

МЕТОДИ ВІЗУАЛІЗАЦІЇ BIG DATA ЗА ТИПАМИ ДАНИХ ТА АНАЛІТИЧНИМИ ЗАВДАННЯМИ

Гарах Ольга Анатоліївна

здобувач другого рівня вищої освіти спеціальності Комп'ютерні науки
Тернопільський національний педагогічний університет імені Володимира Гнатюка
garah_oa@fizmat.tnpu.edu.ua

Карабін Оксана Йосифівна

кандидат педагогічних наук, доцент кафедри інформатики та методики її навчання
Тернопільський національний педагогічний університет імені Володимира Гнатюка
karabin@tnpu.edu.ua

У сучасному цифровому світі обсяги даних зростають значними темпами, що формує нову епоху інформаційного суспільства, де ключову роль відіграє здатність ефективно збирати, зберігати, обробляти та аналізувати величезні масиви даних. Згідно з оцінками компанії Gartner (2022) [3], глобальний обсяг даних подвоюється приблизно кожні два роки, що підкреслює не лише швидкість генерації інформації, але й масштаби, з якими сучасні аналітичні системи повинні працювати. Обумовлений феномен зумовлює суттєві виклики у сфері інформаційних технологій, зокрема в питаннях забезпечення високої продуктивності обчислювальних платформ, ефективності алгоритмів обробки даних, надійності систем зберігання та інтерпретації результатів аналізу.

Швидке збільшення обсягів даних відбувається під впливом кількох факторів: стрімке зростання користувачів цифрових технологій, поширення сенсорних пристроїв та мобільних додатків, розвиток Інтернету речей, а також масове використання соціальних мереж та електронної комерції [1]. Окрім того, наукові дослідження, фінансові транзакції та цифрові медіа продукують великі потоки інформації, що вимагають інтеграції у єдині аналітичні системи. Такий обсяг і швидкість надходження даних перевищують можливості традиційних систем управління базами даних, що створює необхідність у впровадженні нових технологій та методів аналізу, включаючи розподілені обчислювальні платформи, алгоритми машинного навчання та інструменти візуалізації. Зазначені інновації не лише підвищує вимоги до продуктивності серверів та хмарних платформ, але й ставить нові завдання перед аналітиками, які повинні швидко виділяти ключові закономірності та взаємозв'язки в багатовимірних та гетерогенних масивах даних. У цьому контексті особливе значення набуває візуалізація даних, яка дозволяє трансформувати великі й складні набори інформації в доступні, наочні графічні моделі, що полегшують процес прийняття рішень та забезпечують високий рівень інтерпретованості результатів аналізу.

Зауважимо, що нині технології обробки даних спрямовані на подолання обмежень класичних систем управління базами даних і включають розподілені обчислювальні платформи, алгоритми машинного та глибинного навчання, а також інтегровані інструменти візуалізації [2]. Використання таких технологій підвищує продуктивність серверів та хмарних платформ, забезпечує масштабованість обчислень та підтримує аналіз багатовимірних і гетерогенних даних у режимі реального часу. Для аналітиків основним завданням стає своєчасне виділення закономірностей, виявлення взаємозв'язків і аномалій, що робить

візуалізацію даних критично важливою складовою процесу прийняття рішень. Зазначимо, що Big Data визначаються кількома ключовими параметрами:

– обсяг (volume): параметр, що визначає загальну кількість даних, що накопичуються в інформаційних системах. Збільшення обсягу інформації є результатом інтенсивного розвитку цифрових технологій, поширення сенсорних пристроїв, соціальних мереж, медіа-платформ і наукових досліджень. Великий обсяг даних створює додаткові вимоги до інфраструктури зберігання та обробки, змушуючи застосовувати розподілені системи та хмарні обчислення для забезпечення ефективної роботи з масивами інформації, які можуть досягати петабайтних і ексабайтних масштабів;

– швидкість (velocity): характеристика, що визначає темп надходження нових даних у систему. Швидке генерування інформації виникає у режимі реального часу або близькому до нього, наприклад, при обробці поточкових даних із сенсорів, фінансових ринків або соціальних платформ. Висока швидкість надходження даних вимагає впровадження ефективних алгоритмів потокової обробки, здатних забезпечувати миттєве оновлення аналітичних моделей та підтримку прийняття рішень у реальному часі;

– різноманіття (variety): параметр, що відображає багатоформатність і багатоструктурність даних. Воно включає структуровані дані (табличні формати), напівструктуровані (XML, JSON) та неструктуровані (текст, зображення, аудіо, відео). Різноманіття формату та структури даних створює виклики для інтеграції, стандартизації та аналізу інформації, оскільки кожен тип потребує спеціалізованих підходів до обробки та візуалізації для забезпечення повного розуміння закономірностей;

