

**Шевченко Виталий Сергеевич**

*студент, ФГБОУ ВО «Поволжский государственный университет  
телекоммуникаций и информатики»,  
Россия, г. Самара*

Бедняк Светлана Геннадьевна

*к.п.н., доцент,  
ФГБОУ ВПО «Поволжский государственный университет телекоммуникаций  
и информатики»,  
Россия, г. Самара*

## **РАЗРАБОТКА СИСТЕМЫ АНАЛИЗА МНОГОМЕРНЫХ ДАННЫХ НА ОСНОВЕ МЕДИЦИНСКИХ ПОКАЗАНИЙ**

**Аннотация.** В статье рассматриваются преимущества использования MongoDB для создания систем анализа многомерных данных. Представлен обзор основных функций MongoDB, таких как гибкая схема, поддержка различных типов данных и мощные возможности агрегации. Приводится пример анализа зависимости уровня глюкозы от возраста у пациентов с диабетом второго типа. Кроме того, обсуждается роль визуализации данных в медицинских исследованиях и перспективы применения машинного обучения. MongoDB является перспективной технологией для создания систем анализа медицинских данных, чтобы выявлять и прогнозировать развитие заболеваний, а также разрабатывать персонализированные подходы к лечению.

**Abstract.** The article discusses the advantages of using MongoDB to create multidimensional data analysis systems. An overview of MongoDB's main features is provided, such as flexible schema, support for various types of data, and powerful aggregation capabilities. An example of an analysis of the dependence of glucose levels on age in patients with type 2 diabetes is given. In addition, the role of data visualization in medical research and the prospects for the use of machine learning are

discussed. MongoDB is a promising technology for creating medical data analysis systems to detect and predict the development of diseases, as well as develop personalized treatment approaches.

**Ключевые слова:** многомерный анализ, медицинские данные, MongoDB, NoSQL, визуализация данных, машинное обучение, здравоохранение, биоинформатика.

**Keywords:** multidimensional analysis, medical data, MongoDB, NoSQL, data visualization, machine learning, healthcare, bioinformatics.

Актуальность многомерного анализа медицинских данных неуклонно возрастает в связи с усложнением клинических задач, требующих для решения всестороннего анализа больших объемов данных. Несмотря на то, что математические основы обработки данных остаются относительно стабильными на протяжении длительного исторического периода, технологические аспекты данной области претерпевают кардинальные изменения. Традиционные инструменты анализа, такие как ручные вычисления и графические методы, уступают место современным компьютерным системам, обладающим значительно большей вычислительной мощностью и функциональностью. В настоящее время применение компьютерных технологий для анализа данных стало не просто эффективным средством, но и необходимым условием для получения достоверных и обоснованных результатов, особенно в контексте медицинских исследований и клинической практики.

Многомерный анализ данных представляет собой комплекс методов и технологий, направленных на исследование сложных взаимосвязей между различными параметрами, характеризующими объект исследования. В контексте медицины такие методы помогают выявлять скрытые закономерности, прогнозировать развитие заболеваний, оценивать эффективность лечения и разрабатывать персонализированные подходы к терапии. Применение систем анализа многомерных данных на основе медицинских показателей, становится

актуальным благодаря развитию информационных технологий и накоплению больших объемов данных о пациентах.

Одним из ключевых аспектов применения многомерного анализа является интеграция различных источников информации, таких, как данные лабораторных исследований, результаты инструментальных обследований, анамнез пациентов, а также информация о генетических предрасположенностях и образе жизни. С помощью многомерного анализа создаются сложные модели, учитывающие множество факторов, влияющих на состояние здоровья человека. Например, при анализе данных о сердечно-сосудистых заболеваниях могут использоваться показатели артериального давления, уровня холестерина, индекса массы тела, возраста, пола, наличия вредных привычек и других параметров.

Важным направлением развития систем анализа многомерных данных является использование машинного обучения и искусственного интеллекта для автоматизации процесса обработки и интерпретации данных. Данные технологии дают возможность обучать алгоритмы на больших объемах исторических данных и затем применять их для прогнозирования исходов у новых пациентов. Например, нейронные сети используют для выявления ранних признаков онкологических заболеваний на основе результатов МРТ или КТ-исследований.

