

DOI: <https://doi.org/10.36910/6775-2524-0560-2025-58-11>

УДК 004:02

Мороз Борис Іванович, д.т.н., професор

<https://orcid.org/0000-0002-5625-0864>

Шишацький Олександр Олександрович, аспірант

<https://orcid.org/0009-0008-6008-7079>

Національний технічний університет «Дніпровська політехніка», м. Дніпро, Україна

КЛАСИФІКАЦІЯ ВИДІВ ПОМИЛОК В ГЕОПРОСТОРОВІЙ БАЗІ ДАНИХ НА ПРИКЛАДІ ДЕРЖАВНОГО ЗЕМЕЛЬНОГО КАДАСТРУ

Мороз Б. І., Шишацький О.О. Класифікація видів помилок в геопросторовій базі даних на прикладі державного земельного кадастру. У статті досліджено проблему класифікації помилок у геопросторових базах даних на прикладі інформації з Державного земельного кадастру України. Запропоновано розширену класифікацію помилок, яка охоплює атрибутивні, геометричні, топологічні та контекстуальні помилки. Зокрема, розглянуто помилки позиціонування, топологічні порушення, некоректні атрибутивні дані, дублювання та непослідовність даних. Стаття акцентує увагу на важливості автоматизованого виявлення помилок для забезпечення якості геопросторових даних. Зроблено висновок про необхідність подальших досліджень із використанням статистичних методів для ідентифікації аномалій. Запропоновані підходи сприяють підвищенню точності аналізу та підтримують прийняття рішень у геоінформаційних системах.

Ключові слова: геопросторові дані, помилки, класифікація, земельний кадастр, PostGIS.

Moroz B., Shyshatskiy O. Classification of errors in geospatial databases based on the example data of the state land cadastre. The article investigates the issue of error classification in geospatial databases using data of the State Land Cadastre of Ukraine. An extended classification of errors is proposed, including attributive, geometric, topological, and contextual errors. Specifically, it examines positioning errors, topological violations, incorrect attributive data, duplication, and data inconsistency. The article emphasizes the importance of automated error detection to ensure the quality of geospatial data. It concludes with the need for further research using statistical methods for anomaly detection. The proposed approaches enhance analysis accuracy and support decision-making in geographic information systems.

Keywords: geospatial data, errors, classification, land cadastre, PostGIS.

Постановка наукової проблеми. Історично геопросторові дані зберігались у паперовому вигляді, тому люди мало приділяли уваги проблемам, які виникали через неточності та помилки[1]. Так, було відомо що геопросторові набори даних страждають від неточностей, проте спосіб зберігання не давав можливості побачити проблему в цілому. Розвиток комп'ютерних технологій дав поштовх для переходу на цифрові формати зберігання даних. Цей перехід відкрив нові можливості для автоматизації, але також привніс нові виклики.

Внесення даних у геопросторову базу зазвичай включає в себе ручне введення або сканування інформації, що є потенційним джерелом різних помилок через людський фактор, що важко виявити на етапі введення[2]. Цей процес може призвести до появи помилок, таких як дублювання записів, неправильні географічні прив'язки або некоректні атрибути просторових записів. В процесі внесення даних також вносяться і ті помилки, що існували в паперових документах.

Такі помилки в подальшому впливають на прийняття управлінських та юридичних рішень, тому важливо розробити методи для їх автоматизованого виявлення і, за можливості, автоматичного виправлення. Для вирішення цієї наукової проблеми необхідно виконати ряд завдань:

1. Провести аналіз наукової літератури щодо відомих типів помилок і їх причин.
2. Розробити класифікацію помилок за їх типами.
3. Дослідити існуючі методи і алгоритми для автоматизованого пошуку помилок.
4. Створити методи та алгоритми для пошуку помилок.
5. Розробити програму для аналізу геопросторових даних.
6. Провести тестування розробленого програмного забезпечення на реальних даних.

Виконання цих завдань дозволить не лише вирішити окреслену проблему, але й закладе основу для довготривалого забезпечення якості геопросторових даних.

Аналіз останніх досліджень і публікацій. Проблема пошуку помилок в геоінформаційних системах досліджувалась такими науковцями як Dr. Christopher Ndehedehe[3], Mr. Manuel S Pascual [4] та Dr. Chrisman, Nicholas[2]. У своїх дослідженнях автори пропонують розділяти помилки у геопросторових базах даних на кілька категорій:

