УДК 37.012.85
DOI: http://dx.doi.org/10.21686/1818-4243-2023-3-17-26

К.В. Мулюкова[1], И.В. Мулюков[2], В.М. Курейчик[1]

[1] Инженерно-технологическая академия Южного федерального университета, Таганрог, Россия
[2] Южный федеральный университет, Ростов-на-Дону, Россия

Технология интеллектуального анализа больших данных для исследования пространственно-временных тенденций застройки крупных городов

Целью данной работы *является исследование современных проблем и перспектив решения обработки больших данных, содержащих сведения об объектах недвижимости, а так же возможность практической реализации методики обработки подобных массивов данных путем проектирования и наполнения специальной графической абстракции «метадом» на практическом примере.*

Материалы и методы. *Исследование включает обзор библиографических источников по проблемам анализа больших данных и применение их в современной области строительства крупных городов. При исследовании была применена методика представления данных в графической форме – абстракции. Математической основой методики является использование многомерных пространств, где измерения — это характеристики отдельных объектов. Применено компьютерное моделирование практической задачи с помощью языка программирования C#. Хранение больших данных выполнено на основе сервера MongoDB. Для визуализации данных применяется Web-интерфейс на основе HTML и CSS.*

Результаты. *В ходе работы были выделены основные характеристики больших данных, а также описана специфика массивов данных, состоящих из сведений об объектах недвижимости крупного города.*

При обработке данных, состоящих из сведений об обьектах недвижимости крупного города, возникают определенные сложности. В связи с этим были предложены приёмы эффективного решения поставленной практической задачи обработки и поиска закономерностей в большом массиве данных: абстракция «метадом», агрегатор данных.

Были получены табличные данные по крупному городу путем анализа трех миллионов записей, содержащие более 10 групп данных, при базовом наборе параметров: этаж, этажность, цена, площадь, жилая площадь, кухонная площадь, тип, операция. Был создан кластер на MongoDB на несколько компьютеров, каждый из которых занимался собственным набором данных без сведения промежуточных результатов.

Результаты вычислительного эксперимента показали, что при использовании графической формы (векторной) представления больших данных, сократились расходы и время на интерпретацию данных интеллектуального анализа.

Совмещение методов обработки больших данных и их представления через графическую абстракцию — позволяют получить новые результаты по имеющимся наборам данных

Заключение. *В ходе исследования было выявлено, что представление групп полученных данных в графическом изображении обладает рядом преимуществ над табличным представлением данных (векторное изображение легко масштабировать, возможность сравнения без построения графиков).*

Предложенная схема визуализации больших данных путем построения абстрактных векторных изображений является альтернативой традиционным таблицам, позволяя взглянуть иначе на массивы данных и результаты их обработки.

Полученные результаты могут использоваться как в целях первичного изучения технологий обработки больших данных, так и в качестве основы разработки уже реальных приложений в следующих сферах: анализ изменения площадей домов с течением времени, анализ изменения этажности городской застройки, динамика и распределение спроса и предложения и т.д.

Ключевые слова: *Большие данные, недвижимость, анализ данных, графическое представление данных, C#, MongoDB, умные города.*

Ksenia V. Mulyukova[1], Ivan V. Mulyukov[2], Victor M. Kureichik[1]

[1] Engineering-Technological Academy of SFU, Taganrog, Russia
[2] Southern Federal University, Rostov-on-Don, Russia

Big Data Intelligent Analysis Technology for the Study of Spatial and Time Trends in the Development of Large Cities

The purpose of this research *is to study modern problems and prospects for solving the processing of big data containing information about real estate, as well as the possibility of practical implementation of the methodology for processing such data arrays by designing and filling a special graphic abstraction "metahouse" on a practical example.*

Materials and methods. *The study includes a review of bibliographic sources on the problems of big data analysis and their application in the modern field of construction of large cities. During the study, a technique for presenting data in a graphical form - abstraction was used. The mathematical basis of the technique is the use of multidimensional spaces, where measurements are the characteristics of individual objects. Computer simulation of a practical problem was applied using the C# programming language. Big data storage is based on the MongoDB server. To visualize data, a Web interface based on HTML and CSS is used.*

Results. *In the course of the work, the main characteristics of big data were identified, and the specifics of data arrays consisting of information about real estate objects in a large city were described. When processing data consisting of information about real estate objects of a large city, certain difficulties arise. Thereby, methods for effectively solving the set practical task of processing and searching for patterns in a large data array were proposed: "metahouse" abstraction, data aggregator.*

Tabular data were obtained for a large city by analyzing three million records containing more than 10 data groups, with a basic set of parameters: floor, number of floors, price, area, living area, kitchen area, type, operation. A MongoDB cluster was created on several computers, each of which was working with its own data set without intermediate results.
The results of the computational experiment showed that when using the graphical form (vector) of big data representation, the costs and time for interpreting mining data were reduced.
Combining big data processing methods and their presentation through graphical abstraction allows getting new results from existing data sets.
Conclusion. *During the study, it was found that the presentation of groups of the received data in a graphic image has a number of advantages over a tabular presentation of data (a vector image is easy to scale, the ability to compare without plotting).*
The proposed way for visualizing big data by constructing abstract vector images is an alternative to traditional tables, allowing you to take a different look at data arrays and the results of their processing. The results obtained can be used both for the primary study of big data processing technologies and as a basis for the development of real applications in the following areas: analysis of changes in the area of houses over time, analysis of changes in the number of floors of urban development, dynamics and distribution of supply and demand, etc.

