

**МЕТОДИ, АЛГОРИТМИ ТА ПРОГРАМИ
ДЛЯ ВИЯВЛЕННЯ ПОМИЛОК В ГЕОПРОСТОРОВІЙ БАЗІ ДАНИХ
НА ПРИКЛАДІ ДЕРЖАВНОГО ЗЕМЕЛЬНОГО КАДАСТРУ**

Анотація. Ведення великих геопросторових баз даних супроводжується значними ризиками внесення помилок, що можуть негативно впливати на точність даних та подальші процеси їх використання. Основними причинами помилок є людський фактор, технічні неточності при поєднанні даних та розбіжності у проєкціях зберігання. В умовах постійного зростання обсягу даних, ручні перевірки стають малоефективними, що зумовлює необхідність розробки методів автоматизованого пошуку помилок.

*Метою статті є розробка та перевірка алгоритмів для виявлення помилок у на прикладі бази земельних ділянок Державного земельного кадастру України. Результати дослідження показали ефективність застосування просторових запитів, таких як *point-in-polygon* та *polygon-in-polygon*, для виявлення некоректного розташування об'єктів. Запропоновані алгоритми дозволяють зменшити кількість помилок шляхом автоматичного контролю правильності внесення інформації. Застосування автоматизованих механізмів перевірки сприятиме підвищенню якості кадастрових даних, що є важливим для сфери містобудування та управління земельними ресурсами.*

Ключові слова: геопросторові бази даних, аномалії, Державний земельний кадастр, автоматизована перевірка, просторовий аналіз, алгоритми пошуку помилок.

Постановка проблеми. Підтримка великих геопросторових баз даних супроводжується значними ризиками внесення помилок, які можуть негативно впливати на якість даних і, відповідно, на подальші процеси роботи з ними. Основні джерела таких помилок включають людський фактор під час введення або обробки даних, технічні неточності при інтеграції даних з різних джерел (зокрема через розбіжності у системах координат та форматах зберігання). Впровадження автоматизованих механізмів перевірки дозволяє підвищити продуктивність обробки даних та мінімізувати ймовірність помилок, особливо тих, що повторюються і піддаються алгоритмічній перевірці.

Зі збільшенням обсягу інформації ручні методи верифікації стають неефективними через високі часові витрати, що зумовлює необхідність впровадження автоматичного пошуку та класифікації помилок. Для вирішення цієї наукової проблеми необхідно виконати ряд завдань:

- Провести аналіз поширених видів помилок.
- Створити методи та алгоритми для пошуку цих помилок.

- Розробити програму для перевірки геопросторових даних.
- Провести тестування розробленого програмного забезпечення на реальних даних.

Оскільки геопросторові дані можуть мати критичну роль у процесах управління земельними ресурсами, містобудування, екологічного моніторингу та інших сферах, розробка алгоритмів та програм для підтримки достатнього рівня їхньої точності є актуальним завданням.

Аналіз останніх досліджень. Проблема пошуку помилок в геоінформаційних системах у загальному вигляді досліджувалась різними науковцями. У своїх дослідженнях автори підтверджували актуальність проблеми та поділяли помилки у геопросторових базах даних на кілька високорівневих категорій[1][2]:

- *Помилки позиціонування.* Виникають через неточності у визначенні розташування об'єктів на карті.
- *Топологічні помилки.* До цього типу помилок належать геометричні порушення, зокрема перетини, розриви або відсутність зв'язків між елементами.
- *Атрибутивні помилки.* Цей тип похибок виникає при неправильному присвоєнні характеристик геооб'єктам. Наприклад, землі сільськогосподарського призначення можуть бути помилково віднесені до боліт.

На жаль, досліджувані джерела не деталізують помилки та не пропонують конкретні методи і алгоритми їх пошуку.

Мета статті. Метою статті є опис методів пошуку окремих видів помилок геопросторових баз даних, розробка алгоритмів та перевірка їх ефективності на прикладі бази даних земельних ділянок Земельного кадастру України.

