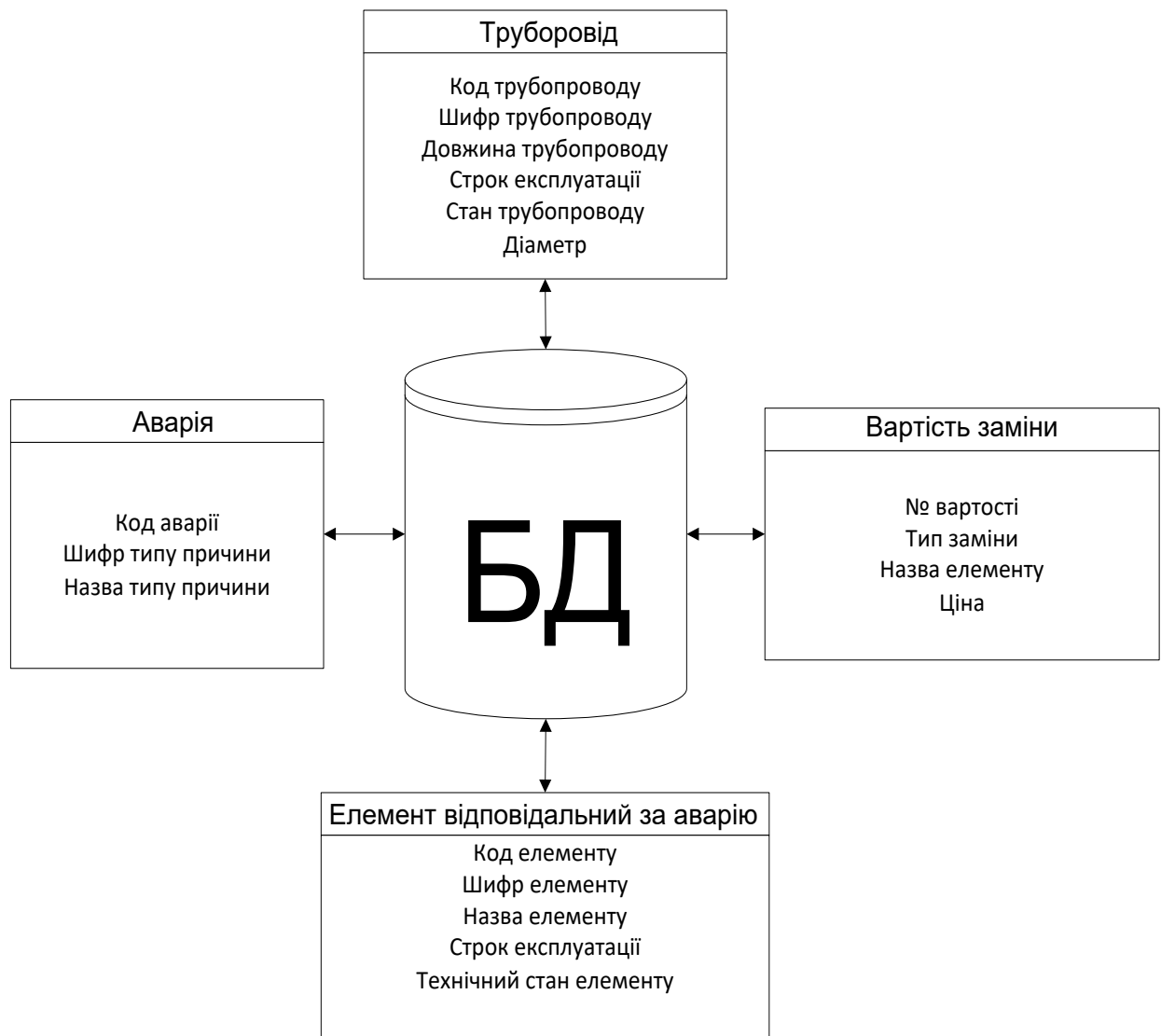


Рисунок 3.2 Схема бази даних

Для кластеризації використовуватимуться такі таблиці та атрибути:

Трубопровід:

- Довжина трубопроводу
- Строк експлуатації
- Стан трубопроводу
- Діаметр

Елементи відповідальні за аварію:

- Строк експлуатації елементу
- Технічний стан елементу

3.3 Нормалізація атрибутів об'єктів кластеризації

Існує багато підходів для нормалізації даних. Для кластеризації даних найкращим методом є нормалізація мінімуму-максимуму. Для даної нормалізації використовується наступна формула:

$$X^* = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (3.1)$$

в даному випадку X^* — це нормалізоване значення, $\min$ і $\max$ — мінімальне і максимальне значення яке приймає атрибут об'єкту.

Приклад нормалізації атрибутів які приймають числові значення зазначені в табл. 3.1:

Таблиця 3.1 - Об'єкти та значення атрибуту

	A1	A2	A3	A4
x1	174	835	372	725

A – об'єкти, x1– атрибут нормалізацію якого необхідно провести для кожного об'єкту.

За формулою 3.1:

$$\frac{372-174}{835-174} = 0.3$$

$$\frac{725-174}{835-174} = 0.83$$

Після нормалізації таблиці 3.1, дані приймають значення зазначені у табл. 3.2:

Таблиця 3.2 - Значення атрибуту об'єктів після нормалізації

	A1	A2	A3	A4
x1	0	1	0.3	0.83

Після проведення нормалізації для кожного атрибуту об'єкту вони вкладаються однаковий вклад до визначення відстані між об'єктами.

Але крім числових значень атрибуту також приймають категоріальні значення, наприклад колір. Для нормалізації таких атрибутів використовують *нечіткі змінні*.

Нехай U – універсальна множина, u – елемент U , а G – деяка властивість. Звичайна (чітка) підмножина A універсальної множини U , елементи якої мають властивість G , визначається як множина впорядкованих пар $\{ \langle \mu_A(u) | u \rangle \}$, де $\mu_A(u)$ – характеристична функція приналежності, що приймає значення 1, якщо u має властивість G , та 0 – у протилежному випадку.

Нечітка підмножина відрізняється від звичайної тим, що для елементів u з U немає однозначної відповіді «ні» або «так» щодо властивості G .

У зв'язку з цим нечітка підмножина A універсальної множини U визначається як множина впорядкованих пар $A = \{ \langle \mu_A(u) | u \rangle \}$, де $\mu_A(u)$ – характеристична функція приналежності (або просто функція приналежності), що приймає значення в деякій цілком впорядкованій множині M (наприклад, $M = [0; 1]$).

Функція приналежності вказує ступінь приналежності елемента u нечіткій підмножині A . Множину M називають множиною приналежностей. Якщо $M=\{0, 1\}$, то нечітка підмножина A може розглядатися як чітка множина. Чітку множину A можна розглядати як граничний випадок нечіткої множини A , функція приналежності якої $\mu_A(u)$ набуває лише бінарних значень.

Приклад представлення нечіткої змінної:

Представити у вигляді нечіткої множини поняття «Середня довжина труби») на універсальній множині $U=\{155,160,165,170,175,180,185,190\}$. Одне з можливих рішень виглядає так:

$$A=(0/155, 0,1/160, 0,3/165, 0,8/170, 1/175, 1/180, 0,5/185, 0/190).$$

Нечітка змінна визначається як $\langle a, U, A \rangle$, де a – найменування змінної, $U=\{u\}$ – область визначення змінної (набір можливих значень u), $A=\{\langle \mu_A(u)|u \rangle\}$ – нечітка множина, що описує обмеження на можливі значення змінної a (семантику).

Нечітка змінна – це теж саме, що і нечітке число, тільки з додаванням імені, яким формалізується поняття, що описується цим числом. Для людини зручніше задавати значення змінної не числами, а словами. Щодня ми приймаємо рішення на основі лінгвістичної інформації типу: «дуже висока температура»; «утомлива поїздка»; «швидка відповідь»; «красивий букет»; «гармонійний смак» і тому подібне.

Також над нечіткими змінними може проводитися процес «Дефаззифікації» (defuzzification) - процедура перетворення нечіткої множини в чітке число.

У теорії нечітких множин процедура дефаззифікації аналогічна знаходження характеристик становища (математичного очікування, моди, медіани) випадкових величин в теорії ймовірності.

Основним для цього процесу є алгоритм дефаззифікації нечіткої множини

$$\tilde{A} = \int_{[u, \bar{u}]} \mu_A(u) / u \quad (3.2)$$

за методом центру ваги, здійснюється за формулою:

$$a = \frac{\int_{\underline{u}}^{\bar{u}} u \cdot \mu_A(u) du}{\int_{\underline{u}}^{\bar{u}} \mu_A(u) du} \quad (3.3)$$

Фізичним аналогом цієї формули є знаходження центра ваги плоскої фігури, обмеженої осями координат і графіком функції приналежності нечіткої множини. У разі дискретної універсальної множини дефаззифікації нечіткої множини

$$\tilde{A} = \sum_{i=1}^k \mu_A(u_i) / u_i \quad (3.4)$$

за методом центру ваги здійснюється за формулою:

$$a = \frac{\sum_{i=1}^k u_i \cdot \mu_A(u_i)}{\sum_{i=1}^k \mu_A(u_i)} \quad (3.5)$$

Приклад процесу дефаззифікації за формулою 3.5:

$$a = \frac{0 \cdot 155 + 0.1 \cdot 160 + 0.3 \cdot 165 + 0.8 \cdot 170 + 1 \cdot 175 + 1 \cdot 180 + 0.5 \cdot 185 + 0 \cdot 190}{0 + 0.1 + 0.3 + 0.8 + 1 + 1 + 0.5 + 0} = 175.4$$

Таким чином кожне значення атрибутів об'єктів можна звести до вигляду де усі змінні прийматимуть значення в діапазоні від 0 до 1. Це забезпечить однаковий вклад змінних до розрахунку відстані між об'єктами, та їх подальшої кластеризації.

3.4 Алгоритм кластеризації об'єктів

Для нечіткої кластеризації даних використовують два основних алгоритми кластеризації: «Нечітких с-середніх» та «Гірської кластеризації».

Їх принципова різниця заключається у тому що для першого необхідна кількість кластерів на яку необхідно розділити вибірку, а для другого умова зупинки кластеризації.

3.4.1 Алгоритм нечітких с-середніх

Крок 1. Встановити параметри алгоритму: c - кількість кластерів; m - експонентний вага; ϵ - Параметр зупину алгоритму.

Крок 2. Випадковим чином згенерувати матрицю нечіткого розбиття F , що задовольняє умови :

$$\sum_{i=1,c} \mu_{ki} = 1, \quad k = \overline{1,M}$$
$$0 < \sum_{k=1,M} \mu_{ki} < N, \quad i = \overline{1,c}$$

Крок 3. Розрахувати центри кластерів:

$$V_i = \frac{\sum_{k=1,N} (\mu_{ki})^m \cdot X_k}{\sum_{k=1,N} (\mu_{ki})^m}, \quad i = \overline{1,c}. \quad (3.6)$$

Крок 4. Розрахувати відстані між об'єктами з X і центрами кластерів:

$$D_{ki} = \sqrt{\|X_k - V_i\|^2}, \quad k = \overline{1,M}, i = \overline{1,c}. \quad (3.7)$$

Крок 5. Перерахувати елементи матриці нечіткого розбиття

$(i = \overline{1, c}, k = \overline{1, M})$:

$$\mu_{ki} = \frac{1}{\left(D_{ik}^2 \cdot \sum_{j=\overline{1, c}} \frac{1}{D_{jk}^2} \right)^{1/(m-1)}} ; \quad \text{якщо } D_{ki} > 0 ; \quad (3.8)$$

$$\mu_{kj} = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases}, \quad j = \overline{1, c} . \quad \text{якщо } D_{ki} = 0 ; \quad (3.9)$$

Крок 6. Перевірити умову $\|F - F^*\|^2 < \varepsilon$, де F^* - матриця нечіткого розбиття на попередній ітерації алгоритму. Якщо "так", то перейти до кроку 7, інакше - до кроку 3.

Крок 7. Кінець.

У наведеному алгоритмі найважливішим параметром є кількість кластерів (c). Правильно вибрати кількість кластерів для реальних завдань без будь-якої апіорної інформації про структури в даних досить складно. Існує два формальних підходу до вибору числа кластерів.

Перший підхід заснований на критерії компактності і розділимості отриманих кластерів. Логічно припустити, що при правильному виборі кількості кластерів дані будуть розбиті на компактні і хороші віддільні один від одного групи. В іншому випадку, кластери, ймовірно, не будуть компактними і добре віддільними.

3.4.2 Основні ідеї методу гірської кластеризації

На першому кроці гірської кластеризації визначають точки, які можуть бути центрами кластерів. На другому кроці для кожної такої точки розраховується значення потенціалу, що показує можливість формування кластера в її околиці. Чим щільніше розташовані об'єкти в околиці потенційного центру кластера, тим вище значення його потенціалу. Після

цього ітераційно вибираються центри кластерів серед точок з максимальними потенціалами.

Алгоритм гірської кластеризації:

На першому кроці необхідно сформувати потенційні центри кластерів. Для алгоритму гірської кластеризації число потенційних центрів кластерів (Q) має бути кінцевим. Ними можуть бути об'єкти кластеризації (рядки матриці $\mathbf{X}$), тоді $Q = M$.