– достовірність (veracity): параметр, що оцінює точність, надійність та правдивість даних. Високий рівень достовірності необхідний для прийняття обґрунтованих рішень, особливо у критично важливих сферах, таких як медицина, фінанси та державне управління. Недостовірні або спотворені дані можуть призвести до хибних висновків, неправильних прогнозів та стратегічних помилок;

– цінність (value): характеристика, що визначає практичну користь даних для прийняття рішень та отримання інсайтів. Дані мають сенс лише у випадку, якщо вони можуть бути трансформовані в інформацію, що забезпечує підвищення ефективності бізнес-процесів, оптимізацію ресурсів або наукові відкриття. Цінність даних безпосередньо пов'язана з можливістю їх аналізу, візуалізації та інтеграції в процеси підтримки рішень.

Варто зазначити, що виділення цих характеристик свідчить про те, що ефективна робота з великими даними вимагає не лише високопродуктивних платформ для зберігання та обробки інформації, але й надійних методів аналізу та інтерпретації результатів, серед яких візуалізація відіграє визначальну роль у забезпеченні зрозумілості та практичної цінності аналітичних висновків. Дані, що опрацьовуються у великих аналітичних системах, відрізняються за структурою і форматом. Структуровані дані зберігаються у табличних форматах і легко піддаються обробці стандартними алгоритмами аналізу. Напівструктуровані дані, наприклад JSON або XML, мають часткову організацію і вимагають спеціалізованих методів для виявлення зв'язків та закономірностей. Неструктуровані дані, такі як текстові документи, зображення, аудіо- та

відеофайли, потребують складних методів обробки, включаючи алгоритми машинного навчання, методи зменшення розмірності та інтерактивну візуалізацію для адекватного представлення інформації. Кожен тип даних визначає специфіку підбору технік візуалізації, що повинні забезпечувати інтуїтивне розуміння інформації та можливість швидкого виділення ключових закономірностей.

Аналітичні завдання у Big Data охоплюють різні рівні складності. Описова аналітика дозволяє визначити, що відбулося у минулому, діагностична аналітика пояснює причини явищ, прогнозна аналітика оцінює потенційні сценарії розвитку подій, а прескриптивна аналітика пропонує оптимальні рішення для управління процесами та мінімізації ризиків. Кожен рівень аналітики потребує специфічних методів візуалізації, що підкреслюють головні аспекти даних, полегшують виявлення аномалій та тенденцій, а також сприяють розумінню взаємозв'язків між різними параметрами.

Класичні підходи до візуалізації, такі як статичні діаграми, гістограми або кругові графіки, в умовах великих даних часто стають неефективними через надмірність обсягу інформації та багатовимірність даних. Обробка мільярдів транзакцій, мільйонів зображень або поточкових даних робить традиційні графіки перевантаженими та непридатними для виявлення прихованих закономірностей. Різні формати даних потребують використання спеціалізованих технік візуалізації: структуровані дані відображаються через таблиці та графіки, напівструктуровані – через мережеві діаграми, Sankey або поточкові графи, а неструктуровані медіа-файли – через інтерактивні карти, спектрограми та алгоритмічне зменшення розмірності (PCA, t-SNE, UMAP) [4].

Ефективне застосування таких методів дозволяє підвищити точність і швидкість аналітичних процесів, забезпечує прозорість даних та робить можливим прийняття обґрунтованих рішень у масштабних системах. Візуалізація виступає не тільки інструментом представлення інформації, але й механізмом інтерпретації складних закономірностей у багатовимірних масивах даних, що робить її невід'ємною складовою сучасного аналізу великих даних.

Таким чином, сучасні підходи до візуалізації Big Data повинні інтегрувати як характеристики даних, так і специфіку аналітичних завдань, що стоять перед користувачем. Крім того, вони повинні забезпечувати можливість швидкого отримання інсайтів, інтерактивності, масштабованості та підтримки прийняття рішень у режимі реального часу. Також ефективна візуалізація сприяє кращому розумінню даних різними групами користувачів, як від фахівців з аналітики до керівників та науковців, забезпечуючи прозорість, обґрунтованість і практичну цінність результатів аналізу.

Список використаних джерел

1. Данильчук Г. Б. Візуалізація даних у бізнес-аналітиці: принципи та інструментарій. *Економіка та суспільство*. 2023. № 48.
2. Карабін О. Й. Особливості сучасного стану інформаційного середовища в вищих навчальних закладах. *Збірник наукових праць Херсонського державного університету. Педагогічні науки*. Херсон, 2009. В. LII, С. 440–443.
3. Key Trends in Data and Analytics (D&A). URL: https://www.gartner.com/en/data-analytics/topics/data-trends?utm_source=chatgpt.com (дата звернення: 12.03.2026).
4. Wilke C. O. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. Sebastopol: O'Reilly Media, 2019. 390 p.