Основний матеріал. У своїй роботі доктор Dr. Chrisman[1] описує один із можливих підходів до перевірки валідності просторових даних на прикладі програми, в яку вноситься розташування річкових буїв. Він наголошує, що об'єкти, які за своєю природою повинні розташовуватися на водній поверхні, повинні мати координати, що потрапляють у визначені межі (рис. 1), в іншому випадку це може свідчити про помилки у даних.



Рисунок 1 - Схематичне зображення річкового буя та визначеної акваторії

Для автоматизованої верифікації цієї умови може бути використаний просторовий запит point-in-polygon, який дозволяє перевірити, чи знаходяться точкові об'єкти (у даному випадку – буї) у межах полігональних об'єктів (межах водних територій). Такий підхід є простим, але ефективним інструментом контролю якості даних і може бути застосований не лише для навігаційних об'єктів, але й у ширшому спектрі завдань,

пов'язаних із перевіркою коректності розташування об'єктів у геопросторових базах даних.

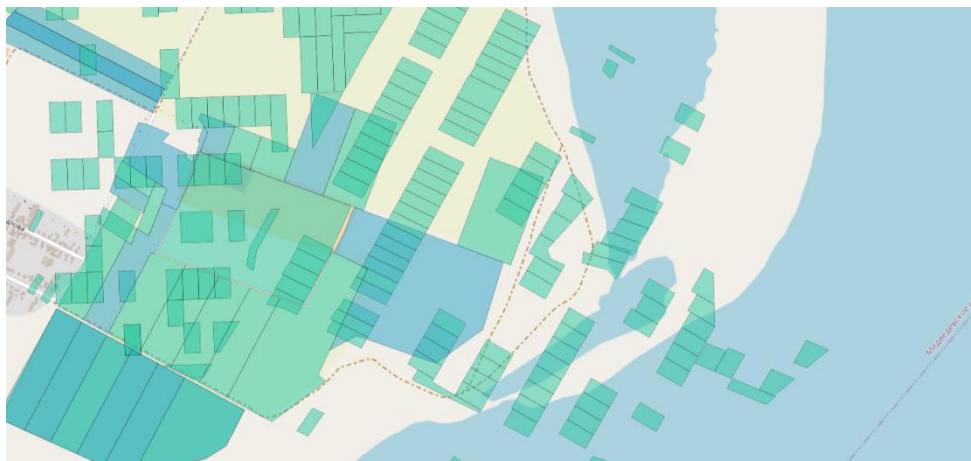


Рисунок 2 - Приклад помилково зареєстрованих земельних ділянок на узбережжі Криму

Цей підхід можна адаптувати для перевірки коректності інформації про земельні ділянки, зокрема щодо їхнього розташування у заборонених зонах або зонах з обмеженням використання.

Перевірка входження ділянки до дозволеної зони.

Прямою аналогією до річкових буїв та акваторій є перевірка того, чи знаходиться земельна ділянка або її частина у межах прибережної захисної смуги або водного об'єкта. Такий вид помилок можна знайти вручну в Державному земельному кадастрі (рис. 2), а отже ця проблема потребує механізмів для її автоматичного розпізнавання.

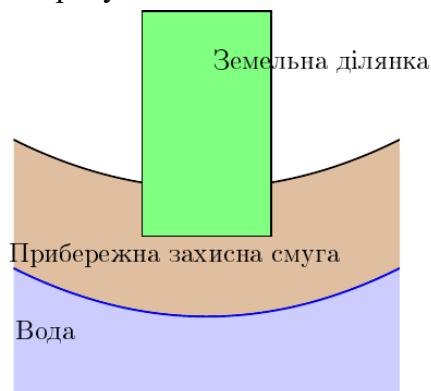


Рисунок 3 - Схематичне зображення земельної ділянки що перетинає прибережну захисну смугу

Необхідність такого виду перевірки визначається положеннями законів України, зокрема Водного кодексу України, які встановлюють особливі обмеження у користуванні та власності таких об'єктів. Стаття 6 Водного кодексу встановлює, що води (водні об'єкти) є виключно власністю Українського народу і надаються тільки у користування [3]. При цьому водним об'єктом вважається будь-який природний або створений

штучно елемент довкілля, в якому зосереджуються води. Це може бути море, лиман, річка, струмок, озеро, водосховище, ставок, канал, тощо.

