

РАЗРАБОТКА МЕТОДИКИ УПРАВЛЕНИЯ КАЧЕСТВОМ ДАННЫХ НА ОСНОВЕ ЛОГИКО-СЕМАНТИЧЕСКОГО ЦИФРОВОГО ДВОЙНИКА

Т.В. Павлович¹, tv_pavlovich@mail.ru

Е.А. Дронь², elena_dron@bk.ru, <https://orcid.org/0009-0004-2487-0876>

Г.Г. Куликов³, grisha@molniya-ufa.ru

¹ МИРЭА – Российский технологический университет, Москва, Россия

² Уфимский университет науки и технологий, Уфа, Россия

³ АО «Уфимское научно-производственное предприятие «Молния», Уфа, Россия

Аннотация. Качество данных – ключевой вопрос в системах поддержки принятия решений. Использование же данных, содержащих одну или несколько логических ошибок, может привести к неправильным решениям и большим потерям для предприятия. **Цель исследования.** Целью исследования является разработка классификации данных, учитывающих профиль производства, типы данных на производстве, формальную структуру бизнес-процессов для каждого типа данных на основе предлагаемой логико-семантической архитектуры цифрового двойника для контроля качества данных. **Теоретико-методологической основой исследования** являются труды отечественных и зарубежных ученых по проблемам управления цифровыми платформами и цифровой трансформацией и цифровых двойников предприятия. **Результаты.** В ходе исследования была обобщена и систематизирована классификация основных логико-семантических факторов некачественного предоставления данных, представлена методика оценки качества данных. Бизнес-процессы сбора, хранения и обработки формализованы для каждого типа данных и определено одно ответственное лицо для соответствующих бизнес-процессов, а также разработаны требования к компетентности для отдельных ответственных лиц и инструмент для контроля качества данных. В результате предлагается подход на основе логико-семантической архитектуры цифрового двойника для контроля качества данных. **Заключение.** В современных реалиях необходима логическая фильтрация огромного количества избыточных данных в системах поддержки принятия решений (ППР). Цифровые двойники (ЦД) в системах поддержки принятия решений могут обеспечить высокое качество исходных цифровых данных для ППР. В этой статье представлен комплексный подход к управлению качеством данных на основе их логико-семантических цифровых двойников при создании эффективной системы поддержки принятия решений на примере нефтегазовых компаний. Управление качеством данных на основе применения логико-семантических цифровых двойников обеспечивает полезность цифровых данных.

Ключевые слова: управление качеством данных, большие данные, цифровые двойники, цифровая трансформация

Для цитирования: Павлович Т.В., Дронь Е.А., Куликов Г.Г. Разработка методики управления качеством данных на основе логико-семантического цифрового двойника // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». 2023. Т. 23, № 2. С. 111–118. DOI: 10.14529/ctcr230210