1. **Помилки позиціонування.** Ці помилки виникають через неточності в положенні даних на карті. Dr. Christopher Ndehedehe зазначає, що такі помилки можуть бути спричинені масштабом карти, методами картографування або людським фактором [3]. Manuel S. Pascual додає, що переформатування даних, наприклад, зміна проекції, може вносити похибки через неоднозначності перетворення координат[4]. Dr. Chrisman акцентує увагу на важливості позиційної точності, зазначаючи, що різниця у точності між джерелами даних може призводити до накопичення похибок при накладанні шарів.
2. **Топологічні помилки.** До топологічних помилок відносяться порушення геометрії, такі як перетини, прогалини або відсутність з'єднань між елементами. Помилки при введенні в дані ГІС можуть призводити як до спотворення самої геометрії, так і до спотворення загальної картини. До спотворення геометрії відносять утворення вузлів у полігонах, незвичних форм полігонів, пошкоджені вихідні дані. До спотворення загальної картини відносять, наприклад, неспівпадіння точок двох дотичних полігонів, що створює перетин або прогалину.
Dr. Chrisman у своїй роботі також окремо виділяє логічну узгодженість. Він стверджує, що сучасні ГІС можуть автоматично перевіряти наявність логічних помилок, таких як відсутні кордони чи непозначені полігони. Такі перевірки можуть виявити помилки, які були б невидимими на рівні атрибутивної чи позиційної точності. Наприклад, простий розрахунок 'point-in-polygon' може уникнути розміщення умовних буйків на суші або річок за межами їх заплав [2].
3. **Неправильні атрибутивні дані.** Атрибутивні помилки виникають у разі неправильного або некоректного присвоєння характеристик об'єктам. Наприклад, сільськогосподарські угіддя можуть бути позначені як болота. Dr. Christopher Ndehedehe підкреслює, що такі помилки часто пов'язані з людським фактором і суб'єктивністю в класифікації даних.

Результати публікацій свідчать про те, що точність і якість даних є критично важливими для проєктів ГІС. Помилки, неточності та невизначеності можуть серйозно вплинути на результати аналізу.

У той же час, аналіз наявних публікацій виявив недолік: попри значний внесок у дослідження проблеми, автори робіт наводять лише основні типи помилок (позиційні, топологічні, атрибутивні). Класифікація виконана на загальному рівні, без деталізації. Для подальших досліджень необхідно деталізувати наведені авторами види помилок, описати їх визначення та методи пошуку.

Метою статті є формулювання авторського визначення поняття "аномалії у геопросторових базах даних" та уточнення і поглиблення відомих класифікацій помилок для подальшого дослідження алгоритмів для їх ідентифікації.

Виклад основного матеріалу. Проблеми з даними та помилки в геопросторових базах даних можна озагальнити словом "аномалія". Аномалія є відхиленням від очікуваних або стандартних характеристик даних[5], яке у випадку з геопросторовими базами даних може бути результатом помилки, викликаного людським фактором, технічними збоями або спадковими проблемами з першоджерел даних.

У контексті геопросторових даних, аномалії можуть проявлятися у вигляді:

- **Помилки позиціонування:** перетину або накладання об'єктів.
- **Топологічних порушень:** наприклад, об'єктів, які знаходяться поза межами дозволеної області;
- **Невірних атрибутивних даних:** невідповідних категорій об'єктів, характеристик, тощо;

Для аналізу та класифікації видів помилок у межах статті було обрано базу даних земельних ділянок України, що опублікована Державною службою України з питань геодезії, картографії та кадастру протягом 2020-2022 рр. Хоча в земельному кадастрі наявні найрізноманітніші дані (від опорних пунктів геодезичної мережі до координат державного кордону), дослідження в межах цієї статті обмежується лише даними про земельні ділянки та їх атрибути.

На першому етапі було підготовано дані для аналізу. З допомогою програмних засобів було завантажено та поєднано два джерела даних:

- Набір даних про зареєстровані земельні ділянки з офіційного сайту Держгеокадастру [6].
- Прикладний програмний інтерфейс для доступу до відомостей про земельну ділянку [7].

Інформація з відкритих джерел була поєднана і завантажена в геопросторову базу даних, що містить таблицю, поля якої зазначені в таблиці №1. Для зберігання даних використано об'єктно-реляційну базу даних PostgreSQL з встановленим додатком PostGIS, який дозволяє зберігати геопросторові дані. Загальна кількість записів у отриманій таблиці - 24,058,660 рядків.

Таблиця 1. Опис створених полів бази даних

Поле	Опис
CAD_NUM	Кадастровий номер земельної ділянки.
OWNERSHIP_CODE	Код форми власності на земельну ділянку.
OWNERSHIP	Назва форми власності земельної ділянки.
PURPOSE_CODE	Код категорії та призначення земель.
CATEGORY	Назва категорії земель за цільовим призначенням.
PURPOSE	Назва цільового призначення земельної ділянки.
UNIT_AREA	Одиниці виміру площі земельної ділянки.
AREA_SIZE	Числове значення площі земельної ділянки.
GEOMETRY	Геометрія земельної ділянки в проекції EPSG:4326

По завершенню підготовчого етапу, проведено дослідження отриманого набору даних на наявність різних видів помилок. Результатом дослідження стала детальна класифікація видів помилок, наведена нижче.

Помилки атрибуції. Атрибутивна інформація охоплює лише ті поля, які не мають прямого зв'язку з геопросторовими даними чи їх похідними. У цьому розділі розглянуто класифікацію основних типів помилок атрибутивних даних за їхніми характеристиками.