Keywords: *Big data, real estate, data analysis, graphical presentation of data, C#, MongoDB, smart cities.*

Введение

Обработка данных средствами вычислительной техники является одной из основных задач большинства информационных систем. Любая информация, структурированная определенным образом, может быть обработана как для получения непосредственных результатов вычислений, так и для подготовки к передаче по каналам связи или дальнейшей обработки. По мере развития средств хранения и коммуникаций, объемы информации возрастают нелинейно. В соответствии с законами диалектики, количественное изменение массива обрабатываемой информации переходит в качественное новое состояние – большие данные.

Под большими данными обычно понимают такие массивы данных, которые не могут быть эффективно обработаны стандартными средствами вычислений [1] и хранения, в силу качественно отличающего размера информационных блоков. Это вынуждает нас искать и реализовывать принципиально новые приёмы работы с подобными данными.

До 2005 года большая часть данных была локальна и сосредоточена. В 2005 году появилась концепция Веб 2.0., которая кардинально меняет подход к проектированию интернет-сайтов. Сайты становятся динамической веб-системой с постоянно растущей базой данных, где информация распределена между различными узлами. Также 2005 год отмечен появлением мобильной и облачной компьютеризации. Данные факторы и послужили началом эпохи больших данных.

В 2008 году Клиффорд Линч в своей статье ввел понятие «большие данные». Это и стало толчком к формированию больших данных, как научного направления. Уже в 2010 году Марц Натан и Уоррен Джеймс в своей книги дают рекомендации по практической работе с большими данными [2].

В наше время большие данные используются во многих областях: банковское дело, медицина, электронная коммерция, веб-аналитика и градостроительство.

Большие данные в градостроительстве [3] позволяют увидеть город под другим углом. Они дают возможность эффективно решать проблемы доступности качественного жилья, социальной комфортности, оптимизации городского траффика движения, территориальное зонирование [4] и многие другие вопросы.

Один из известных исследователей в области больших данных в градостроительстве – профессор Роб Китчин. Он является одним из создателей проекта «Программируемый город». Ведущее направление проекта – работа с большими данными, касающихся людей, объектов недвижимости, транзакций и территорий.

Так же в 2014 году было создано Европейское инновационное партнерство умных городов и общин (Smart Cities and Communities, EIP-SCC) [5]. А с 2018 года и в России [6] реализуется проект цифровизации «Умный город», где одним из аспектов проекта является интеллектуальная аналитика больших данных.

Из этого следует, что работа с большими данными в градостроительстве является востребованной и актуальной на сегодняшний день.

Предметом исследования является набор больших данных (Big Data), состоящий из сведений об объектах недвижимости крупного города, получаемых из открытых источников, имеющих определенную структуру и достаточно большой объем, делающий привычные способы анализа данных – неточными или бесполезными [7].

Основные отличия таких данных от линейных наборов записей:

1. объем;
2. независимое формирование.

Для построения достоверной аналитики (за период 2–3 года) для крупного города, которая содержит сведения о продаже, постройке и аренде и т.д, объем данных будет превышать десятки миллионов записей. Речь идет об объемах данных, которые существенно выше типового объема базы данных отдельной компании, реестра или выборки сайта аналитики рынка. Традиционные методы получения новых данных, вроде выборки средних значений, медианы или

группировки по признакам – дают либо неточные, либо лишенные смысла результаты.

Целью работы является исследование современных проблем и перспектив решения обработки больших данных, содержащих сведения об объектах недвижимости, а так же практическая реализация методики обработки подобных массивов данных путем проектирования и наполнения специальной графической абстракции «метадом».

Актуальность выбранной темы обусловлена тем, что работа с большими и неточными данными, содержащими сведения об объектах недвижимости, изменяется в лучшую сторону, если представить большие данные в графической форме вместо табличных и численных представлений. Для этого необходимо получить наборы данных из открытых источников, представить их в традиционном виде, а потом в графической форме на основе разработанной абстракции

1. Специфика больших данных в градостроительстве

Под большими данными обычно подразумеваются обобщенные наборы, включающие в себя структурированные и неструктурированные данные, существенные по объему и разнообразные по структуре [8].

Существует множество характеристик для больших данных. Но мы сформулируем основные характерные признаки, которые позволили бы не количественно, но качественно отличить их от обычных, традиционных массивов данных. Как и любое нечеткое понятие, определяемое чаще всего по факту, оно может быть описано несколькими признаками, дополняющими друг друга. Чем больше признаков соответствует исследуемому набору данных, тем более вероятно, что массив относится к большим данным. Выделим более подробно основные признаки:

Объем – большие данные включают в себя масштабные записи, которые содержат данные о заказах в интернет-магазинах [9] или клинические данные и медицинские изображения [10].