На другому кроці алгоритму розраховується потенціал центрів кластерів за такою формулою:

$$P(Z_h) = \sum_{k=1, \overline{M}} \exp(-\alpha \cdot D(Z_h, X_k)) \quad , \quad h = \overline{1, Q} \quad (3.10)$$

де $Z_h = (z_{1,h}, z_{2,h}, \dots, z_{n,h})$ - потенційний центр h -го кластера, $h = \overline{1, Q}$;

α - Позитивна константа

$D(Z_h, X_k)$ - Відстань між потенційним центром кластера (Z_h) і об'єктом кластеризації (X_k). В евклідовому просторі це відстань розраховується за формулою:

$$D(Z_h, X_k) = \sqrt{\|Z_h - X_k\|^2} \quad (3.11)$$

У випадку, коли об'єкти кластеризації задані двома ознаками ($n = 2$), графічне зображення розподілу потенціалу буде являти собою поверхню, що нагадує гірський рельєф. Звідси й назва - гірський метод кластеризації.

На третьому кроці алгоритму в якості центрів кластерів вибирають координати "гірських" вершин. Для цього, центром першого кластера призначають точку з найбільшим потенціалом. Зазвичай, найвища вершина оточена декількома досить високими піками.

Тому призначення центром наступного кластера точки з максимальним потенціалом серед решти вершин призвело б до виділення великої кількості близько розташованих центрів кластерів. Щоб вибрати наступний центр кластера необхідно спочатку виключити вплив тільки що знайденого кластеру.

Для цього значення потенціалу для решти можливих центрів кластерів перераховується таким чином: від поточних значень потенціалу віднімають внесок центру тільки що знайденого кластера (тому кластеризацію за цим методом іноді називають субтрактивною).

Перерахунок потенціалу відбувається за формулою:

$$P_2(Z_h) = P_1(Z_h) - P_1(V_1) \cdot \exp(-\beta \cdot D(Z_h, V_1)) , \quad (3.12)$$

де $P_1()$ - потенціал на 1-й ітерації;

$P_2()$ - Потенціал на 2-й ітерації;

V_1 - Центр першого знайденого кластера:

$$V_1 = \arg \max_{Z_1, Z_2, \dots, Z_Q} (P_1(Z_1), P_1(Z_2), \dots, P_1(Z_Q)) \quad (3.13)$$

β - Позитивна константа.

Центр другого кластера визначається за максимальним значенням оновленого потенціалу:

$$V_2 = \arg \max_{Z_1, Z_2, \dots, Z_Q} (P_2(Z_1), P_2(Z_2), \dots, P_2(Z_Q)) \quad (3.14)$$

Потім знову перераховується значення потенціалів:

$$P_3(Z_h) = P_2(Z_h) - P_2(V_2) \cdot \exp(-\beta \cdot D(Z_h, V_2)) \quad (3.15)$$

Ітераційна процедура перерахунку потенціалів і виділення центрів кластерів продовжується до тих пір, поки максимальне значення потенціалу перевищує деякий поріг.

Оскільки при сортуванні об'єктів можлива величезна кількість кластерів доцільніше використовувати розбиття на невідому заздалегідь їх кількість, щоб уникнути використання для розрахунку прогнозу даних, що можуть призвести до похибок у результатах.

В результаті аналізу методів, для кластеризації об'єктів даних про аварійні ситуації на системах водопостачання міста був обраний метод «Гірської кластеризації».

3.4 Приклад роботи алгоритму «Гірської кластеризації» об'єктів

Далі представлений приклад роботи алгоритму «Гірської кластеризації». Проводитиметься кластеризація об'єктів атрибути яких представлених у таблиці 3.3:

Таблиця 3.3 - Типові об'єкти та їх атрибути

Номер об'єкту	Довжина трубопроводу	Строк експлуатації трубопроводу	Стан трубопроводу	Діаметр трубопроводу	Строк експлуатації елементу	Технічний стан елементу
1	12	17	задовільно	50	13	незадовільно
2	52	14	незадовільно	100	7	задовільно

Продовження таблиці 3.3

3	26	18	Близьке до задовільного	200	10	Близьке до задовільного
4	15	15	незадовільно	100	15	Близьке до задовільного
5	26	17	задовільно	50	15	незадовільно
6	73	17	Близьке до задовільного	500	12	Близьке до задовільного
7	46	18	Близьке до	50	6	задовільно

			задовільного			
8	51	19	Близьке до незадовільно	100	16	задовільно
9	24	26	незадовільно	50	25	Близьке до незадовільно
10	14	28	задовільно	500	25	незадовільно
11	17	24	незадовільно	500	16	задовільно
12	37	22	Близьке до задовільного	200	12	Близьке до незадовільно
13	72	18	задовільно	100	13	Близьке до задовільного
14	52	25	Близьке до незадовільно	50	21	задовільно
15	35	31	задовільно	500	31	незадовільно
16	65	12	задовільно	500	11	Близьке до задовільного
17	35	31	Близьке до незадовільно	100	21	Близьке до незадовільно
18	42	29	Близьке до незадовільно	50	24	задовільно
19	48	34	задовільно	200	21	незадовільно
20	27	14	незадовільно	50	7	задовільно

Першим кроком при кластеризації даних є нормалізація змінних, для цього ми розраховуємо нову таблицю. Для нечітких атрибутів ми приймаємо значення:

- Задовільно – 0
- Близьке до задовільного – 0.3
- Близьке до незадовільно – 0.7
- Незадовільно – 1

Атрибути що приймають числові значення ми розраховуємо за формулою 3.1.

Довжина трубопроводу:

Максимальне значення – 73, мінімальне – 12.

- | | |
|------------------------------------|------------------------------------|
| 1. $(12 - 12) / (73 - 12) = 0$ | 11. $(17 - 12) / (73 - 12) = 0.08$ |
| 2. $(52 - 12) / (73 - 12) = 0.66$ | 12. $(37 - 12) / (73 - 12) = 0.41$ |
| 3. $(26 - 12) / (73 - 12) = 0.23$ | 13. $(72 - 12) / (73 - 12) = 0.98$ |
| 4. $(15 - 12) / (73 - 12) = 0.05$ | 14. $(52 - 12) / (73 - 12) = 0.66$ |
| 5. $(26 - 12) / (73 - 12) = 0.23$ | 15. $(35 - 12) / (73 - 12) = 0.38$ |
| 6. $(73 - 12) / (73 - 12) = 1$ | 16. $(65 - 12) / (73 - 12) = 0.87$ |
| 7. $(46 - 12) / (73 - 12) = 0.56$ | 17. $(35 - 12) / (73 - 12) = 0.38$ |
| 8. $(51 - 12) / (73 - 12) = 0.64$ | 18. $(42 - 12) / (73 - 12) = 0.49$ |
| 9. $(24 - 12) / (73 - 12) = 0.2$ | 19. $(48 - 12) / (73 - 12) = 0.59$ |
| 10. $(14 - 12) / (73 - 12) = 0.03$ | 20. $(37 - 12) / (73 - 12) = 0.37$ |

Строк експлуатації трубопроводу:

Максимальне значення – 34, мінімальне – 12.

- | | |
|-----------------------------------|------------------------------------|
| 1. $(17 - 12) / (34 - 12) = 0.23$ | 9. $(26 - 12) / (34 - 12) = 0.64$ |
| 2. $(14 - 12) / (34 - 12) = 0.09$ | 10. $(28 - 12) / (34 - 12) = 0.73$ |
| 3. $(18 - 12) / (34 - 12) = 0.27$ | 11. $(24 - 12) / (34 - 12) = 0.55$ |
| 4. $(15 - 12) / (34 - 12) = 0.14$ | 12. $(22 - 12) / (34 - 12) = 0.46$ |
| 5. $(17 - 12) / (34 - 12) = 0.23$ | 13. $(18 - 12) / (34 - 12) = 0.27$ |
| 6. $(17 - 12) / (34 - 12) = 0.23$ | 14. $(25 - 12) / (34 - 12) = 0.59$ |
| 7. $(18 - 12) / (34 - 12) = 0.27$ | 15. $(31 - 12) / (34 - 12) = 0.86$ |
| 8. $(19 - 12) / (34 - 12) = 0.32$ | 16. $(12 - 12) / (34 - 12) = 0$ |

$$17. (31 - 12) / (34 - 12) = 0.86$$

$$19. (34 - 12) / (34 - 12) = 1$$

$$18. (29 - 12) / (34 - 12) = 0.77$$

$$20. (14 - 12) / (34 - 12) = 0.09$$

Діаметр трубопроводу:

Максимальне значення – 500, мінімальне – 50.

$$1. (50 - 50) / (500 - 50) = 0$$

$$11. (500 - 50) / (500 - 50) = 1$$

$$2. (100 - 50) / (500 - 50) = 0.11$$

$$12. (200 - 50) / (500 - 50) = 0.33$$

$$3. (200 - 50) / (500 - 50) = 0.33$$

$$13. (100 - 50) / (500 - 50) = 0.11$$

$$4. (100 - 50) / (500 - 50) = 0.11$$

$$14. (50 - 50) / (500 - 50) = 0$$

$$5. (50 - 50) / (500 - 50) = 0$$

$$15. (500 - 50) / (500 - 50) = 1$$

$$6. (500 - 50) / (500 - 50) = 1$$

$$16. (500 - 50) / (500 - 50) = 1$$

$$7. (50 - 50) / (500 - 50) = 0$$

$$17. (100 - 50) / (500 - 50) = 0.11$$

$$8. (100 - 50) / (500 - 50) = 0.11$$

$$18. (50 - 50) / (500 - 50) = 0$$

$$9. (50 - 50) / (500 - 50) = 0$$

$$19. (200 - 50) / (500 - 50) = 0.33$$

$$10. (500 - 50) / (500 - 50) = 1$$

$$20. (50 - 50) / (500 - 50) = 0$$

Строк експлуатації елементу:

Максимальне значення – 31, мінімальне – 6.

$$1. (13 - 6) / (31 - 6) = 0.28$$

$$5. (15 - 6) / (31 - 6) = 0.36$$

$$2. (7 - 6) / (31 - 6) = 0.04$$

$$6. (12 - 6) / (31 - 6) = 0.24$$

$$3. (10 - 6) / (31 - 6) = 0.16$$

$$7. (6 - 6) / (31 - 6) = 0$$

$$4. (15 - 6) / (31 - 6) = 0.36$$

$$8. (16 - 6) / (31 - 6) = 0.4$$

$$9. (25 - 6) / (31 - 6) = 0.76$$

$$15. (31 - 6) / (31 - 6) = 1$$

$$10. (25 - 6) / (31 - 6) = 0.76$$

$$16. (11 - 6) / (31 - 6) = 0.2$$

$$11. (16 - 6) / (31 - 6) = 0.4$$

$$17. (21 - 6) / (31 - 6) = 0.6$$

$$12. (12 - 6) / (31 - 6) = 0.24$$

$$18. (24 - 6) / (31 - 6) = 0.72$$

$$13. (13 - 6) / (31 - 6) = 0.28$$

$$19. (21 - 6) / (31 - 6) = 0.6$$

$$14. (21 - 6) / (31 - 6) = 0.6$$

$$20. (7 - 6) / (31 - 6) = 0.04$$

В результаті розрахунку маємо нормалізовані дані представлені у табл. 3.4:

Таблиця 3.4 – Нормалізовані дані

Номер об'єкту	Довжина трубопроводу	Строк експлуатації трубопроводу	Стан трубопроводу	Діаметр трубопроводу	Строк експлуатації елементу	Технічний стан елементу
1	0	0.23	0	0	0.28	1
2	0.66	0.09	1	0.11	0.04	0
3	0.32	0.27	0.3	0.33	0.16	0.3
4	0.05	0.14	1	0.11	0.36	0.3
5	0.23	0.23	0	0	0.36	1
6	1	0.23	0.3	1	0.24	0.3
7	0.56	0.27	0.3	0	0	0
8	0.64	0.32	0.7	0.11	0.4	0
9	0.2	0.64	1	0	0.76	0.7
10	0.03	0.73	0	1	0.76	1

11	0.08	0.55	1	1	0.4	0
12	0.41	0.46	0.3	0.33	0.24	0.7
13	0.98	0.27	0	0.11	0.28	0.3
14	0.66	0.59	0.7	0	0.6	0
15	0.38	0.86	0	1	1	1
16	0.87	0	0	1	0.2	0.3
17	0.38	0.86	0.7	0.11	0.6	0.7
18	0.49	0.77	0.7	0	0.72	0
19	0.59	1	0	0.33	0.6	1
20	0.37	0.09	1	0	0.04	0

Другим етапом кластеризації є розрахунок «відстані» між об'єктами, цей розрахунок проводитиметься за формулою розрахунку Евклідової відстані.

$$\rho(x, x') = \sqrt{\sum_i^n (x_i - x'_i)^2} \quad (3.16)$$

В результаті отримаємо симетричну матрицю відстаней між об'єктами, з 0 на головній діагоналі.

Для 1го об'єкту представлений приклад розрахунку, такий розрахунок проводитиметься для кожного об'єкту по аналогії.

