

ПРАКТИЧЕСКИЕ МЕТОДЫ СТРУКТУРИРОВАНИЯ И ВИЗУАЛИЗАЦИИ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ АВТОМАТИЗИРОВАННЫХ ИНФОРМАЦИОННЫХ РЕШЕНИЙ

PRACTICAL METHODS OF STRUCTURING AND VISUALIZING DATA USING AUTOMATED INFORMATION SOLUTIONS

*E. Salnikov
L. Demidov*

Summary. This article discusses the promising application of information systems for structuring and visualizing data. The existing methods and technologies underlying such systems are analyzed, as well as the possibilities, problems and prospects of their application are considered.

Keywords: information systems, structured data, data visualization tools and methods.

Сальников Евгений Анатольевич

Доцент-практик, Финансовый университет при
правительстве РФ
EASalnikov@fa.ru

Демидов Лев Николаевич

Канд. техн. наук, доцент, Финансовый университет
при правительстве РФ
LDemidov@fa.ru

Аннотация. В данной статье рассматривается перспективное применение информационных систем для структурирования и визуализации данных. Анализируются существующие методы и технологии, находящиеся в основе таких систем, а также рассматриваются возможности, проблемы и перспективы их применения.

Ключевые слова: информационные системы, структурированные данные, средства и методы визуализации данных.

Введение

В последние десятилетия информационные системы стали неотъемлемой частью повседневной жизни человека, их применение распространилось на многие отрасли, включая экономику, науку, образование и здравоохранение. Одной из ключевых областей применения информационных систем является структурирование и визуализация данных. Использование таких систем позволяет эффективно организовать большие объемы данных, а также визуально представить информацию, что облегчает ее понимание и анализ.

В условиях развития цифровых технологий и современной промышленности объемы информации, которые генерируются производственными процессами, становятся колоссальными. Для того, чтобы эффективно управлять производством и оптимизировать его, необходимо иметь возможность собирать, обрабатывать и структурировать эти данные. Существует множество различных методов и инструментов, позволяющих работать с данными и получать при этом ценную информацию.

В данной статье рассмотрим некоторые из методы структурирования и обработки данных, которые ис-

пользуются в промышленных информационных решениях.

Основная часть

Структурирование и визуализация данных являются ключевыми методами, которые позволяют упорядочить данные и предоставить их удобно для анализа. В этой статье мы рассмотрим различные методы структурирования и обработки данных и их применение в научных исследованиях.

Структурирование данных — это процесс преобразования информации в такой формат, в котором данные легко вычисляются и анализируются. Необходимость в этом возникает, когда, например, данные поступают в неструктурированном виде, например, в текстовом формате или изображении.

Одним из методов структурирования данных является использование базы данных.

Базы данных позволяют хранить и регулировать большие объемы информации, обеспечивать быстрый доступ к ним и возможность их анализа. Базы данных представляют собой собрание связанных между собой данных, которые хранятся в организованном виде.

Базы данных могут быть реляционными и нереляционными. Реляционные базы данных организованы в виде таблиц, где каждая строка представляет отдельную запись, а каждый столбец — отдельное свойство этой записи. Нереляционные базы данных представлены, например, в виде документов или графов. Существует множество систем управления базами данных, таких как MySQL, PostgreSQL и Oracle.

Еще одним методом структурирования является использование средств языков программирования, таких как Python или R. Например, библиотека pandas на языке Python позволяет работать с данными в формате таблиц, удобных для анализа и визуализации.

Информационные системы предоставляют множество инструментов для обработки данных. Один из основных инструментов — это анализ данных, позволяющий извлекать ценную информацию из больших объемов информации и использовать ее для принятия решений. Одним из методов обработки данных является машинное обучение.

Машинное обучение — это метод, который позволяет компьютерной программе изучать данные и делать прогнозы на их основе. Машинное обучение используется для решения задач классификации, регрессии, кластеризации, обработки изображений и текстов и т.д. В промышленности машинное обучение может использоваться для оптимизации производственных процессов, определения неисправностей и т.д. Одним из примеров использования машинного обучения в промышленных информационных решениях является прогнозирование отказов оборудования. Для этого используется модель машинного обучения, которая основывается на исторических данных об отказах оборудования и предсказывает вероятность отказа в будущем. Таким образом, можно предпринять меры по предотвращению отказов остановок оборудования и снижению рисков.