Многообразие – большие данные включают в себя тысячи различных структур [11]: структурированной, неструктурированной, квази структурированной, полуструктированной информации различных форматов. Часть из этих данных даже не имеет описания данных на уровне хранилища, что требует перестраивать обработку данных по мере выявления новых структур.

Распределенность – большие данные расположены на многих кластерах и носителях.

Временной интервал – большие данные содержат информацию за огромный период времени, что требует введения особых методов вычислений, если расчеты касаются временной динамики [12].

Приведём сравнительную таблицу, где укажем отличие больших данных от обычных массивов информации. Для более доступного восприятия данной информации была создана следующая таблица

Все вышеперечисленные признаки относятся к большим данным в обобщённой форме.

При использовании больших данных, состоящих из сведений об объектах недвижимости крупного города, перечисленные признаки осложняются спецификой предметной области:

1. Отсутствие гарантий на соответствие одной записи одному объекту реального мира – объект может быть внесен разными людьми, из разных источников, разными способами, наконец, в разное время, порой весьма продолжительное. Таким образом, взяв среднее значение по записям, мы не получим реальное среднее значение в реальном мире. Если же какие-то источники имеют больший объем

Таблица 1 (Table 1)

Отличие больших данных и обычных массивов информации
The difference between big data and regular information arrays

Признак	Обычные данные	Большие данные
Объем	В большинстве случаев локализованы как частные базы данных.	Размер набора данных выходит за предел возможностей программных средств классической СУБД.
Распределенность	Традиционно конкретные данные содержатся на одном-двух носителях, обрабатываются в едином адресном пространстве приложения и не превышают размеров файловой системы.	Распределены на кластерах из сотен и тысяч узлов.
Многообразие	Классические структуры данных оформляются в виде кортежей «поля-записи» в случае реляционных таблиц, или в виде хранилища типовых документов в документо-ориентированных системах.	Содержат различные типы данных: структурированные, полуструктурированые, квази структурированной, неструктурированные.
Временной интервал	Типовые массивы данных содержат информацию за небольшой период времени – текущий месяц, квартал, отчетный год. Этого достаточно для задач учета и первичного анализа.	Содержат информацию за большой период времени

данных по конкретному признаку объекта – то вычисление будет заведомо неверным, и новые данные окажутся далеки от фактических показателей.

2. Возможная неполнота у многих записей – сведения, получаемые из разнообразных источников, чаще всего не обладают жестко заданной структурой. Это накладывает дополнительные требования для анализа: необходимо пропускать отсутствующие данные, при этом, учитывая те, что заданы явно.

Для реализации задачи обработки опишем модель данных наиболее подходящим для анализа способом и реализуем собственный алгоритм обработки больших данных с целью получения новых знаний из входящей информации.

Хотя данная статья не предполагает полноценной разработки программного решения, мы опишем отдельные части системы на языке C# 4.0 в виде классов. Хранение больших данных выполняется на основе сервера MongoDB [13], который подходит как для документов-ориентированных записей, так и для нечетких данных [14]. Для визуализации данных и интерфейса доступа пользователя применяется Web-интерфейс на основе HTML и CSS, позволяющий использовать систему на любом HTTP-сервере, как в локальной сети, так и на удаленном выделенном сервере.

2. Использование технологий: абстракции «метадом» и агрегатора разнородных данных. Пример практической реализации технологий

2.1. Разработка алгоритма решения

В основе разработанного алгоритма анализа больших данных, состоящих из сведений об объектах недвижимости крупного города лежат две идеи:

1) Абстракция «метадом»;

2) агрегатор [15] данных разнородных источников, заполняющий абстракцию «метадом» на основе входящих записей.

«Метадом» – это абстрактный объект, не существующий в реальном мире, но наиболее похожий на типичный дом, квартиру или другой объект недвижимости в заданном наборе среди больших данных.

Агрегатор разделяет данные записей на две части: критерии построения и свойства метадома. Также агрегатор вычисляет эти свойства на основе разработанного алгоритма: представим все записи массива больших данных как точки в многомерном пространстве размерности N, где измерения $0...K$ – это параметры построения, а измерения $K+1...N-1$ – это свойства записей. Тогда «метадом» будет обладать следующим свойством: он располагается так, что сумма расстояний от него до всех записей в пространстве свойств является минимальной среди возможных. За счет этого мы снижаем влияние погрешностей исходного набора больших данных, повторные и неточные данные выполняют сдвиг координат «метадома», но в меньшей степени, чем при обычных вычислительных методах вроде вычисления среднего или медианного значения.

2.2. Архитектура решения

Архитектура решения для обработки больших данных состоит из трех блоков:

1) Сбор записей из источников в хранилище путем приведения к формату конкретного класса;

2) проведение этих данных через агрегатор, для получения набора заполненных «метадомов», в соответствии с настройками анализа;

3) представление обработанных данных в табличной и графической форме, для пользователя системы анализа больших данных.