Відстань між 1м та 2м об'єктом:

$$\sqrt{(0.66 - 0)^2 + (0.09 - 0.23)^2 + (1 - 0)^2 + (0.11 - 0)^2 + (0.04 - 0.28)^2 + (0 - 1)^2} = 1.59$$

Відстань між 1м та 3м об'єктом:

$$\sqrt{(0.32 - 0)^2 + (0.27 - 0.23)^2 + (0.3 - 0)^2 + (0.33 - 0)^2 + (0.16 - 0.28)^2 + (0.3 - 1)^2} = 0.9$$

Відстань між 1м та 4м об'єктом:

$$\sqrt{(0.05 - 0)^2 + (0.14 - 0.23)^2 + (1 - 0)^2 + (0.11 - 0)^2 + (0.36 - 0.28)^2 + (0.3 - 1)^2} = 1.23$$

Відстань між 1м та 5м об'єктом:

$$\sqrt{(0.23 - 0)^2 + (0.23 - 0.23)^2 + (0 - 0)^2 + (0 - 0)^2 + (0.36 - 0.28)^2 + (1 - 1)^2} = 0.24$$

Відстань між 1м та 6м об'єктом:

$$\sqrt{(1 - 0)^2 + (0.23 - 0.23)^2 + (0.3 - 0)^2 + (1 - 0)^2 + (0.24 - 0.28)^2 + (0.3 - 1)^2} = 1.6$$

Відстань між 1м та 7м об'єктом:

$$\sqrt{(0.56 - 0)^2 + (0.27 - 0.23)^2 + (0.3 - 0)^2 + (0 - 0)^2 + (0 - 0.28)^2 + (0 - 1)^2} = 1.2$$

Відстань між 1м та 8м об'єктом:

$$\sqrt{(0.64 - 0)^2 + (0.32 - 0.23)^2 + (0.7 - 0)^2 + (0.11 - 0)^2 + (0.04 - 0.28)^2 + (0 - 1)^2} = 1.4$$

Відстань між 1м та 9м об'єктом:

$$\sqrt{(0.2 - 0)^2 + (0.64 - 0.23)^2 + (1 - 0)^2 + (0 - 0)^2 + (0.76 - 0.28)^2 + (0.7 - 1)^2} = 1.24$$

Відстань між 1м та 10м об'єктом:

$$\sqrt{(0.03 - 0)^2 + (0.73 - 0.23)^2 + (0 - 0)^2 + (1 - 0)^2 + (0.76 - 0.28)^2 + (1 - 1)^2} = 1.22$$

Відстань між 1м та 11м об'єктом:

$$\sqrt{(0.08 - 0)^2 + (0.55 - 0.23)^2 + (1 - 0)^2 + (1 - 0)^2 + (0.4 - 0.28)^2 + (0 - 1)^2} = 1.77$$

Відстань між 1м та 12м об'єктом:

$$\sqrt{(0.41 - 0)^2 + (0.46 - 0.23)^2 + (0.3 - 0)^2 + (0.33 - 0)^2 + (0.24 - 0.28)^2 + (0.7 - 1)^2} = 0.72$$

Відстань між 1м та 13м об'єктом:

$$\sqrt{(0.98 - 0)^2 + (0.27 - 0.23)^2 + (0 - 0)^2 + (0.11 - 0)^2 + (0.28 - 0.28)^2 + (0.3 - 1)^2} = 1.21$$

Відстань між 1м та 14м об'єктом:

$$\sqrt{(0.66 - 0)^2 + (0.59 - 0.23)^2 + (0.7 - 0)^2 + (0 - 0)^2 + (0.6 - 0.28)^2 + (0 - 1)^2} = 1.47$$

Відстань між 1м та 15м об'єктом:

$$\sqrt{(0.38 - 0)^2 + (0.86 - 0.23)^2 + (0 - 0)^2 + (1 - 0)^2 + (1 - 0.28)^2 + (1 - 1)^2} = 1.44$$

Відстань між 1м та 16м об'єктом:

$$\sqrt{(0.87 - 0)^2 + (0 - 0.23)^2 + (0 - 0)^2 + (1 - 0)^2 + (0.2 - 0.28)^2 + (0.3 - 1)^2} = 1.52$$

Відстань між 1м та 17м об'єктом:

$$\sqrt{(0.38 - 0)^2 + (0.86 - 0.23)^2 + (0.7 - 0)^2 + (0.11 - 0)^2 + (0.6 - 0.28)^2 + (0.7 - 1)^2} = 1.11$$

Відстань між 1м та 18м об'єктом:

$$\sqrt{(0.49 - 0)^2 + (0.77 - 0.23)^2 + (0.7 - 0)^2 + (0 - 0)^2 + (0.72 - 0.28)^2 + (0 - 1)^2} = 1.49$$

Відстань між 1м та 19м об'єктом:

$$\sqrt{(0.59 - 0)^2 + (1 - 0.23)^2 + (0 - 0)^2 + (0.33 - 0)^2 + (0.6 - 0.28)^2 + (1 - 1)^2} = 1.1$$

Відстань між 1м та 20м об'єктом:

$$\sqrt{(0.37 - 0)^2 + (0.09 - 0.23)^2 + (1 - 0)^2 + (0 - 0)^2 + (0.04 - 0.28)^2 + (0 - 1)^2} = 1.49$$

Матриця відстаней зображена на рис 3.3.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	1.59	0.9	1.23	0.24	1.61	1.22	1.39	1.24	1.22	1.77	0.72	1.21	1.47	1.44	1.52	1.11	1.49	1.07	1.49
2	1.59	0	0.89	0.75	1.52	1.24	0.74	0.52	1.24	2.03	1.21	1.13	1.13	0.82	2.09	1.4	1.25	1.03	1.79	0.31
3	0.9	0.89	0	0.82	0.86	0.96	0.53	0.68	1.13	1.3	1.11	0.46	0.77	0.88	1.44	0.96	0.96	0.98	1.17	0.86
4	1.23	0.75	0.82	0	1.24	1.49	1	0.75	0.78	1.67	1.03	0.97	1.37	0.91	1.82	1.58	0.97	0.96	1.62	0.56
5	0.24	1.52	0.86	1.24	0	1.48	1.15	1.3	1.19	1.2	1.77	0.62	1.04	1.36	1.35	1.41	1.03	1.41	0.94	1.46
6	1.61	1.24	0.96	1.49	1.48	0	1.16	1.1	1.65	1.43	1.25	1	0.94	1.27	1.39	0.4	1.42	1.43	1.39	1.43
7	1.22	0.74	0.53	1	1.15	1.16	0	0.58	1.35	1.78	1.4	0.85	0.67	0.8	1.86	1.18	1.18	0.97	1.45	0.75
8	1.39	0.52	0.68	0.75	1.3	1.1	0.58	0	1.01	1.72	1.12	0.89	0.84	0.35	1.73	1.25	0.94	0.58	1.43	0.6
9	1.24	1.24	1.13	0.78	1.19	1.65	1.35	1.01	0	1.46	1.28	0.97	1.47	0.91	1.49	1.83	0.46	0.83	1.23	1.16
10	1.22	2.03	1.3	1.67	1.2	1.43	1.78	1.72	1.46	0	1.47	1.06	1.62	1.71	0.44	1.43	1.24	1.64	0.93	2.01
11	1.77	1.21	1.11	1.03	1.77	1.25	1.4	1.12	1.28	1.47	0	1.25	1.67	1.21	1.6	1.43	1.26	1.19	1.72	1.19
12	0.72	1.13	0.46	0.97	0.62	1	0.85	0.89	0.97	1.06	1.25	0	0.81	0.98	1.17	1.06	0.71	1.04	0.8	1.13
13	1.21	1.13	0.77	1.37	1.04	0.94	0.67	0.84	1.47	1.62	1.67	0.81	0	0.95	1.58	0.94	1.21	1.13	1.15	1.25
14	1.47	0.82	0.88	0.91	1.36	1.27	0.8	0.35	0.91	1.71	1.21	0.98	0.95	0	1.67	1.46	0.81	0.28	1.33	0.86
15	1.44	2.09	1.44	1.82	1.35	1.39	1.86	1.73	1.49	0.44	1.6	1.17	1.58	1.67	0	1.45	1.24	1.61	0.82	2.12
16	1.52	1.4	0.96	1.58	1.41	0.4	1.18	1.25	1.83	1.43	1.43	1.06	0.94	1.46	1.45	0	1.61	1.61	1.48	1.54
17	1.11	1.25	0.96	0.97	1.03	1.42	1.18	0.94	0.46	1.24	1.26	0.71	1.21	0.81	1.24	1.61	0	0.73	0.83	1.22
18	1.49	1.03	0.98	0.96	1.41	1.43	0.97	0.58	0.83	1.64	1.19	1.04	1.13	0.28	1.61	1.61	0.73	0	1.29	1.01
19	1.07	1.79	1.17	1.62	0.94	1.39	1.45	1.43	1.23	0.93	1.72	0.8	1.15	1.33	0.82	1.48	0.83	1.29	0	1.82
20	1.49	0.31	0.86	0.56	1.46	1.43	0.75	0.6	1.16	2.01	1.19	1.13	1.25	0.86	2.12	1.54	1.22	1.01	1.82	0

Рисунок 3.3 Матриця відстаней.

На третьому етапі необхідно знайти потенціал кожної вершини як центра кластеру. Цей потенціал необхідно розрахувати для кожного об'єкту, й він розраховується за формулою 3.10. Значення констант альфа та бета приймаємо як одиниця поділена на середнє значення відстані між об'єктами.

$$\alpha = 1/1.1236 = 0.89$$

Приклад розрахунку потенціалу представлений для першого об'єкту, потенціали інших об'єктів розраховуються по аналогії.

Об'єкт 1:

$$P_1 = \exp(-0.89 * 1.59) + \exp(-0.89 * 0.90) + \exp(-0.89 * 1.23) + \exp(-0.89 * 0.24) + \exp(-0.89 * 1.61) + \exp(-0.89 * 1.22) + \exp(-0.89 * 1.39) + \exp(-0.89 * 1.24) + \exp(-0.89 * 1.22) + \exp(-0.89 * 1.77) + \exp(-0.89 * 0.72) + \exp(-0.89 * 1.21) + \exp(-0.89 * 1.47) + \exp(-0.89 * 1.44) + \exp(-0.89 * 1.52) + \exp(-0.89 * 1.11) + \exp(-0.89 * 1.49) + \exp(-0.89 * 1.07) + \exp(-0.89 * 1.49) = 6.54$$

$$\text{Об'єкт 2: } P_2 = 6.89$$

$$\text{Об'єкт 12: } P_{12} = 7.93$$

$$\text{Об'єкт 3: } P_3 = 8.04$$

$$\text{Об'єкт 13: } P_{13} = 6.74$$

$$\text{Об'єкт 4: } P_4 = 6.94$$

$$\text{Об'єкт 14: } P_{14} = 7.62$$

$$\text{Об'єкт 5: } P_5 = 6.14$$

$$\text{Об'єкт 15: } P_{15} = 5.11$$

$$\text{Об'єкт 6: } P_6 = 6.15$$

$$\text{Об'єкт 16: } P_{16} = 5.75$$

$$\text{Об'єкт 7: } P_7 = 7.27$$

$$\text{Об'єкт 17: } P_{17} = 7.23$$

$$\text{Об'єкт 8: } P_8 = 8.07$$

$$\text{Об'єкт 18: } P_{18} = 7.15$$

$$\text{Об'єкт 9: } P_9 = 6.51$$

$$\text{Об'єкт 19: } P_{19} = 5.96$$

$$\text{Об'єкт 10: } P_{10} = 5.26$$

$$\text{Об'єкт 20: } P_{20} = 6.86$$

$$\text{Об'єкт 11: } P_{11} = 5.55$$

На четвертому етапі визначається об'єкт з найбільшим потенціалом центра кластеру. Після цього необхідно видалити вплив цієї вершини на інші об'єкти цей перерахунок потенціалів відбувається за формулою 3.12. Це відбувається до тих пір доки максимальне значення потенціалу перевищує деякий поріг. У нашому випадку приймемо поріг що рівняється половині максимального значення потенціалу.

Далі представлений приклад перерахунку:

$$\text{Об'єкт 1: } P_1 = (6.54 - 8.07 * 0.29) = 4.2$$

$$\text{Об'єкт 2: } P_2 = (6.89 - 8.07 * 0.63) = 1.81$$

$$\text{Об'єкт 3: } P_3 = (8.04 - 8.07 * 0.55) = 3.63$$

$$\text{Об'єкт 4: } P_4 = (6.94 - 8.07 * 0.51) = 2.80$$

$$\text{Об'єкт 5: } P_5 = (6.14 - 8.07 * 0.31) = 3.61$$

$$\text{Об'єкт 6: } P_6 = (6.15 - 8.07 * 0.38) = 3.12$$

$$\text{Об'єкт 7: } P_7 = (7.27 - 8.07 * 0.60) = 2.46$$

$$\text{Об'єкт 9: } P_9 = (6.51 - 8.07 * 0.41) = 3.22$$

$$\text{Об'єкт 10: } P_{10} = (5.26 - 8.07 * 0.22) = 3.52$$

$$\text{Об'єкт 11: } P_{11} = (5.55 - 8.07 * 0.37) = 2.57$$

$$\text{Об'єкт 12: } P_{12} = (7.93 - 8.07 * 0.45) = 4.28$$

$$\text{Об'єкт 13: } P_{13} = (6.74 - 8.07 * 0.47) = 2.92$$

$$\text{Об'єкт 14: } P_{14} = (7.62 - 8.07 * 0.73) = 1.71$$

$$\text{Об'єкт 15: } P_{15} = (5.11 - 8.07 * 0.21) = 3.38$$

$$\text{Об'єкт 16: } P_{16} = (5.75 - 8.07 * 0.33) = 3.10$$

$$\text{Об'єкт 17: } P_{17} = (7.23 - 8.07 * 0.43) = 3.74$$

$$\text{Об'єкт 18: } P_{18} = (7.15 - 8.07 * 0.60) = 2.34$$

$$\text{Об'єкт 19: } P_{19} = (5.96 - 8.07 * 0.28) = 3.70$$

$$\text{Об'єкт 20: } P_{20} = (6.86 - 8.07 * 0.59) = 2.13$$

Новим центром кластеру є об'єкт номер 12. Його значення - 4.28. Це значення більше ніж $8.07/2 = 4.035$. У цьому випадку проводиться перерахунок ще одного центру кластеру.

$$\text{Об'єкт 1: } P_1 = 3.31$$

$$\text{Об'єкт 2: } P_2 = 0.35$$

$$\text{Об'єкт 13: } P_{13} = 1.95$$

$$\text{Об'єкт 3: } P_3 = 2.04$$

$$\text{Об'єкт 14: } P_{14} = 0.25$$

$$\text{Об'єкт 4: } P_4 = 1.09$$

$$\text{Об'єкт 15: } P_{15} = 2.35$$

$$\text{Об'єкт 5: } P_5 = 2.72$$

$$\text{Об'єкт 16: } P_{16} = 1.90$$

$$\text{Об'єкт 6: } P_6 = 1.72$$

$$\text{Об'єкт 17: } P_{17} = 2.34$$

$$\text{Об'єкт 7: } P_7 = 1.23$$

$$\text{Об'єкт 18: } P_{18} = 0.85$$

$$\text{Об'єкт 9: } P_9 = 1.85$$

$$\text{Об'єкт 19: } P_{19} = 2.77$$

$$\text{Об'єкт 10: } P_{10} = 2.36$$

$$\text{Об'єкт 20: } P_{20} = 0.65$$

$$\text{Об'єкт 12: } P_{12} = 2.87$$

Новим центром кластеру є об'єкт номер 1. Його значення – 3.31. Це значення менше мінімального порогу. В результаті роботи алгоритму ми маємо 3 об'єкти - центри кластерів.