Еще одним перспективным направлением является применение методов визуализации многомерных данных, чтобы представить информацию в удобной для восприятия форме. Такие инструменты помогают врачам и исследователям лучше понимать структуру данных и выявлять важные паттерны, которые могут быть незаметными при традиционном анализе. Примером может служить создание интерактивных графиков и диаграмм, отображающих динамику изменения показателей здоровья пациента во времени.

О. В. Марухина, О. Г. Берестнева, Е. Е. Мокина исследовали возможность использования методов визуализации и анализа многомерных данных для оценки состояния биосистемы, особенно применительно к пациентам с

бронхиальной астмой. Современная медицина активно использует методы визуализации для анализа большого объема данных о состоянии здоровья пациентов, для лучшего понимания динамики изменений и выявления скрытых закономерностей. Основное внимание уделено разработке программного обеспечения на базе пиктографиков «Лица Чернова», которое помогает отображать функциональное состояние пациента через графический интерфейс. Разработка программы на основе пиктограмм «Лицо Чернова» для создания графического образа состояния пациента помогают отслеживать изменения в состоянии больного во времени, особенно при лечении различных форм астмы. Авторы утверждают, что визуализация и использование интегрированных индикаторов значительно улучшают процесс диагностики и мониторинга состояния пациентов [6]. Разработанный программный продукт может существенно облегчить работу врачей, предоставляя удобный инструмент для отслеживания динамики состояния пациентов с бронхиальной астмой.

И.А. Осадчая., О.Г. Берестнева, Е.В. Немеров считают, что визуальные методы обработки данных становятся очень важными в медицинской практике благодаря способности быстро и наглядно представлять сложные данные [7].

Таким образом, демонстрируется важность интеграции современных технологий визуализации в медицинскую практику для повышения эффективности диагностики и лечения заболеваний.

В исследовательской работе S.K. Goyal, S. Das, M. Lakshmi, подчеркивается растущее значение больших данных в здравоохранении, особенно в управлении медицинскими картами пациентов. Традиционные базы данных SQL, несмотря на свою мощность, часто не могут эффективно обрабатывать огромные, неструктурированные и быстро растущие объемы медицинских данных. Такое ограничение требует внедрения гибких и масштабируемых технологий баз данных, таких как NoSQL. По мнению автора MongoDB, популярная база данных NoSQL, предлагает отличное решение для решения задач, связанных с управлением медицинскими данными [1].

MongoDB - высокопроизводительная, документоориентированная база данных, которая предоставляет идеальную основу для разработки сложных систем анализа многомерных медицинских данных. Ее гибкая схема, масштабируемость и высокая производительность делают ее незаменимым инструментом для хранения и обработки больших объемов разнородных данных, характерных для медицинской области [2].

BSON (Binary JSON) - эффективный двоичный формат для хранения данных, который обеспечивает высокую скорость сериализации и десериализации, а также поддерживает спектр данных, включая числа, строки, массивы, объекты и даже двоичные данные. Также дает возможность хранить в MongoDB не только структурированные данные, но и медиафайлы, такие как медицинские изображения или аудиозаписи.

Модель документа в MongoDB естественным образом представляет сложные медицинские данные, такие как результаты анализов, истории болезни, изображения медицинских исследований [3]. Каждый пациент может быть представлен отдельным документом, содержащим множество вложенных документов, описывающих различные аспекты его здоровья. Такая организация данных упрощает поиск, фильтрацию и агрегацию информации, что особенно важно для анализа временных рядов, выявления корреляций между различными показателями и построения прогнозных моделей.

Выделим основные преимущества MongoDB для разработки систем анализа медицинских данных, которые указаны в табл. 1.