Жирним курсивом наведено коротку назву виду помилок, за яким детально описано визначення помилки та розглянуто механізм виявлення:

1. **Вихід значень атрибуту поза задані обмеження.** Атрибут має обмежений набір допустимих значень або фіксований діапазон, проте фактичні значення атрибуту виходять за межі цих обмежень. Виявлення такого типу помилок виконується поетапно. Розглянемо процес пошуку цих помилок детальніше на прикладі атрибуту OWNERSHIP_CODE у досліджуваному наборі даних.

Запит:		
<pre>SELECT ownership, ownershipcode, count(id) as count FROM parcels_2023 GROUP BY ownership, ownershipcode ORDER BY count DESC;</pre>		
Результати:		
OWNERSHIP	OWNERSHIPCODE	Кількість ділянок
"Приватна власність"	"100"	20255499
"Не визначено"	"0"	1751902
"Комунальна власність"	"200"	1324549
"Державна власність"	"300"	726648
"Не визначено"	"4"	35
"Не визначено"	"1"	12
"Не визначено"	"12"	6
...
"Не визначено"	"21"	1

Рис. 1. Інформація про OWNERSHIPCODE у зрізі за 2022 рік.

Першим етапом перевірки є визначення обмежень. У випадку з земельними ділянками, джерелом таких правил є Положення "Про Порядок ведення Державного земельного кадастру" яке встановлює фіксований набір значень, яке може приймати поле

OWNERSHIP_CODE. Згідно з Додатком 3 Положення [8], можливі значення мають відповідності як зазначено в таблиці №2.

Другим етапом є порівняння отриманих значень із встановленими обмеженнями та виявлення відхилень. Для проведення перевірки було сформовано SQL-запит, який отримує з бази даних усі унікальні значення атрибуту OWNERSHIP_CODE разом із кількістю повторень. Результати виконання запиту наведено на рисунку №1. Отримані результати показують, що у кадастрових даних за 2022 рік атрибут OWNERSHIP_CODE містить 62 записи зі значеннями, які не входять до переліку допустимих.

Таблиця 2. Форми власності земельних ділянок [8].

Код	Форма власності
100	Приватна власність
200	Комунальна власність
300	Державна власність

Крім того, значна кількість записів (1,751,902 або 7.28% від загальної кількості) не містить інформації про форму власності. Для позначення відсутності даних використано код 0, який не передбачено затвердженою документацією, що свідчить про застосування неформального підходу до позначення відсутності інформації.

2. **Відсутні дані.** Атрибути мають порожні значення там, де повинна бути інформація. Перевірку на відсутність даних можливо здійснити шляхом SQL-запиту, який виявляє порожні значення в обраному атрибуті. Розглянемо цей тип помилок на прикладі атрибуту CATEGORY у досліджуваному наборі даних.

SQL-запит для аналізу:

```
SELECT category, count(*)  
FROM public.parcels_2022  
WHERE category IN (NULL, '')  
GROUP BY category;
```

Результати:

CATEGORY	Кількість ділянок
"	79

Рис. 2. Інформація про відсутній атрибут CATEGORY у зрізі за 2022 рік.

Як показано на рисунку № 2, у базі даних виявлено 79 записів, у яких атрибут CATEGORY має порожнє значення. Відсутність даних у цьому полі порушує вимоги щодо повноти зазначення інформації[9].

3. **Неправильний тип даних.** Атрибут містить значення з невідповідним типом даних, наприклад, текст замість числа.

Для виявлення цього виду помилок складено SQL-запит, який перевіряє можливість конвертації значень у числовий формат та сортує їх у порядку спадання. Для перевірки обрано атрибут AREA_SIZE, який за своїм визначенням має бути дробовим числом, однак у досліджуваних наборах відкритих даних він представлений у вигляді текстового типу.

Як показано на рисунку № 3, поле AREA_SIZE містить записи, які не можуть бути перетворені в числовий формат, через що повертається значення NULL. Усього виявлено 21 запис із такими помилками.

Запит:

```
SELECT
  cadnum, area, (
  CASE
    WHEN trim(area) ~ '^-?\d+(\.\d+)?\$$'
      THEN CAST(area AS DECIMAL)
    ELSE NULL
  END) AS area_as_decimal
FROM public.parcels_2022
ORDER BY area_as_decimal DESC
LIMIT 20;
```

Результати:

CADNUM	AREA_SIZE	AREA_SIZE_DECIMAL
"4412700000:12:005:0038"		null
"0122787500:07:001:0280"		null
...
"4411400000:16:003:0056"		null

Рис. 3. Недостовірні значення AREA_SIZE за 2022 рік (усього 21 запис).

4. **Орфографічні помилки.** Дублювання записів, що мають однакове логічне значення, проте внесені в базу даних в різному форматі через орфографічні помилки.

Помилки в написанні можуть ускладнювати пошук та групування даних. У досліджуваному наборі даних це стосується, наприклад, поля PURPOSE, яке відображає призначення земельної ділянки.