На наступному і останньому кроці кластерного аналізу об'єкти переносяться до кластерів. Об'єкт вважається приналежним до кластеру якщо відстань До центру цього кластеру менша ніж відстань до інших центрів кластерів.

Згідно цього розташовуємо об'єкти у кластерах:

Перший кластер з центром - Об'єкт№ 1 вміщує об'єкти:

- 1, 5;

Другий кластер з центром - Об'єкт№ 8 вміщує об'єкти:

- 8, 2, 4, 7, 11, 14, 18, 20;

Третій кластер з центром - Об'єкт№ 12 вміщує об'єкти:

- 12, 3, 6, 9, 10, 13, 15, 16, 17, 19;

В результаті розрахунку, було визначено що алгоритм надає точні дані про кластери та має високу швидкість сходження. Це задовольняє вимогам які були представлені до алгоритму кластеризації.

4.1 Аналіз технології збору, передачі та обробки інформації

Опис технології збору, передачі та обробки інформації інформаційної системи зображено у вигляді DFD – діаграми, на рис 4.1

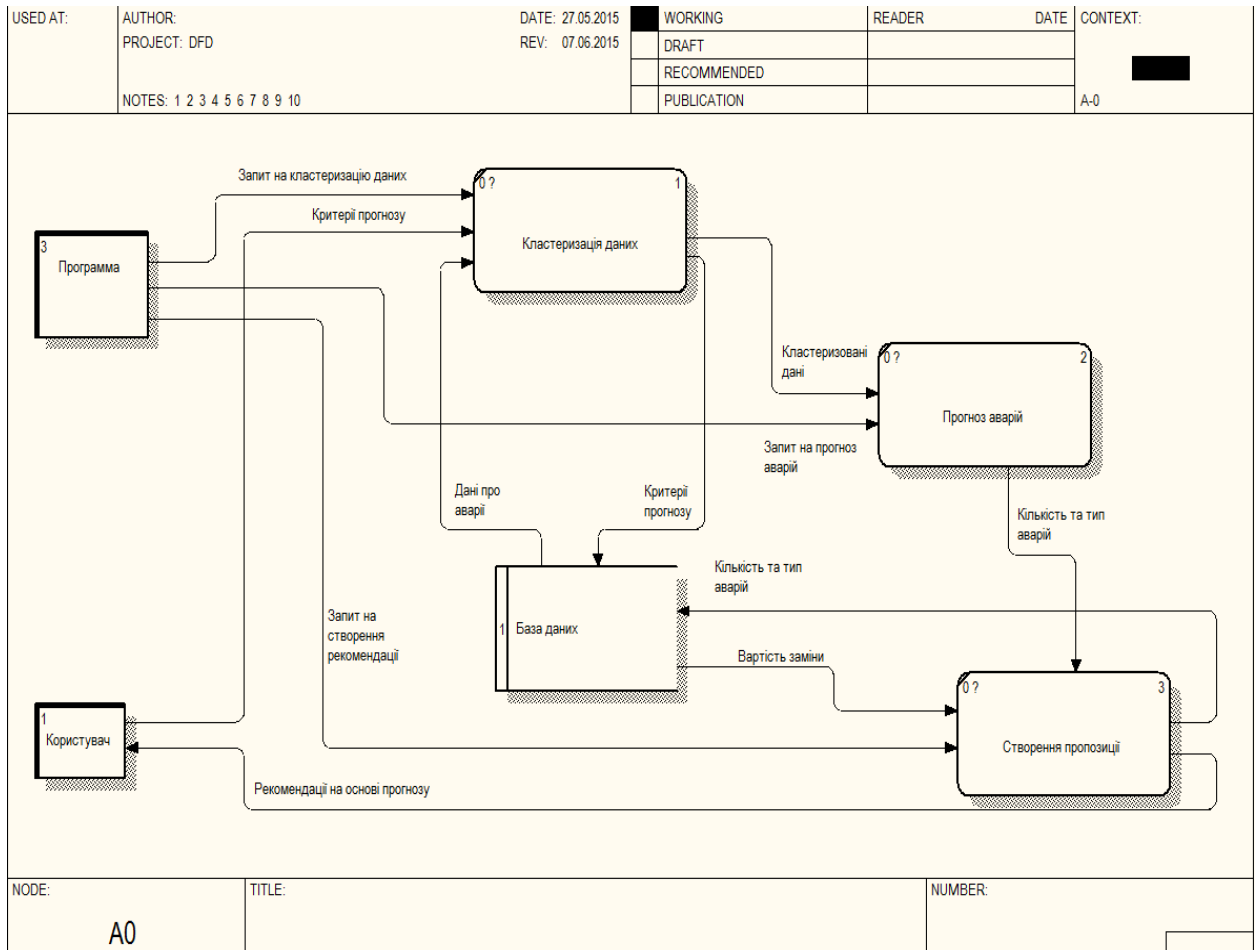


Рисунок 4.1 Діаграма потоків даних

4.2 Опис вхідної інформації

Вхідна інформація необхідна для роботи програми буде представлена у вигляді таблиць даних, які будуть розташовані у базі даних. За допомогою запитів будуть відібрані необхідні поля даних. Усі дані зведені до наступної таблиці інформаційних реквізитів із зазначенням типів та джерел.

Таблиця 4.1 Зведена таблиця інформаційних реквізитів

№ п/п	Назва атрибуту	Назва джерела	Характеристика (тип)
1	Шифр типу причини	Стандартизований список аварій та їх причин	числовий
2	Назва типу причини	Стандартизований список аварій та їх причин	текстовий
3	Тип заміни	Звіт про ліквідацію аварії	текстовий
4	Назва елементу	Стандартизований список даних про елементи трубопроводу	текстовий
5	Ціна	Бухгалтерські дані	числовий
6	Шифр трубопроводу	Стандартизований список даних про трубопроводи	числовий
7	Довжина трубопроводу	Стандартизований список даних про трубопроводи	числовий
8	Строк експлуатації трубопроводу	Стандартизований список даних про трубопроводи	числовий
9	Стан трубопроводу	Стандартизований список даних про трубопроводи	текстовий
10	Діаметр	Стандартизований список даних про трубопроводи	числовий
11	Шифр елементу	Стандартизований список даних про елементи трубопроводу	числовий
12	Назва елементу	Стандартизований список даних про елементи трубопроводу	текстовий
13	Строк експлуатації	Стандартизований список даних про елементи трубопроводу	числовий

продовження таблиці 4.1

14	Технічний стан елементу	Стандартизований список даних про елементи трубопроводу	текстовий
15	Адреса виникнення аварії	Звіт про ліквідацію аварії	текстовий
16	Дата виникнення аварії	Звіт про ліквідацію аварії	дата

4.3 Концептуальна модель бази даних

Концептуальна модель бази даних у вигляді діаграми Чена зображена на рис 4.2.

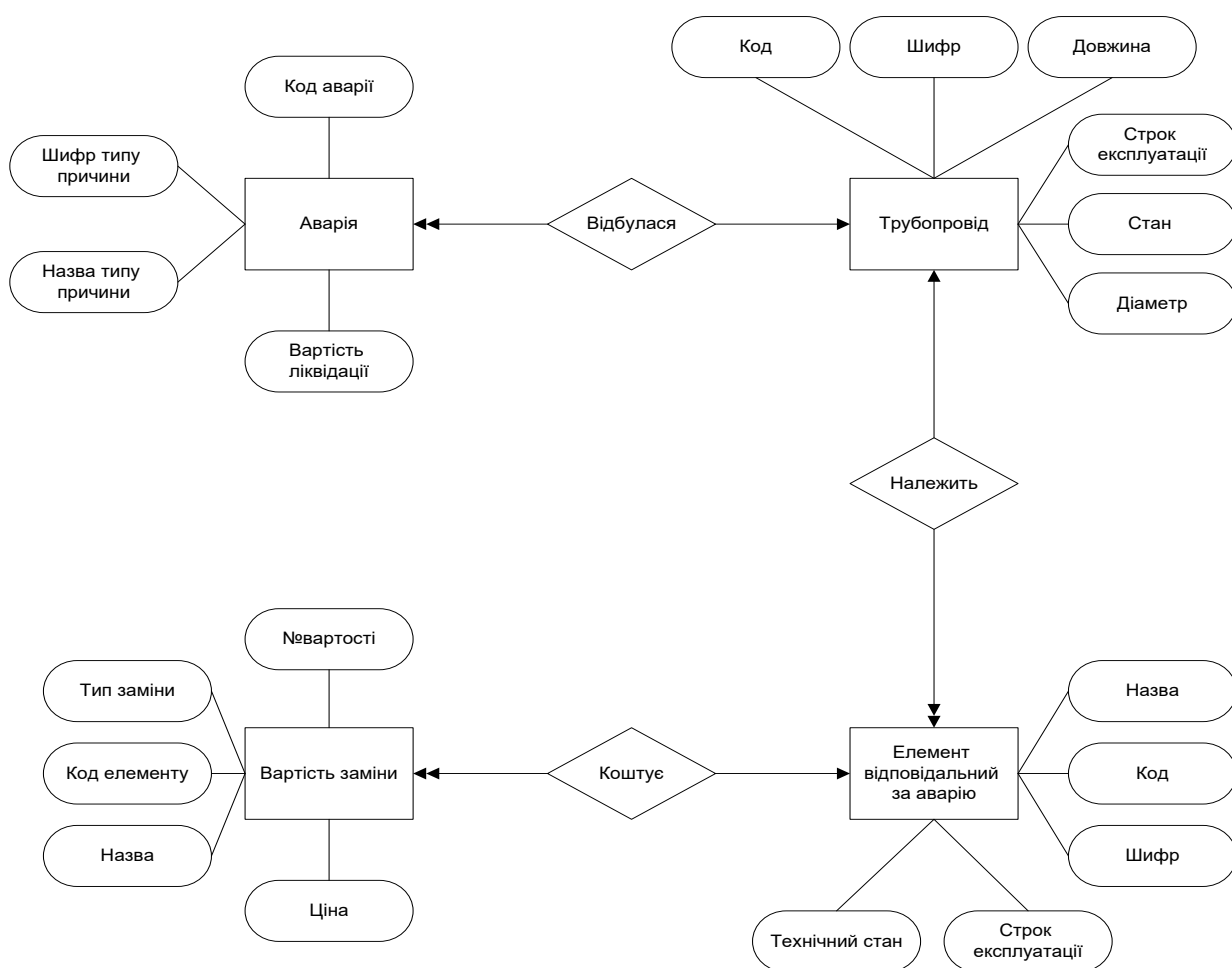


Рисунок 4.2 Концептуальна модель бази даних. Діаграма Чена.

Розглянемо типи зв'язків між сутностями:

Аварія – Трубопровід

Зв'язок між цими сутностями один до одного. Це зумовлено тим що, аварія розглядається окремо для кожного трубопроводу. Для кожної окремої аварії створюється звіт з даними про трубопровід на якому відбулася аварія.

Трубопровід – Елемент відповідальний за аварію

Зв'язок між цими сутностями один до багатьох. Це зумовлено тим що, при аварії на трубопроводі відповідальними можуть виявитись декілька елементів цього трубопроводу.

Елемент відповідальний за аварію – Вартість заміни

Зв'язок між цими сутностями один до багатьох. Це зумовлено тим що, кожен елемент може бути замінений та відремонтований. Крім того ціна може бути різною для різних ступенів поломки елемента.

Даталогічна модель бази даних зображена на рис 4.3.

