



УДК 004.75/.62

М. В. Арzubов, А. Є. Батюк

Національний університет «Львівська політехніка», м. Львів, Україна

АНАЛІЗ РІШЕНЬ ТА ПІДХОДІВ КЛАСТЕРИЗАЦІЇ ГЕОПРОСТОРОВИХ ДАНИХ ДЛЯ ОПТИМІЗАЦІЇ ПРОДУКТИВНОСТІ ВЕБКАРТИ ТА ВЗАЄМОДІЇ КОРИСТУВАЧА

У сучасну епоху зростає значення управління та візуалізації геопросторової інформації у веббраузерах. Вебкарти стали незамінними інструментами в різних областях, таких як туризм, доставка товарів чи екологія. Важлива також широка підтримка веббраузерів на різних пристроях, що робить використання геоданих у веб доступнішим для різних користувачів. Але постійне збільшення геопросторової інформації створює нові виклики в ефективному відображенні даних та навігації по цих даних на вебкартах. Тому в роботі з геоданими важлива їх кластеризація. Різні методи кластеризації по-різному впливають на продуктивність чи візуальну зрозумілість вебкарт.

У роботі виконано ґрунтовний огляд типів даних та методів кластеризації. Проаналізовано інструменти та бібліотеки, які спеціалізуються на кластеризації геоданих у вебкартах. Також досліджено різні типи геоданих і підходи до роботи з ними. Описано таке поняття, як напівстатичні дані, і яке місце вони займають разом зі статичними і динамічними типами даних.

Під час аналізу виявлено, у яких випадках краще використовувати певні методи кластеризації або коли варто застосовувати підхід із кластеризацією на серверній стороні. Також зроблено висновок про те, який підхід варто вибрати для роботи з великим обсягом статичних чи напівстатичних геоданих, а саме використання кластеризації на серверній стороні з кешуванням.

У підсумку проаналізовано різні підходи кластеризації у вебкартах як на клієнтській стороні, так і на серверній. Також описано переваги і недоліки обох підходів і коли краще той чи інший метод використовувати. Виявлено відсутність чітких підходів у кластеризації великих геоданих для зображення на вебкартах, що зумовлює актуальність досліджень у цьому напрямі.

Ключові слова: геодані; геопросторові, великі геодані; статичні геодані; напівстатичні геодані; динамічні геодані; кластеризація; вебкарти.

Вступ / Introduction

Зростання обсягу географічних даних та їх використання у різних вебдодатках потребує їх ефективного відображення та навігації по цих даних на вебкартах.

Користувачі наявних систем стикаються з трьома основними проблемами: візуалізація великого обсягу даних; повільність передавання даних через інтернет; відсутність підтримки динамічних запитів [19].

Тому одним із важливих аспектів у роботі з геоданими є їх кластеризація. Кластеризація дає можливість групувати об'єкти за вказаним критерієм схожості, спрощуючи сприйняття та збереження загального відображення. У вебдодатках, де ми маємо справу з великою кількістю маркерів на вебкарті, кластеризація стає необхідним інструментом, забезпечуючи ефективне використання обчислювальних ресурсів і плавну анімацію під час переміщення та масштабування карти, швидко інтерактивну взаємодію з користувачем.

У статті здійснено огляд відомих рішень та методів кластеризації геоданих у вебкартах. Проаналізовано різні підходи, їх переваги та обмеження, а також порівняно популярні бібліотеки, які надають функціональність кластеризації геоданих.

Об'єкт дослідження – процеси кластеризації для відображення геоданих на вебкартах.

Предмет дослідження – методи, моделі та засоби кластеризації геоданих на вебкарті.

Мета роботи – підвищення ефективності використання наявних рішень кластеризації геоданих у вебкартах, та надання інформації для прийняття обґрунтованих рішень щодо вибору відповідного підходу кластеризації, враховуючи конкретні потреби та вимоги проекту.

Для досягнення вказаної мети визначено такі основні завдання дослідження:

- аналіз актуальних досліджень та публікацій у галузі геопросторових даних, вебкарт та кластеризації геоданих;
- дослідження впливу методів кластеризації на продуктивність вебкарт. Важливо визначити, які методи найефективніші для досягнення поставленої мети;
- дослідження типів геоданих та підходів кластеризації відповідного типу даних;
- вибір оптимальних підходів та методів для вебкарт із великими статичними та напівстатичними геоданими. Це передбачає розгляд

різних підходів до кластеризації геопросторових даних та вибір найефективніших для поставленого завдання.

У наступних розділах статті розглянемо основні підходи до кластеризації геоданих у вебкартах, зосереджуючись на K-means, density-based spatial clustering of applications with noise (DBSCAN) та ієрархічній кластеризації. Для кожного підходу надано описання методу, його переваги та обмеження, а також приклади відповідних бібліотек та інструментів, які можна використовувати для здійснення кластеризації у вебдодатках.

Матеріали і методи дослідження. У роботі використано аналітичні методи та інструменти для проведення дослідження та аналізу кластеризації геопросторових даних у вебкартах. Окрім цього, проаналізуємо різні типи даних і особливості підходів кластеризації відповідних типів.

Для кластеризації напівстатичних геоданих запропоновано застосувати метод серверної кластеризації. Цей підхід дає змогу підготовлювати кластери геоданих на сервері й не витрачати на це час на стороні браузера.

Аналіз останніх досліджень та публікацій. З метою покращення користувацького досвіду та оптимізації використання обчислювальних ресурсів кластеризація геоданих стає необхідним інструментом для опрацювання великих обсягів маркерів на карті. Актуальність вирішення проблеми полягає у пошуку ефективних рішень кластеризації, які дають можливість забезпечити зручну інтерактивність та швидко опрацювання геоданих на вебкартах.

Огляд наявних рішень кластеризації геоданих на вебкартах є важливим кроком для розуміння і вибору оптимальних методів опрацювання та візуалізації геоданих. Це дасть змогу розробникам та дослідникам зосередитись на важливих аспектах, таких як швидкість, точність, гнучкість та масштабованість під час розроблення вебдодатків, пов'язаних з геоданими.

Дослідження різних підходів та інструментів кластеризації геоданих на веб картах також має практичне значення для розробників, оскільки вони можуть вибрати оптимальний метод для своїх проєктів. Враховуючи різні вимоги та обмеження, можна забезпечити оптимальну продуктивність та якість візуалізації геоданих на вебкартах.

