

ГЕОМЕТРИЧНІ МОДЕЛІ БАГАТОВИМІРНИХ БАЗ ДАНИХ

Київський національний університет будівництва і архітектури

У даній публікації окреслені задачі аналізу існуючих розробок в сфері візуалізації багатовимірних інформаційних образів, а також напрямки досліджень інформаційних моделей, їх систематизація і розробка комбінованих геометричних моделей багатовимірних баз даних. Цей напрямок досліджень є гостроактуальним в наш вік суцільної діджиталізації і тенденції візуалізації ефектів в процесі взаємодії людини з навколишнім світом.

Візуалізація даних - завдання, з яким стикається в своїй роботі будь-який дослідник. Традиційні інструменти в цій галузі - графіки та діаграми, погано справляються із завданням візуалізації, коли виникає необхідність візуалізувати багатовимірну базу даних. Робота в цьому напрямку вимагає аналізу природи даних і систематизації або розробки способів формалізації даних, видів інформаційних моделей, способів геометричного відображення даних і способів трансформації, дослідження геометричних моделей багатовимірних баз даних.

Ключові слова: інформація, інформаційні процеси, візуалізація багатовимірних моделей, бази даних, геометричні образи.

Постановка проблеми. Будь-яка діяльність людини являє собою процес збору та переробки інформації, прийняття рішень на її основі і їх виконання. За даними Каліфорнійського університету в Берклі щорічний приріст інформації в світі становить 1 мільйон терабайт [1]. Агрегування, узагальнення та візуалізація даних є одним з основних напрямків роботи з інформаційними даними. Візуалізація даних це потужний інструмент роботи з даними завдяки своїй ефективності в формалізації інформації, розумінні та інтерпретації результатів. Йдеться про створення інформативних візуальних образів. Аналіз і вивчення візуальних образів інформаційних моделей дозволяє робити не тільки аналітичні висновки, а й творчі або наукові прогнози і ідеї, доступні виключно людському інтелекту.

Однак візуалізація багатовимірних наборів даних, які зазвичай мають більше двох параметрів, викликає необхідність використання спеціальних інформаційних інструментів, оскільки середовище аналізу і обміну даними людиною зазвичай обмежена двома вимірами. Що обумовлено, в тому числі, фізіологічними і антропологічними можливостями сучасної людини. А також у зв'язку з різноманітною природою даних виникає необхідність формалізації, оцифрування та інтерпретації інформаційних даних

візуальними геометричними образами.

Формулювання цілей та завдання статті. Візуальний аналіз даних дозволяє в процесі дослідження і трансформації даних подати великі обсяги даних в такій формі, де людина могла б побачити те, що важко обробити аналітично. Це не завжди можливо в найпростіших графічних двовимірних видах представлення даних, таких як діаграми, гістограми, двовимірні графіки, графи, блок-схеми, тощо. У зв'язку з цим виникає необхідність в більш складних засобах відображення інформації і результатів аналізу. Робота в цьому напрямку вимагає аналізу природи даних і систематизації або розробки способів формалізації даних, видів інформаційних моделей, способів геометричного відображення даних і способів трансформації, дослідження геометричних моделей багатовимірних баз даних. Цей напрямок досліджень є гостроактуальними в наше століття суцільної діджиталізації і тенденції візуалізації ефектів в процесі сприйняття навколишнього світу людиною. Метою даної публікації є визначити шляхи аналізу існуючих розробок, окреслити напрямки досліджень інформаційних моделей і систематизувати, розробити методи створення та взаємодії з геометричними моделями багатовимірних баз даних.

Аналіз основних досліджень і публікацій. Основна мета методів візуалізації багатопараметричних даних - зробити подання геометричних моделей і виявити зв'язку між даними [8]. Побудова візуальної концепції для подання багатовимірних даних зазвичай зводиться до підходів, що дозволяє зменшити розмірність досліджуваної моделі до звичної двовимірної або тривимірної.

Візуальний аналіз даних має дві основні переваги перед аналітичними методами:

- візуальний аналіз дозволяє легко працювати з неоднорідними і зашумленими даними, в той час як не всі автоматичні методи можуть працювати з такими даними і давати задовільні результати;
- візуальний аналіз інтуїтивно зрозумілий і не вимагає складних математичних або статистичних алгоритмів.