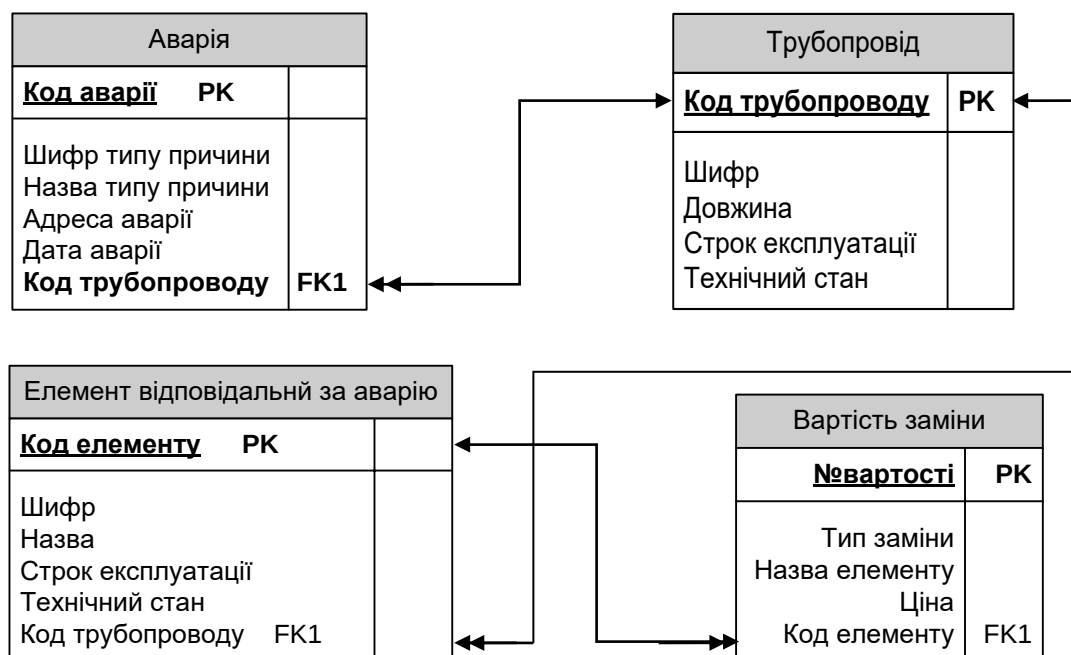


Рисунок 4.3 Даталогічна модель бази даних

4.4 Фізична модель бази даних

У наступних таблицях представлена фізична модель бази даних.

У таблиці 4.2 представлена фізична структура таблиці Аварія.

Таблиця 4.2 Фізична структура таблиці Аварія

Назва поля	Ідентифікатор	Тип	Призначення
1	2	3	4
Код аварії	KODAVAR	ціле	Первинний ключ
Шифр типу причини	SHIFRTIPAPRICHINU	ціле	
Назва типу причини	NAZVTIPAPRICHINU	текст(35)	
Адреса виникнення	ADRES	текст(35)	
Дата виникнення	DATAAVAR	дата	
Код трубопроводу	F_KODTRUBI	Ціле	Вторинний ключ

Код аварії – унікальний ідентифікатор аварії. Приймає цілі значення вище 0.

Шифр типу причини – Приймає цілі значення вище 0. Введений для полегшення розпізнавання програмою причини аварії.

Назва типу причини – приймає текстове значення розмірністю до 35 символів. Являє собою текстове позначення причини аварії. Введений для полегшення розуміння бази даних оператором.

Адреса виникнення аварії – текстове значення, використовується для відслідковування аварій за адресою виникнення.

Дата виникнення аварії – поле приймає значення у вигляді дат. Відображає дату в яку виникла аварія. Використовується для прогнозу.

У таблиці 4.3 представлена фізична структура таблиці Трубопровід.

Таблиця 4.3 Фізична структура таблиці Трубопровід

Назва поля	Ідентифікатор	Тип	Призначення
1	2	3	4
Код трубопроводу	KODTRUBI	Ціле	Первинний ключ
Шифр трубопроводу	SHIFRTRUBI	Ціле	
Довжина трубопроводу	DOVJTRUBI	Дійсне	
Строк експлуатації	STROKEXPL	ціле	
Стан трубопроводу	TECKSTAN	текст(35)	
Діаметр	DIAMTRUBI	Ціле	

Код трубопроводу – унікальний ідентифікатор трубопроводу на якому виникла аварія. Приймає цілі значення вище 0.

Шифр трубопроводу - Приймає цілі значення вище 0. Введений для полегшення розпізнавання програмою трубопроводів.

Довжина трубопроводу – Приймає цілі значення, довжина труби у метрах. Використовується для кластеризації.

Строк експлуатації - приймає цілі значення, строк експлуатації труби у роках. Використовується для кластеризації.

Стан трубопроводу – приймає текстове значення. А саме 4 позиції: задовільний, близький до задовільного, близький до незадовільного, незадовільний. Використовується для кластеризації.

Діаметр – приймає цілі значення, стандартні діаметри труби: 50, 100, 200, 500. Використовується для кластеризації.

Код аварії – унікальний вторинний ключ, використовується для зв'язування таблиць «аварія» та «трубопровід».

У таблиці 4.4 представлена фізична структура таблиці Елемент відповідальний за аварію.

Таблиця 4.4 Фізична структура таблиці Елемент відповідальний за аварію

Назва поля	Ідентифікатор	Тип	Призначення
1	2	3	4
Код елемента	KODELEM	ціле	Первинний ключ
Шифр елемента	SHIFRELEM	ціле	
Назва елемента	STROKEKSP	текст(35)	
Строк експлуатації	NAZVELEM	ціле	
Технічний стан	TECSTAN	текст(35)	
Код трубопроводу	F_KODTRUBI	ціле	Вторинний ключ

Код елемента – унікальний ідентифікатор трубопроводу на якому виникла аварія. Приймає цілі значення вище 0.

Шифр елемента - Приймає цілі значення вище 0. Введений для полегшення розпізнавання програмою елементів.

Назва елементу – приймає текстове значення розмірністю до 35 символів. Являє собою текстове позначення елементу відповідального за аварію. Введений для полегшення розуміння бази даних оператором.

Строк експлуатації - приймає цілі значення, строк експлуатації елементу у роках. Використовується для кластеризації.

Технічний стан - приймає текстове значення. А саме 4 позиції: задовільний, близький до задовільного, близький до незадовільного, незадовільний. Використовується для кластеризації.

Код трубопроводу - вторинний ключ, використовується для зв'язування таблиць «трубопровід» та «елемент відповідальний за аварію».

У таблиці 4.5 представлена фізична структура таблиці Вартість заміни.

Таблиця 4.5 Фізична структура таблиці Вартість заміни

Назва поля	Ідентифікатор	Тип	Призначення
1	2	3	4
№ вартості	NOMVART	ціле	Первинний ключ
Тип заміни	TIPZAM	текст(35)	
Назва елементу	NAZVELEM	текст(35)	
Ціна	CENA	ціле	
Код елементу	F_KODELEM	ціле	Вторинний ключ

№ вартості - унікальний ідентифікатор вартості на заміну елемента. Приймає цілі значення вище 0.

Тип заміни – приймає текстове значення. А саме 2 позиції: повна заміна та ремонт.

Назва елементу – приймає текстове значення розмірністю до 35 символів. Являє собою текстове позначення елементу для якого проводиться заміна. Введений для полегшення розуміння бази даних оператором.

Ціна – приймає цілі значення. Відповідає ціні на заміну елемента в залежності від типу заміни.

Код елементу - вторинний ключ, використовується для зв'язування таблиць «елемент відповідальний за аварію» та «Вартість заміни».

4.5 Опис основних алгоритмів

Далі представлені основні алгоритми:

4.5.1 Алгоритм ініціалізації діалогового вікна

```
ComboBox1->Items->Add("Загальний розрахунок");
for(int i = 0; i < DBGrid5->FieldCount; i++)
{
    bool bCrItem = true;
    for(int j = 0; j < ComboBox1->Items->Count; j++)
    {
        if(DBGrid5->Fields[3]->AsString == ComboBox1->Items->Strings[j])
        {
            bCrItem = false;
        }
    }
    if(bCrItem == true)
    {
        ComboBox1->Items->Add(DBGrid5->Fields[3]->AsString);
    }
    DBGrid5->DataSource->DataSet->Next();
}
ComboBox1->ItemIndex = 0;
```

В даному алгоритмі реалізована система яка записує усі адреси, які записані до бази даних, за якими траплялися аварії до списку. Цей список дозволяє проводити кластеризацію лише для адреси яка обрана у списку.

4.5.2 Алгоритм доступу до редагування бази даних

```
if(Edit1->Text == "1")
{
    Form1->EnableEditing();
    Form2->Visible = false;
    Form1->Enabled = true;
}
else
{
    ShowMessage("Невірний пароль");
}
```

В даному алгоритмі реалізована система запиту паролю для доступу до редагування бази даних. Головне вікно переводиться у неактивний стан, вікно запиту паролю чекає введення паролю, та перевіряє його правильність, якщо пароль введений правильно надається доступ до редагування, у противному випадку з'являється повідомлення.

4.5.3 Алгоритм кластеризації

Функція кластеризації – основний алгоритм системи. Для більшої наочності роботи алгоритму, зобразимо його у вигляді блок схеми яка зображена на рис. 4.4. Вхідними даними до кластеризації є вибірка атрибутів з бази даних, за обраною адресою. Також при виборі адреси передбачений пункт для кластеризації по всім записам бази даних.

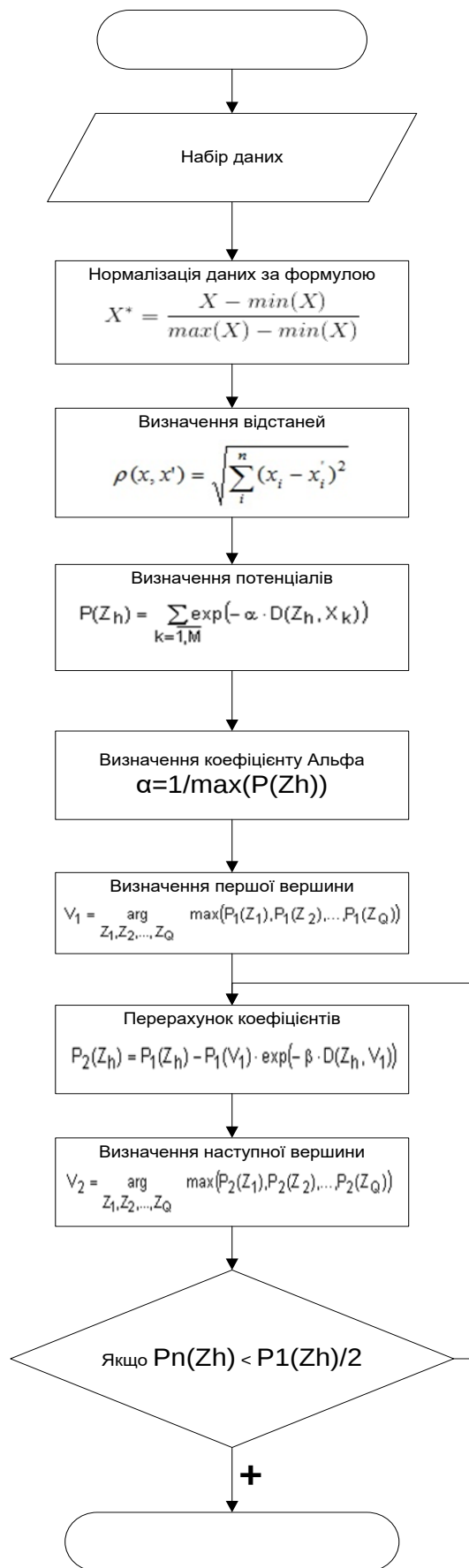


Рисунок 4.4 Схеми алгоритму кластеризації

Далі представлений код алгоритму кластеризації, кожний етап цього алгоритму коментуватиметься.

```
if(ComboBox1->ItemIndex == 0)

{Query1->Active = false;

    Query1->SQL->Clear();

    Query1->SQL->Text = "SELECT A.KODAVAR, T.DOVJTRUBI, T.STROKEEXPL, T.TECKSTAN,
T.DIAMTRUBI, E.STROKEKSP, E.TECKSTAN FROM AVARIA A, TRUBOPROVID T, ELEMVIDPZAAVAR E
\WHERE A.KODAVAR = T.F_KODAVAR AND T.KODTRUBI = E.F_KODTRUBI";

    Query1->Active = true;}

else

{Query1->Active = false;

    Query1->SQL->Clear();

    Query1->SQL->Text = "SELECT A.KODAVAR, T.DOVJTRUBI, T.STROKEEXPL,
T.TECKSTAN, T.DIAMTRUBI, E.STROKEKSP, E.TECKSTAN FROM AVARIA A, TRUBOPROVID T,
ELEMVIDPZAAVAR E WHERE A.KODAVAR = T.F_KODAVAR AND A.ADRES = '"+ComboBox1->Text+"'
AND T.KODTRUBI = E.F_KODTRUBI";

    Query1->Active = true;}
```

Дана частина алгоритму відповідає за вибірку даних за адресою. Створюється SQL запит до компоненту Query1 який у свою чергу відповідає за заповнення використовуваної таблиці.

```
for(int i = 0; i < iNumberOfRec; i++)

{

    iCurMax = (iCurMax > m_pKlastMas[i].iDovjTrub) ? iCurMax : m_pKlastMas[i].iDovjTrub;

    iCurMin = (iCurMin < m_pKlastMas[i].iDovjTrub) ? iCurMin : m_pKlastMas[i].iDovjTrub;}

for(int i = 0; i < iNumberOfRec; i++)

{

    m_pKlastNormMas[i].dDovjTrub = (((double)m_pKlastMas[i].iDovjTrub - (double)iCurMin) /
((double)iCurMax - (double)iCurMin));}
```