Обобщенная схема архитектуры решения для сбора и анализа больших данных показана на рисунке 1

2.3. Агрегация входных данных

Опишем структуру входных данных для кластерного решения обработки больших данных на MongoDB [16]. Ка-

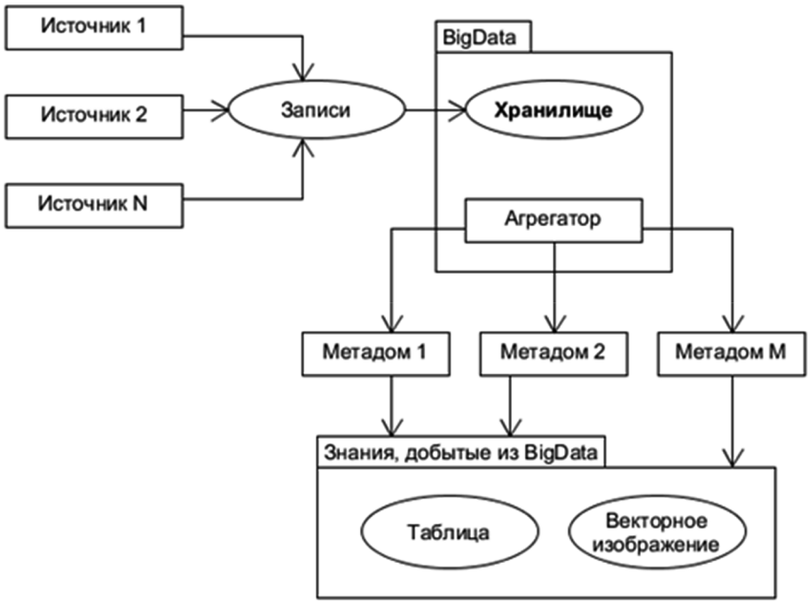

Рис. 1. Обобщенная схема архитектуры решения
Fig. 1. Generalized scheme of the solution architecture

ждая запись на рисунке 2 – это JSON-структура [17], независимая от прочих.

```
{
        "datetime":"",
        "Floor":"",
        "FloorAll":"",
        "Price":"",
        "S":"",
        "Sk":"",
        "Sh":"",
        "TypeCode":"",
        "OperCode":"",
        "Lat":"",
        "Lon":""
    }
```

Рис. 2. Структуру входных данных

Fig. 2. The structure of the input data

Отсутствующие или неполные данные представлены кодом null, чтобы избежать путаницы с нулевыми значениями, которые в отличие от неполных данных – имеют другой математический смысл.

Агрегация классов может выполняться в реляционной базе данных, файловом хранилище, облачном формате [18] или в документ-ориентированной базе, как выбрано в нашей реализации. Важно лишь то, что в силу заявленного объема данных, нет возможности содержать их в обычных коллекциях вроде вектора или списка [19]. Накопление больших данных происходит в постоянном режиме, поступление данных идет путем анализа открытых источников, сбора данных с интернет-сайтов и подключения архивных данных организаций. Условия контроля достоверности, точности и актуальности данных – здесь не заданы, качественный переход достигается за счет количества данных, даже если часть информации будет содержать неточные или заведомо неверные сведения.

Класс агрегатора на рисунке 3 имеет всего два метода – установка параметров и поступление на вход нового объекта. Также он содержит ссылку на объект, который по структуре кода является элементом, но по архитектуре – это метадом, абстрактный объект, обладающий наиболее типичными характеристиками для заданных параметров после проведения всех записей.

```
public class Agregator
    {
        public void setParameter(String code, int value);
        public void addElement(Element el);
        public Element result;
```

Рис. 3. Класс агрегатор

Fig. 3. Aggregator class

```
public class TableBuilder
    {
        public String getTableView(JSONObject el);
    }
```

Рис. 4. Класс, формирующий таблицу

Fig. 4. The class that forms the table

Такой упрощенный подход к архитектуре [20], тем не менее, позволяет решать поставленную задачу в полной мере. Проход всех записей через агрегатор позволяет получить выходные данные и может быть повторен с разными наборами параметров для анализа пространственно-временных состояний объектов.

2.4. Представления входных данных

Мы разделили выходные данные на две части:

1. Табличная составляющая [21] – мы получаем некие числовые данные, которые представляют собой новые знания о массиве городской недвижимости, получаемые из всей совокупности данных за весь период;

2. визуальная составляющая – позволяет не только получить числа, но и увидеть их в наглядной форме.

Табличная составляющая представляет собой обычные таблицы по характеристикам объекта, в виде пар «ключ-значение». Эти таблицы могут быть загружены в дополнительные программы и в свою очередь использоваться как источник данных для более глубокого анализа. В таблице 2 показан пример выходных данных в табличной форме – «метадом» с конкретными характеристиками.