Таблица 1 - Основные преимущества MongoDB для разработки систем анализа медицинских данных [1,2,3,4]

Характеристика	Описание	Применение
Гибкая схема	Отсутствие строгой схемы легко адаптирует структуру базы данных под изменяющиеся требования медицинских исследований	Возможность добавлять новые поля и типы данных без миграции всей базы, что актуально при

	и клинической практики.	появлении новых медицинских технологий или стандартов.
Масштабируемость	MongoDB легко горизонтально масштабируется, чтобы обрабатывать огромные объемы данных, генерируемых современными медицинскими устройствами.	Эффективное хранение и анализ больших объемов геномных данных, медицинских изображений и данных электронных медицинских карт.
Высокая производительность	Оптимизированные алгоритмы поиска и индексирования обеспечивают быстрый доступ к данным, что критично для интерактивного анализа и визуализации.	Быстрая обработка запросов для поддержки принятия решений в режиме реального времени, например, при мониторинге состояния пациентов в интенсивной терапии.
Поддержка геопространственных данных	MongoDB имеет встроенные возможности для работы с геопространственными данными, чтобы анализировать данные о местоположении	Анализ распространения заболеваний, изучение факторов риска, связанных с географическим положением, и

	пациентов, эпидемиологические данные и т.д.	оптимизация размещения медицинских учреждений.
Агрегация данных	Механизм агрегации выполняет сложные расчеты, такие как среднее значение, стандартное отклонение, корреляция и другие статистические показатели, непосредственно в базе данных.	Вычисление статистических показателей для больших когорт пациентов, построение гистограмм и диаграмм рассеяния для визуализации данных.
Интеграция с другими системами	MongoDB легко интегрируется с другими системами и инструментами, чтобы создавать комплексные решения для анализа медицинских данных.	Интеграция с системами электронной истории болезни, лабораторными информационными системами, системами медицинского изображения и другими приложениями.
Поддержка JSON	MongoDB использует формат BSON (Binary JSON), что упрощает интеграцию с другими	Легкая интеграция с современными фронтенд-

	системами, использующими JSON, такими как веб-сервисы и приложения на JavaScript.	фреймворками и инструментами визуализации данных.
Безопасность	MongoDB предоставляет различные механизмы безопасности, такие как аутентификация, авторизация, шифрование данных, чтобы защитить конфиденциальность медицинских данных.	Соответствие требованиям регуляторных органов (например, HIPAA) и защита персональных данных пациентов.
Репликация и отказоустойчивость	MongoDB поддерживает репликацию и автоматическое восстановление при сбоях, что обеспечивает высокую доступность данных.	Минимизация рисков потери данных и обеспечение непрерывности работы медицинских систем.

Необходимо отметить, что MongoDB упрощает поиск и анализ данных, поскольку нет необходимости в сложных объединениях нескольких таблиц, как часто бывает в базах данных SQL [4]. Кроме того, бессхемный характер MongoDB обеспечивает большую гибкость при адаптации к изменяющимся структурам данных. По мере появления новых медицинских данных и изменения требований к данным схема базы данных может быть адаптирована без нарушения существующих приложений или целостности данных. Такая адаптивность имеет решающее значение в динамично развивающейся сфере здравоохранения, где постоянно внедряются новые медицинские технологии и методы лечения [5].

Чтобы обеспечить безопасность конфиденциальной информации о пациентах, авторы предлагают подход к токенизации. Путем преобразования конфиденциальных учетных данных, таких как имена пользователей, пароли и контактные данные, в значения ASCII, исходные данные затемняются, что снижает риск несанкционированного доступа и утечек данных. Процесс токенизации реализован с использованием мощных возможностей MongoDB по обработке запросов и агрегации. Способность MongoDB обрабатывать большие объемы неструктурированных данных, его гибкая схема и надежные средства защиты делают его перспективной технологией для разработки сложных систем анализа многомерных медицинских данных. Используя возможности MongoDB, организации здравоохранения могут получать ценную информацию из своих данных, улучшать уход за пациентами и продвигать медицинские исследования.

Итак, есть база данных MongoDB, содержащая медицинские записи пациентов. Каждая запись представляет собой документ, включающий следующие поля:

- patientID - уникальный идентификатор пациента;
- age - возраст пациента;
- gender - пол пациента;
- diagnosis – диагноз;
- medications - список принимаемых лекарств;
- lab\_results - результаты лабораторных анализов (например, уровень глюкозы, холестерина);
- visit\_dates (даты посещения врача).