Другим методом обработки данных является обработка естественного языка (Natural Language Processing, NLP), которая позволяет компьютеру анализировать тексты на естественном языке и извлекать из них информацию. В промышленности NLP может использоваться для анализа отзывов пользователей о продукции, мониторинга социальных сетей на предмет упоминаний о бренде и т.д.

Визуализация данных — это процесс представления данных в графическом виде для более наглядного и понятного анализа. Визуализация данных позволяет проводить наблюдения, выявлять тенденции и аномалии в данных, что делает ее незаменимым инструментом для научных исследований.

Существует множество методов визуализации с целью дальнейшего анализа данных. Гистограммы, диаграммы рассеяния, круговые диаграммы и линейные графики — это только некоторые из наиболее часто встречающихся результатов визуализации данных. Каждый из них имеет свои преимущества и может быть использован в зависимости от типа данных и целей исследования. Например, гистограммы — это графическое представление данных по универсальному признаку. Они часто используются для анализа частоты и дисперсии данных. Диаграммы рассеяния, в свою очередь, используются для отображения своей взаимосвязи между двумя переменными и позволяют выявлять корреляцию между ними.

Для визуализации данных существует множество инструментов, которые позволяют создавать интерактивные дашборды и отчеты, позволяющие быстро и удобно анализировать данные.

Новые методы исследования данных, такие как виртуальная и дополненная реальность, также получают все большее распространение в науке. Они позволяют создавать интерактивные данные, которые требуют более глубокого восприятия и взаимосвязи между данными.

Структурирование и визуализация данных являются частью современной науки и бизнеса. Они позволяют упорядочить, отрегулировать и проанализировать огромные объемы данных, что делает их более доступными и регламентированными для принятия решений. Развитие искусственного интеллекта и машинного обучения при постоянном увеличении объемов информации требует новых инструментов и методов для анализа и выявления данных.

В настоящее время существует множество информационных систем, которые используются для структурирования и визуализации данных. Одним из примеров таких систем является Microsoft Excel. Эта система используется для организации и обработки данных, а также для создания графиков и диаграмм. Однако, Excel не всегда является эффективным инструментом для работы с данными, особенно если речь идет о больших объемах информации.

Для работы с большими объемами данных больше подходят специализированные информационные системы, например, Tableau или QlikView. Эти системы позволяют быстро и удобно организовать данные и создать интерактивные диаграммы и визуализации, облегчающие анализ и понимание информации. Они также позволяют проводить анализ данных в режиме реального времени и получать быстрые результаты.

Кроме того, существуют системы, специализирующиеся на конкретных типах данных. Например, система GenBank используется для хранения генетических данных, а система PubMed — для организации медицинских публикаций. Эти системы предоставляют удобный интерфейс для работы с соответствующими типами данных и облегчают их организацию и практическое использование.

В последние годы стали популярными системы машинного обучения, о которых упоминалось выше. Эти системы позволяют автоматически обрабатывать большие объемы информации и выявлять скрытые закономерности в данных. Примерами таких систем являются TensorFlow и Apache Spark. TensorFlow — это открытая программная библиотека для глубокого машинного обучения. Она разработана компанией Google Brain и предоставляет инструменты для создания и обучения нейронных сетей. TensorFlow может применяться для решения различных задач, таких как классификация, кластеризация и регрессия. Одним из преимуществ использования TensorFlow является его гибкость и масштабируемость. TensorFlow может работать на различных платформах, включая процессоры и графические процессоры (GPU). TensorFlow может быть также использован для создания различных моделей нейронных сетей, включая сверточные нейронные сети, рекуррентные нейронные сети и раскрытие нейронных сетей.

Apache Spark — это высокопроизводительная система обработки данных с получением исходного кода. API-интерфейс Spark работает с большими объемами данных, которые распределены по нескольким узлам кластера. Spark поддерживает различные языки программирования, такие как Java, Scala, Python и R. Одно из преимуществ использования Apache Spark — его высокая вероятность увеличения объема данных на кластерах из локальных узлов. Spark может использоваться для решения различных задач обработки данных, таких как сортировка, объединение и агрегация. Spark также поддерживает множество структур данных, таких как RDD (Resilient Distributed Datasets), DataFrame и DataSet.