Вирізняють такі типи методів візуалізації [7]:

- стандартні 2D / 3D-образи - гістограми, діаграми, лінійні графіки, тощо;
- геометричні перетворення - діаграма розкиду даних (діаграма розсіювання), афінні перетворення, тощо;
- відображення іконок - лінійні зображення;
- методи, орієнтовані на пікселі - рекурсивні шаблони, циклічні сегменти, тощо;
- ієрархічні образи - деревовидні карти, графи, блок-схеми.

Основна частина. Візуалізація даних як процес подання інформаційної моделі, дружньої до зору покликана вирішувати такі прикладні задачі:

- Швидкий і простий спосіб передачі інформації
- Допомога в прийнятті рішень
- Творчий і науковий прогноз моделі
- Маніпулювання та динамічний вплив.

Інформаційна модель являє собою сукупність знань про об'єкти, факти, процеси, явища, якими можуть обмінюватися люди в рамках конкретного контексту [2]. Сукупність послідовних дій, що виконуються з інформацією (інформаційною моделлю, даними), з метою отримання результату - це інформаційний процес в загальному сенсі слова. Серед всіх інформаційних процесів можна виділити найбільш загальні. До них відносяться передача, зберігання і обробка інформаційних даних.

Інформацію за способом сприйняття можна розділити на наступні види: візуальна, аудіальна, смакова, нюхова і тактильна [7]. Інформація визначається даними. У простому розумінні дані - це змінні, атрибути або аспекти об'єкту, процесу або явища. Багатовимірна модель означає наявність декількох змінних або атрибутів, що характеризують або виділяють даний об'єкт, явище чи процес серед інших.

Для візуалізації даних і дослідженні геометричних моделей даних важлива така характеристика, як щільність даних. Для аналогових або континуальних моделей ця характеристика легко керована, а для дискретних моделей є необхідність вирішувати задачі геометричних перетворень для забезпечення необхідної щільності даних [3], [4]. Методи геометричних перетворень візуальних образів спрямовані на трансформацію багатовимірних наборів даних з метою відображення їх у декартовому і в недекартовому геометричних просторах, а також керування зображенням і вирішення задач аналізу даних.

Найпростіше зображення має один вимір. Це можна розглядати як точку. Графік двох змінних представляється двома координатами. Три виміри можна уявити, додавши додаткову вісь (z). Але в цьому випадку ми отримуємо двовимірне зображення тривимірної моделі (в геометричному сенсі це аксонометрична проекція).

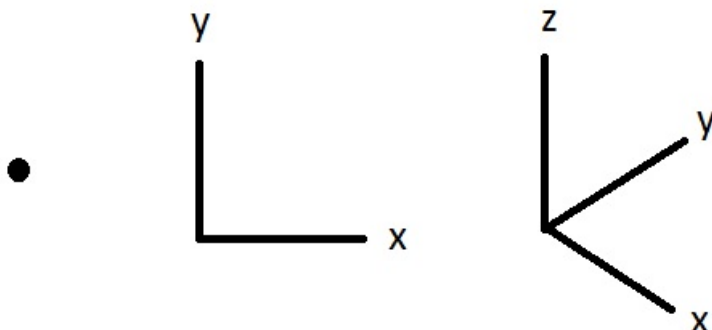


Рис. 1. Графіки одно-, дво-, три-вимірному простору

Інформаційні моделі з часового фактору і фактору взаємодії з

користувачем можна систематизувати наступним чином:



Інтерактивна візуалізація даних дозволяє користувачеві керувати параметрами зображення. При цьому також оновлюється нова вхідна інформація. Цей постійний зв'язок між користувачем і процесом візуалізації даних є сутністю інтерактивної моделі даних.

Вирізняють наступні види даних [8], з якими можуть працювати засоби візуалізації:

- одномірні дані - одномірні масиви, тимчасові ряди, тощо;
- двовимірні дані - точки двовимірних графіків, географічні координати, тощо;
- багатовимірні дані - фінансові показники, результати експериментів, тощо;
- тексти і гіпертексти - інформаційні статті, веб-документи та інше;
- ієрархічні і пов'язані дані - структура підпорядкованості, тощо;
- алгоритми і програми - інформаційні потоки, налагоджувальні операції, тощо.