Далее проводится анализ для выявления зависимости между уровнем глюкозы в крови и возрастом пациентов, страдающих диабетом второго типа.

Реализация на языке Python с использованием PyMongo

Python

```
from pymongo import MongoClient
```

```
# Подключение к базе данных MongoDB
```

```
client = MongoClient("mongodb://localhost:27017/")
```

```

db = client["medical_db"]
collection = db["patients"]
# Поиск пациентов с диагнозом "диабет второго типа"
pipeline = [
    {"$match": {"diagnosis": "диабет второго типа"}},
    {"$project": {"age": 1, "lab_results.glucose": 1}}
]
results = list(collection.aggregate(pipeline))
# Импорт необходимых библиотек для анализа данных
import pandas as pd
import matplotlib.pyplot as plt
# Преобразование результатов в DataFrame для удобства анализа
df = pd.DataFrame(results)
# Визуализация зависимости уровня глюкозы от возраста
plt.scatter(df['age'], df['lab_results.glucose'])
plt.xlabel('Возраст')
plt.ylabel('Уровень глюкозы')
plt.title('Зависимость уровня глюкозы от возраста у пациентов с диабетом
2 типа')
plt.show()
# Расчет коэффициента корреляции
correlation = df['age'].corr(df['lab_results.glucose'])
print("Коэффициент корреляции:", correlation)

```

Устанавливается соединение с базой данных MongoDB. Создается конвейер агрегации для поиска документов, соответствующих заданным критериям (пациенты с диагнозом "диабет второго типа").

Из выбранных документов извлекаются только необходимые поля: возраст и уровень глюкозы. Результаты агрегации преобразуются в DataFrame библиотеки Pandas для удобной обработки данных. Далее строится график рассеяния для визуального представления зависимости между возрастом и

уровнем глюкозы. Вычисляется коэффициент корреляции Пирсона для количественной оценки связи между двумя переменными.

MongoDB является инструментом для разработки систем анализа многомерных медицинских данных, предоставляя гибкость, производительность и масштабируемость, необходимые для решения сложных задач в области здравоохранения. Ее использование помогает создавать инновационные решения, способствующие улучшению диагностики, лечения и профилактики заболеваний.

В заключение следует отметить, что системы анализа многомерных данных играют ключевую роль в развитии современной медицины, обеспечивая возможность понимания механизмов возникновения и развития заболеваний, а также создания эффективных стратегий профилактики и лечения. Продолжающееся совершенствование таких технологий открывает новые возможности в области медицины и улучшения качества медицинской помощи.

#### **Источники:**

1. Goyal S.K., Das S., Lakshmi M. Effective handling of patient health records using NoSQL: MONGODB. Research Journal of Pharmaceutical, Biological and Chemical Sciences. 2016. Vol. 7. 134-138
2. Microsoft SQL Server [Электронный ресурс]. URL: <https://www.microsoft.com/en-us/sql-server/sql-server-2022>
3. MongoDB Features [Электронный ресурс]. URL: <https://www.mongodb.com/what-is-mongodb/features>
4. Why Use MongoDB and When to Use It? [Электронный ресурс]. URL: <https://www.mongodb.com/why-use-mongodb>
5. Иванов С.И. Программное обеспечение для учета и хранения клинической и социодемографической информации о больных / С.И. Иванов, Н.В. Тарутина, М.А. Голубчиков, Р.Р. Сафаров // Программные продукты и системы. 2015. №3 (111). С. 220-225.

6. Марухина, О. В. Визуализация и анализ многомерных экспериментальных данных / О. В. Марухина, О. Г. Берестнева, Е. Е. Мокина // Информационные и математические технологии в науке и управлении. – 2016. – № 4-2. – С. 90-95.

7. Осадчая И.А., Берестнева О.Г., Немеров Е.В. Анализ многомерных медицинских данных с помощью пиктографиков «Лица Чернова» // Бюллетень сибирской медицины. 2014. №4. С. 89-93.