Класс на рисунке 4, осуществляющий заданное построение, имеет один метод, формирующий таблицу по объекту «метадома»

Таблица 2 (Table 2)

Табличное представление «метадома»

Tabular representation of the "metahouse"

Этаж	2.5
Этажность	5.7
Цена	2530000
Площадь	60.89
Площадь жилая	30.12
Площадь кухни	7.51
Тип	Квартира
Операция	Предложение

Визуализация позволяет вывести информацию о «метадоме» в графическом виде [22], так, что пользователь сможет быстро понять смысл полученных результатов и использовать их для извлечения практической пользы, например, анализа динамики застройки города или тенденции изменения площадей домов.

На программном уровне, визуализация реализуется как вывод векторного схематичного изображения одного или нескольких домов, параметры которых (высота, ширина, толщина линии, тип изображения) отвечают за представление отдельных характеристик. Изображения подписываются

критериями и группируются по одному или нескольким параметрам. Пример визуализации одного «метадома» типа «дом» показана на рисунке 5.

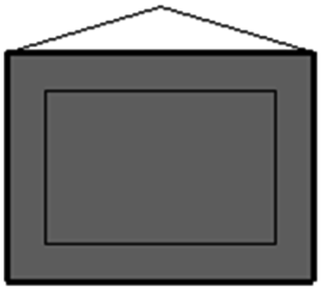

Рис. 5. Пример визуализации «метадома» векторным изображением

Fig. 5. An example of the "metahouse" visualization with a vector image

Конкретные характеристики визуализации «метадомов»:

Площадь – отображается как основание векторной схемы. Чем больше площадь, тем шире основание. Зависимость от площади вычисляется, как квадратный корень.

Этажность – отображается как высота векторной схемы. Чем выше этаж, тем выше схема. Зависимость от этажа – линейная.

Цена – отображается как толщина линии векторной схемы. Чем выше цена, тем толще линии. Зависимость от цены – логарифмическая.

Спрос и предложение – задается цветом заливки фона векторной схемы, где белый цвет кода 0xFFFFFF– это абсолютный спрос, а серый цвет кода 0x808080 – это абсолютное предложение.

Разброс параметров – представлен внутренней рамкой, которая показывает, насколько велик разброс внутри данного массива больших данных. Чем дальше внутренняя рамка от внешней, тем больший разброс параметров был обнаружен при построении данного «метадома».

Таким образом, на рисунке 5 показан «метадом» типа «дом» с большой площадью, средней этажностью, преобладанием предложения над спросом, высокой ценой и умеренным разбросом параметров.

Приведем еще несколько выходных визуализаций для сравнения относительно изменений параметров. На рисунке 6 показан «метадом» с меньшим разбросом параметром, меньшей ценой и преобладанием спроса над предложением.

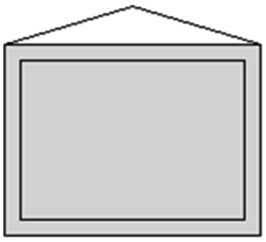

Рис. 6. Дополнительный пример визуализации «метадома» векторным изображением

Fig. 6. An additional example of the "metahouse" visualization with a vector image

На рисунке 7 показан «метадом» типа «квартира» с высокой этажностью, малой площадью, равновесием спроса и предложения, а также большим разбросом параметров.

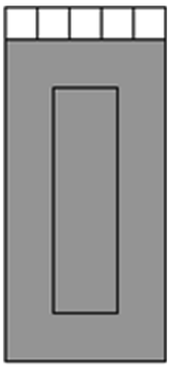

Рис. 7. Дополнительный пример визуализации «метадома» векторным изображением

Fig. 7. An additional example of the "metahouse" visualization with a vector image

Аналогичный «метадом», но с абсолютным преобладанием спроса над предложением, высокой ценой и минимальным разбросом параметров, выглядит так, как представлено на рисунке 8.

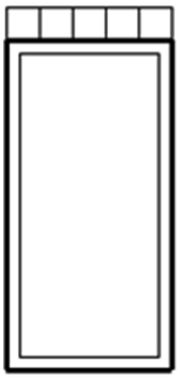

Рис. 8. Дополнительный пример визуализации «метадома» векторным изображение

Fig. 8. An additional example of the "metahouse" visualization with a vector image

Класс на рисунке 9, осуществляющий графическое представление, аналогично текстовому, имеет метод для создания изображения по входящему объекту

```
public class GraphicBuilder
    {
        public Bitmap getGraphicView(Element el);
    }
```

Рис. 9. Класс для создания изображения по входящему объекту

Fig. 9. Class for creating an image for an incoming object

3. Результаты практической реализации технологий: абстракции «метадом» и агрегатора разнородных данных

Работа с большими данными (Big data) всё еще является достаточно сложной с практической точки зрения и требующей дополнительного изучения в теории**.**

В процессе исследования была выдвинута гипотеза о том, что представление больших данных в графической форме обладает рядом преимуществ над традиционным представлением больших данных в табличной форме. Для доказательства гипотезы была спроектирована путем обработки больших данных абстракция «метадом».

Также были получены табличные данные по крупному

городу путем анализа трех миллионов записей, содержащие более 10 групп данных, при базовом наборе параметров: этаж, этажность, цена, площадь, жилая площадь, кухонная площадь, тип, операция. Для решения задачи обработки такого массива данных, был создан кластер на MongoDB на несколько компьютеров, каждый из которых занимался собственным набором данных без сведения промежуточных результатов.