Дана частина алгоритму відповідає за нормалізацію числових даних. Це один з блоків, подібний блок використовується для кожного атрибуту.

```
for(int i = 0; i < iNumberOfRec; i++)

{

    for(int j = 0; j < iNumberOfRec; j++)

    {

        dpDistMatr[i][j] = sqrt(pow((m_pKlastNormMas[i].dDovjTrub-
m_pKlastNormMas[j].dDovjTrub),2)+pow((m_pKlastNormMas[i].dStrokExpTr-
m_pKlastNormMas[j].dStrokExpTr),2)+pow((m_pKlastNormMas[i].dTeckStanTr-
```

```

m_pKlastNormMas[j].dTeckStanTr,2)+pow((m_pKlastNormMas[i].dDiamTr-
m_pKlastNormMas[j].dDiamTr),2)+pow((m_pKlastNormMas[i].dStrExpEl-
m_pKlastNormMas[j].dStrExpEl),2)+pow((m_pKlastNormMas[i].dTeckStanEl-
m_pKlastNormMas[j].dTeckStanEl),2));}}

```

Дана частина алгоритму відповідає за створення та заповнення матриці віддалей об'єктів.

```

double dAlfa(0);

for(int i = 0; i < iNumberOfRec; i++)

{for(int j = 0; j < iNumberOfRec; j++)

    {dAlfa = dAlfa + dpDistMatr[i][j];}}

dAlfa = dAlfa/(iNumberOfRec*iNumberOfRec);

dAlfa = 1/dAlfa;

```

Дана частина алгоритму відповідає за визначення коефіцієнту Альфа.

```

double *dpPotencMas = new double[iNumberOfRec];

for(int i = 0; i < iNumberOfRec; i++)

{for(int j = 0; j < iNumberOfRec; j++)

    {if(i != j)

        dpPotencMas[i] = dpPotencMas[i]+exp((-1*dAlfa)*dpDistMatr[i][j]);}}

```

Дана частина алгоритму відповідає за розрахунок потенціалів кожного об'єкту як вершини кластеру.

```

double dMaxPotenc(0);

int iPotIndex;

for(int i = 0; i < iNumberOfRec; i++)

{if(dMaxPotenc < dpPotencMas[i])

    {dMaxPotenc = dpPotencMas[i];

    iPotIndex = i;}}

```

Дана частина алгоритму відповідає за визначення першої вершини кластеру.

```

double dTemp = dMaxPotenc/2;

while(dMaxPotenc > dTemp)

{for(int i = 0; i < iNumberOfRec; i++)

    {bool bOk = false;

    for(int j = 0; j < iNumberofVersh; j++)

        {if(ipVershKlastInd[j] != i)

```

```

        {bOk = true;
        goto Ok;}}

Ok: if(bOk == true)

        {dpPotencMas[i] = dpPotencMas[i] - dMaxPotenc * (exp((-
1*dAlfa)*dpDistMatr[iPotIndex][i]));}

        Else dpPotencMas[i] = 0;}

dMaxPotenc = 0;

iPotIndex = 0;

for(int i = 0; i < iNumberOfRec; i++)

{if(dMaxPotenc < dpPotencMas[i])

    {dMaxPotenc = dpPotencMas[i];

    iPotIndex = i;}}

iNumberOfVersh++;

ipVershKlastInd[iNumberOfVersh-1] = iPotIndex;

ShowMessage("Вершиною кластеру визначено об'єкт №"

+ AnsiString("")+ iPotIndex + AnsiString("")+ \

" його значення потенціалу = " + AnsiString("")+ dMaxPotenc + AnsiString("")+ "");}

```

Дана частина алгоритму – цикл, який відповідає за знаходження усіх наступних вершин кластерів, умовою виходу з циклу є зменшення потенціалів знаходжуваних вершин до половини величини потенціалу першої вершини кластеру.

```

for(int i = 0; i < iNumberOfRec; i++)

    {iAccesInd = 0;

    double dMin;

    for(int j = 1; j < iNumberOfVersh; j++)

        {dMin = dpDistMatr[i][ipVershKlastInd[j-1]];

        if(dMin > dpDistMatr[i][ipVershKlastInd[j]])

            {dMin = dpDistMatr[i][ipVershKlastInd[j]];

            iAccesInd = j;}}

    ipVershVKlast[iAccesInd][ipNumbOfVershInKlast[iAccesInd]] = i;

    ipNumbOfVershInKlast[iAccesInd]++;}

```

Дана частина алгоритму відповідає за визначення приналежності кожного з об'єктів кластеризації до одного з кластерів. Це відношення визначається за матрицею відстаней, об'єкт відноситься до того кластеру до вершини якого його відстань найменша.

5.1 Введення початкових даних для контрольного прикладу

Далі представлені скріншоти програми які ілюструють систему редагування бази даних.

Для доступу до редагування бази даних необхідно ввести пароль. Діалогове вікно яке відповідає за перевірку паролю зображене на рис 5.1.

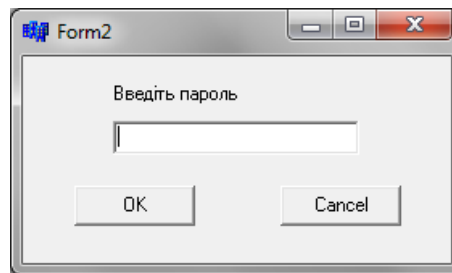


Рисунок 5.1 Діалогове вікно введення паролю

Далі представлені таблиці баз даних, з елементами керування введенням даних.

Таблиця «Аварія» зображена на рис 5.2.

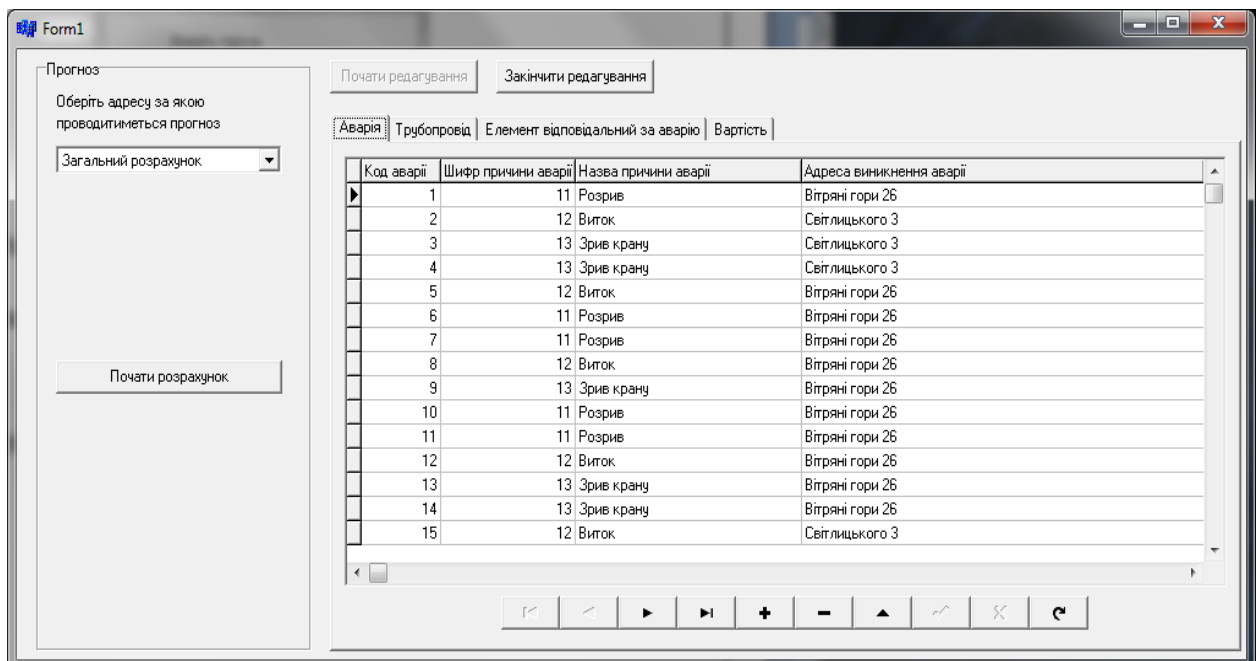


Рисунок 5.2 Діалогове вікно редагування таблиці «Аварія»

Таблиця «Трубопровід» зображена на рис 5.3.

Form1

Прогноз

Оберіть адресу за якою проводитиметься прогноз

Загальний розрахунок

Почати розрахунок

Почати редагування Закінчити редагування

Аварія Трубопровід Елемент відповідальний за аварію Вартість

Код труби	Шифр труби	Довжина труби	Строк експлуатації труби	Технічний стан труби	Діаметр труби	Код аварії
1	111	12	17	Задовільно	50	1
2	112	52	14	Незадовільно	100	2
3	113	26	18	Близьке до задовільного	200	3
4	114	15	15	Незадовільно	100	4
5	115	26	17	Задовільно	50	5
6	116	73	17	Близьке до задовільного	500	6
7	117	46	18	Близьке до задовільного	50	7
8	118	51	19	Близьке до незадовільного	100	8
9	119	24	26	Незадовільно	50	9
10	120	14	28	Задовільно	500	10
11	121	17	24	Незадовільно	500	11
12	122	37	22	Близьке до задовільного	200	12
13	123	72	18	Задовільно	100	13
14	124	52	25	Близьке до незадовільного	50	14
15	125	35	31	Задовільно	500	15
16	126	65	12	Задовільно	500	16

Рисунок 5.3 Діалогове вікно редагування таблиці «Трубопровід»

Таблиця «Елемент відповідальний за аварію» зображена на рис 5.4.

Form1

Прогноз

Оберіть адресу за якою проводитиметься прогноз

Загальний розрахунок

Почати розрахунок

Почати редагування Закінчити редагування

Аварія Трубопровід Елемент відповідальний за аварію Вартість

Код елемента	Шифр елемента	Назва елемента	Строк експлуатації елемента	Технічний стан елемента
1	1111	Тіло труби	13	Незадовільно
2	1113	Колодязь	7	Задовільно
3	1112	Кран	10	Близьке до задовільного
4	1114	Кран	15	Близьке до задовільного
5	1115	Кран	15	Незадовільно
6	1116	Кран	12	Близьке до задовільного
7	1117	Кран	6	Задовільно
8	1118	Кран	16	Задовільно
9	1119	Кран	25	Близьке до незадовільного
10	1112	Кран	25	Незадовільно
11	1112	Кран	16	Задовільно
12	1112	Кран	12	Близьке до незадовільного
13	1112	Кран	13	Близьке до задовільного
14	1112	Кран	21	Задовільно
15	1112	Кран	31	Незадовільно

Рисунок 5.4 Діалогове вікно редагування таблиці «Елемент відповідальний за аварію»

Таблиця «Вартість» зображена на рис 5.5.

Form1

Прогноз

Оберіть адресу за якою проводитиметься прогноз

Загальний розрахунок

Почати розрахунок

Почати редагування Закінчити редагування

Аварія Трубопровід Елемент відповідальний за аварію Вартість

Номер вартості	Назва елементу	Тип заміни	Ціна	Код елементу
1	Тіло труби	Ремонт	200	1
2	Тіло труби	Повна заміна	300	1
3	Колодязь	Ремонт	700	2
4	Колодязь	Повна заміна	1000	2
5	Кран	Ремонт	20	3
6	Кран	Повна заміна	50	3

Navigation buttons: back, forward, first, last, add, delete, up, down, undo, redo, refresh

Рисунок 5.5 Діалогове вікно редагування таблиці «Вартість»

Після закінчення редагування необхідно натиснути кнопку «Закінчити редагування», після цього діалог редагування таблиць стане не активним.

Вигляд неактивного діалогу зображено на рис 5.6.

Form1

Прогноз

Оберіть адресу за якою проводитиметься прогноз

Загальний розрахунок

Почати розрахунок

Почати редагування Закінчити редагування

Аварія Трубопровід Елемент відповідальний за аварію Вартість

Рисунок 5.6 Неактивне діалогове вікно редагування

5.2 Розрахунок контрольного прикладу

Далі представлені скріншоти які ілюструють роботу програми.

Перед початком розрахунку необхідно обрати адресу, дані про аварії на якій будуть кластеризовані. Розгорнутий список вибору адреси зображено на рис. 5.7.