Для візуалізації перерахованих типів даних використовуються різні візуальні образи і методи їх створення. Основна вимога до них - це наочність і зручність аналізу даних, які вони представляють. Методи візуалізації можуть бути як найпростіші (лінійні графіки, діаграми, гістограми тощо), так і більш складні, засновані на складному математичному апараті. Крім того, при візуалізації можуть використовуватися комбінації різних методів.

До простих методів візуалізації відносяться графіки, діаграми, гістограми тощо коли можна обмежитися візуалізацією 2-мірних моделей, оскільки саме в такому вигляді людина сприймає геометричні структури найприродніше і відносини між об'єктами виглядають найбільш наочно. Причому для 2-мірних моделей досить просто можна додати додаткові вимірювання даних за рахунок введення в модель кольору або типу лінії двовимірного графіка, іконок, та інше. У моделях такого типу додаткові виміри даних можуть мати не чисельну природу або формально оцифровану. Формально оцифрована інформація зазвичай адекватна реальній інформації за окремими ознаками, які саме характерні або інформаційно затребувані для конкретної моделі даних.

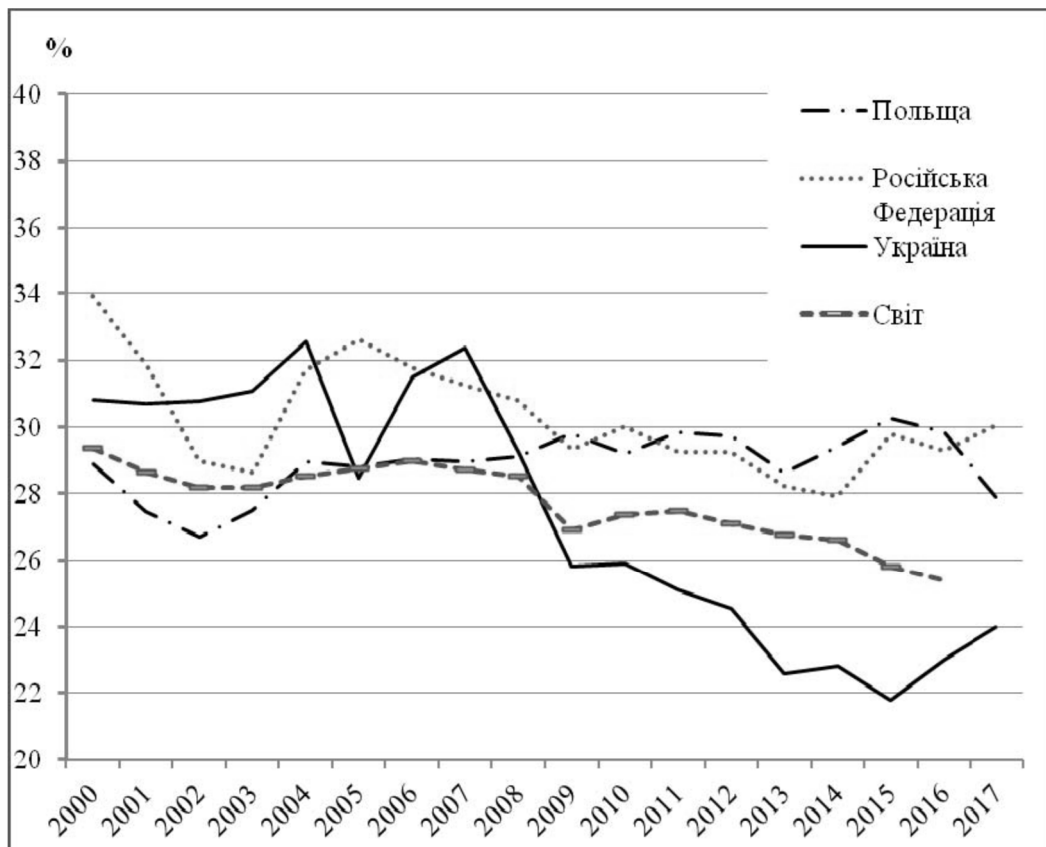


Рис.2. «Частка доданої вартості промисловості у ВВП» [5]- приклад двовимірної моделі з трьома видами даних, третій вимір позначено різновидом лінії двовимірного графіка.