DEVELOPMENT OF A DATA QUALITY MANAGEMENT METHODOLOGY BASED ON LOGICAL-SEMANTIC DIGITAL DOUBLE

T.V. Pavlovich¹, tv_pavlovich@mail.ru

E.A. Dron², elena_dron@bk.ru, <https://orcid.org/0009-0004-2487-0876>

G.G. Kulikov³, grisha@molniya-ufa.ru

¹ MIREA – Russian Technological University, Moscow, Russia

² Ufa University of Science and Technology, Ufa, Russia

³ JSC “Ufa Scientific and Production Enterprise “Molniya”, Ufa, Russia

Abstract. Data quality is a key issue in decision support systems. The urgency of the problem is due to the fact that the introduction of corporate information systems, digital technologies and digital twins requires strict synchronization and updating of a large amount of data. Management decisions depend on big data. The use of data containing one or more logical errors can lead to incorrect decisions and large losses for the enterprise. **Aim.** The aim of the study is to develop a classification of data that takes into account the production profile, types of data in production, the formal structure of business processes for each type of data based on the proposed logical-semantic architecture of the digital twin for data quality control. **The theoretical and methodological basis of the study** are the works of domestic and foreign scientists on the problems of managing digital platforms and digital transformation and digital twins of the enterprise. **Results.** In the course of the study, the classification of the main logical-semantic factors of poor-quality data provision was generalized and systematized, and a methodology for assessing data quality was presented. Business processes for collecting, storing and processing are formalized for each type of data and one responsible person is identified for the relevant business processes, as well as competency requirements for individual responsible persons and a tool for data quality control are developed. As a result, an approach based on the logical-semantic architecture of the digital twin for data quality control is proposed. **Conclusion.** In modern realities, logical filtering of a huge amount of redundant data in decision support systems (DSS) is necessary. Digital twins (DTs) in decision support systems can provide high quality of initial digital data for PPR. This article presents an integrated approach to data quality management based on their logical-semantic digital twins when creating an effective decision support system using the example of oil and gas companies. Data quality management based on the use of logical-semantic digital twins ensures the usefulness of digital data.

Keywords: data quality management, big data, digital twins, digital transformation

For citation: Pavlovich T.V., Dron E.A., Kulikov G.G. Development of a data quality management methodology based on logical-semantic digital double. *Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics*. 2023;23(2):111–118. (In Russ.) DOI: 10.14529/ctcr230210

Введение

Цифровые двойники – это цифровые копии предприятий, бизнес-процессов или отдельных физических объектов. Цифровой двойник – это инструмент для решения проблем управления эффективностью бизнеса, выбора наилучших способов развития предприятия для обеспечения выживания и благополучия компаний на высоко конкурентных рынках [1, 2].

Качество данных напрямую определяет ценность применения цифрового двойника и качество бизнес-решений.

В этой статье представлен комплексный подход к управлению качеством данных на основе их логико-семантических цифровых двойников при создании эффективной системы поддержки принятия решений на примере нефтегазовых компаний.

Анализ основных факторов качества данных

Согласно стандарту ISO 9000: 2015 основными критериями качества данных являются их полнота, надежность, точность, согласованность, доступность и своевременность.

В дальнейшем бизнес-процессы сбора, хранения и обработки необходимо формализовать для каждого типа данных и определить одно ответственное лицо (SPA) для соответствующих бизнес-процессов, а также разработать требования к компетентности для отдельных ответственных лиц и инструмент для контроля качества данных (рис. 2).



Рис. 2. Этапы внедрения методики управления качеством данных
Fig. 2. Stages of implementing data quality management techniques



Рис. 3. Архитектура целостности элементов качества данных
Fig. 3. Data quality element integrity architecture

Этот инструмент представляет собой контрольный список для отдельных ответственных лиц, который включает проверку пяти факторов. Эти факторы были определены как противоположность ошибок данных и недостатков данных.

Контрольный список для проверки качества данных показан на рис. 3. Один ответственный человек сравнивает полученные данные по пяти факторам. В результате аудита может оказаться, что отсутствует один или несколько факторов. В этом случае уточняются недостающие факторы. Только после прояснения каждого фактора единственное ответственное лицо имеет право передавать эти проверенные данные вышестоящему руководству для принятия деловых решений.

Методика управления качеством данных с применением цифровых двойников в системах поддержки принятия решений в нефтегазовых компаниях

Использование умных скважин повысило точность прогноза дебита скважин на 15–20 % с учетом меняющихся условий добычи. Кроме того, виртуальное тестирование режимов работы

скважин позволяет избежать экспериментов с реальными объектами. Это также позволяет специалистам создать несколько сценариев и выбрать лучший по заданным критериям. Вместе с тем в ряде случаев наблюдались серьезные ошибки прогнозов дебитов нефти и выбора режимов работы погружного оборудования. Эти ошибки увеличили эксплуатационные расходы и затраты на подъем, затраты на геологические и технические мероприятия, а также на ремонт и техническое обслуживание оборудования.