Популярні алгоритми для кластеризації геоданих на вебкартах. Grid-based Clustering – метод кластеризації по сітці, один із найпоширеніших підходів до кластеризації геоданих на вебкартах. Він ґрунтується на розподі-

лі простору карти на регулярну сітку, де кожна клітинка представляє потенційний кластер (рис. 1). Геооб'єкти, які містяться в межах однієї клітинки, приєднуються до одного кластера. Цей метод ефективний для великих обсягів геоданих, оскільки дає можливість швидко згрупувати близькі геооб'єкти [2].

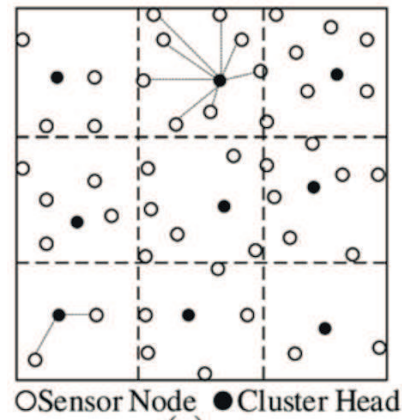


Рис. 1. Метод кластеризації по сітці / Grid-based clustering

Distance-based Clustering – це метод кластеризації, який ґрунтується на визначенні відстані між об'єктами та групуванні об'єктів, розташованих на обчисленій відстані один від одного. Найпоширенішим алгоритмом є алгоритм k-means, який спробує розділити дані на кластери, мінімізуючи відстані між об'єктами в кожному кластері та максимізуючи відстані між кластерами (рис. 2) [6], [9], [13], [15].

Hierarchical Clustering (Ієрархічна кластеризація) використовується для створення деревоподібної структури кластерів, де кожен вузол представляє окремий кластер або групу об'єктів (рис. 3) [6], [12], [16], [18]. Є два підходи до ієрархічної кластеризації: агломеративі (об'єднувальні) й розділювальні.

В агломеративному підході кожен об'єкт спочатку розглядається як окремий кластер, а потім послідовно об'єднується з найближчими кластерами, що приводить до створення більших кластерів. Цей процес продовжується доти, доки не буде сформована одна велика кластерна структура.

У розділювальному підході весь набір даних розглядається як один кластер, а потім поділяється на менші підкластери за допомогою рекурсивного розділення. Процес продовжується до досягнення необхідного рівня деталізації або до заданої кількості кластерів.

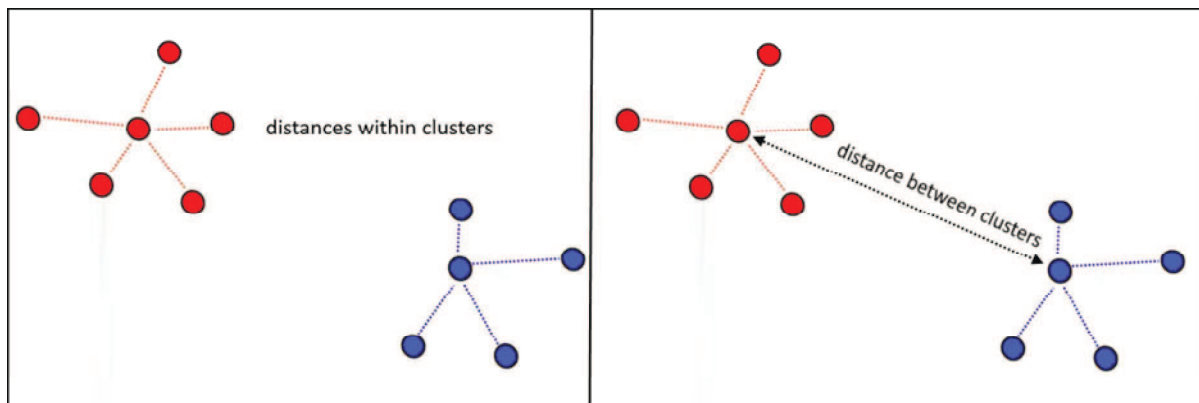


Рис. 2. Метод K-means / K-means method

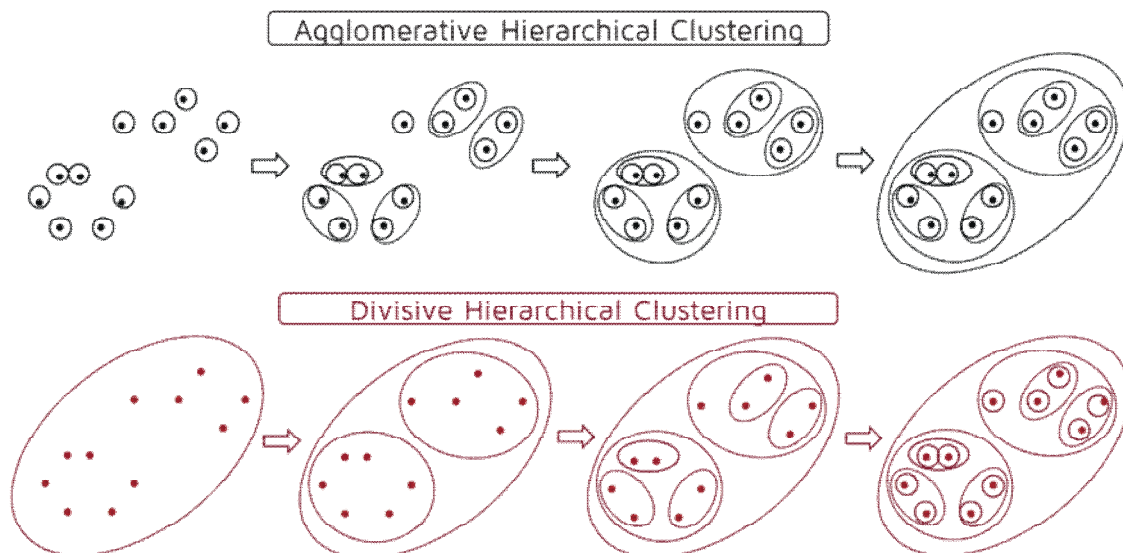


Рис. 3. Метод ієрархічного кластерингу / Hierarchical clustering method

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) – метод кластеризації, який ґрунтується на щільності точок у просторі. Визначає кластери як області з високою щільністю точок, розділені областями із низькою щільністю (рис. 4). DBSCAN виявляє кластери різної форми та розміру і має здатність виявляти шум [20].