Окрім обмежень безпосередньо на території водних об'єктів, Водний кодекс також визначає поняття прибережної захисної смуги[3] як частини водоохоронної зони відповідної ширини вздовж водойм, на якій встановлено більш суворий режим господарської діяльності, ніж на решті території. Такі зони встановлюються з метою охорони поверхневих водних ресурсів від забруднення та засмічення, а також для збереження їхньої водності. Відповідно до законодавчих норм, у межах таких смуг можуть діяти певні обмеження на використання земель, включаючи заборону на будівництво, ведення сільськогосподарської діяльності або зміну ландшафту.

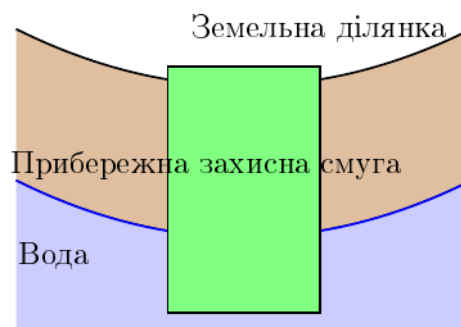


Рисунок 4 - Схематичне зображення земельної ділянки що перетинає водну поверхню

Обмеження господарської діяльності, встановлені Водним кодексом, мають бути внесені у базу даних земельного кадастру при реєстрації ділянки, тобто є окремим об'єктом відомостей кадастру[4], що є потенційним місцем виникнення помилок через людський фактор.

Для автоматизованої перевірки відповідності розташування земельних ділянок цим нормативним вимогам може бути використаний просторовий аналіз із запитом *polygon-in-polygon* або *intersects*, який дозволяє визначити, чи перетинається контур ділянки з межами водної поверхні (рис. 4). Такий підхід дозволяє виявити потенційні проблеми ще на етапі реєстрації ділянки або під час моніторингу змін у кадастрових даних.

Оскільки у прибережних смугах встановлюються обмеження у використанні земельних ділянок[5], доцільно також автоматично перевірити потрапляння земельної до меж прибережної смуги (див. рис. 3), що унеможливить відсутність внесених у базу даних обмежень для конкретної ділянки.