Анализ причин этих ошибок показал проблемы с качеством данных. Данные, которые не соответствуют хотя бы одному из пяти элементов качества данных, являются некондиционными, или мы можем сказать, что эти данные содержат минусы. Все это создало предпосылки для уточнения методики управления качеством данных применительно к цифровым аналогам и интеллектуальным полям с целью повышения эффективности поддержки принятия решений. В основе методики лежит следующая идея: качество анализа данных зависит от знания и понимания идеальной картины разработки месторождения, а также целевых показателей (например, по снижению темпов падения добычи). Это означает, что ответственный руководитель должен знать, как должна выглядеть сфера деятельности, в данном случае развитие сферы с рациональной и логической точки зрения [3–6].

Если это условие выполнено, то необходимо выполнить следующие два шага: анализ данных, анализ текущей ситуации.

В методологии управления качеством данных ситуация – это общая картина с определенным объемом данных [7, 8]. Данные – это факты, диаграммы, утверждения, решения, действия, описания, которые предлагаются как правильные [9, 10]. Минусы – это любые отдельные данные, которые предлагаются как верные, но оказываются нелогичными в результате их сравнения с 5 элементами. Плюсы – это достоверные данные, которые оказываются верными в результате сравнения их с теми же 5 элементами.

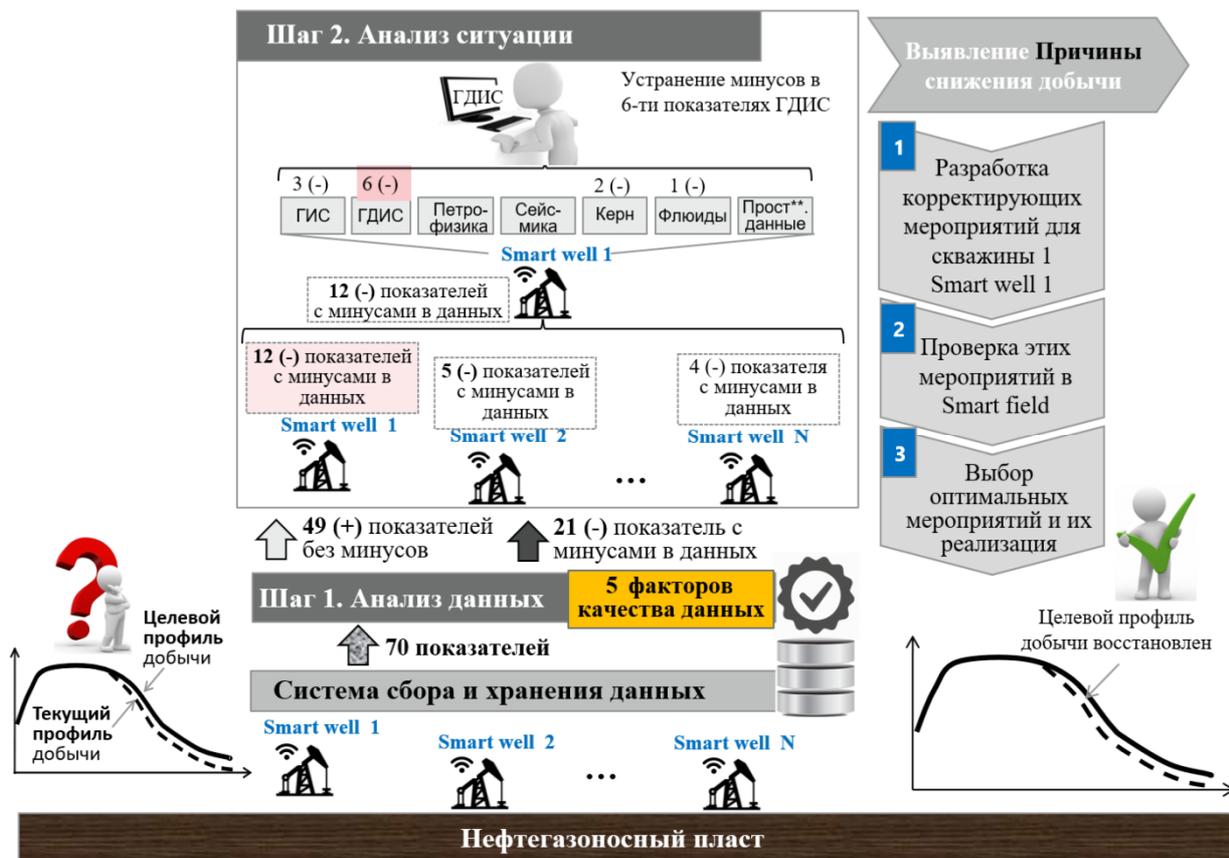


Рис. 4. Ключевые этапы методологии управления качеством данных с применением факторов логико-семантического цифрового двойника
Fig. 4. Key stages of the data quality management methodology using the factors of the logical-semantic digital twin

На первом этапе качество данных оценивается путем сравнения его с 5 элементами. В результате этого шага могут быть обнаружены данные, содержащие минусы, т. е. некачественные данные.

Анализ текущей ситуации на шаге два заключается в том, что данные с минусами сортируются по участкам месторождения или по компонентам скважины. Область с наибольшим количеством данных с минусами становится основной целью коррекции (рис. 4).

Ситуационный анализ относится к распределению данных с минусами по единицам целого [11]. В результате ситуационного анализа становится понятно, какой участок необходимо исправить и в чем причина недостижения запланированных показателей и целей [12]. Таким образом, созданная система поддержки принятия решений на первом этапе выдает ответственному специалисту по разработке месторождения список данных с минусами или ошибками. В результате реализации второго шага ответственное лицо информируется о проблемной зоне, которой соответствует наибольшее количество данных с минусами.