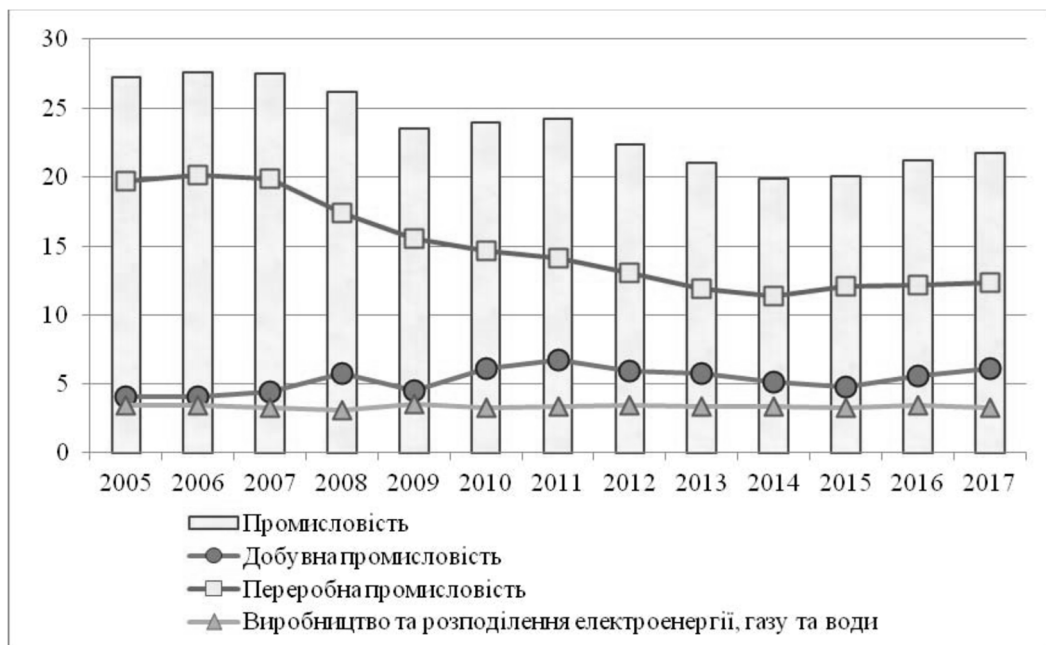


Рис. 3. «Частка основних видів економічної діяльності промисловості у ВВП України у 2005–2017 рр.» [5] – приклад двовимірної моделі з трьома різновидами даних, третій вимір позначено іконками.

Основним недоліком цих методів візуалізації є неможливість прийнятної візуалізації даних складної природи і великої кількості даних.

Інший вид візуалізації - діаграма розкиду даних (розсіювання). Діаграма розсіювання - простий графік дискретних даних. Використання більше трьох вимірів на двовимірному плоскому носії (аркуш паперу, екран монітора) не дозволяє забезпечити наочність зображення. Але є можливість емулювати багатовимірність простору для візуалізації додаткових вимірів даних за допомогою кольору або інтенсивності кольору, розміру або форми.