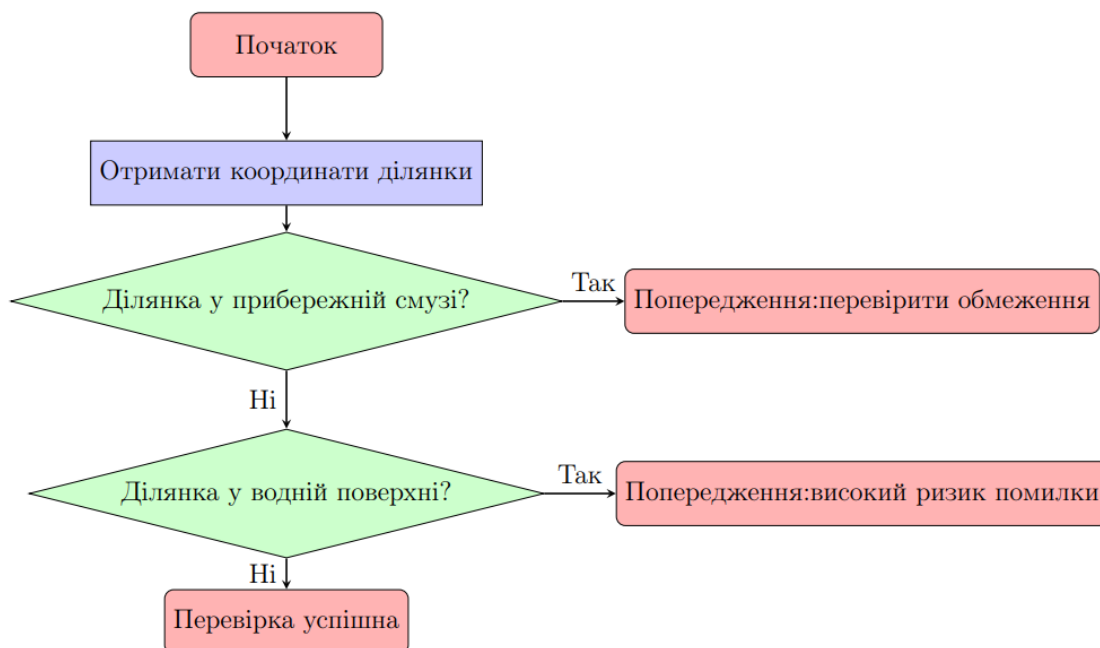


Рисунок 5 - Блок-схема перевірки входження ділянки у прибережну смугу та водну поверхню

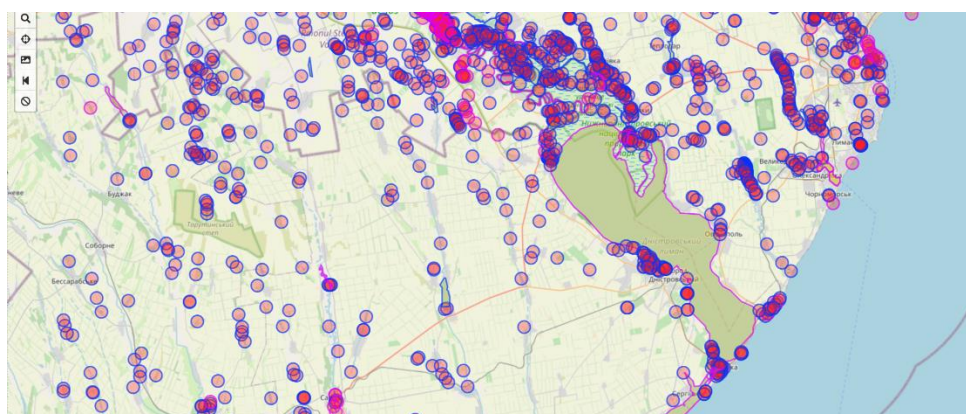


Рисунок 6 - Результати пошуку геометрії водних об'єктів в OpenStreetMap з допомогою overpass turbo

За наявною інформацією складемо алгоритм перевірки місцеположення ділянок:

1. Початок
2. Отримати координати земельної ділянки
3. Перевірити належність земельної ділянки до прибережної захисної смуги:
 - а. Якщо ділянка входить у прибережну захисну смугу → Вивести попередження "Перевірити обмеження"
 - б. Інакше → Перейти до кроку 4
4. Перевірити, чи ділянка розташована на водній поверхні:
 - а. Якщо ділянка потрапляє у межі водної поверхні → Вивести попередження "Високий ризик помилки"

в. Інакше → Перейти до кроку 5

5. Кінець

Схематично алгоритм представлений на рисунку №5.

Для практичної перевірки алгоритму з відкритого джерела OpenStreetMap завантажено масив полігонів, які позначають межі водних об'єктів в межах Одеського району Одеської області, а з Державного земельного кадастру - масив геометрій земельних ділянок. Завантаження меж водних об'єктів здійснено інструментом `overpass turbo`, який дозволяє робити запити за допомогою інструменту формування запитів (рис. 6). Для завантаження меж земельних ділянок використано ресурс `kadastr.live`.

В результаті отримано два набори даних - 2268 меж водних об'єктів та 417573 межі земельних ділянок. Створено програму для вибірки з наборів даних тих об'єктів, які підпадають під умову розробленого алгоритму. В результаті виконання програми знайдено 1378 записів в земельному кадастрі, які перетинають водну поверхню (рис. 7).

