

# ВЛИЯНИЕ ЗАДВОЕННОСТИ ЗАПИСЕЙ НА КАЧЕСТВО ИНФОРМАЦИИ В ПРОФЕССИОНАЛЬНЫХ СИСТЕМАХ

**Лобанов Олег Андреевич**

Аспирант, преподаватель,  
Московский государственный технологический  
университет СТАНКИН, г. Москва  
lobanov98oleg@mail.ru

## THE IMPACT OF DUPLICATE RECORDS ON INFORMATION QUALITY IN PROFESSIONAL SYSTEMS

**O. Lobanov**

*Summary.* The article investigates the problem of unintentional data duplication, known in professional circles as duplicate records. Unlike managed duplication, which is used to ensure fault tolerance, duplicate records are a destructive artifact that arises during the stages of data collection, integration, and processing. The primary focus is on the consequences of duplication for information quality management in professional fields. Duplicate records can lead to decreased information quality, increased storage and processing costs, and errors in analytics and decision support. With the growing operational volume of data processing in complex systems, existing solutions fail to fully address the occurrence of such problems and have not offered a universal solution that would guaranteed eliminate this phenomenon or reduce the associated risks to a negligible level. Practical case studies demonstrate how this anomaly leads to significant operational costs, systemic errors in analytical processes, and the risk of incorrect management decisions. It is emphasized that existing approaches are reactive and do not ensure complete problem resolution. The necessity of developing proactive approaches to prevent this anomaly at the systemic level is substantiated.

*Keywords:* database, duplicate records, data duplication, information quality, information reliability, data quality.

*Аннотация.* В статье исследуется проблема непреднамеренного дублирования данных, известная в профессиональной среде как задвоенность записей. В отличие от управляемого дублирования, используемого для обеспечения отказоустойчивости, задвоенность является деструктивным артефактом, возникающим на этапах сбора, интеграции и обработки информации. Основное внимание уделено последствиям задвоенности для управления качеством информации в сферах профессиональной деятельности. Задвоенные записи могут приводить к снижению качества информации, увеличению затрат на хранение и обработку данных, а также к ошибкам в аналитике и поддержке принятия решений. С увеличением операционного объема обработки данных в сложных системах, существующие решения не позволяют полностью справиться с возникновением таких проблем и не предложили универсального решения, которое бы гарантированно устранило данный феномен или снизило сопутствующие риски до уровня «можно пренебречь». На практических кейсах продемонстрировано, как данная аномалия приводит к существенным операционным затратам, системным ошибкам в аналитических процессах и рискам принятия неверных управленческих решений. Подчеркивается, что существующие подходы носят реактивный характер и не обеспечивают полного устранения проблемы. Обоснована необходимость разработки проактивных подходов к предотвращению аномалии на системном уровне.

*Ключевые слова:* база данных, задвоенность записей, дублирование данных, качество информации, достоверность информации, качество данных.

### Введение

В современной научной и профессиональной литературе отсутствует единство терминов для описания феномена непреднамеренного повторения данных. В рамках данной статьи предлагается следующая терминологическая дифференциация, основанная на критериях функциональности:

- Дублирование данных — управляемая избыточность, целенаправленно создаваемая для повышения отказоустойчивости и производительности (например, репликация, резервное копирование).
- Задвоенность данных — непреднамеренный артефакт, возникающий вследствие ошибок интеграции, сбоев процессов или человеческого фактора, приводящий к снижению качества информации.

- Дубликат — факт наличия повторяющейся записи, нейтральный термин.

Данное исследование фокусируется на проблеме задвоенности как деструктивном явлении, нарушающем целостность данных.

С увеличением объема и сложности баз данных вероятность возникновения задвоенной записи только возрастает. Несмотря на многочисленные попытки устранить эту проблему поиском уже образованных дублей, эта проблема продолжает существовать и наносить ущерб бизнес-процессам. В материалах исследования приводятся ссылки на статьи со сферами деятельности с упоминанием задвоенной записи, некорректности работы с этими записями и появлением ошибок. Задвоенность может возникать на этапах сбора, интеграции

и хранения информации. Данная аномалия негативно сказывается на качестве данных, что влияет на формирование достоверных аналитических отчетов и на эффективность бизнес-процессов.