Основні підходи до пошуку орфографічних помилок у текстових полях базуються на алгоритмах пошуку схожих текстових рядків: перевірі ключових колізій та методу найближчих сусідів.

Перевірка ключових колізій — передбачає створення альтернативного представлення рядка у вигляді ключа, наприклад, після приведення його у нижній регістр, очищення від розділових знаків, видалення зайвих пробілів, тощо. Усі рядки, які мають однаковий ключ, об'єднуються в одну групу. Цей метод дуже швидкий, але не гнучкий.

Результати:

PURPOSE	Кількість записів
Для ведення особистого селянського господарства	12713
для ведення особистого селянського господарства	9425
для ведення особистого селянського господарства	137
Для ведення особистого селянського господарства.	39
Для ведення особистого селянського господарства	28
Для ведення особистого селянського господарства	17
Для ведення особистого селянського господарства,	10

Рис. 4. Приклади варіацій атрибуту "PURPOSE" знайдені методом Key Colision.

Метод найближчих сусідів — використовує параметр, що визначає поріг відстані: будь-які рядки, відстань між якими менша за задане значення, будуть включені в одну групу разом. Цей метод дозволяє точніше регулювати відмінності між рядками, але потребує великих обчислювальних ресурсів, оскільки потрібно порівнювати всі пари рядків, що уповільнює процес для великих наборів даних.

Результати:

PURPOSE	Кількість записів
Для ведення особистого селянського господарства	12713
для ведення особистого селянського господарства	9425
для ведення особистого селянського господарства	137
Для ведення особистого селянського господарства.	39
Для ведення особистого селянського господарства,	10
"Для ведення особистого селянського господарства"	2
- Для ведення особистого селянського господарства	2
.Для ведення особистого селянського господарства	1
Для ведення особистогоселянського господарства	1
для ведення особистого селянськогогосподарства	1
для ведення особистогоселянського господарства	1
для веденняособистого селянського господарства	1

Рис. 5. Приклади варіацій атрибуту "PURPOSE" знайдені методом Nearest Neighbor.

На рисунках №№ 4, 5, продемонстрована перевага алгоритму найближчих сусідів над ключовими колізіями у пошуку орфографічних помилок: завдяки гнучкості параметризованого радіусу вдалося знайти додатково такі помилки:

- Використання символів ' та '- для обрамлення запису.
- Помилки у написанні слів (відсутність розділового знаку): "особистогоселянського" та "селянськогогосподарства".

5. **Непослідовність даних.** Атрибути, що повинні бути узгодженими, мають різні значення. Для перевірки непослідовності даних створимо SQL-запит, який витягне з бази даних інформацію про всі пари значень атрибутів OWNERSHIP та OWNERSHIPCODE, а також кількість таких записів.

Запит:

```
SELECT ownership, ownershipcode, count(id) as count
FROM parcels_2023
GROUP BY ownership, ownershipcode
ORDER BY count DESC;
```

Результати:

OWNERSHIP	OWNERSHIPCODE	Кількість ділянок
"Приватна власність"	"100"	20255499
"Не визначено"	"0"	1751902
"Комунальна власність"	"200"	1324549
"Державна власність"	"300"	726648
"Не визначено"	"4"	35
"Не визначено"	"1"	12
"Не визначено"	"12"	6
...
"Не визначено"	"21"	1

Рис. 6. Інформація про OWNERSHIP та OWNERSHIPCODE.

На рисунку № 6 показано, що основна кількість ділянок має однозначне зіставлення між полем OWNERSHIP та OWNERSHIPCODE, проте, існують невеликі групи записів з різними значеннями OWNERSHIPCODE, але з OWNERSHIP = "Не визначено". Ці записи не мають чіткої ідентифікації типу власності і скоріше за все були занесені помилково.

6. **Неправильні значення.** Атрибути мають значення, які виходять за межі допустимих або не відповідають реальності. Наприклад, від'ємна площа.

Поле AREA_SIZE за своїм визначенням є дробовим числом, проте у наборах відкритих даних представлено у вигляді строкового типу. Зважаючи на це, складемо SQL запит таким чином, аби він одночасно конвертував значення в число та сортував значення від більшого до меншого.

Запит:

```

SELECT
  cadnum, area, (
    CASE
      WHEN trim(area) ~ '^-?\d+(\.\d+)?\$$'
      THEN CAST(area AS DECIMAL)
      ELSE NULL
    END) AS area_as_decimal
FROM public.parcels_2022
ORDER BY area_as_decimal DESC
LIMIT 20;

```

Результати:

CADNUM	AREA_SIZE	AREA_SIZE DECIMAL
"2320681002:01:003:0001"	"1410819481.2743"	1410819481.2743
"2324580600:01:004:0045"	"1385675089.7042"	1385675089.7042
"3222487200:04:006:0094"	"937517730.8986"	937517730.8986
"3222487200:04:001:0040"	"937365803.7717"	937365803.7717
"1822383800:03:000:0135"	"916102795.6748"	916102795.6748

Рис. 7. Аномально великі значення AREA_SIZE.