Результати:				
cadnum	geometry
0	5110800000:03:001:0478	...	MULTIPOLYGON	
1	5110800000:03:001:0478	...	MULTIPOLYGON	
2	5120484200:01:002:1057	...	POLYGON	
3	5121010100:01:001:0130	...	MULTIPOLYGON	
4	5121084100:01:001:1460	...	POLYGON	
...	
1373	5121085600:01:001:0586	...	POLYGON	
1374	5110300000:02:004:0266	...	POLYGON	
1375	5110300000:02:015:0006	...	POLYGON	
1376	5110300000:01:001:0657	...	POLYGON	
1377	5110300000:02:001:0427	...	POLYGON	

[1378 rows x 88 columns]

Рисунок 7 - Перелік ділянок з перетином водної поверхні

Ручна перевірка результатів показує, що більшість результатів - лише попередження про входження ділянки у захисну зону, проте є і приклади ділянок, які розташовані, наприклад, посеред Хаджибейського лиману (рис. 8). Впровадження автоматизованих методів перевірки на ранніх етапах внесення даних може допомогти в подальшому уникнути такої проблеми.

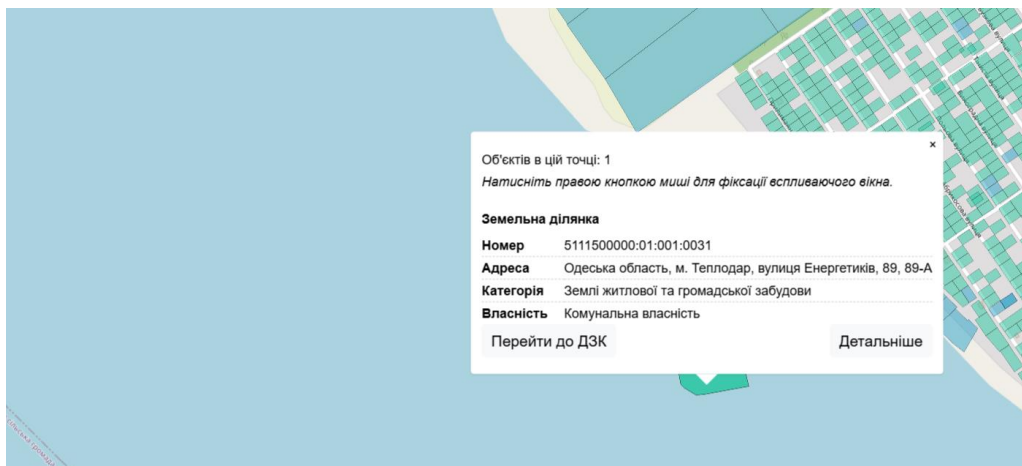


Рисунок 8 - Приклад ділянки зареєстрованої посеред Хаджибейського лиману

Перевірка відповідності атрибутивної інформації до зонування

Не завжди заборонені зони відносяться до усіх об'єктів одного класу. Наприклад, у об'єктах Природно-Заповідного фонду можуть створюватись земельні ділянки, власне, заповідного фонду, або на території озер - ділянки для ведення рибного господарства. Важливо враховувати контекст атрибутивних значень при перевірці. Контекст може враховуватись як алгоритмічно, так і з використанням методів обробки природної мови[6].

НКЗ : НКК : НЗД

Рисунок 9 - Формат кадастрового номера

Розглянемо детальніше це твердження на прикладі кадастрового номеру. Формат кадастрового номера має чітку структуру (рис. 9), що складається з трьох компонентів[7]:

1. 12 цифр НКЗ — номер кадастрової зони, в якому останні дві цифри відокремлюються від перших десяти двокрапкою;
2. 3 цифри НКК — номер кадастрового кварталу в межах кадастрової зони;
3. 4 цифри НЗД — порядковий номер земельної ділянки в межах кадастрового кварталу

Розбиття кадастрового номера на складові проілюстровано на рис. №10.