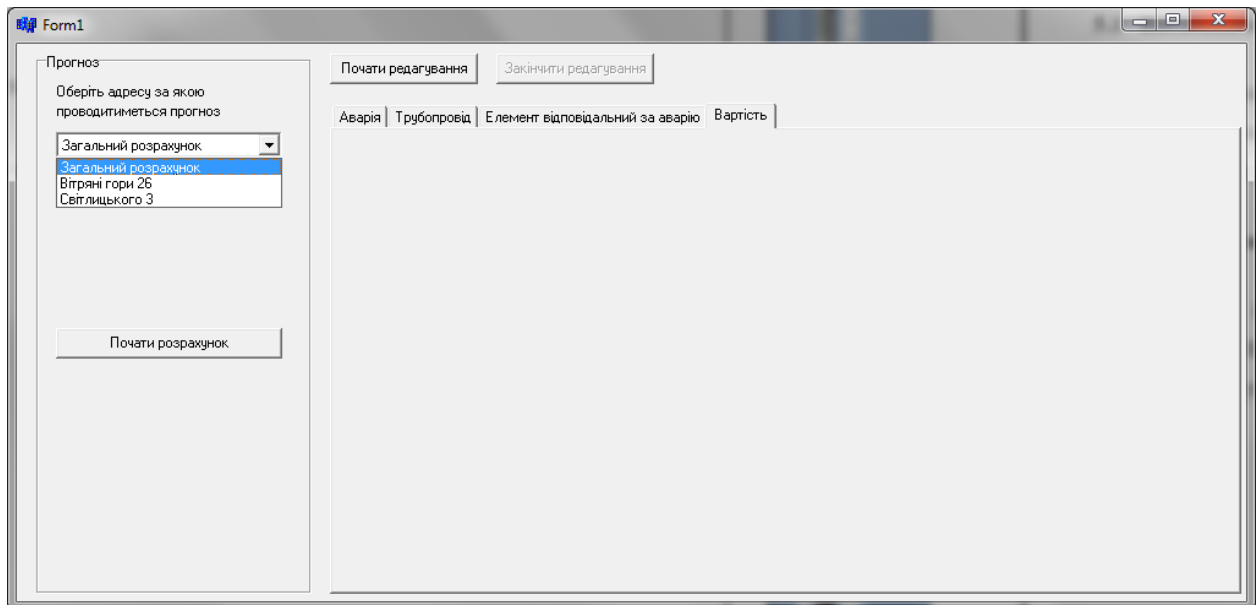


Рисунок 5.7 Розгорнутий список вибору адреси

Для першого тестового прикладу обираємо пункт «Загальний розрахунок». Після цього натискаємо кнопку «Почати розрахунок». Алгоритм програми знаходить усі необхідні для кластеризації дані для кожного з випадків аварії які відбулися за даною адресою. Після цього ці дані вносяться до масиву. Для наглядності у алгоритмі передбачений механізм виводу повідомлень з етапами роботи програми, та даними які на цьому етапі були отримані. Перше повідомлення сповіщає користувача про кількість знайдених записів у базі даних по даній адресі. Це діалогове вікно зображене на рис 5.8. У тестовому прикладі було знайдено 20 записів про аварії.

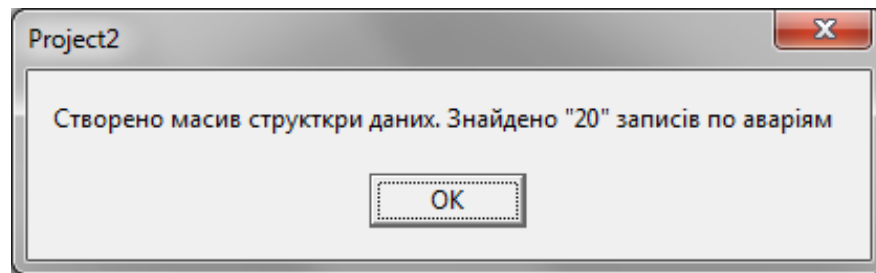


Рисунок 5.8 Повідомлення про кількість записів

Оскільки дані, які були отримані з бази даних знаходяться у різних шкалах вимірювання, вони вноситимуть різний вклад у «схожість» аварій. Щоб цього уникнути програма проводить нормалізацію усіх атрибутів, і приводить їх значення до шкали від 0 до 1. Для наглядності виводиться повідомлення яке сповіщає про створення матриці нормалізованих даних. Та представляє вигляд даних для першого об'єкту. Це повідомлення зображено на рис 5.9.

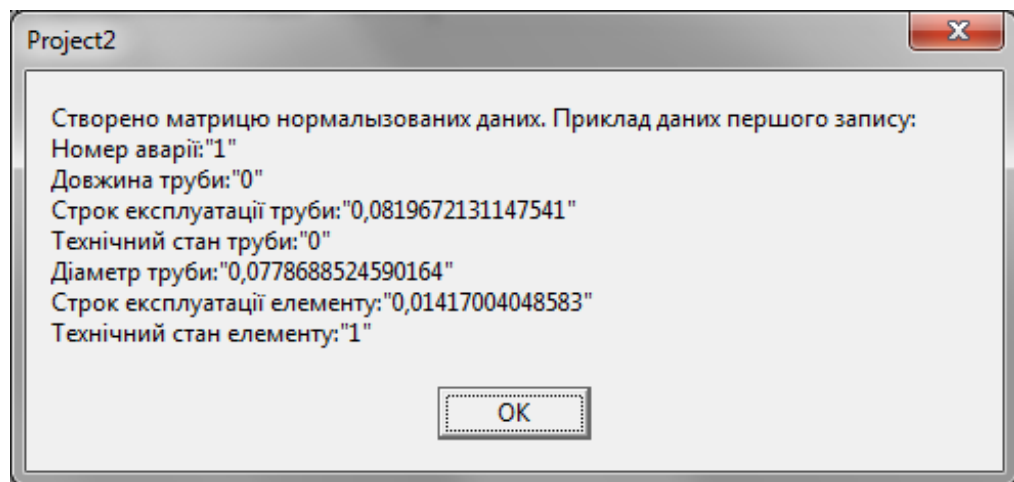


Рисунок 5.9 Повідомлення про створення матриці нормалізованих даних

Наступним кроком алгоритму є визначення віддалей між об'єктами кластеризації, тобто аваріями. Ця відстань визначається за формулою Чебишова. Оскільки матриця відстаней зазвичай величезного розміру, а саме кількість знайдених аварій в квадраті, вивід даних навіть за першим об'єктом

кластеризації не є доцільним. Тому наступне повідомлення сповіщає лише про створення та розрахунок матриці відстаней. Крім того повідомлення сповіщає про початок розрахунку коефіцієнту Альфа. Це повідомлення зображено на рис 5.10.

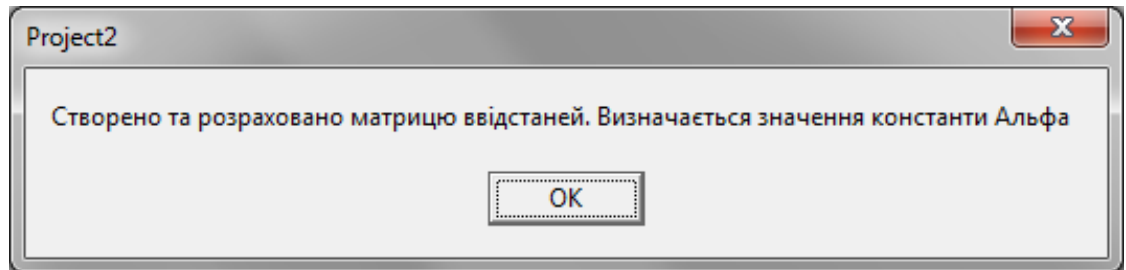


Рисунок 5.10 Повідомлення про створення матриці відстаней

Далі необхідно розрахувати коефіцієнт який відповідає за швидкість сходження алгоритму, після проведення аналізу цього методу кластеризації було визначено що оптимальним коефіцієнтом сходження є одиниця поділена на середнє значення відстані між об'єктами кластеризації (аваріями). Наступне повідомлення сповіщає значення розрахованого коефіцієнту Альфа. У даному прикладі розрахунку значення Альфа = 1,026.

Це повідомлення зображено на рис 5.11.

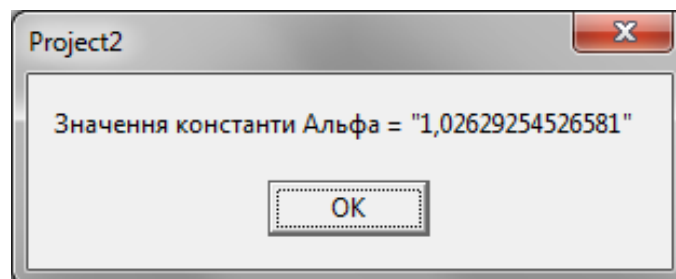


Рисунок 5.11 Повідомлення про значення коефіцієнту Альфа

Наступним кроком роботи програми є визначення потенціалу кожного об'єкту кластеризації як вершин кластеру. Після розрахунку потенціалів визначається максимальне значення. Об'єкт з максимальним значенням потенціалу і є вершиною першого кластеру. Для наглядності виводиться повідомлення яке сповіщає про знаходження першої вершини кластеру, у нашому випадку це вершина №11 значення її потенціалу 8,463. Це повідомлення зображено на рис 5.12.

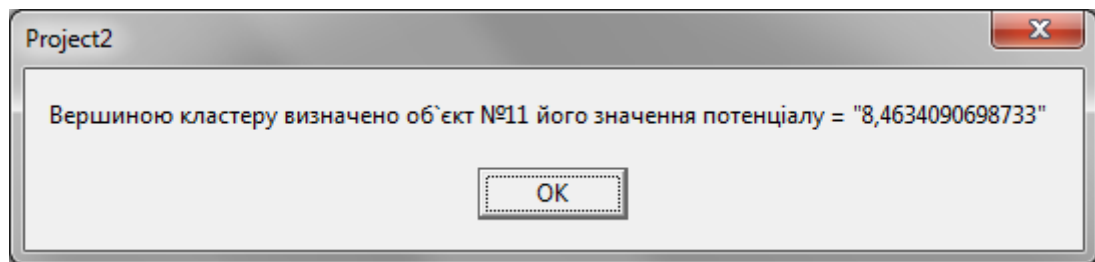


Рисунок 5.12 Повідомлення про знаходження вершин кластеру

Далі програма перераховуватиме значення потенціалу кожного об'єкту для того щоб, усунути вплив знайденої вершини кластеру. Знаходитиметься нова вершина кластеру, й цей цикл повторюватиметься до тих пір доки потенціал вершини знайденого кластеру буде більшим за половину значення потенціалу першої вершини. Для наглядності у алгоритмі передбачено виведення повідомлень що сповіщають про знаходження наступної вершини. Такі діалоги з'являтимуться до тих пір доки алгоритм знаходження вершин не закінчить роботу. Ці повідомлення зображені на рис 5.13 та рис 5.14. У тестовому прикладі було знайдено ще 2 вершини. №7 зі значенням потенціалу 4,97. Та №3 зі значенням потенціалу 1,64.

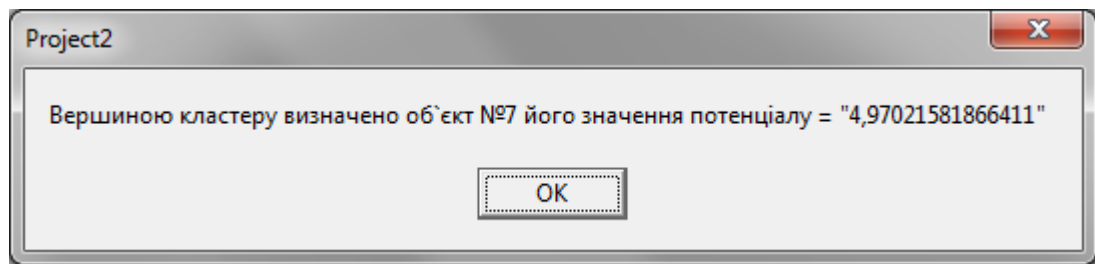


Рисунок 5.13 Повідомлення про знаходження вершин кластеру

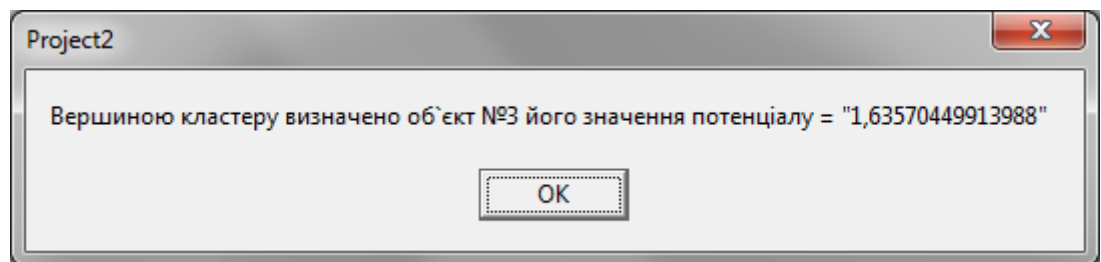


Рисунок 5.14 Повідомлення про знаходження вершин кластеру

Для наглядності виведеться повідомлення яке сповіщає про кількість кластерів, у тестовому прикладі їх було знайдено 3, та номери їх вершин, тобто об'єктів (аварій) які є центрами цих кластерів. У даному випадку це об'єкти 11, 7 та 3. Це повідомлення зображено на рис 5.15.

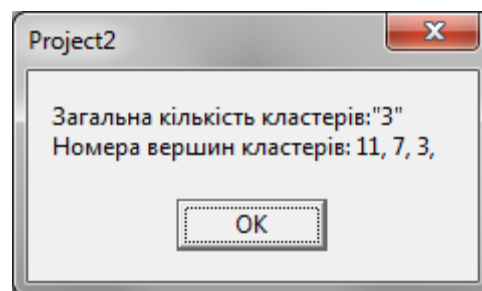


Рисунок 5.15 Повідомлення про кількість знайдених кластерів

Далі програма за допомогою матриці відстаней розраховує приналежність кожного об'єкту до одного з кластерів. Об'єкт відноситься до того кластеру, відстань до вершини якого у об'єкта найменша. Для наглядності алгоритм передбачає вивід повідомлення що сповіщають про знаходження приналежності об'єктів до кластерів. Ці повідомлення зображені на рис 5.16, рис 5.17 та рис 5.18.