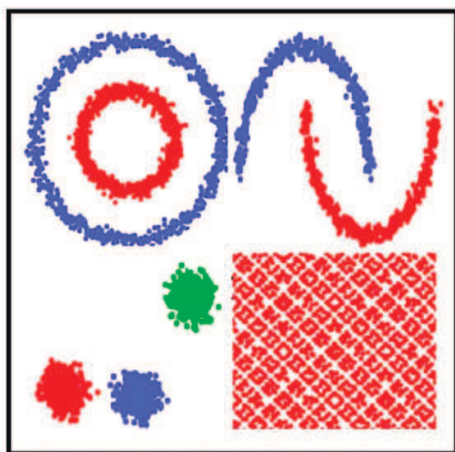


Рис. 4. Метод DBSCAN / DBSCAN method



Рис. 5. Метод OPTICS / OPTICS method

OPTICS (Ordering Points To Identify the Clustering Structure) – інший метод кластеризації, оснований на

щільності точок, але відрізняється від DBSCAN тим, що створює упорядковану структуру точок (рис. 5). OPTICS надає інформацію про градієнти щільності, щоб виявити як густі, так і розріджені кластери, а також границі між кластерами [3].

Кожен із цих алгоритмів має певні особливості та переваги в різних сценаріях застосування.

Метод кластеризації по сітці є простим і ресурсоефективним для опрацювання великого обсягу геоданих. Він надає можливість швидко розділити дані на кластери, використовуючи простий алгоритм розподілу простору на клітинки.

Метод кластеризації за відстанню, зокрема алгоритм k-means, – популярний і гнучкий підхід. Він надає можливість групувати дані на основі географічних відстаней між ними. Цей метод добре підходить для випадків кластерів сферичної форми й однакового розміру.

Ієрархічна кластеризація дає можливість створювати структуру кластерів з різними рівнями деталізації. Вона може виявляти як великі, так і малі кластери, а також границі між ними. Цей підхід корисний, коли важлива ієрархічна організація даних.

Методи DBSCAN і OPTICS використовують щільність точок для виявлення кластерів. Вони можуть працювати з кластерами різної форми і розміру, а також виявляти шумові точки. Ці алгоритми особливо корисні для аналітики даних.

Однак варто зазначити, що немає універсального рішення для кластеризації геоданих у вебкартах, оскільки кожен проект може мати унікальні вимоги та обмеження. Вибираючи алгоритм кластеризації, необхідно враховувати такі фактори, як розмір даних, швидкість, форма та розмір кластерів, чутливість до шуму та інші вимоги, зусовнені конкретним завданням.

Бібліотеки для фронтенд роботи з кластеризацією у вебкартах. Існує низка потужних бібліотек та інструментів для сучасних вебзробок, які спеціалізуються на кластеризації геоданих на вебкартах [11]. Вони надають розширені функціональні можливості та зручний інтерфейс для роботи з кластерами на фронтенді. Нижче наведено кілька прикладів таких бібліотек:

Leaflet.markercluster – одна із найпопулярніших бібліотек для кластеризації маркерів на вебкартах. В цій бібліотеці для побудови кластерів імплементовано метод кластеризації по сітці та за відстанню. Вона підтримує Leaflet – потужну бібліотеку для роботи з інтерактивними картами. Leaflet.markercluster надає широкий набір налаштувань для кластеризації, урахування кластерні іконки, зміну масштабу кластерів, опрацювання подій та багато іншого. На жаль, бібліотека не підтримує відображення більше ніж 100 тис. маркерів [17]. Ця бібліотека швидка і легка у використанні, але якщо у вас Big Data, то це цей варіант може не справитись з поставленими завданнями.

OpenLayers – бібліотека з відкритим вихідним кодом для відображення інтерактивних карт. Вона не має конкретної прив'язки до конкретного провайдера карти, так само як і у Leaflet, тому можна використовувати карту з будь-якого джерела, такого як OSM, Bing, Mapbox тощо. OpenLayers має вбудований метод кластеризації, який забезпечує достатньо високий рівень продуктивності. Із використанням 250 тис. точок усереднений час відображення становив ~8 с, але вже за 500 тис. зріс утричі й досяг ~24 с [17]. На жаль, бібліотека не підтримує відображення більше ніж 500 тис. маркерів [17]. Це також потрібно враховувати, вибираючи бібліотеку відповідно до поставлених завдань.

Supercluster – високопродуктивна бібліотека для кластеризації геоданих на основі піксельних географічних індексів. Вона спеціально розроблена для роботи з великими наборами даних та підтримує оновлення у реальному часі. Supercluster використовує оптимізований алгоритм ієрархічної жадної кластеризації, який дає можливість швидко обробляти та відображати тисячі маркерів на вебкарті. Бібліотека підтримує різні налаштування кластеризації, зокрема радіус кластерів, максимальну кількість об'єктів у кластері, стилізацію та фільтрацію даних.

MarkerClustererPlus – бібліотека для кластеризації маркерів у Google Maps JavaScript API. У ній для побудови кластерів можна вибрати один із двох методів – по сітці або ієрархічний. Але, насправді, ієрархічний метод – та сама бібліотека Supercluster, тому продуктивність буде фактично такою самою. MarkerClustererPlus пропонує простий інтерфейс для кластеризації маркерів для Google Maps, урахування налаштування іконок, стилізацію та взаємодію з кластерами. Оскільки вона використовує бібліотеку Supercluster, то автоматично підтримує ефективне керування великими наборами маркерів і надає можливість динамічного оновлення кластерів у разі зміни даних. Бонусом у цьому рішенні є те, що можна передати свій метод кластеризації, що робить цю бібліотеку дуже гнучкою і розробник не залежить від конкретного вбудованого методу, як у інших.