Одним из примеров применения информационных систем для структурирования и визуализации данных является использование системы Google Analytics для анализа веб-трафика и поведения пользователей на сайте. Эта система позволяет визуально представлять данные о количестве и социальном статусе посетителей, времени пребывания на сайте, их поведении и многом другом. Таким образом, система помогает оптимизировать веб-сайт и улучшить пользовательский опыт.

Еще один пример — система SAP Business Objects, которая предназначена для создания отчетов и аналитических приложений для бизнес-анализа. Эта система позволяет управлять данными из различных источников, объединять их и создавать отчеты и диаграммы, которые помогают принимать бизнес-решения на основе данных.

Использование информационных систем для структурирования и визуализации данных имеет несколько преимуществ. Прежде всего, это позволяет облегчить работу с большими объемами данных, которые необходимо организовать и проанализировать. Во-вторых, это позволяет представлять данные в удобной и понятной форме, что облегчает их понимание и анализ. В-третьих, это позволяет автоматизировать процессы обработки данных и ускорить получение результатов анализа.

Однако, использование информационных систем для структурирования и визуализации данных не лишено ограничений.

Одним из основных ограничений является необходимость определения структуры данных заранее. Когда структура данных известна, то трудностей с настройкой информационной системы для структурирования и визуализации данных нет. Однако, если структура данных неизвестна или изменяется со временем, то это может создать проблемы в обработке и визуализации данных.

Еще одним ограничением визуализации могут быть слишком большие объемы данных для обработки в режиме online.

Многие пользователи не имеют знаний и опыта работы с БД, SQL запросами и другими технологиями, необходимыми для работы с информационными системами. Необходимость в специальной подготовке персонала также является одним из основных ограничений для использования, описанных выше систем и методов для структурирования и визуализации данных.

Кроме того, для создания эффективной информационной системы требуется значительное количество времени и ресурсов. Настройка системы может занять несколько месяцев, что является существенным ограничением для организаций, нуждающихся в быстром доступе к своим данным.

Еще одним ограничением в использовании информационных систем является необходимость действовать в соответствии с законодательными и регуляторными требованиями. Некоторые организации имеют

специфические требования к хранению и обработке данных, такие как HIPAA, PCI DSS, и другие. Применение информационных систем в соответствии этим требованиям может привести к дополнительным расходам.

Несмотря на эти ограничения, информационные системы могут быть очень полезными для организаций, нуждающихся в структурировании и визуализации данных, и широко применяются на практике.

Решения и примеры

В современном мире, где объемы данных растут с каждым днем, структурирование и визуализация данных являются важными аспектами для обеспечения эффективного управления информацией и, как следствие, всеми процессами, в которых эта информация используется. Существует множество инструментов и технологий для структурирования и визуализации данных, таких как базы данных, графические инструменты, BI-системы, визуализаторы данных и многое другое.

Кроме рассмотренных выше примеров с базами данных, широко применяемыми являются графические инструменты, такие как Microsoft Visio, которые позволяют создавать диаграммы и схемы для визуализации данных. Они часто используются для создания планов зданий, схем производственных процессов и т.д.

BI-системы (бизнес-интеллект) также широко используются для структурирования и визуализации данных. Они позволяют компаниям и организациям анализировать свои данные и принимать более эффективные решения на их основе. Примерами BI-систем являются Tableau, Power BI и Qlik.

Наконец, визуализаторы данных, такие как D3.js и Plotly, дают возможность создавать интерактивные графики и визуализации больших объемов данных и использовать их для анализа.

Для упрощения процесса структурирования и обработки данных могут применяться промышленные информационные решения.

Промышленные информационные решения — это программное обеспечение, которое предназначено для управления, мониторинга и анализа данных в промышленных средах. Это программное обеспечение может выполнять такие функции, как управление производственными процессами, мониторинг энергопотребления и управление качеством продукции.

Одним из наиболее распространенных промышленных информационных решений является система

управления производственными процессами (SCADA). Это программное обеспечение позволяет операторам контролировать производственные процессы, мониторить параметры производства и получать данные о состоянии оборудования.

Другим примером промышленного информационного решения является система управления энергопотреблением (EMS). EMS используется в промышленных средах. Это программное обеспечение позволяет контролировать и оптимизировать потребление энергии, управлять нагрузками и мониторить потребление энергии в реальном времени.