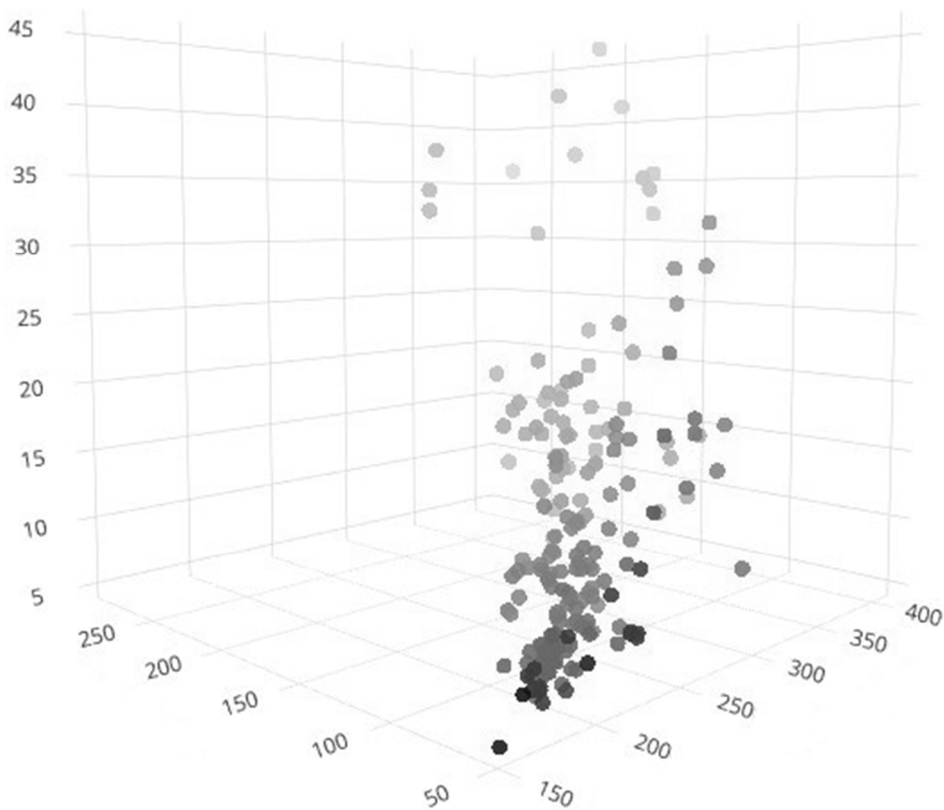


Рис. 4. Приклад графіка дискретних даних чотирьох вимірів. Три виміри візуалізовані на двовимірному носії за рахунок зображення тривимірної аксонометричної моделі простору, четвертий вимір візуалізується за рахунок інтенсивності кольору

Одним з методів візуалізації даних є методи відображення іконок. Їх основною ідеєю є відображення значень елементів багатовимірних даних в якості образів. Такі образи можуть являти собою: геометричні елементарні об'єкти, стрілки, елементи інфографіки, тощо. Візуалізація генерується відображенням атрибутів елементів даних в якості образів. Такі образи можна групувати для цілісного аналізу даних. Результуюча візуалізація включає шаблони текстур, які мають відмінності, що відповідають

характеристикам даних [9], [10].

З точки зору геометричного моделювання даних становить інтерес інструмент зображення інформації, прив'язаної до географічної сітки координат - технології ГІС (геоінформаційні системи). Моделі ГІС базуються на задачі побудови двовимірної «підкладки» для довільного набору даних. Всі інші дані різної природи можуть нанизуватися у вигляді інформаційних шарів на «підкладку». Таким чином можна візуалізувати дані, одночасно зазначаючи на підкладці супроводжуючу інформацію (підписи, анотації, атрибути, інформаційні розмальовки). Підкладка, що називається картою даних, може бути заміною географічній карті в тих моделях, де її просто не існує. Принципова відмінність полягає лише в тому, що на географічній карті сусідні об'єкти мають близькі географічні координати, на карті даних близькі об'єкти мають близькі властивості або виміри [7].

Висновки та перспективи. Існує досить багато методів візуалізації даних, але всі вони мають переваги та недоліки. Основна ідея комбінування останніх полягає в об'єднанні різних методів візуалізації для нівелювання недоліків одних методів перевагами інших. Наприклад, різні проєкції розсіювання точок можуть бути скомбіновані з методами фарбування або компонування точок у всіх проєкціях, іконками, трансформацією окремих масивів даних, накладенням шарів на карту даних, тощо.

Будь-який засіб візуалізації може бути класифікований за всіма параметрами: за часовим фактором, можливістю взаємодії користувача з інформаційною моделлю, виглядом даних, візуальними образами, за можливістю взаємодії з цими візуальними образами. В ідеалі необхідно, щоб метод візуалізації підтримував різні види даних, різні візуальні образи і різні способи взаємодії з образами.

Література

1. *Peter Lyman and Hal R. Varian.* How much information. Technical report, UC Berkeley, 2000. <http://www.sims.berkeley.edu/research/projects/how-much-info/>.
2. ISO/IEC 10746-2:2009 [ISO/IEC 10746-2:2009]
3. *Іванова Л.С.* Умови конструювання складених дискретно поданих кривих (ДПК) на множині рівновіддалених значень аргументу. // Прикладна геометрія та інж.графіка.- К: КДТУБА, 2001, Вип.68.
4. *Іванова Л.С.* Згладжуюча апроксимація ДПК на рівномірній сітці методом коригування зовнішніх сил. // Прикладна геометрія та інж.графіка.- К: КДТУБА, 2003, Вип.72.
5. Розвиток промисловості для забезпечення зростання та оновлення української економіки : науково-аналітична доповідь / за ред. д-ра екон. Наук Дейнеко Л.В. ; НАН України, ДУ «Ін-т екон. та прогнозув. НАН України». – К., 2018. – 158 с.