Mapbox GL JS – це бібліотека для інтерактивних карт, яка має вбудовані методи кластеризації. Але, як і MarkerClustererPlus, вона також використовує Supercluster для побудови кластерів. Тому продуктивність на рівні, аналогічному з конкурентною бібліотекою для GoogleMaps, за винятком того, що Mapbox GL JS використовує технологію WebGL для відображення вмісту карти, яка своєю чергою використовує графічний

процесор. З першого погляду очікується, що продуктивність буде на рівні Supercluster з Leaflet, але у разі відображення 100 тис. точок бібліотека для Mapbox була повільнішою, проте зі збільшенням кількості даних стає помітнішим використання WebGL [17]. Під час тестування з 500 тис. точок Mapbox GL JS була швидшою на ~4 с, а відображаючи 3 млн точок, взагалі удвічі швидшою, ніж Supercluster з Leaflet [17].

PruneCluster – проєкт від норвезької організації SINTEF (Stiftelsen for industriell og teknisk forskning), який дає можливість кластеризувати геодані у вебплікаціях, що використовують бібліотеку Leaflet. Фактично це є плагіном до неї, як альтернатива до MarkerCluster. У цій бібліотеці метод кластеризації ґрунтується на алгоритмі Sweep and prune (знайти і відрізати), який використовують у завданнях пошуку зіткнень між об'єктами. Плагін може працювати з великою кількістю даних і достатньо швидко. Наприклад, PruneCluster зміг завантажити близько 500 тис. точок і відобразити кластери за ~23 с [17]. Хоча й проєкт не оновлюється активно з 2017 р., люди його продовжують використовувати, тому можна сказати, що ця бібліотека є хорошою альтернативою для MarkerCluster.

Ці бібліотеки є лише деякими з численних інструментів, доступних для роботи з кластеризацією геоданих на вебкартах. Вибір конкретної бібліотеки залежить від ваших потреб та вимог проєкту, а також від вашого досвіду та вподобань. Перед вибором потрібно оцінити її функціональні можливості, продуктивність, легкість використання та підтримку спільнотою розробників. Крім того, не треба забувати про документацію та приклади використання, які допоможуть ефективно використати вибрану бібліотеку у проєкті.

Бібліотеки для бекенд-роботи з кластеризацією на вебкартах. Щоб виконувати кластеризацію геоданих на вебкартах, важливо також мати потужні інструменти на бекенді для опрацювання та кластеризації великих обсягів даних. Нижче наведено декілька прикладів бібліотек, які допомагають управляти цими завданнями:

GeoSpark – це бібліотека геопросторової аналітики та опрацювання даних, спеціально розроблена для роботи з великими масивами геоданих [21]. Вона надає потужні методи кластеризації, такі як Spatial K-Means, DBSCAN та інші. GeoSpark підтримує широкий спектр операцій з геоданими та надає можливість ефективно виконувати запити до геоданих у розподіленому середовищі. Ця бібліотека особливо корисна для оброблення великих обсягів геоданих у реальному часі.

Scikit-learn – одна з найпопулярніших бібліотек машинного навчання для Python, її також можна використати для кластеризації геоданих [10]. Перевагою Scikit-learn є простота використання, широкий вибір методів та інструменти для оцінювання результатів роботи. Ця бібліотека містить реалізації популярних методів кластеризації, таких як K-means, DBSCAN тощо.

PostGIS – розширення географічного об'єкта для бази даних PostgreSQL, яке надає можливість зберігати та обробляти геодані [1]. Підтримує різні географічні операції, урахування кластеризацію геоданих. За допомогою PostGIS можна виконувати запити до бази даних, що містить геодані, й використовувати різні алгоритми кластеризації, такі як K-means або DBSCAN,

для створення кластерів на основі географічних характеристик. Після кластеризації можна використовувати результати для візуалізації на вебкарті або для подальшого аналізу.

MongoDB – одна з найпопулярніших NoSQL баз даних, яку можна успішно використати для збереження та оброблення геоданих, урахувавши їх кластеризацію. MongoDB має вбудовану підтримку геопросторових запитів і операцій, що дає можливість виконувати різноманітні географічні запити безпосередньо на рівні бази даних, наприклад, можна використовувати різні алгоритми кластеризації, такі як K-means, DBSCAN тощо [1].

Elasticsearch – розподілена система пошуку та аналізу даних, яка також може бути використана для кластеризації геоданих. Elasticsearch має підтримку географічних типів даних та операцій з ними, що надає можливість виконувати кластеризацію на основі географічних ознак. Система забезпечує розширені можливості для пошуку та аналізу геоданих, а також може бути інтегрована з вебдодатками для відображення кластерів на вебкартах [8].

Ці бібліотеки для бекенд-роботи з кластеризацією геоданих на вебкартах надають потужні інструменти для опрацювання та аналізу великих обсягів геоданих. Вони дають змогу ефективно виконувати кластеризацію, що є важливим етапом у візуалізації та аналізі геоданих на вебкартах. Під час вибору бібліотеки слід враховувати її сумісність з вибраною базою даних та іншими компонентами системи, а також потужність та швидкодію для опрацювання великих обсягів даних. Використовуючи бекенд-бібліотеки, потрібно враховувати затримку в передаванні даних до клієнта.

Результати дослідження та їх обговорення / Research results and their discussion

Враховуючи широкий спектр застосувань геоданих, від геолокації та туризму до геоаналітики та маркетингу, важливо мати доступ до ефективних рішень кластеризації, які забезпечують швидке опрацювання та відображення геоданих на вебкартах так, щоб забезпечити користувачам зручний та інтуїтивний доступ до інформації. Правильний вибір методу кластеризації позитивно впливає на продуктивність, візуальну зрозумілість та корисність вебдодатків, що використовують геодані.

Огляд виконаних досліджень виявив, що у контексті кластеризації геоданих на вебкартах важливо розрізняти статичні та динамічні дані. Це впливає на вибір підходів та алгоритмів кластеризації, а також на принципи побудови та оновлення кластерів у реальному часі.

Статичні дані відображають інформацію про об'єкти на карті, яка змінюється рідко (не частіше ніж один раз на місяць чи квартал) або не змінюється взагалі [14]. Це можуть бути, наприклад, географічні мітки для певних місць, маркери підприємств або точок інтересу, кордони адміністративних одиниць та інші статичні об'єкти (рис. 6).

Під час кластеризації статичних даних основну увагу зосереджують на побудові оптимальних кластерних структур, що відображають просторові згуртованості об'єктів [7]. Застосування методів, таких як K-means, DBSCAN, може бути ефективним для розділення статичних даних на кластери, забезпечуючи логічну та зручну візуалізацію на вебкарті.