Еще одним примером промышленного информационного решения является система управления качеством продукции (QMS). Это программное обеспечение позволяет контролировать процессы производства, мониторить параметры качества продукции и управлять процессами исправления и предупреждения дефектов.

Промышленные информационные решения могут использоваться для упрощения процесса структурирования и обработки данных в промышленных средах и предоставляют операторам и менеджерам инструменты для контроля и управления всеми производственными процессами, мониторинга энергопотребления и управления качеством продукции.

Примером использования информационных систем для структурирования и визуализации данных является Google Analytics. Этот инструмент позволяет отслеживать посещаемость сайта, поведение пользователей и другие параметры, используя различные типы данных, такие как текстовые, графические и числовые. Google Analytics также предоставляет возможность анализировать данные и строить графики и диаграммы, что упрощает процесс принятия решений.

Для примера рассмотрим применение BI-системы Tableau для структурирования и визуализации данных. Tableau позволяет быстро создавать интерактивные дашборды, графики и отчеты на основе данных из различных источников, таких как базы данных, Excel-файлы, файлы CSV и многое другое. Подобные системы используются для представления данных в виде графиков, диаграмм и других форм, которые позволяют проще понимать и анализировать информацию. Также с помощью Tableau пользователи могут строить графики, диаграммы и другие визуальные представления данных, а также создавать интерактивные дашборды, которые обеспечивают являются более удобным и понятным способом анализа данных.

Информационные системы для структурирования и визуализации данных широко применяются также в области социальных наук. В данном случае, с помощью информационных систем исследователи могут анализировать такие большие объемы данных как опросы и социологические исследования, и создавать их визуализации, чтобы лучше понимать результаты исследований.

Например, для этого можно использовать инструмент Power BI. Визуализация данных на его основе поможет выявлению связей между различными факторами исследования.

Одним из наиболее интересных примеров использования информационных систем для структурирования и визуализации данных является исследование геномов различных видов. Например, база данных NCBI GenBank содержит миллионы последовательностей геномов, доступных для общественного использования. С помощью специализированных информационных систем и инструментов, таких как BLAST и CLC Genomics Workbench, исследователи могут быстро и эффективно проводить анализ геномных данных и создавать их визуализации. Также существует возможность анализировать мутации генома и изучать их взаимодействие с окружающей средой, например, с помощью программного обеспечения R.

В настоящее время информационные системы для структурирования и визуализации данных широко используются в научных исследованиях, финансовых анализах, маркетинге, медицине и многих других областях деятельности. Они могут быть полезными в различных областях, включая науку и образование, например, для анализа результатов научных исследований, визуализации данных в преподавании различных предметов и для обеспечения доступа к большим объемам знаний для студентов и ученых.

Однако, применение информационных систем в науке и образовании также связано с рядом проблем. Одной из основных проблем является отсутствие стандартизации данных. Различные научные и образовательные учреждения часто используют различные форматы данных, что затрудняет их совместное использование и обработку.

Кроме того, многие информационные системы для структурирования и визуализации данных требуют больших современных вычислительных ресурсов и специальных знаний для их разработки и использования. На данный момент это является препятствием для их широкого распространения и использования в науке и образовании.

Перспективы развития

В ближайшем будущем ожидается дальнейшее развитие информационных систем для структурирования и визуализации данных.

Одним из основных направлений развития является разработка систем, которые могут работать с огромным объемом данных и обеспечивать быстрый доступ к информации. Ожидается также, что информационные системы для структурирования и визуализации данных будут все больше использоваться в науке и образовании. Это будет способствовать развитию новых подходов к анализу данных и новым методам обучения.

Современные информационные системы для структурирования и визуализации данных продолжают развиваться и улучшаться, чтобы удовлетворять все более сложные требования пользователей.

Вот некоторые из перспективных направлений развития:

- ◆ Использование искусственного интеллекта для автоматической обработки и анализа данных.
- ◆ Развитие технологий виртуальной и дополненной реальности для создания более удобных и интерактивных способов визуализации данных.
- ◆ Интеграция различных видов данных из разных источников, включая данные из интернета вещей (IoT) и социальных медиа.
- ◆ Развитие систем, которые позволяют пользователям работать с данными из разных мест, включая удаленный доступ и работу с мобильных устройств.