6. Кендалл М., Стюарт А. Статистические выводы и связи. - М.: Наука, 1973.-900 с.
7. Зиновьев А.Ю. Визуализация многомерных данных. - Интернет-издание, 2000. — 180 с.
8. Блум Ф., Лейзерсон А., Хофстедтер Л. Мозг, разум и поведение, М., Мир, 1988.
9. Кохонен Т. Ассоциативные запоминающие устройства. – М.: Мир, 1982. – 383 с.
10. Kohonen T. Self-organization and Associative Memory. Springer-Verlag: New York, 1997. 428 p.

GEOMETRIC MODELS OF MULTIDIMENSIONAL DATABASES

Larysa Ivanova

Kyiv National University of Construction and Architecture, Ukraine

This publication outlines the tasks of analyzing existing developments in the field of visualization of multidimensional information images, as well as the directions of research of information models, their systematization and development of combined geometric models of multidimensional databases. These researches are topical now days, at the time of continuous digitalization and trends of the visualization of effects in the process of human perception of the world.

Data visualization is a task that any researcher faces in his work. The problem of data visualization is reduced to the problem of visualizing experimental data, theoretical research results, statistical data, etc. Traditional instruments in this area - charts and graphs do not cope with the imaging task when there is a need to visualize multidimensional database. In addition, due to the diverse nature of data, there is a need for formalization, digitization and interpretation of information data by visual geometric images. This work requires an analysis of the nature of data and the systematization or development of methods for formalizing data, types of information models, methods for geometric display of data and methods for transformation, and studying geometric models of multidimensional databases.

The main goal of multi-parameter data visualization methods is to make representations of geometric models in a form where a person could see what is difficult to process analytically and identify the relationships between the data.

The visualization method can be classified in all respects: by the time factor, by the possibilities of user interaction with the information model, by the type of data with which it works, by the visual images that it can provide, and by the possibilities of interaction with these visual images. Ideally, it is necessary that the visualization method support different types of data, different visual images, and different ways of interacting with images.

Keywords: information, information processes, visualization of multidimensional models, databases, geometric images.

ГЕОМЕТРИЧЕСКИЕ МОДЕЛИ МНОГОМЕРНЫХ БАЗ ДАННЫХ

Л. С. Иванова

Киевский национальный университет строительства и архитектуры

В данной публикации намечены задачи анализа существующих разработок в сфере визуализации многомерных информационных образов, а также направления исследований информационных моделей, их систематизация и разработка комбинированных геометрических моделей многомерных баз данных. Это направление исследований является остроактуальным в наш век сплошной диджитализации и тенденции визуализации эффектов в процессе взаимодействия человека с окружающим миром.

Визуализация данных – задача, с которой сталкивается в своей работе любой исследователь. К задаче визуализации данных сводится проблема представления в наглядной форме данных эксперимента, результатов теоретического исследования, статистических данных и т.п. Традиционные инструменты в этой области – графики и диаграммы, плохо справляются с задачей визуализации, когда возникает необходимость визуализировать многомерную базу данных. А также, в связи с многообразной природой данных, возникает необходимость формализации, оцифровывания и интерпретации информационных данных визуальными геометрическими образами. Работа в этом направлении требует анализа природы данных и систематизации или разработки способов формализации данных, видов информационных моделей, способов геометрического отображения данных и способов трансформации, исследования геометрических моделей многомерных баз данных.

Основная цель методов визуализации многопараметрических данных - сделать представления геометрических моделей в такой форме, где человек мог бы увидеть то, что трудно обработать аналитически, и выявить связи между данными.

Средства визуализации могут быть классифицированы по разным параметрам: по временному фактору, возможностям взаимодействия пользователя с информационной моделью, по виду данных, по визуальным образам, по возможностям взаимодействия с этими визуальными образами. В идеале необходимо, чтобы метод визуализации поддерживал разные виды данных, разные визуальные образы и разные способы взаимодействия с образами.

Ключевые слова: информация, информационные процессы, визуализация многомерных моделей, базы данных, геометрические образы.