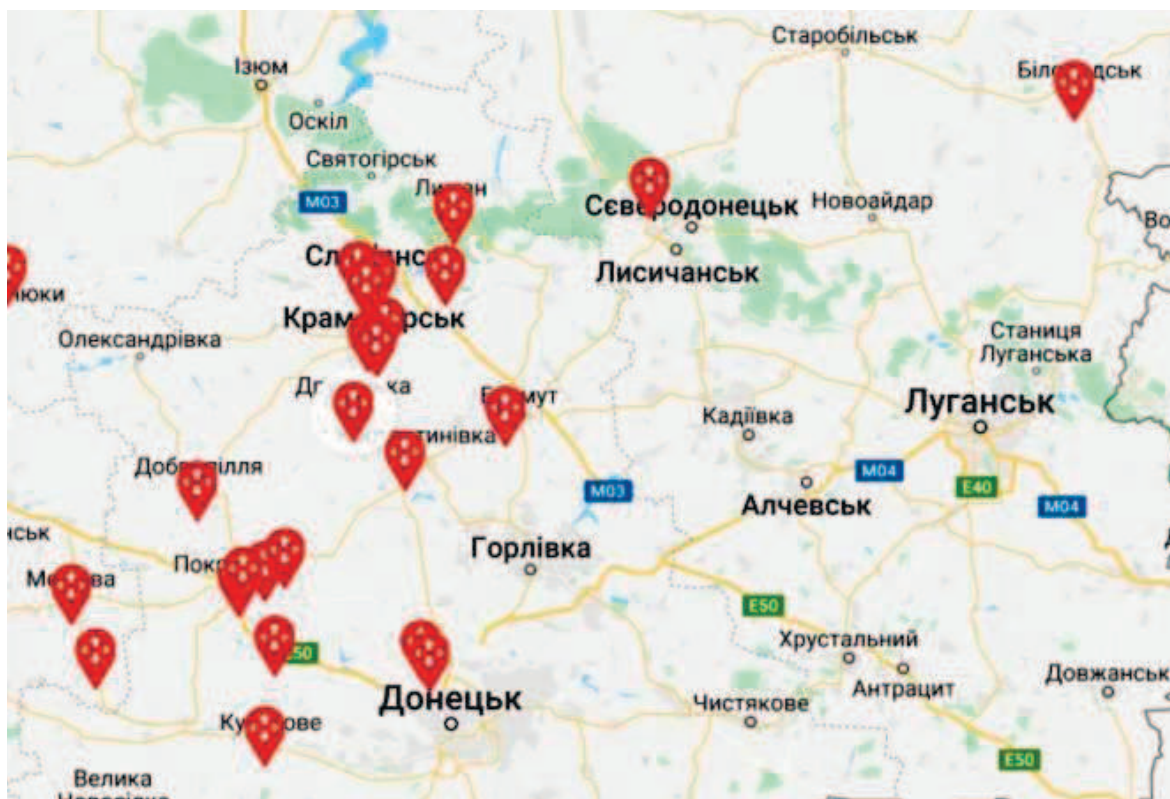


Рис. 6. Приклад відображення статичних даних на карті / Example of rendering static data on the map

Динамічні дані відображають інформацію, яка змінюється у реальному часі. Наприклад, це можуть бути рухомі об'єкти, такі як транспортні засоби (рис. 7), люди або метеорологічні дані [4], [5]. В таких випадках кластеризація динамічних даних потребує постійного оновлення кластерів та адаптації до змін у розташуванні або стані об'єктів. Важливо враховувати часовий аспект та динаміку даних під час побудови кластерів.

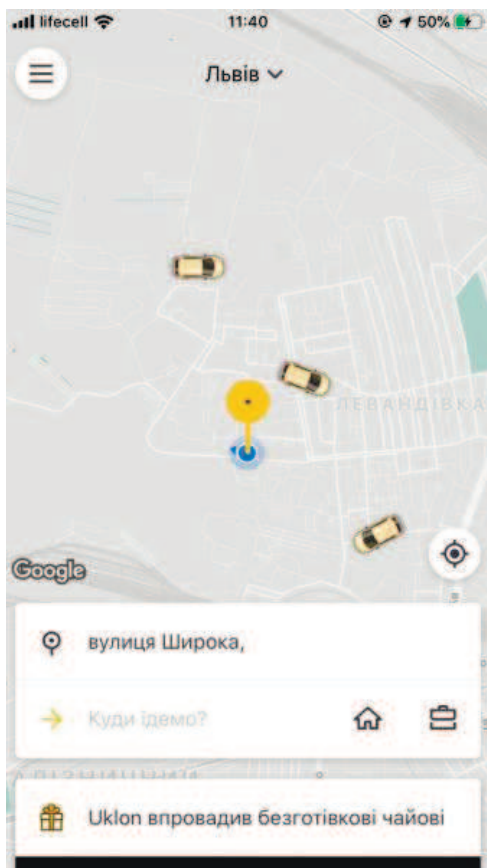


Рис. 7. Приклад відображення динамічних даних на карті / Example of rendering dynamic data on the map

Для кластеризації динамічних даних на вебкартах можна використовувати спеціальні алгоритми, що враховують змінність даних у часі. Наприклад, алгоритмічний підхід може використовувати такий метод, як експоненціальне згладжування, для оновлення кластерів та врахування останніх змін у даних.

Іншим підходом може бути використання технологій потокового опрацювання даних, таких як Apache Kafka або Apache Flink. Ці системи надають можливість обробляти великі потоки даних у реальному часі й розподіляти їх на кластери, враховуючи зміни у положенні об'єктів.

Крім того, для ефективної кластеризації динамічних даних можна застосовувати техніки прогнозування та передбачення майбутніх змін, наприклад, методи прогнозування руху об'єктів на основі їх попереднього руху та шаблонів поведінки.

Вибираючи підходи для кластеризації динамічних даних на вебкартах, необхідно враховувати швидкодію та масштабованість алгоритмів, а також потреби в реалізації вебінтерфейсу, що забезпечує візуалізацію та взаємодію із динамічними кластерами на вебкарті.

Напівстатичні геодані містяться між статичними та динамічними типами даних за частотою їх оновлення (від однієї години і до місяця часу). Інформація може залишатися постійною протягом днів, тижнів або навіть місяця, перш ніж оновиться. Наприклад, набір даних щодо доступних паркувальних місць у місті (рис. 8) може бути здебільшого статичним щодня. Однак іноді його оновлюють кожні кілька тижнів або місяців, щоб відобразити нове будівництво або зміни в правилах паркування.