Современные информационные системы для структурирования и визуализации данных уже достаточно развиты и широко применяются в различных областях. Однако с развитием технологий появляются новые возможности для оптимизации и совершенствования этих систем. В ближайшем будущем можно ожидать дальнейшего развития информационных систем для структурирования и визуализации данных.

Развитие технологий искусственного интеллекта позволит автоматизировать процессы обработки данных. Еще одной из перспективных областей развития является использование искусственного интеллекта и машинного обучения для автоматической классификации и анализа данных. Это позволит ускорить процесс обработки и анализа больших объемов данных, а также повысить точность получаемых результатов.

Другой перспективой является улучшение визуализации данных. Сегодня уже существует достаточно мно-

го различных инструментов, которые позволяют визуализировать данные, но в будущем можно ожидать еще большего развития этой области, и не только самой визуализации данных, но и их интерпретации.

Еще одной перспективной областью является разработка более удобных и эргономичных интерфейсов для работы с информационными системами. Новые технологии, такие как виртуальная и дополненная реальность, могут значительно улучшить визуализацию данных и упростить их анализ.

Еще одной перспективной областью является использование информационных систем для создания интеллектуальных аналитических инструментов, которые могут использоваться в различных сферах деятельности, таких как медицина, банковское дело, наука и т.д. Эти инструменты будут обладать высокой степенью точности и могут помочь в принятии важных решений.

Заключение

Применение информационных систем для структурирования и визуализации данных является перспективной областью развития в ближайшем будущем. Это связано с появлением и дальнейшей разработкой новых технологий, таких как искусственный интеллект

и машинное обучение, виртуальная и дополненная реальность

Информационные системы для структурирования и визуализации данных становятся все более важным инструментом для работы с большим объемом данных, которые генерируются в современном мире. Они необходимы в науке, бизнесе, образовании и всех прочих сферах деятельности. Их развитие позволит получить новые методы анализа данных и обучения.

Однако их применение также связано с рядом проблем, включая отсутствие стандартизации данных и требование больших вычислительных ресурсов. Стоит отметить, что успешное применение информационных систем для структурирования и визуализации данных требует не только развития технологий, но также и стандартизации данных, обучения пользователей и организации доступа к информации. Для достижения этих целей требуется сотрудничество различных научных и образовательных учреждений, а также государственных органов.

В целом, информационные системы для структурирования и визуализации данных имеют значительный потенциал для науки и образования, и их дальнейшее развитие может привести к новым открытиям и достижениям в этих областях.

ЛИТЕРАТУРЫ

1. Авдеев А.Ю. «Автоматизация информационного обеспечения управления предприятиями». Москва: Инфра-М, 2017.
2. Кобзарь, А.И. Введение в анализ данных с использованием R. М.: ООО «Бином», 2015.
3. Богданова Н.В. «Методы анализа и интерпретации данных». Москва: Издательский дом «ЛКИ», 2018.
4. Ларионова Е.Ю., Блинов В.И. «Программное обеспечение для визуализации данных: обзор рынка и анализ существующих решений». Информатика и ее применения, 2019, т. 13, № 4, с. 76–83.
5. Смольянинова Ю.Ю. «Использование методов структурирования и визуализации данных в управлении проектами». Экономические науки, 2018, т. 23, № 2, с. 86–92.
6. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer, 2016.
7. Abdi, H. "Visualization of Multivariate Data Using R and Exploratory Data Analysis (EDA)". CRC Press, 2018.
8. Keim, D.A., Andrienko, G., Fekete, J.D., et al. "Visual Analytics: Definition, Process, and Challenges". Springer, 2008.
9. Witten, I.H., Frank, E., Hall, M.A. "Data Mining: Practical Machine Learning Tools and Techniques". Morgan Kaufmann, 2016.
10. K. Cios, W. Pedrycz, and R. Swiniarski, Data Mining Methods for Knowledge Discovery, Springer, 2018.
11. J. Heer and M. Bostock, "Designing and Building Data Visualizations," ACM Transactions on Computer-Human Interaction, vol. 20, no. 3, article 21, Sep. 2013.
12. Tableau Software, "Tableau: Business Intelligence and Analytics Software," [Online]. Available: <https://www.tableau.com/>. [Accessed: 11-Mar-2023].