

Введение. В современном мире, где управленческие решения в сфере государственного управления всё чаще принимаются на основе анализа больших объёмов данных, качество этих данных становится критически важным фактором. Статья посвящена комплексному анализу методов оценки качества данных в секторе государственного управления, выявлению основных проблем и предложению путей их решения. Основное внимание уделено аспектам, таким как точность, актуальность, полнота и доступность данных. Рассматриваются как теоретические подходы к оценке качества данных, так и практические аспекты их применения в государственных органах.

Основная часть. В рамках исследования рассмотрена декларативная система оценки качества данных на основе стандартов ISO (BIGQA), которая облегчает планирование и выполнение операций по оценке качества больших данных для бизнес-пользователей на любом этапе жизненного цикла данных. Были оценены результаты экспериментов на двух различных наборах данных, демонстрирующие способность BIGQA генерировать планы оценки качества, которые могут быть эффективно реализованы в любом контексте больших данных и на любой параллельной вычислительной платформе для создания настраиваемых отчетов о качестве данных. Эксперименты показали, что производительность оценки качества данных 1 ГБ файла улучшилась на 71% при параллельном выполнении на одной машине и на 75% при оценке 25 ГБ файла в распределенной среде. Кроме того, было продемонстрировано влияние инкрементной оценки качества данных, показывающее значительное снижение времени и потребления ресурсов, тем самым решая проблемы больших объёмов данных [1].

В другой статье [2] рассматривается оценка качества данных с использованием методологии ГОСТ и основанных на них метрик. Исследование охватывает анализ мер, связанных с внутренним и контекстным качеством данных, и подчеркивает их значимость в процессах оценки в различных областях. Представлены наиболее распространенные индикаторы качества, и предлагаются рекомендации по их применению для повышения эффективности управления данными в организациях и предприятиях.

Р.А. Ешенко и Г.А. Гурвиц рассматривают профайлинг данных как средство для комплексной оценки качества информации [3]. Анализируются типичные ошибки в данных, их влияние на процессы принятия решений и методы их устранения. Подчеркивается значимость происхождения данных и их структуры, а также представлены способы выявления и коррекции "грязных" данных с помощью визуальных методов и автоматизированных инструментов профайлинга. Особое внимание уделено визуализации ошибок и анализу распределения значений для идентификации аномалий и несоответствий. Работа подчеркивает необходимость глубокого понимания данных для повышения их качества и обеспечения надежности бизнес-анализа.

Выводы. Разработана система оценки качества данных, которая учитывает ограничения сектора государственного управления в России, в том числе требования ГОСТ, и может быть использована для повышения эффективности государственного управления за счёт улучшения процессов обработки и анализа данных.

Список использованных источников:

1. Fadlallah H., Kilany R., Dhayne H., El Haddad R. BIGQA: Declarative Big Data Quality Assessment // Journal of Data and Information Quality – 2023. – № 15(3). – С. 24–29.

2. Байшев А.В. Характеристики качества данных // Вестник Тувинского государственного университета. Технические и физико-математические науки. – 2023. – № 1 – С. 3–7.

3. Ешенко Р.А., Гурвиц Г.А. Оценка качества данных по происхождению, профайлинг данных // Вестник Хабаровского государственного университета экономики и права. – 2017. – № 6. – С. 2–5.