Обговорення результатів дослідження. Отримані результати підтверджують важливу роль у візуалізації та аналізі географічних даних кластеризації геоданих на вебкартах. Врахування відмінностей між статичними, напівстатичними та динамічними даними є ключовим аспектом у виборі підходів та алгоритмів кластеризації. За умови належного розуміння та використання різноманітних методів та технологій можна досягти оптимальних результатів у побудові кластерів та відображенні геоданих на вебкартах.

Незважаючи на істотний прогрес у розробленні рішень для кластеризації геоданих на вебкартах, розробники та користувачі можуть зіткнутися із деякими проблемами, які перераховано нижче:

1. Масштабованість. Однією із найважливіших проблем є масштабованість алгоритмів кластеризації для опрацювання великих обсягів геоданих. Під час роботи з великими наборами даних можуть виникати проблеми з продуктивністю та часом відповіді. Здатність деяких алгоритмів ефективно працювати з великими обсягами даних обмежена, тому важливо вибрати рішення, яке можна масштабувати під ваші потреби.
2. Динамічне оновлення кластерів. Ще однією проблемою є динамічна побудова кластерів для геоданих, що постійно змінюються у часі. В таких випадках важливо мати ефективні механізми оновлення кластерів після змін даних. Кластеризація динамічних даних іноді потребує додаткових обчислювальних ресурсів та оптимізованих алгоритмів, щоб забезпечити швидке та точне оновлення кластерів.
3. Різноманітність даних. Здатність деяких рішень для кластеризації геоданих працювати з різноманітними типами даних обмежена. Наприклад, якщо дані містять як числові, так і текстові атрибути, необхідно мати змогу враховувати цю різноманітність під час кластеризації. Деякі алгоритми можуть працювати лише з числовими атрибутами, тоді як інші підтримуватимуть текстові атрибути або об'єднуюватимуть їх у спільну модель. Вибір рішення залежить від потреб в конкретному завданні та типів даних, які необхідно кластеризувати.
4. Відсутність контексту. Іноді кластеризація геоданих у вебкартах обмежена відсутністю контексту або додаткової інформації про дані. Наприклад, під час відображення кластерів на карті можуть виникати ситуації, коли користувачі не мають достатньої інформації про кластери, подані на карті. Додавання додаткового контексту та інформації може зробити кластеризацію зрозумілішою та цікавішою для користувачів.

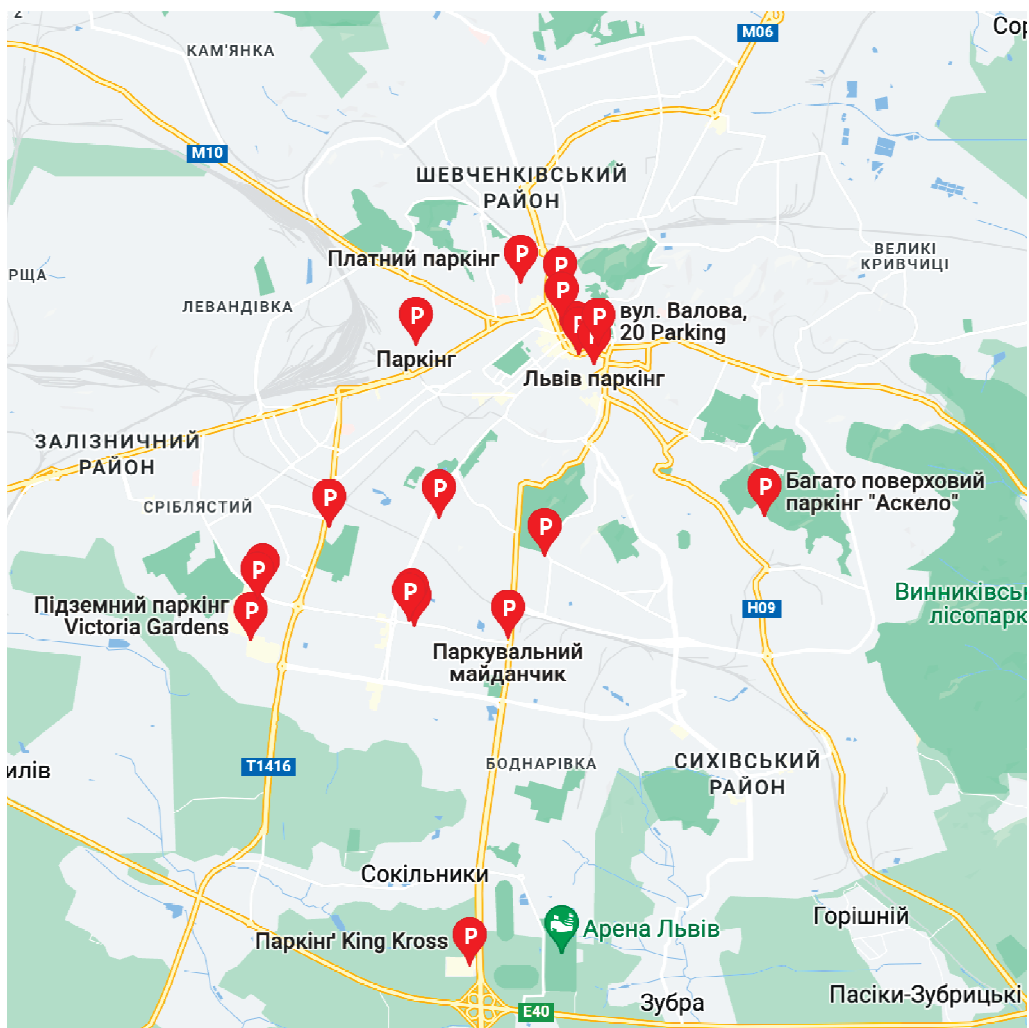


Рис. 8. Приклад відображення напівстатичних даних на карті / Example of rendering semistatic data on the map

Один з викликів полягає у виборі оптимального методу кластеризації для конкретного завдання. Існує широкий спектр методів, кожен з яких має певні переваги та обмеження. Вибір відповідного методу залежить від розмірів даних, їх характеристик, масштабованості, потреб у реалізації та інших факторів.