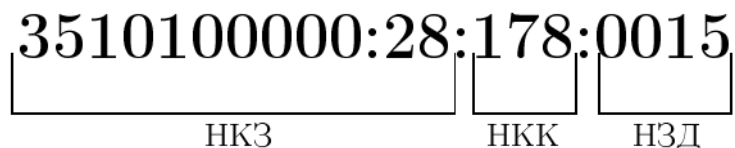


Рисунок 10 - Розбиття кадастрового номера на складові частини

Інформація про межі зон та кварталів зберігається в індексній кадастровій карті і містить дані про геометрію кожної зони та кварталу, це дозволяє провести логічну перевірку: усі земельні ділянки повинні просторово належати відповідним кадастровим зонам і кварталам, зазначеним у їх номерах. Схематично це проілюстровано на малюнку №11.



Рисунок 11 - Кадастрова зона та земельна ділянка в третьому кварталі

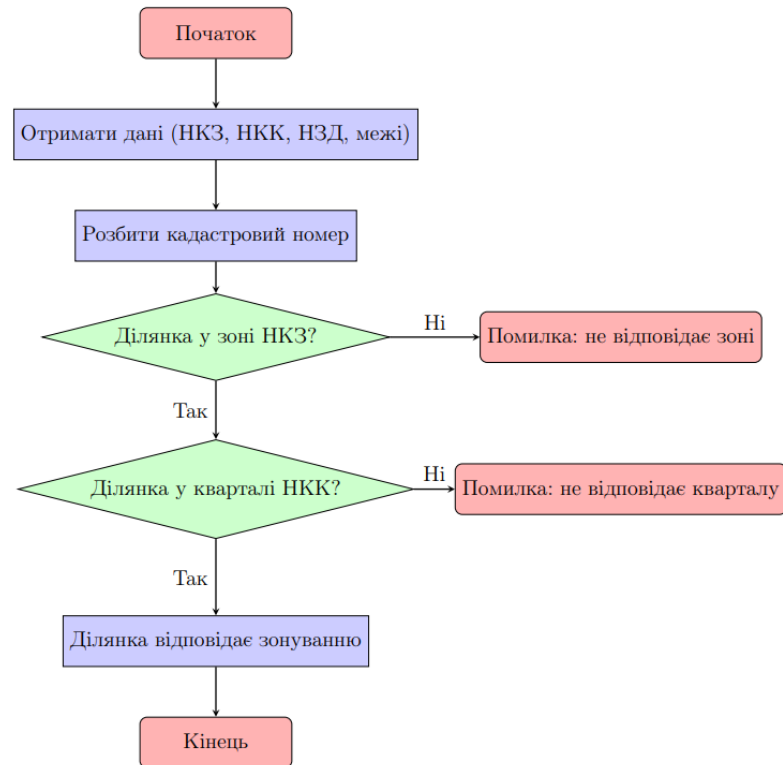


Рисунок 12 - Блок-схема перевірки відповідності земельної ділянки індексній карті

За наявною інформацією побудуємо алгоритм перевірки відповідності ділянки кадастровій зоні та кварталу:

1. Початок
2. Отримати дані:
 - a. Кадастровий номер земельної ділянки (НКЗ, НКК, НЗД)
 - b. Геометричні межі ділянки
 - c. Геометричні межі кадастрових зон і кварталів
3. Розбити кадастровий номер на складові:
 - a. Виділити НКЗ (номер кадастрової зони)
 - b. Виділити НКК (номер кадастрового кварталу)
 - c. Виділити НЗД (номер земельної ділянки)
4. Перевірити належність земельної ділянки до відповідної кадастрової зони:
 - a. Якщо межі ділянки входять у межі зони НКЗ → Перейти до кроку 5
 - b. Інакше → Вивести помилку "Ділянка не відповідає кадастровій зоні"
5. Перевірити належність земельної ділянки до відповідного кадастрового кварталу:
 - a. Якщо межі ділянки входять у межі кварталу НКК → Перейти до кроку 6
 - b. Інакше → Вивести помилку "Ділянка не відповідає кадастровому кварталу"
6. Результат перевірки:

- а. Якщо всі перевірки пройдені успішно → Вивести "Ділянка відповідає зонуванню"

7. Кінець

У вигляді блок-схеми описаний алгоритм показано на рис. 12.