Ненадежные данные приводят к неправильным решениям и сводят на нет эффект от использования интеллектуальных систем, таких как интеллектуальный анализ данных, машинное обучение, прогнозная аналитика и т. д. [13–15]. Управление качеством данных на основе применения логико-семантических цифровых двойников обеспечивает полезность цифровых данных. Ключевым фактором успеха цифровой трансформации компаний являются подходы, обеспечивающие качество корпоративных данных. Для этого компании внедряют новые бизнес-процессы, назначают ответственных людей и обучают персонал управлению качеством данных, поскольку надежность данных, их доступность и согласованность в различных ИТ-системах требуют комплексного подхода.

Список литературы

1. Антонов В.В., Куликов Г.Г. Семантико-математический язык описания структуры интеллектуальной системы на основе нечеткой логики // Программные продукты и системы. 2011. № 3. С. 33–35.
2. Юсупова Н.И., Еникеева К.Р. Интеллектуальная информационная поддержка принятия решений при анализе рисков чрезвычайных ситуаций и управлении ими. М.: Машиностроение, 2014. 206 с.
3. Миронов М.В., Головкин Ю.Б., Юсупова Н.И. Об алгоритмах управления по состоянию ситуации // Управление сложными техническими системами: межвуз. науч. сб. Уфа: УАИ, 1985. № 8. С. 107–111.
4. Антонов В.В., Куликов Г.Г. Метод построения математической модели предметной области // Вестник Самарского государственного экономического университета. 2010. № 5 (67). С. 10–14.
5. Подход к применению концепции цифровых двойников для трансформации корпоративной информационной системы под требования INDUSTRY 4.0 (на примере создания единого информационного пространства «вуз – предприятие») / Куликов Г.Г., Сапожников А.Ю., Кузнецов А.А. и др. // Вестник УГАТУ. 2019. Т. 23, № 4 (86). С. 154–160.
6. Концепция системного представления предметной области при формировании цифрового двойника производственного процесса машиностроительного предприятия / А.В. Речалов, А.В. Артюхов, Г.Г. Куликов, В.Н. Новиков // Вестник УГАТУ. 2022. Т. 26, № 1 (95). С. 120–135. DOI: 10.54708/19926502_2022_26195120
7. Юсупова Н.И., Богданова Д.Р., Бойко М.В. Обработка слабоструктурированной информации на основе методов искусственного интеллекта. М.: Инновационное машиностроение, 2016. 256 с.
8. ПНСТ 429–2020. Умное производство. Двойники цифровые производства. Часть 1. Общие положения. М.: Стандартиформ, 2020. 8 с.
9. Никулина Н.О., Шилина М.А., Кондратьева О.В. Проектирование автоматизированных информационных систем управления процессами: учеб. пособие. Уфа: УГАТУ, 2018. 375 с.
10. Методы и технологии проектирования программных средств для организации обработки данных / В.В. Антонов, Г.Г. Куликов, М.А. Шилина и др. Уфа: УГАТУ, 2018. 219 с.
11. Структурирование информационного пространства с использованием процессного подхода и семантической идентификации / Г.В. Старцев, М.А. Шилина, А.А. Бармин, О.В. Бармина // Экономика, статистика и информатика. Вестник УМО. 2014. № 3. С. 203–207.