Итогом данного интеллектуального анализа данных является таблица размером 8 на 10 ячеек, которая имеет ряд недостатков:

1. таблицу сложно уменьшить без потери удобства восприятия числовых показателей. Такое визуальное восприятие неудобно для пользователя;

2. для восприятия отличий разных групп данных по различным показателям необходимо строить графики. Это в свою очередь, приводит к дополнительным экономическим и временным затратам.

Применение концепции «метадома» и представление групп полученных данных в графическом (вектором) изображении обладает рядом преимуществ над табличным представлением данных:

1. векторное изображение легко масштабировать;

2. представление различных показателей разными частями изображения позволяет их сравнить без построения графиков – высота, ширина, цвет и толщина линий сразу же дают наглядное и видимое сравнение для пользователя.

Предложенная схема визуализации больших данных путем построения абстрактных векторных изображений с изменяемыми характеристиками является альтернативой традиционным таблицам или диаграммам, позволяет взглянуть под другим углом на массивы данных и результаты их обработки.

Результаты вычислительного эксперимента показали, что при использовании графической формы (векторной) представления больших данных, сократились расходы и время на интерпретацию данных интеллектуального анализа.

Совмещение методов обработки больших данных и их представления через графическую абстракцию – позволяют получить новые результаты по имеющимся наборам данных.

Заключение

Полученные результаты могут использоваться как в целях первичного изучения технологий обработки больших данных, так и в качестве основы разработки реальных приложений в следующих сферах:

1. Анализ изменения этажности городской застройки по районам города – прописав в коде параметры агрегатора больших данных на координаты, можно получить таблицу с метадомами, числа этажности которых будут отображать реальное состояние этажей городской застройки и то, как она изменяется в пространстве. Такие показатели высотности застройки важны как для спасательных служб, так и для оптимизации городского траффика движения.

2. Анализ изменения площадей домов с течением времени – прописав параметры агрегатора больших данных на время, можно получить таблицу с «метадомами», характеристики площадей которых будут отображать реальное состояние площадей городской застройки и то, как она изменяется во времени. Такие показатели позволяют определять более благополучно застроенные районы, а при анализе временных рамок – оценивать скорость решения проблем доступности качественного жилья.

3. Динамика и распределение спроса и предложения – прописав в коде параметры агрегатора больших данных на время или координаты, можно получить визуализацию, из которой будет следовать соотношение спроса и предложения в разных районах в различные годы. Такая динамика крайне важна для определения растущих или «депрессивных» зон, а также предсказания кризиса или роста рынка на основе стабильно изменяющихся совокупных показателей.

4. Анализ динамики цен по времени, районам и отдельным типам объектов – хотя цены и не являются объективными показателями, в отличие от площади или этажности, всё же визуализация распределенных рыночных цен позволяет получить общую картину городской застройки, выделяя более активные районы среди прочих и наблюдая за изменением состояния рынка длительного периода.

При добавлении к разработанной системе инструментов предоставления публичных API и документаций разработчика, можно реализовать эффективный инструмент для организаций и частных лиц, которым необходимы достоверные данные о городской застройке.

Литература

1. Valeev S.S., Kondratyeva N.V. Aviation industry stochastic model based on big data concept // Ученые записки Казанского университета. Серия Физико-математические науки. 2018. № 2(160). С. 392–398.

2. Марц Н., Уоррен Д. Большие данные. Принципы и практика построения масштабируемых систем обработки данных в реальном времени. М.: Вильямс, 2017. 368 с.

3. Honarvar A.R., Sami A. Towards Sustainable Smart City by Particulate Matter Prediction Using Urban Big Data, Excluding Expensive Air Pollution Infrastructures // Big Data Research. 2019. Т. 17. № 22. С. 222–226. DOI: 10.1016/j.bdr.2018.05.006.

4. Xiao X., Chao X. Rational planning and urban governance based on smart cities and big data // Environmental Technology & Innovation. 2021. Т. 21. С. 65–76. DOI: 10.1016/j.eti.2021.101381.

5. Ivanov N., Gnevanov M. Big data: perspectives of using in urban Planning and management // Business Technologies for Sustainable Urban Development, December 20–22 2017, St. Petersburg, Russia. DOI: 10.1051/matecconf/201817001107.

6. Умный город. Ведомственный проект Минстроя России [Электрон. ресурс] // Проектная дирекция Минстроя России. 2023. Режим доступа: https://russiasmartcity.ru/. (Дата обращения: 13.03.2023).

7. Благирев А.П., Хапаева Н. Big Data простым языком. М.: АСТ, 2019. 256 с.

8. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP. 2 изд. СПб.: БХВ-Петербург, 2007. 384 с.

9. Хаханов В.И., Обризан В.И., Мищенко А.С., Tamer B.A. Метрика для анализа big data // Радиоэлектроника и информатика. 2014. № 2 (65). С. 26–29.