Для перевірки роботи алгоритму було завантажено масив земельних ділянок у межах м. Києва, що налічує 53 849 об'єктів. На основі розробленого алгоритму створено програму, яка здійснює перевірку відповідності зазначеної у кадастровому номері зони та кварталу фактичному місцю розташування ділянки. Як вихідні дані використовувалися офіційні відомості про земельні ділянки та індексна кадастрова карта, опубліковані Державною службою геодезії, картографії та кадастру у 2020–2023 рр.

Для обробки геопросторових даних та реалізації алгоритму використано мову програмування Python із залученням бібліотеки GeoPandas. Поєднання векторних даних із кадастрової карти та їх перевірка здійснювалися шляхом пошуку перетинів, що дозволило автоматично виявляти невідповідності між записами кадастру та реальним місцем розташування земельних ділянок.

```
Результати:  
> python check_data.py  
Warning: cadnum 8000000000:75:741:0028 does not match zona and kvart  
Warning: cadnum 8000000000:72:184:0054 does not match zona and kvart  
Warning: cadnum 8000000000:75:741:0130 does not match zona and kvart  
Warning: cadnum 8000000000:82:011:0020 does not match zona and kvart  
Warning: cadnum 8000000000:72:252:0236 does not match zona and kvart  
Warning: cadnum 8000000000:85:311:0029 does not match zona and kvart  
Warning: cadnum 8000000000:79:117:0003 does not match zona and kvart
```

Рисунок 13 - Результати перевірки кадастрових номерів м. Києва

Результати виконання програми наведені на рис. 13 та показують, що у 7 ділянках неправильно зазначені кадастрові зони або квартали, що може свідчити про помилку на етапі формування кадастрової ділянки або внесення змін до неї.

Додаткові матеріали. З вихідними кодами розроблених програм можна ознайомитись за посиланням: <https://github.com/oshyshatskyi-phd/parcels-check-programs/>

Висновки. Описані в статті методи та алгоритми дозволяють здійснювати комплексну перевірку як просторових, так і атрибутивних даних. Наведені приклади перевірки належності об'єктів до визначених географічних зон:

- земельних ділянок, що входять до прибережних захисних смуг
- відповідності кадастрових номерів фактичному розташуванню

Результати виконання розроблених програм демонструють ефективність підходу. Використання просторових запитів типу *point-in-polygon* та *polygon-in-polygon* дозволяє швидко ідентифікувати потенційні помилки на ранніх етапах обробки даних. Практична перевірка алгоритмів на реальних даних підтвердила ефективність у виявленні невідповідностей.

Отже, застосування розроблених методів може підвищити загальну точність та надійність інформації, яка є основою для прийняття важливих управлінських рішень у сфері містобудування, земельного кадастру та екологічного моніторингу. Подальші

дослідження можуть бути спрямовані на розширення функціоналу алгоритмів з метою виявлення більш специфічних помилок, пошук універсальних методів аналізу простої інформації.