12. Основы концепции онтологического моделирования бизнес-процессов для задач принятия решений / К.А. Конев, В.В. Антонов, Д.А. Ризванов и др. // Современные наукоемкие технологии. 2020. № 12-1. С. 71–77. DOI: 10.17513/snt.38413

13. Конев К.А. Принятие решений в сложных социально-экономических системах // Методы менеджмента качества. 2018. № 1. С. 30–36.

14. Использование BI систем в оперативном управлении производством / Т.К. Гиндуллина, М.С. Демченко, В.Н. Капустин, А.Н. Гаффарова // 90 лет УГАТУ на службе науке, образованию и бизнесу: материалы Всерос. науч.-практ. конф. Уфа: УУНиТ, 2022. С. 21–23.

15. Гелисханов И.З., Юдина Т.Н., Бабкин А.В. Цифровые платформы в экономике: сущность, модели, тенденции развития // Научно-технические ведомости СПбГПУ. Экономические науки. 2018. Т. 11, № 6. С. 22–36. DOI: 10.18721/JE.11602

References

1. Antonov V.V., Kulikov G.G. Semantical-mathematical language of the description of structure of intellectual system on the basis of fuzzy logic. *Software & Systems*. 2011;3:33–35. (In Russ.)

2. Yusupova N.I., Enikeeva K.R. *Intellektual'naya informatsionnaya podderzhka prinyatiya resheniy pri analize riskov chrezvychaynykh situatsiy i upravlenii imi* [Intellectual information support for decision-making in the analysis of risks of emergency situations and their management]. Moscow: Mashinostroyeniye; 2014. 206 p. (In Russ.)

3. Mironov V.V., Golovkin Yu.B., Yusupova N.I. [On control algorithms based on the state of the situation]. *Upravleniye slozhnymi tekhnicheskimi sistemami: mezhvuzovskiy nauchnyy sbornik* [Control of complex technical systems: interuniversity scientific collection]. Ufa: Ufa Aviation Institute. 1985;8:107–115. (In Russ.)

4. Antonov V.V., Kulikov G.G. The method of constructing the mathematical model of subject domain. *Vestnik Samara state university of economics*. 2010;5(67):10–14. (In Russ.)

5. Kulikov G.G., Sapozhnikov A.Yu., Kuznetsov A.A., Mavrina A.S., Zagidullin D.I. Software implementation of the expert system functions to control of design documentation in UIS engineering enterprises. *Vestnik UGATU*. 2019;23(4(86)):154–160. (In Russ.)

6. Rechkalov A.V., Artuhov A.V., Kulikov G.G., Novikov V.N. The concept of transformation of the model of planning and management processes based on the digital twin of the production system in the industrial model of a machine-building enterprise. *Vestnik UGATU*. 2022;26(1(95)):120–135. (In Russ.) DOI: 10.54708/19926502_2022_26195120

7. Yusupova N.I., Bogdanova D.R., Boyko M.V. *Obrabotka slabostrukturirovannoy informatsii na osnove metodov iskusstvennogo intellekta* [Processing of weakly structured information based on artificial intelligence methods]. Moscow: Innovatsionnoye mashinostroyeniye; 2016. 256 p. (In Russ.)