До кластеру вершиною якого є об'єкт 11 відносяться об'єкти:
2,5,11,12,14,15,18.

До кластеру вершиною якого є об'єкт 7 відносяться об'єкти:
1,6,7,13,17.

До кластеру вершиною якого є об'єкт 3 відносяться об'єкти:
0,3,4,8,9,10,16,19.

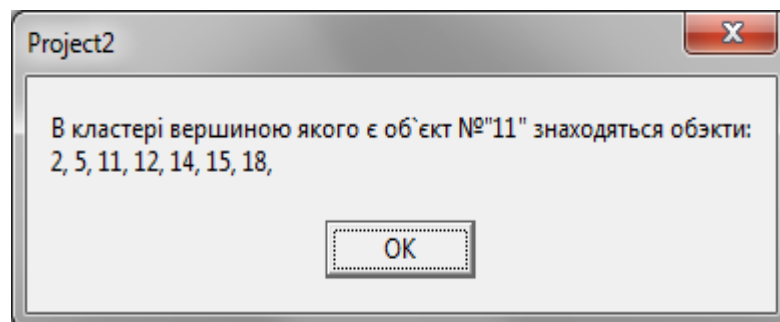


Рисунок 5.16 Повідомлення про об'єкти які знаходяться в кластері

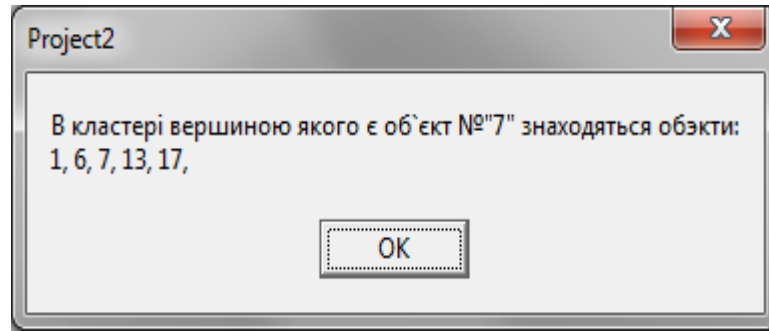


Рисунок 5.17 Повідомлення про об'єкти які знаходяться в кластері

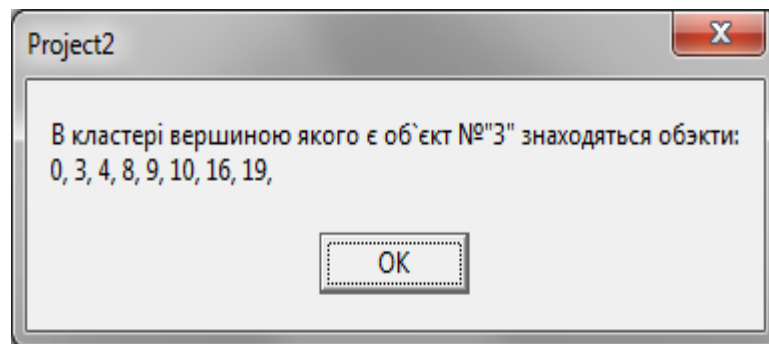


Рисунок 5.18 Повідомлення про об'єкти які знаходяться в кластері

У результаті роботи алгоритму створюються текстові файли, з назвами «Klast1», «Klast2»,... «Klast_», у відповідність до кожного знайденого кластеру. У файлі зазначений місяць та кількість аварій які трапилися у цей місяць. На рис 5.19 зображено типовий файл який буде створено після кластеризації.

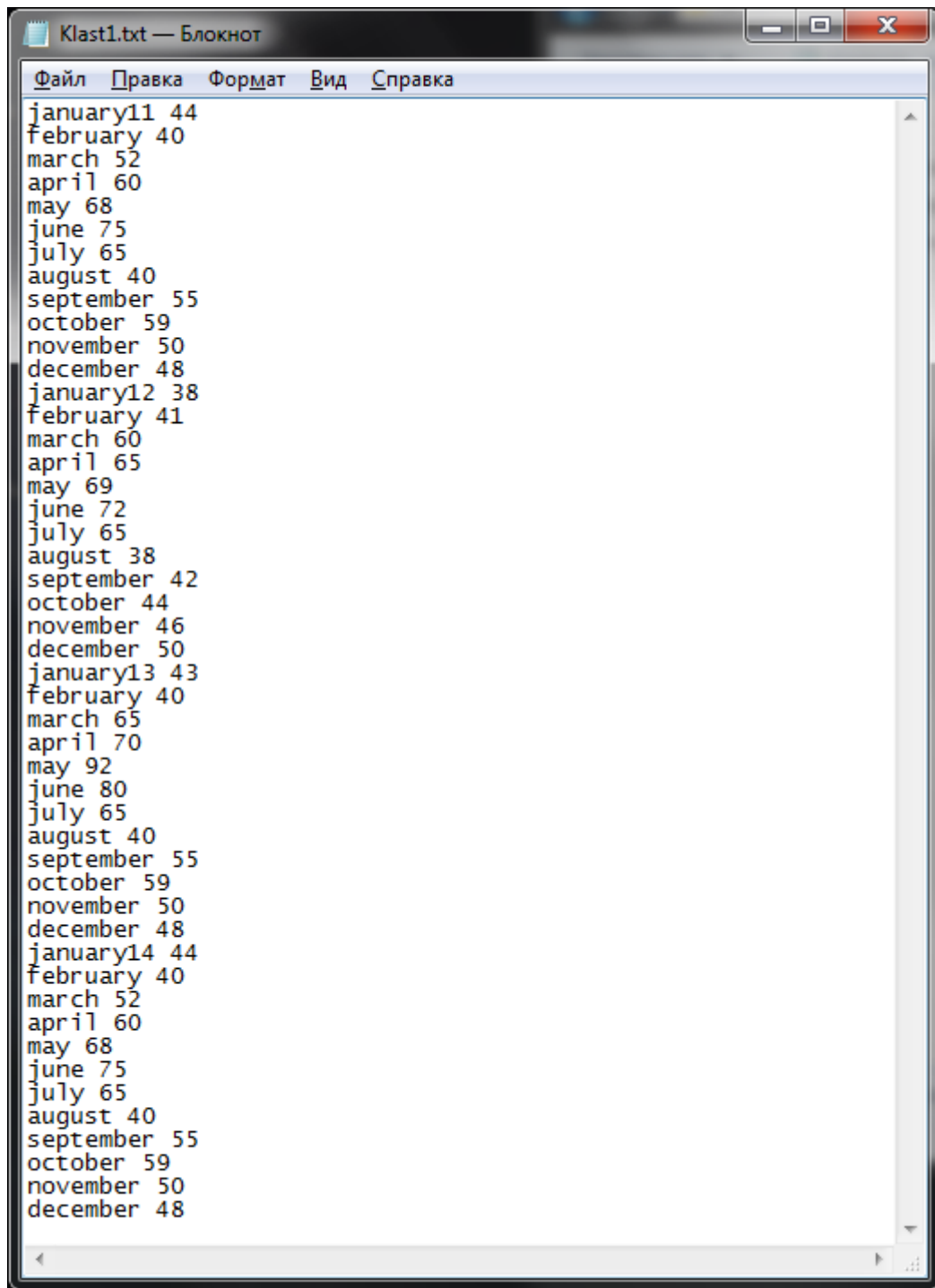


Рисунок 5.19 Типовий файл даних по кластеру

В результаті розрахунку тестового прикладу було визначено 3 кластери вершинами яких є об'єкти (аварії) №11, №7, №3. Програма проводить правильні розрахунки, тому може в майбутньому використовуватися для підготовки даних до прогнозування аварійних ситуацій на водопровідних мережах міста.

СПИСОК ЛІТЕРАТУРИ

1. Наказ N 30 від 05.07.95. Про затвердження Правил технічної експлуатації систем водопостачання та каналізації населених пунктів України.
2. Дипломне проектування: методичні вказівки / уклад.: В.Б. Задоров, С.В. Цюцюра, Г.А. Гайна, В.В. Демченко, І.М. Доманецька, О.В. Ізмайлова, Г.В. Красовська, К.М. Красовський, С.О. Свинар. - К.: КНУБА, 2010.
3. Цюцюра С.В., Свинар С.О. Водопостачання, як складна динамічна система – 2003. – № 2 (17). – С. 52-58.
4. Інженерні мережі та комунікації. Частина І. Водопостачання. Конспект лекцій/ Укладачі: О.А. Петухова, С.А. Горносталь, А.М. Чернуха. – Х.: УЦЗУ, 2008.
5. "Системи водопостачання населених пунктів" Т.В.Бабенко, д-р. техн. наук, проф., Ю.В. Почта. ISSN 2071-2227, Науковий вісник НГУ, 2012.
6. Водний кодекс України №2120-14 від 07.12.2000 р.
7. Закон України №1809-04 “Про захист населення і території від НС техногенного та природного характеру”.
8. Розробка і дослідження алгоритмів кластеризації для великих колекцій документів / [Стех Ю.В., Файсал М.Е. Сардіх, Лобур М.В., М.С. Домброва М.С., Арцибасов В.Є.] // Моделювання та інформаційні технології. Зб. наук. пр. ІППМЕ ім.Г.Є. Пухова НАН України. – Київ, 2010, – Вип.58. – С. 283-290.
9. Алгоритм пошуку оптимальної кількості кластерів / [Стех Ю. В., Файсал М.Е. Сардіх, Лобур М. В., Керницький А.Б.] // Вісник НУ “Львівська політехніка” : Комп’ютерні системи проектування. Теорія і практика. – 2009. – № 651. – С. 129-131.
10. Кластерний аналіз // Вікіпедія [Електронний ресурс]. URL: http://uk.wikipedia.org/wiki/Кластерний_аналіз

11. Data Mining // Вікіпедія [Електронний ресурс]. URL: https://ru.wikipedia.org/wiki/Data_mining
12. Савчук Т. О. Використання ієрархічних методів кластеризації для аналізу надзвичайних ситуацій на залізничному транспорті / Т. О. Савчук, С. І. Петришин // Міжнародний науково-технічний журнал «Вимірювальна та обчислювальна техніка в технологічних процесах» (м. Хмельницький) – 2009. – №1 – С. 193– 198.
13. Метричний простір // Вікіпедія [Електронний ресурс]. URL: https://uk.wikipedia.org/wiki/Метричний_простір
14. ДСТУ ГОСТ 7.1:2006 «Бібліографічний запис. Бібліографічний опис. Загальні вимоги та правила складання (ГОСТ 7.1-2003, IDT)» . - К. : Держстандарт України, 2007.
15. Баррет Д., Ливингстон Д., Браун М. JavaScript Web-професіоналам [Текст]: учеб. пер. с англ. /– К.: Изд.група BHV, 2001. – 240 с.
16. Баррет Я. JavaScript Web-професіоналам [Текст]: учеб. / - К.: BHV-Київ, 2001. – 352 с.
17. Гужва В.М. Інформаційні системи і технології на підприємствах / В.М. Гужва. - К., 2001.-400 с.
18. Дідковська Л.Г. Менеджмент : навч. посіб. / Л.Г. Дідковська, П.Л. Гордієнко. - К. : Алеута ; КНТ, 2007. -516 с.
19. Дубчак Л.В. Інформаційні системи та технології на підприємствах / Л.В. Дубчак. -Ірпінь : Нац. акад. ДПС України, 2005. - 84 с.
20. Інформаційні системи в менеджменті : навч. посіб. / Ю.М. Мінаєв, В.П. Бочарников, О.Ю. Філімонова, М.М. Гузій. - К. : Європ. ун-т, 2003. -191с.
21. Інформаційні технології у маркетингу і рекламі / В.Л. Пілюшенко, Е.І. Словенко, Л.П. Полякова та ін. - Донецьк : ДонДУУ, 2005. - 204 с.

22. Open MPI Documentation [Електронний ресурс] – Режим доступу: <https://www.open-mpi.org/doc/> , вільний.
23. Open MPI Documentation [Електронний ресурс] – Режим доступу: <https://www.open-mpi.org/doc/> , вільний.
24. Плєскач В.Л. Технології електронного бізнесу : монографія / В.Л. Плєскач. - К. : Київ. нац. торг.-екон. ун-т, 2004. - 223 с.
25. П'ятницька Г.Т. Управління підприємством в епоху глобалізму : монографія / Г.Т. П'ятницька. - К. : Логос, 2006. - 568 с.
26. Пінчук Н.С. Інформаційні системи і технології в маркетингу / Н.С. Пінчук, Г.П. Грузинський, Н.С. Орленко. - К., 2003. - 352 с.
27. Системи управління якістю. Основні положення і словник : ДСТУ ISO 9000-2001. -[Чинний від 2001-01-01]. - К. : Держстандарт України, 2001. - 27 с.
28. Сорока К. Основи теорії систем і системного аналізу: Навчальний посібник/Мін-во освіти і науки України, Харківська нац. ак-я міського господарства. - 2-ге вид. - Харків: Видавець Тимченко А. Н., 2005. - 286 с.
29. Чигринська О. С. Теорія економічного аналізу: Навчальний посібник/ О. С. Чигринська, Т. М. Власюк; Мін-во освіти і науки України, Київський нац. ун-т технологій та дизайну . - К.: Центр навчальної літератури, 2006. - 231 с.
30. Старіш О. Системологія: Підручник/ Олександр Старіш,; Мін-во освіти і науки України, Нац. ун-т "Острозька академія". - К.: Центр навчальної літератури, 2005. - 231 с.