ЛІТЕРАТУРА

1. Chrisman, N. R. The error component in spatial data / N. R. Chrisman // *Geographical information systems*. – 1991. – № 1.12. – С. 165–174.
2. Pascual, M. S. GIS Data: A Look at Accuracy, Precision, and Types of Errors [Електронний ресурс] / M. S. Pascual. – URL: <https://www.geographyrealm.com/gis-data-a-look-at-accuracy-precision-and-types-of-errors/>.
3. Верховна Рада України. Водний кодекс України [Електронний ресурс]. – URL: <https://zakon.rada.gov.ua/laws/show/213/95-%D0%B2%D1%80#Text>.
4. Верховна Рада України. Закон України «Про Державний земельний кадастр» [Електронний ресурс]. – URL: <https://zakon.rada.gov.ua/laws/show/3613-17#Text>.
5. Верховна Рада України. Земельний кодекс України [Електронний ресурс]. – Режим доступу: <https://zakon.rada.gov.ua/laws/show/2768-14#Text>.
6. Євланов М. В., Мороз Б. І., Мороз Д. М., Лучицький В. В. Інформаційна технологія виявлення термінів та артефактів проєкту у вимогах до інформаційної системи // АСУ та прилади автоматики. – 2024. – № 1(182). – С. 73–93. – DOI: <https://doi.org/10.30837/0135-1710.2024.182.073>
7. Верховна Рада України. Структура номера земельної ділянки [Електронний ресурс]. – URL: <https://zakon.rada.gov.ua/laws/show/1051-2012-%D0%BF#n244>.

REFERENCES

1. Chrisman, N. R. (1991). The error component in spatial data. *Geographical Information Systems*, 1(12), 165–174.
2. Pascual, M. S. (n.d.). GIS data: A look at accuracy, precision, and types of errors. Retrieved from <https://www.geographyrealm.com/gis-data-a-look-at-accuracy-precision-and-types-of-errors/>
3. Verkhovna Rada Ukrainy. (1995). *Water Code of Ukraine* (No. 213/95-VR). Retrieved from <https://zakon.rada.gov.ua/laws/show/213/95-%D0%B2%D1%80#Text>
4. Verkhovna Rada Ukrainy. (2011). *Law on the State Land Cadastre* (No. 3613-17). Retrieved from <https://zakon.rada.gov.ua/laws/show/3613-17#Text>
5. Verkhovna Rada Ukrainy. (2001). *Land Code of Ukraine* (No. 2768-14). Retrieved from <https://zakon.rada.gov.ua/laws/show/2768-14#Text>
6. Yevlanov, M. V., Moroz, B. I., Moroz, D. M., & Luchytskyi, V. V. (2024). Information technology for detecting project terms and artifacts in information system requirements. *ASU and Automation Devices*, 1(182), 73–93. <https://doi.org/10.30837/0135-1710.2024.182.073>.
7. Verkhovna Rada Ukrainy. (2012). *Structure of the cadastral number of a land plot* (Resolution No. 1051-2012-P). Retrieved from <https://zakon.rada.gov.ua/laws/show/1051-2012-%D0%BF#n244>

Received 12.03.2025.
Accepted 14.03.2025.

***Methods, algorithms, and programs for error detection
in a geospatial database: a case study of the state land cadastre***

Maintaining large geospatial databases involves significant risks of errors, which can negatively impact data accuracy and subsequent usage processes. The primary causes of errors include human factors, technical inaccuracies in data integration, and discrepancies in storage projections. With the continuous growth of data volume, manual verification methods become inefficient, necessitating the development of automated error detection methods.

The aim of this study is to develop and validate algorithms for error detection using the land parcel database of the State Land Cadastre of Ukraine as a case study. The research results demonstrate the effectiveness of spatial queries, such as point-in-polygon and polygon-in-polygon, in detecting incorrect object placement. The proposed algorithms help reduce the number of errors by automatically controlling the accuracy of data entry. Implementing automated verification mechanisms will enhance the quality of cadastral data, which is crucial for urban planning and land resource management.

Keywords: geospatial databases, anomalies, State Land Cadastre, automated verification, spatial analysis, error detection algorithms.

Мороз Борис Іванович – доктор технічних наук, професор кафедри програмного забезпечення комп'ютерних систем Національний технічний університет «Дніпровська політехніка».

Шишацький Олександр Олександрович – аспірант кафедри програмного забезпечення комп'ютерних систем Національний технічний університет «Дніпровська політехніка».

Moroz Borys – Doctor of Technical Sciences, Professor at the PZKS, National Technical University "Dnipro Polytechnic".

Shyshatskyi Oleksandr – PhD student at the PZKS, National Technical University "Dnipro Polytechnic".