Проблеми стосовно кластеризації геоданих на вебкартах спонукають до додаткових досліджень та розробок для їх вирішення. Розуміння цих проблем допоможе розробникам та дослідникам вдосконалити відомі рішення та створити нові методи та підходи для ефективної кластеризації геоданих на вебкартах. Це своєю чергою сприятиме оптимізації вебдодатків, що використовують геодані, й забезпечить користувачам зручніші та ефективніші інтерактивні можливості на мапах.

Для кожного типу геоданих можна краще визначити відповідні підходи оброблення даних та стратегії роботи з ними, розуміючи діапазон часу для оновлень. Наприклад, для напівстатичних даних можна використати гібридний підхід, коли на стороні сервера кластеризовані дані періодично оновлюються відносно змін у вихідних даних.

Подальші дослідження у цій галузі можуть сприяти створенню нових алгоритмів та методів і покращенню наявних бібліотек та інструментів. Застосування кращих практик у кластеризації геоданих допоможе

вирішити проблеми, пов'язані з продуктивністю, масштабованістю та ефективністю роботи із великими обсягами геоданих.

Підсумуємо: кластеризація геоданих на вебкартах – важливе завдання, що відкриває широкі можливості для візуалізації та аналізу географічної інформації. Розуміння проблем та рішень у цій галузі допоможе розробникам та дослідникам поліпшити та розширити можливості технології, забезпечуючи користувачам ефективніші та потужніші вебдодатки на основі геоданих.

Отже, за результатами досліджень можна сформулювати наукову новизну та практичну цінність роботи.

Наукова новизна отриманих результатів дослідження – розвинено модель кластеризації напівстатичних геопросторових даних на основі поєднання методів і моделей кластеризації “статичних” та “динамічних” даних.

Практична значущість результатів дослідження – результати дослідження, викладені у статті, відображають удосконалення класифікації типів геоданих із додаванням напівстатичного типу. Практичне застосування цих результатів дослідження полягає у впровадженні серверної кластеризації для великих обсягів напівстатичних геоданих, що сприяє підвищенню продуктивності вебкарт.

Висновки / Conclusions

У результаті аналізу розвинуто модель кластеризації напівстатичних геоданих. Це розширює можливості застосування методів кластеризації, забезпечуючи точніше використання цих методів для напівстатичних даних. Використання серверної кластеризації може істотно підвищити швидкість оброблення даних та відображення інформації на картографічних сервісах, що має велике значення для оперативного реагування на зміни в геоданих і для надання актуальної інформації користувачам вебкарт. В результаті такого підходу підвищується точність і поліпшується якість аналітичних моделей, основаних на геоданих.

Отримані результати дослідження засвідчують важливість кластеризації геоданих у контексті вебкарт, яка має значення для різноманітних застосувань, урахуваючи геолокацію, туризм, геоаналітику та маркетинг. Проблеми із масштабованістю, динамічним оновленням кластерів, різноманітністю даних та відсутністю контексту – серед викликів, які потребують додаткових досліджень і розроблення.

У майбутніх дослідженнях необхідно звернути увагу на розроблення нових, вдосконалення наявних та подальший розвиток моделей, методів кластеризації, враховуючи особливості геоданих та вимоги вебдодатків. Запропоновано використовувати підхід із кластеризацією на серверній стороні для подолання певних обмежень у випадку великих статичних чи напівстатичних геоданих. Але, працюючи із напівстатичним типом даних, потрібно результати періодично оновлювати залежно від того, як часто змінюються вихідні дані та які вимоги у системі. Важливо продовжувати працювати над підвищенням продуктивності, масштабованості та візуалізації для кластеризації геоданих на вебкартах.