8. *PNST 429–2020. Umnoye proizvodstvo. Dvoyniki tsifrovyye proizvodstva. Chast' 1. Obshchiye polozheniya* [Preliminary National Standard 429–2020. Smart manufacturing. Digital manufacturing twins. Part 1. General principles]. Moscow: Standartinform; 2020. 8 p. (In Russ.)

9. Nikulina N.O., Shilina M.A., Kondrat'yeva O.V. *Proyektirovaniye avtomatizirovannykh informatsionnykh sistem upravleniya protsessami: ucheb. posobiye* [Designing secure information systems for process management: textbook]. Ufa: USATU; 2018. 375 p. (In Russ.)

10. Antonov V.V., Kulikov G.G., Shilina M.A. et al *Metody i tekhnologii proyektirovaniya programnykh sredstv dlya organizatsii obrabotki dannykh* [Methods and technologies for designing software tools for organizing data processing]. Ufa: USATU; 2018. 219 p. (In Russ.)

11. Startsev G.V., Shilina M.A., Barmin A.A., Barmina O.V. Structuring of information space of technical university using process approach and semantic identification. *Ekonomika, statistika i informatika. Vestnik UMO*. 2014;3:203–207. (In Russ.)

12. Konev K.A., Antonov V.V., Rizvanov D.A., Suleymanov A.K., Bakusova N.S. Fundamentals of the concept of ontological modeling of business processes for decision-making tasks. *Modern high technologies*. 2020;12-1:71–77. (In Russ.) DOI: 10.17513/snt.38413

13. Konev K.A. [Decision making in complex socio-economic systems]. *Methods of Quality Management*. 2018;1:30–36. (In Russ.)

14. Gindullina T.K., Demchenko M.S., Kapustin V.N., Gaffarova A.N. [The use of BI systems in the operational management of production]. In: *90 let UGATU na sluzhbe nauke, obrazovaniyu i*

business: materialy Vserossiyskoy nauchno-prakticheskoy konferentsii [90 years of USATU in the service of science, education and business: Proceedings of the All-Russian scientific and practical conference]. Ufa: Ufa University of Science and Technology; 2022. P. 21–23. (In Russ.)

15. Geliskhanov I.Z., Yudina T.N., Babkin A.V. Digital platforms in economics: essence, models, development trends. *St. Petersburg State Polytechnical University Journal. Economics*. 2018;11(6):22–36. (In Russ.) DOI: 10.18721/JE.11602

Информация об авторах

Павлович Татьяна Вячеславовна, канд. техн. наук, доц. кафедры бизнес-информатики, МИРЭА – Российский технологический университет, Москва, Россия; tv_pavlovich@mail.ru.

Дронь Елена Анатольевна, канд. техн. наук, доц. кафедры автоматизированных систем управления, Уфимский университет науки и технологий, Уфа, Россия; elena_dron@bk.ru.

Куликов Григорий Геннадьевич, технический директор, АО «Уфимское научно-производственное предприятие «Молния», Уфа, Россия; grisha@molniya-ufa.ru.

Information about the authors

Tatyana V. Pavlovich, Cand. Sci. (Eng.), Ass. Prof. of the Department of Business Informatics, MIREA – Russian Technological University, Moscow, Russia; tv_pavlovich@mail.ru.

Elena A. Dron, Cand. Sci. (Eng.), Ass. Prof. of the Department of Automated Control Systems, Ufa University of Science and Technology, Ufa, Russia; elena_dron@bk.ru.

Grigory G. Kulikov, Technical Director, JSC “Ufa Scientific and Production Enterprise “Molniya”, Ufa, Russia; grisha@molniya-ufa.ru.

Статья поступила в редакцию 20.09.2022

The article was submitted 20.09.2022