На рисунку № 7 наведено приклад значень AREA_SIZE, які є аномально великими площами, на кшталт 1410819481.2743 га. Число 1,410,819,481 (або близько 1.41 мільярда гектар) є надзвичайно великим. Для порівняння: загальна площа України становить приблизно 603,500 квадратних кілометрів, що еквівалентно 60.35 мільйона гектар. Таким чином, 1.41 мільярда гектар перевищує площу України більш ніж у 20 разів.

Помилки геометрії. У геопросторових базах даних можуть також виникати різноманітні топологічні помилки, пов'язані з геометрією. У цьому розділі розглянуто помилки, що стосуються полігональних об'єктів, адже саме такий вид об'єктів використовується у базі даних земельних ділянок.

Індивідуальні помилки. Цей вид помилок відносяться до тих, що можуть бути визначені для індивідуального об'єкту окремо від контексту, тобто без врахування його сусідів або атрибутів. Такі помилки можуть суттєво впливати на точність і надійність геопросторового аналізу, тому важливо проводити валідацію та корекцію топології в базах даних.

Класифікація індивідуальних помилок полігональних об'єктів. Як і в попередньому розділі, **жирним курсивом** наведено коротку назву виду помилок, за яким детально описано визначення помилки та розглянуто механізм виявлення:

1. **Самоперетин.** Контур одного полігону або лінії перетинається сам із собою, як показано на рис. 8. Може бути наслідком помилки при введенні даних, спрощення або похибки при конвертації між різними проекціями.

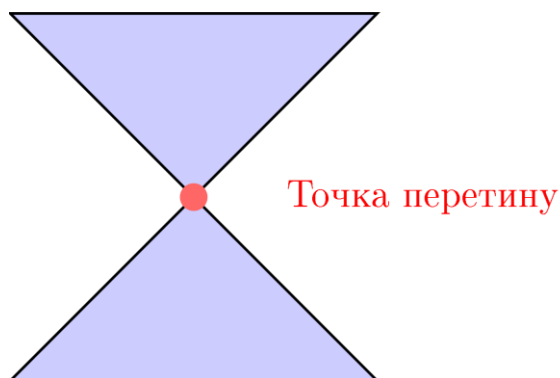


Рис. 8. Приклад самоперетину полігону.

2. **Невірне замикання.** Полігони мають незамкнені контури, що порушує їхню цілісність, як показано на рис. 9. Виникає, коли координати кінцевих точок полігона не співпадають з його початковими точками, що призводить до розриву контурів.

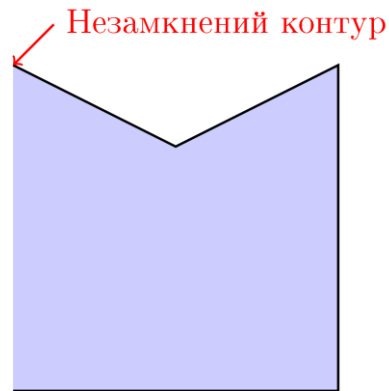


Рис. 9. Приклад незамкнутого контуру полігону

3. **Дублікати ребер.** Полігони мають задубльовані ребра. Ребра дублюються, що спотворює представлення полігону, як показано на рис. 10.

Дубльоване ребро

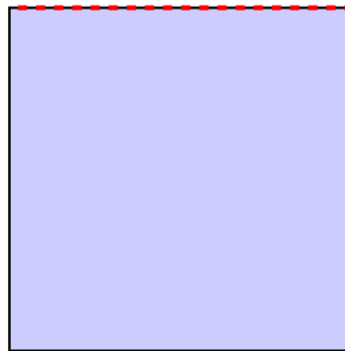
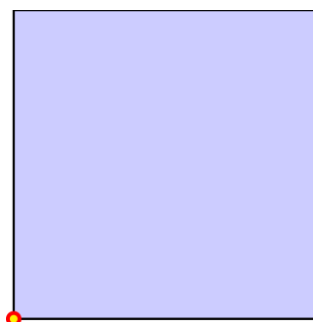


Рис. 10. Дублікати ребер.

Ця проблема може виникати через невірне визначення або обробку координат, коли одні й ті самі ребра заново вводяться або не правильно об'єднуються.

4. **Дублікати вершин.** Полігони мають задубльовані вершини. Ця проблема може спотворювати представлення геометрії, утруднювати аналіз та перевірку топології полігонів, як показано на рис. 11. Задубльовані вершини виникають через помилки в обробці даних, наприклад, при некоректній конверсії форматів або редагуванні геометрії.



Дублікат вершини

Рис. 11. Приклад дублікату вершини.

5. **Отвори поза контуром.** Полігони мають отвори, що розташовані поза основним контуром полігону, як показано на рисунку. 12. Виникає, коли внутрішні області полігона помилково визначені поза його межами.