### Задвоенная запись в различных сферах деятельности

Задвоенная запись (задвоенность) проявляется при учете материально-производственных запасов на предприятии, например, стивидорного бизнеса. Стивидорный бизнес — деятельность компаний, которые занимаются погрузкой и разгрузкой товаров в портах, а также размещением грузов на судне. Функциональные простои в компании ведут к крупным финансовым потерям. Доктор экономических наук, профессор А.Ю. Алексеев приводит в своей статье, что риск задвоенности на участке внутрифирменного перемещения запасов остатков уменьшается за счет учета запасов, а также оборотов по ним [1].

Как отмечает А.Н. Михайлов, одной из причин хаоса в персональных данных является проблема с задвоенностью. Он предложил проводить контроль дублирования допустимых столбцов и значений, а также предлагает внедрить общие стандарты и прозрачность процессов, чтобы каждый участник знал, где найти нужную информацию [2].

Также про системы управления взаимоотношениями с клиентами пишет в своей статье доктор технических наук, профессор В.А. Докучаев. Он, рассматривая проблемы актуализации данных в CRM-системах, заявляет, что для уменьшения ошибок из-за задвоенности требуется регулярно проводить проверки и очистки данных [3].

В государственном лесном реестре и едином государственном реестре недвижимости порождается ряд ошибок из-за задвоенных записей в базе данных. В первую очередь это приводит к неактуальности такой информации о лесном фонде страны. В едином государственном реестре недвижимости и государственном лесном реестре существует разобщенность сведений, в результате возникает опасность отчуждения земель лесного фонда в земли иных категорий. Разница этих сведений составляет 300 миллионов гектаров, некоторые эти несовпадения вызваны, в том числе из-за дублирования данных [4].

Проблемы с задвоением информации в базах данных встречаются во всех крупных организациях, в том числе и в государственных системах. Одним из таких примеров можно привести: возникновение дублей в Едином государственном реестре прав на недвижимость (ЕГРН). Они проявились на этапе объединения двух сложных государственных систем ЕГРН и Государственного кадастра недвижимости, в том числе из-за ошибок в документах-основаниях, из которых были получены данные [5].

В статье А.А. Алимова «Преимущества и недостатки программного обеспечения 1С бухгалтерия» приводится некорректная работа программ 1С, в том числе среди них можно обнаружить задвоенность информации, которая приводит к неверным результатам. В процессе генерации бухгалтерского баланса в программном комплексе 1С могут возникать случаи задвоения счетов, приводящие к некорректности формируемой отчетности. Для устранения данного недостатка необходимо выполнение дополнительных ручных проверок и корректировок, что существенно увеличивает трудозатраты [6].

В программе 1С, как и в любой системе управления баз данных, задвоенности могут возникнуть при получении документов из разных источников, при изменении у поставщиков названий номенклатур в документе, при орфографической небрежности записи и других причин. Проблема может возникать даже из-за несоответствия номенклатуры от разных поставщиков, причиной этому — отсутствие обязательной законодательной регламентации стандартов на корректность изначальной записи о продукте.

Также при машинном обучении для качественной работы интеллектуальной системы нужно в первую очередь проанализировать данные, чтобы исключить задвоенную информацию, которая может повлиять на конечное решение обученной модели.

### Качество информации при задвоенности

Одним из показателей низкого качества данных является их задвоенность, ведущая в итоге к ошибочной интерпретации одного и того же объекта как несколько разных [7].

Задвоенные записи приводят к противоречивым данным, затрудняющим определение актуальной информации, увеличению нагрузки на систему из-за выполнения лишних операций. При этом анализ данных с дубликатом снижает качество полученных статистических данных, а также возникает вопрос: насколько такая информация достоверна, и можно ли на ее основе предлагать дальнейшие действия или использовать эти выводы. Невольное использование ошибочных и недостоверных данных в аналитических процессах создает существенные риски принятия неоптимальных управленческих решений, что может повлечь за собой к серьезным операционным издержкам и стратегическим просчетам.