10. Hong L., Luo M., Wang R., Lu P., Lu W., Lu L. Big Data in Health Care: Applications and Challenges // Data and Information Management. 2018. Т. 2. № 3. С. 175–197. DOI: 10.2478/dim-2018-0014.

11. Рыцарев И.А., Кирш Д.В., Куприянов А.В. Кластеризация медиа-контента из социальных сетей с использованием технологии Big Data // Компьютерная оптика. 2018. № 5 (42). С. 921–927.

12. Мулюкова К.В., Курейчик В.М. Проблема анализа больших веб-данных и использование технологии Data Mining для обработки и поиска закономерностей в большом массиве веб-данных на практическом примере // Открытое образование. 2019. № 23 (2). С. 42–49.

13. The most popular database for modern apps [Электрон. ресурс] // MongoDB. 2023. Режим доступа: https://www.mongodb.com/. (Дата обращения: 26.03.2023).

14. Сытник А.А., Шульга Т.Э., Данилов Н.А., Гвоздюк И.В. Математическая модель активности пользователей программного обеспечения // Программные продукты и системы. 2018. № 1 (31). С. 79–84.

15. Григораш А.С., Курейчик В.М., Курейчик В.В. Программный комплекс решения задачи кластеризации // Программные продукты и системы. 2017. № 2 (30). С. 261–269.

16. Heripracoyo S., Kurniawan R. Big Data Analysis with MongoDB for Decision Support System // Telkomnika. 2020. Т. 14. № 3. С. 1083–1089. DOI: 10.12928/TELKOMNIKA.v14i3.3115.

17. Celesti A., Fazio M, Villari, M.A Study on Join Operations in MongoDB Preserving Collections Data Models for Future Internet Applications // Future Internet. 2019. Т. 11. № 83. С. 1–17. DOI: 10.3390/fi11040083.

18. Yang C, Huang Q, Li Z, Liu K, Hu F. Big Data and cloud computing: innovation opportunities and challenges // International Journal of Digital Earth. 2017. Т. 10. № 1. С. 13–53. DOI: 10.1080/17538947.2016.1239771.

19. Novikova G.M., Azofeifa E.J. Semantics of big data in corporate management systems // Discrete and Continuous Models and Applied Computational Science. 2018. Т. 4. № 26. С. 383–392. DOI: 10.22363/2312-9735-2018-26-4-383-392.

20. Ignatova E, Zotkin S, Zotkina I. The extraction and processing of BIM data // IOP Conference Series: Materials Science and Engineering. 2019. Т. 365. № 6. С. 1–9. DOI: 10.1088/1757-899X/365/6/062033.

21. Панков А.В., Крибель А.М., Лаута О.С., Васильев Н.А. Метод по совершенствованию информационно-аналитической работы на основе комплексирования результатов распознавания состояний объектов контроля с использованием методов машинного обучения // Наукоемкие технологии в космических исследованиях Земли. 2022. № 2 (14). С. 27–35.

22. Белов В.А., Никульчев Е.В. Оценка временной эффективности форматов хранения больших данных в динамике роста объема данных // Современные информационные технологии и ИТ-образование. 2021. № 4 (17). С. 889–895.

References

1. Valeev S.S., Kondratyeva N.V. Aviation industry stochastic model based on big data concept. Uchenyye zapiski Kazanskogo universiteta. Seriya Fiziko-matematicheskiye nauki = Scientific notes of Kazan University. Series Physical and Mathematical Sciences. 2018; 2(160): 392–398. (In Russ.)

2. Marts N., Uorren D. Bol'shiye dannyye. Printsipy i praktika postroyeniya masshtabiruyemykh sistem obrabotki dannykh v real'nom vremeni = Big data. Principles and practice of building scalable real-time data processing systems. Moscow: Williams; 2017. 368 p. (In Russ.)

3. Honarvar A.R., Sami A. Towards Sustainable Smart City by Particulate Matter Prediction Using Urban Big Data, Excluding Expensive Air Pollution Infrastructures. Big Data Research. 2019; 17; 22: 222-226. DOI: 10.1016/j.bdr.2018.05.006.

4. Xiao X., Chao X. Rational planning and urban governance based on smart cities and big data. Environmental Technology & Innovation. 2021; 21: 65-76. DOI: 10.1016/j.eti.2021.101381.

5. Ivanov N., Gnevanov M. Big data: perspectives of using in urban Planning and management. Business Technologies for Sustainable Urban Development, December 20-22 2017, Saint Petersburg, Russia. DOI: 10.1051/matecconf/201817001107. (In Russ.)

6. Umnyy gorod. Vedomstvennyy proyekt Minstroya Rossii = Smart city. Departmental project of the Ministry of Construction of Russia [Internet]. Design Directorate of the Ministry of Construction of Russia. 2023. Available from: https://russiasmartcity.ru/. (cited 13.03.2023). (In Russ.)

7. Blagirev A.P., Khapayeva N. Big Data prostym yazykom = Big Data in plain language. Moscow: AST; 2019. 256 p. (In Russ.)