References

- [1] Agarwal, S., & Rajan, K. S. (2016). Performance analysis of MongoDB versus PostGIS/PostgreSQL databases for line intersection and point containment spatial queries. *Spat. Inf. Res.*, 24, 671–677. <https://doi.org/10.1007/s41324-016-0059-1>
- [2] Amini, A., Wah, T. Y., Saybani, M. R., & Yazdi, S. R. A. S. (2011, July). A study of density-grid based clustering algorithms on data streams. In 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), vol. 3, 1652–1656. IEEE. <https://doi.org/10.1109/FSKD.2011.6019867>
- [3] Ankerst, M., Breunig, M. M., Kriegel, H. P., & Sander, J. (1999). OPTICS: Ordering points to identify the clustering structure. *ACM Sigmod record*, 28(2), 49–60. <https://doi.org/10.1145/304181.304187>
- [4] Cekule, M., Mitrofanovs, I., & Cabs, K. (2023). Information technology for real-time monitoring and visualization of load in urban public open spaces based on spatial and statistical data analyses of human behaviour. *International Multidisciplinary Scientific GeoConference: SGEM*, 23(2.1), 89–96. <https://doi.org/10.5593/sgem2023/2.1/s07.12>
- [5] Choi, S., & Bae, B. (2015). The Real-Time Monitoring System of Social Big Data for Disaster Management. In: Park, J., Stojmenovic, I., Jeong, H., Yi, G. (eds) *Computer Science and its Applications. Lecture Notes in Electrical Engineering, Vol. 330*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-45402-2_115
- [6] Doroshenko A. (2020). Analysis of the Distribution of COVID-19 in Italy Using Clustering Algorithms, 2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 325–328. <https://doi.org/10.1109/DSMP47368.2020.9204202>
- [7] ElHaj, K., Alshamsi, D. & Aldahan, A. (2023). GeoZ: a Region-Based Visualization of Clustering Algorithms. *J. geovis spat. anal.*, 7, 15. <https://doi.org/10.1007/s41651-023-00146-0>
- [8] Guo, D., & Onstein, E. (2020). State-of-the-Art Geospatial Information Processing in NoSQL Databases. *ISPRS Int. J. Geo-Inf.*, 9, 331. <https://doi.org/10.3390/ijgi9050331>
- [9] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100–108. <https://doi.org/10.2307/2346830>
- [10] Kramer, O. (2016). Scikit-Learn. In: *Machine Learning for Evolution Strategies. Studies in Big Data*, Vol. 20. Springer, Cham. https://doi.org/10.1007/978-3-319-33383-0_5
- [11] Kulawiak, M., Dawidowicz, A., & Pacholczyk, M. E. (2019). Analysis of server-side and client-side Web-GIS data processing methods on the example of JTS and JSTS using open data from OSM and geoportal. *Computers & Geosciences*, 129, 26–37. <https://doi.org/10.1016/j.cageo.2019.04.011>
- [12] Laasasenaho, K., Lensu, A., Lauhanen, R., & Rintala, J. (2019). GIS-data related route optimization, hierarchical clustering, location optimization, and kernel density methods are useful for promoting distributed bioenergy plant planning in rural areas. *Sustainable Energy Technologies and Assessments*, 32, 47–57. <https://doi.org/10.1016/j.seta.2019.01.006>
- [13] Levus, Ye. V., & Vasyliuk, R. B. (2022). Recommendation algorithm using data clustering. *Ukrainian Journal of Information Technology*, 4(2), 18–24. <https://doi.org/10.23939/ujit2022.02.018>
- [14] Lytvyn, V., Uhryn, D., Ushenko, Y., Masikevych, A., & Bairachnyi, V. (2023). The Method of Clustering Geoinformation Data for Stationary Sectoral Geoinformation Systems Using Swarm Intelligence Methods. In: Cioboată, D. D. (eds) *International Conference on Reliable Systems Engineering (ICoRSE) – 2023. ICoRSE 2023. Lecture Notes in Networks and Systems*, Vol. 762. Springer, Cham. https://doi.org/10.1007/978-3-031-40628-7_44
- [15] Muenchow, J., Schäfer, S., & Krüger, E. (2019). Reviewing qualitative GIS research – Toward a wider usage of open source GIS and reproducible research practices. *Geography Compass*, 13(6), e12441. <https://doi.org/10.1111/gec3.12441>
- [16] Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1), 86–97. <https://doi.org/10.1002/widm.53>
- [17] Netek, R., Brus, J., & Tomecka, O. (2019). Performance Testing on Marker Clustering and Heatmap Visualization Techniques: A Comparative Study on JavaScript Mapping Libraries. *ISPRS Int. J. Geo-Inf.*, 8, 348. <https://doi.org/10.3390/ijgi8080348>
- [18] Praene, J. P., Malet-Damour, B., Radanielina, M. H., Fontaine, L., & Riviere, G. (2019). GIS-based approach to identify climatic zoning: A hierarchical clustering on principal component analysis. *Building and Environment*, 164, 106330. <https://doi.org/10.1016/j.buildenv.2019.106330>
- [19] Rezaei, M., & Franti, P. (2018). Real-time clustering of large Geo-referenced data for visualizing on map. *Adv. Electr. Comput. En.*, 18(4), 63–74, Nov. 2018. <https://doi.org/10.4316/AECE.2018.04008>
- [20] Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)*, 42(3), 1–21. <https://doi.org/10.1145/3068335>
- [21] Yu, J., Wu, J., Sarwat, M. (2015). GeoSpark: a cluster computing framework for processing large-scale spatial data. In: *Proceedings of the ACM SIGSPATIAL GIS*, USA. <https://doi.org/10.1145/2820783.2820860>

SOLUTIONS AND APPROACHES ANALYSIS FOR GEOSPATIAL DATA CLUSTERING TO OPTIMIZE PERFORMANCE AND USER EXPERIENCE OF WEB MAPS

In the contemporary epoch, the management and visualization of geospatial information in web browsers have gained substantial importance. Web maps are indispensable tools across various tourism, goods delivery, and ecology sectors. Furthermore, the extensive support of web browsers on diverse devices enhances the accessibility of geospatial data on the web for various users. However, the incessant increment of geospatial information poses new challenges in efficiently displaying data and navigation through these data on web maps. Therefore, the clustering of geospatial data is crucial in dealing with them. Different clustering methods may affect the performance or visual clarity of web maps.

To improve the user experience and optimize the use of computing resources, geodata clustering becomes a necessary tool for processing large volumes of markers on the map. Despite significant progress in the development of geodata clustering solutions in web maps, there are some challenges that developers and users may encounter. In this article, challenges with scaling, dynamic cluster data, and heterogeneity of data are described. Existing problems in geodata clustering in web maps require additional research and development. Understanding these issues will help developers and researchers improve existing solutions and create new methods and approaches for efficient clustering of geodata in web maps. The urgency of solving the problem lies in the search for effective clustering solutions that provide an opportunity to ensure convenient interactivity and fast processing of geodata in web maps.

This study provides a comprehensive review of data types and clustering methods. Tools and libraries for geodata clustering in web maps are analyzed. Different types of geodata and approaches to working with them were also studied. Concepts such as semi-static data and their positions alongside static and dynamic data types are elucidated.

Through the analysis, optimal scenarios for applying specific clustering methods or the utilization of server-side clustering approaches have been identified. Conclusions have also been drawn on the preferred approach when handling extensive volumes of static or semi-static geospatial data, particularly advocating for the application of server-side clustering with caching.

In conclusion, various clustering approaches in web maps, both client-side and server-side, have been scrutinized. The advantages and disadvantages of both approaches, along with recommendations on when to apply each method, have been delineated. A noticeable absence of explicit approaches in clustering vast geospatial data for web map representation underpins the relevance and necessity of research in this direction.

Keywords: geodata; geospatial, big geodata; static geodata; semi-static geodata; dynamic geodata; clustering; web map.

Інформація про авторів:

Арzubов Максим Валерійович, асистент, кафедра автоматизованих систем управління.

Email: m.arzubov@gmail.com; <https://orcid.org/0009-0008-9978-7072>

Батюк Анатолій Євгенович, канд. техн. наук, доцент, кафедра автоматизованих систем управління.

Email: abatyuk@gmail.com; <https://orcid.org/0000-0001-7650-7383>

Цитування за ДСТУ: Арzubов М. В., Батюк А. Є. Аналіз рішень та підходів кластеризації геопросторових даних для оптимізації продуктивності вебкарти та взаємодії користувача. *Український журнал інформаційних технологій*. 2023. Т. 5, № 2. С. 88–96.

Citation APA: Arzubov, M. V., & Batiuk, A. Y. (2023). Solutions and approaches analysis for geospatial data clustering to optimize performance and user experience of web maps. *Ukrainian Journal of Information Technology*, 5(2), 88–96.

<https://doi.org/10.23939/ujit2023.02.088>