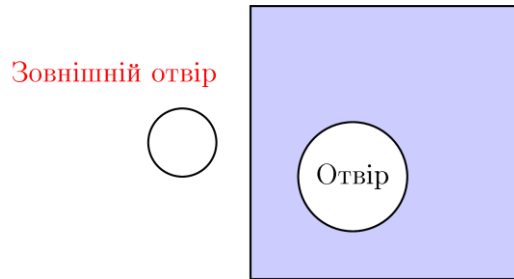


Рис. 12 . Приклад отвору поза контуром полігона.

6. **Роз'єднані частини.** Полігони мають отвори, які розрізають полігон на декілька не з'єднаних частин, як показано на рисунку № 13.

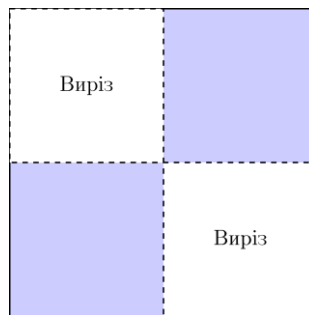


Рис. 13. Роз'єднані частини

Одним із методів для аналізу коректності індивідуальних полігональних об'єктів в середовищі PostGIS є використання функції ST_IsValidDetail. Ця функція дозволяє отримати детальну інформацію про причину невалідності геометрії.

Для виконання такого аналізу можна скористатися SQL-запитом. Результати можна згрупувати за причинами невалідності геометрії, таким чином визначивши основні типи помилок та їх поширення, що і продемонстровано на рисунку № 14.

Результати свідчать, що понад 29 000 земельних ділянок мають помилки типу Self-intersection, що становить майже 90% усіх випадків. Значна частина помилок також пов'язана з перетинами кільцевих структур (Ring Self-intersection), що часто свідчить про проблеми під час створення або обробки геометрії.

SQL-запит для аналізу геометрії:

```
SELECT reason(ST_IsValidDetail(geometry::geometry)) AS reason, count(*)
FROM public.parcels_2022
GROUP BY reason;
```

Результати:

Причина помилки	Кількість ділянок
Self-intersection	29587
Ring Self-intersection	2901
Duplicate Rings	232
Interior is disconnected	7
Hole lies outside shell	1

Рис. 14. Статистика помилок геометрії земельних ділянок.

Серед рідкісних помилок можна виділити проблеми типу Interior is disconnected (7 випадків) та Hole lies outside shell (1 випадок), які свідчать про серйозні порушення цілісності геометрії. Водночас, кількість геометрії з помилками відносно загального об'єму бази даних становить лише 0.12%.

Важливо зазначити, що перевірка геометрії в PostGIS базується виключно на евклідовій геометрії, без урахування еліпсоїдальної форми Землі, що може впливати на результати для даних глобального масштабу. Для таких випадків необхідний додатковий аналіз із використанням геодезичних функцій.

Контекстуальні помилки. Цей вид помилок може бути визначений лише врахувавши контекст об'єкту, тобто його сусідів, атрибутів, відношення до інших об'єктів. Основні види помилок наведені нижче.

1. **Перетин з сусідніми об'єктами.** Полігони або лінії, що мають бути суміжними, перетинаються, як показано на рис. 15. Помилка виникає коли координати суміжних точок не використовують повторно, а визначають окремими і незалежними вимірюваннями, що призводить до внесення похибки в кожному з вимірювань.

Крім того, помилка може бути результатом того, що координати для сусідніх об'єктів визначено в різних системах координат. У такому випадку спроба перетворення координат у єдиний формат спричиняє похибки при конвертації, що призводить до утворення перетинів або розривів між об'єктами.

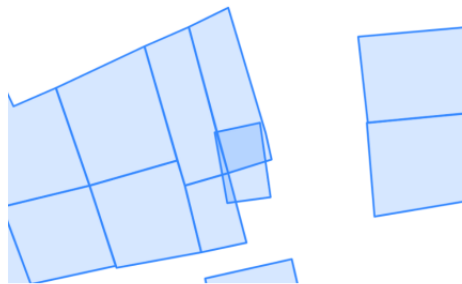


Рис. 15. Приклад накладання полігону на сусіда.

2. **Розриви.** Проміжки між полігонами, які повинні щільно прилягати один до одного, як показано на рис. 16.

Помилка виникає, коли координати суміжних об'єктів, таких як полігони, визначаються окремо та незалежно для кожного з сусідніх об'єктів, що призводить до виникнення похибок.

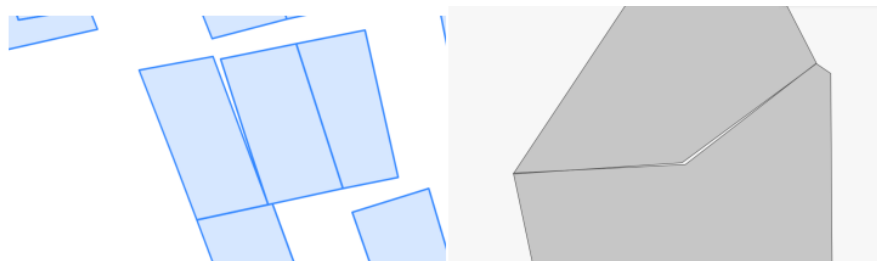


Рис. 16. Приклади розривів між суміжними полігонами.