8. Barsegyan A.A., Kupriyanov M.S., Stepanenko V.V., Kholod I.I. Tekhnologii analiza dannykh: Data Mining, Visual Mining, Text Mining, OLAP. 2 izd = Data analysis technologies: Data Mining, Visual Mining, Text Mining, OLAP. 2nd ed. Saint Petersburg: BHV-Peterburg; 2007. 384 p. (In Russ.)

9. Khakhanov V.I., Obrizan V.I., Mishchenko A.S., Tamer B.A. Metrics for big data analysis. Radioelektronika i informatika = Radioelectronics and Informatics. 2014; 2(65): 26-29. (In Russ.)

10. Hong L., Luo M., Wang R., Lu P., Lu W., Lu L. Big Data in Health Care: Applications and Challenges. Data and Information Management. 2018; 2; 3: 175-197. DOI: 10.2478/dim-2018-0014.

11. Rytsarev I.A., Kirsh D.V., Kupriyanov A.V. Clustering media content from social networks using Big Data technology. Komp'yuternaya optika = Computer Optics. 2018; 5(42): 921-927. (In Russ.)

12. Mulyukova K.V., Kureychik V.M. The problem of analyzing big web data and using Data Mining technology to process and search for patterns in a large array of web data using a practical example. Otkrytoye obrazovaniye = Open Education. 2019; 23(2): 42–49. (In Russ.)

13. The most popular database for modern apps [Internet]. MongoDB. 2023. Available from: https://www.mongodb.com/. (cited 26.03.2023).

14. Sytnik A.A., Shul'ga T.E., Danilov N.A., Gvozdyuk I.V. Mathematical model of software user activity. Programmnyye produkty i sistemy = Software products and systems. 2018; 1(31): 79-84. (In Russ.)

15. Grigorash A.S., Kureychik V.M., Kureychik V.V. Software complex for solving the clustering problem. Programmnyye produkty i sistemy = Software products and systems. 2017; 2(30): 261–269. (In Russ.)

16. Heripracoyo S., Kurniawan R. Big Data Analysis with MongoDB for Decision Support System. Telkomnika. 2020; 14; 3: 1083-1089. DOI: 10.12928/TELKOMNIKA.v14i3.3115.

17. Celesti A., Fazio M, Villari, M.A Study on Join Operations in MongoDB Preserving Collections Data Models for Future Internet Applications. Future Internet. 2019; 11; 83: 1-17. DOI: 10.3390/fi11040083.

18. Yang C, Huang Q, Li Z, Liu K, Hu F. Big Data and cloud computing: innovation opportunities and challenges. International Journal of Digital Earth. 2017; 10; 1: 13-53. DOI: 10.1080/17538947.2016.1239771.

19. Novikova G.M., Azofeifa E.J. Semantics of big data in corporate management systems. Discrete and Continuous Models and Applied Computational Science. 2018; 4; 26: 383–392. DOI: 10.22363/2312-9735-2018-26-4-383-392.

20. Ignatova E, Zotkin S, Zotkina I. The extraction and processing of BIM data. IOP Conference Series: Materials Science and Engineering. 2019; 365; 6: 1-9. DOI: 10.1088/1757-899X/365/6/062033.

21. Pankov A.V., Kribel' A. M., Lauta O.S., Vasil'yev N.A. A method for improving information and analytical work based on the integration of the results of recognition of the states of control objects using machine learning methods. Naukoyemkiye tekhnologii v kosmicheskikh issledovaniyakh Zemli = Science-intensive technologies in space research of the Earth. 2022; 2 (14): 27–35. (In Russ.)

22. Belov V.A., Nikul'chev Ye.V. Estimation of the time efficiency of big data storage formats in the dynamics of data volume growth. Sovremennyye informatsionnyye tekhnologii i IT- obrazovaniye = Modern information technologies and IT education. 2021; 4(17): 889-895. (In Russ.)

Сведения об авторах

Ксения Валериановна Мулюкова
Аспирант, Кафедра «Систем автоматического управления»
Инженерно-технологическая академия Южного федерального университета, Таганрог, Россия
Эл. почта: mu.ksusha@yandex.ru

Иван Валерианович Мулюков
Студент бакалавриата
Южный федеральный университет, Ростов-на-Дону, Россия
Эл. почта: vanya.muliukoff@yandex.ru

Виктор Михайлович Курейчик
Д.т.н., профессор, Кафедра «Систем автоматического управления»
Инженерно-технологическая академия ЮФУ, Таганрог, Россия
Эл. почта: vmkureychik@sfedu.ru

Information about the authors

Ksenia V. Mulyukova
Postgraduate Student, Department of Automatic Control Systems
Engineering-Technological Academy of SFU, Taganrog, Russia
E-mail: mu.ksusha@yandex.ru

Ivan V. Mulyukov
Undergraduate student
Southern Federal University, Rostov-on-Don, Russia
E-mail: vanya.muliukoff@yandex.ru

Victor M. Kureichik
Dr. Sci. (Engineering), Professor, Department of Automatic Control Systems
Engineering-Technological Academy of SFU, Taganrog, Russia
E-mail: vmkureychik@sfedu.ru