Проблему качества информации в задвоенных записях поднимал доктор технических наук, профессор В.В. Бураков в научной работе на тему: «Контекстно зависимый способ поиска нечетких дубликатов в реляционных базах данных» и писал, что одним из критериев низкого качества данных является дубли в базе данных,

что приводит к ошибочной интерпретации одного и того же объекта, как несколько разных [7].

Задвоенность записи нарушает требования большинства критериев качества информации, такие как: соответствие реальным значениям и отсутствие расхождения с текущим состоянием в настоящее время, отсутствие конфликтующих, противоречивых данных и событий, а также нарушение принципа единственной записи — уникальности данных.

### Существующее решение в 1С

Яркой иллюстрацией ограниченности современных подходов служит встроенный инструментарий в системах семейства 1С. Он предоставляет механизм для исправления задвоенности после её обнаружения пользователем, но не включает алгоритмов для предотвращения аномалии на этапах ввода и интеграции первичных данных. Это подтверждает тезис о том, что существующие решения носят большей частью реактивный, борясь со следствиями, а не с причинами возникновения дефектов данных.

### Выводы

Несмотря на множественные попытки решить проблему задвоенности записи, она существует во всех предметных областях, где собирается информация для применения в дальнейшей работе. В любой систе-

ме управления базой данных пользователям рано или поздно приходится сталкиваться с задвоенными записями, что может привести к некорректным решениям для любой поставленной задачи.

Важно отметить, что термин задвоенность, не будучи формально закрепленным в нормативной документации, утвердился в профессиональном лексиконе как яркий пример эволюции языка снизу. Его метафорическая природа (от «двоиться в глазах») и четкая негативная коннотация делают его эффективным инструментом коммуникации среди практиков

Сейчас проблему с задвоенностью решают только после ее обнаружения с помощью специализированных модулей. Система управления базами данных должна не допускать дубли записей и связей на этапе ввода новой информации или изменений предыдущих записей. Для этого необходимо проводить контроль дублей допустимых столбцов и значений, а также внедрить общие стандарты и прозрачность процессов, чтобы каждый участник знал, где найти нужную информацию.

Полноценная и достоверная информация нужна для всех областей, начиная от крупного производства, научных познаний и заканчивая малым бизнесом. В основном из-за того, что везде используются технологии хранения данных при помощи существующих систем управления базами данных, которые самостоятельно никак не решают проблему возникновения аномалий.

### ЛИТЕРАТУРА

1. Жминько С.И., Алексеенко А.Ю. Бухгалтерский учет материально-производственных запасов на предприятиях стивидорного бизнеса // Все для бухгалтера. 2010. №1.
2. Михайлов А.Н. Dataops в CRM: как наладить поток данных для ML и персонализации без хаоса // Вестник науки. 2025. №4 (85).
3. Докучаев В.А., Маклачкова В.В., Бойко А.А. Проблема актуализации данных в CRM-системах // Экономика и качество систем связи. 2025. №35.
4. Бородин О.Б., Чуксин И.В., Фомина А.В., Желонкина Е.Э. Государственный лесной реестр как информационная база в системе устойчивого управления землями лесного фонда Российской Федерации // Московский экономический журнал. 2022. №4.
5. Вавулинская Д.Д., Миклашевская О.В. Информационное обеспечение деятельности Росреестра по вопросам верификации и гармонизации сведений ФГИС ЕГРН // Пространственные данные: наука и технологии. 2022. № 13. С. 5–18. DOI:10.30533/scidata-2022-13-01.
6. Алимов А.А. Преимущества и недостатки программного обеспечения «1С бухгалтерия» // Форум молодых ученых. 2019. №6 (34).
7. Тарасов С.В., Бураков В.В. Контекстно зависимый способ поиска нечётких дубликатов в реляционных базах данных // Информационно-управляющие системы. 2015. №2 (75).

© Лобанов Олег Андреевич (lobanov98oleg@mail.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»