3. **Перетин із забороненими областями.** Геометричні об'єкти, які перетинають області, де вони не повинні перебувати, як показано на рис. 17.

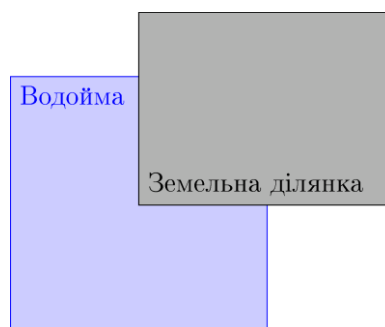


Рис. 17. Ілюстрація ділянки, що перетинає водойму.

Помилка виникає, коли полігони або інші геометричні елементи, такі як дороги, будівлі чи водойми, мають неправильне просторове розташування відносно інших об'єктів або шарів даних. Помилка часто є наслідком неточності в координатах або відсутності топологічної узгодженості між різними наборами даних.

Висновки та перспективи подальших досліджень. У результаті проведеного аналізу відкритих даних Державного земельного кадастру було виявлено кілька категорій помилок, що можуть суттєво впливати на точність, достовірність і практичну цінність інформації в геопросторових базах даних.

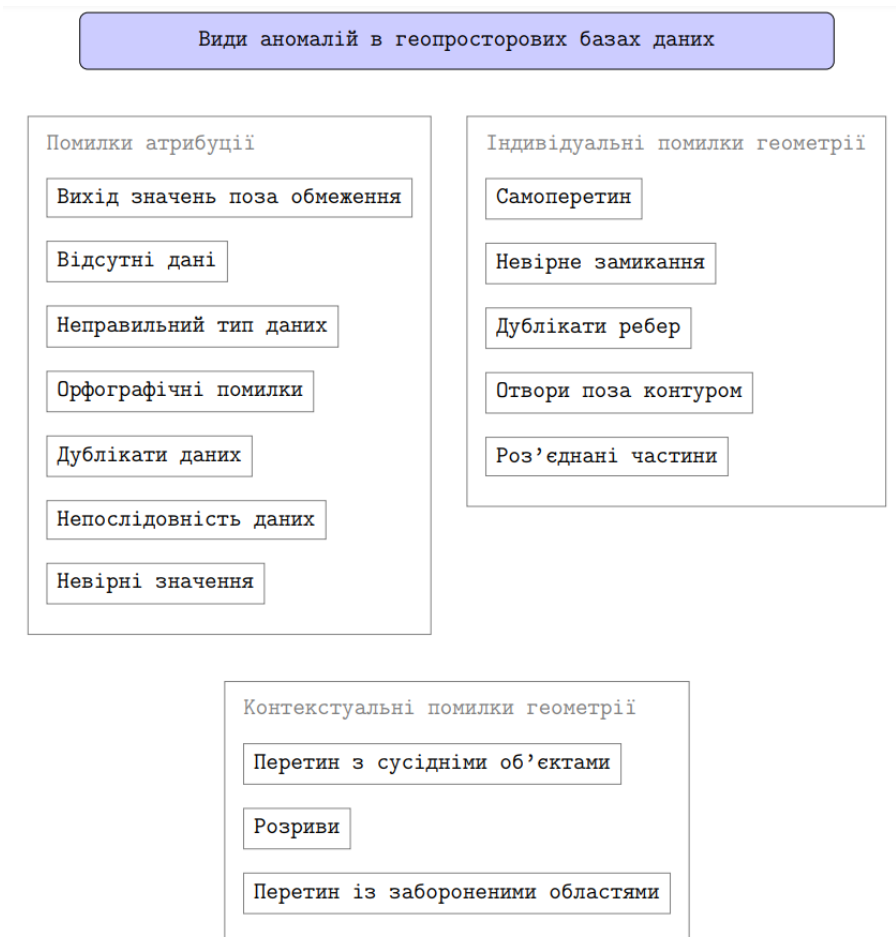


Рис. 18. Візуальна класифікація видів аномалій.

Основними досягненнями дослідження є такі:

1. **Розширення класифікації помилок:** На основі аналізу існуючих досліджень та результатів роботи з даними було створено розширену класифікацію помилок, яка наведена нижче в скороченій формі у вигляді ілюстрації представленої на рисунку № 18.
2. **Ідентифікація недоліків існуючих підходів:** Аналіз літератури та практичних прикладів показав, що існуючі класифікації не є універсальними, а рівень деталізації часто недостатній для вирішення практичних задач.
3. **Візуалізація помилок:** Для покращення розуміння проблеми було створено низку графічних ілюстрацій, що демонструють поширені типи помилок, такі як самоперетини, незамкнені контури, дублікати вершин і отворів поза межами полігонів.

Класифікація помилок є лише першим етапом у більш широкому процесі аналізу геопросторових даних, що спрямований на виявлення аномалій. Пошук аномалій — це наступний крок, який призначений для виявлення нетипових значень, які можуть бути результатом неточностей у даних. Подальше дослідження буде зосереджено на застосуванні статистичних методів для виявлення аномалій. Зокрема, буде використовуватись статистичний аналіз, включаючи методи класифікації,

кластеризації, а також виявлення залежностей між різними атрибутами. Статистичні методи дозволять виявити незвичайні патерни в даних, що можуть свідчити про неточності або проблеми з якістю інформації, не завжди очевидні через традиційні підходи до перевірки помилок. Завдяки застосуванню таких методів, можна буде виявити не тільки очевидні помилки, а й більш складні аномалії, що не мають явних ознак помилок, але можуть суттєво впливати на точність аналізу та прийняття рішень. Це дозволить значно покращити якість геопросторових баз даних.

Список бібліографічного опису.

1. Tsorlini A., Gavrilis D., Makridis S. An example of studying the evolution of a local geographic milieu in early 20th century Greece: Generalkarte (1900-1904) vs National mapping (1917) representations // Proceedings of 5th International Workshop Digital Approaches to Cartographic Heritage, Vienna, Austria. 2010. С. 22–24.
2. Chrisman N. R. The error component in spatial data // Geographical information systems. 1991. Т. 1. №12. С. 165–174.
3. Ndehedehe C. Understanding Errors and their Measurement in Geoinformation // Journal of Environmental Sciences and Resources Management. 2013.
4. Pascual M. S. GIS Data: A Look at Accuracy, Precision, and Types of Errors. URL: <https://www.geographyrealm.com/gis-data-a-look-at-accuracy-precision-and-types-of-errors/> (дата звернення: 10.11.2024).
5. Академічний тлумачний словник Української мови. URL: <https://sum.in.ua/s/anomalija> (дата звернення 10.11.2024).
6. Державна служба України з питань геодезії, картографії та кадастру. Набір даних про зареєстровані земельні ділянки. 2022. URL: <https://web.archive.org/web/20220204115835/https://data.land.gov.ua/> (дата звернення: 10.11.2024).
7. Державна служба України з питань геодезії, картографії та кадастру. API Е-сервіса Держгеокадастру. 2021. URL: <https://web.archive.org/web/20210713075141/https://e.land.gov.ua/api/doc> (дата звернення: 10.11.2024).
8. Верховна Рада України. Додаток 18 до Порядку ведення Державного земельного кадастру. Код форми власності на земельну ділянку. URL: <https://zakon.rada.gov.ua/laws/show/1051a-2012-p/ed20230411#n54> (дата звернення: 10.11.2024).
9. Верховна Рада України. Додаток 3 до Порядку ведення Державного земельного кадастру. Категорії земель. URL: <https://zakon.rada.gov.ua/laws/show/1051b-2012-p/ed20230208#n6> (дата звернення: 10.11.2024).

References:

1. Tsorlini A., Gavrilis D., Makridis S. An example of studying the evolution of a local geographic milieu in early 20th century Greece: Generalkarte (1900-1904) vs National mapping (1917) representations. Proceedings of 5th International Workshop Digital Approaches to Cartographic Heritage, Vienna, Austria. 2010. pp. 22–24.
2. Chrisman N. R. The error component in spatial data. Geographical information systems. 1991. Vol. 1, No. 12, pp. 165–174.
3. Ndehedehe C. Understanding Errors and their Measurement in Geoinformation. Journal of Environmental Sciences and Resources Management. 2013.
4. Pascual M. S. GIS Data: A Look at Accuracy, Precision, and Types of Errors. URL: <https://www.geographyrealm.com/gis-data-a-look-at-accuracy-precision-and-types-of-errors/> (access date: 10.11.2024).
5. Akademichnij tлумachnij slovník. URL: <https://sum.in.ua/s/anomalija> (access date: 10.11.2024).
6. Derzhavna Sluzhba Ukrainy z Pytan Heodezii, Kartohrafii ta Kadastru. Nabor danykh pro zareiestrovani zemelni dilianky. 2022. URL: <https://web.archive.org/web/20220204115835/https://data.land.gov.ua/> (access date: 10.11.2024).
7. Derzhavna Sluzhba Ukrainy z Pytan Heodezii, Kartohrafii ta Kadastru. API E-servisa Derzhheokadastru. 2021. URL: <https://web.archive.org/web/20210713075141/https://e.land.gov.ua/api/doc> (access date: 10.11.2024).
8. Verkhovna Rada Ukrainy. Dodatek 18 do Poriadku vedennia Derzhavnoho zemelnogo kadastru. Kod formy vlasnosti na zemelnu dilianku. URL: <https://zakon.rada.gov.ua/laws/show/1051a-2012-p/ed20230411#n54> (access date: 10.11.2024).
9. Verkhovna Rada Ukrainy. Dodatek 3 do Poriadku vedennia Derzhavnoho zemelnogo kadastru. Katehorii zemel. URL: <https://zakon.rada.gov.ua/laws/show/1051b-2012-p/ed20230208#n6> (access date: 10.11.2024).