



УДК 004.89

DOI 10.52575/2687-0932-2022-49-3-566-574

## **Использование онтологического подхода для извлечения ожиданий к качеству данных корпоративных хранилищ**

**Афанасьев А.А., Кудинов В.А.**

Курский государственный университет,  
Россия, 305000, Курская область, г. Курск, ул. Радищева, д. 33  
E-mail: afanasyevkursk@gmail.com, kudinovva@yandex.ru

**Аннотация.** Важную роль в процессах интеллектуального анализа данных, построения корпоративных хранилищ данных и разработки информационно-аналитических систем играет качество данных. Статья посвящена постановке и обоснованию актуальности проблемы извлечения ожиданий к качеству данных корпоративных хранилищ. Несмотря на многочисленность публикаций, посвященных проблеме извлечения экспертных знаний, в области качества данных (Data quality) и управления данными в целом (Data governance) данная проблема слабо развита. Целью исследования является определение возможности расширения существующей проблемы извлечения экспертных знаний предлагаемой для дальнейших исследований проблемой извлечения ожиданий к качеству данных корпоративных хранилищ, а также разработка онтологической модели извлечения ожиданий к качеству данных корпоративных хранилищ. Исследование проводилось на основе онтологического подхода. В результате исследования была обоснована необходимость централизованного сбора требований к качеству данных организации, поставлена проблема извлечения ожиданий к качеству данных корпоративных хранилищ путем расширения существующей проблемы извлечения экспертных знаний, а также разработана онтологическая модель извлечения ожиданий к качеству данных корпоративных хранилищ.

**Ключевые слова:** онтология предприятия, онтологическое моделирование, экспертные системы, проблема извлечения знаний, качество данных

**Для цитирования:** Афанасьев А.А., Кудинов В.А. 2022. Использование онтологического подхода для извлечения ожиданий к качеству данных корпоративных хранилищ. Экономика. Информатика, 49(3): 566–574. DOI 10.52575/2687-0932-2022-49-3-566-574

---

## **Using an Ontological Approach to Extract Expectations About the Data Quality of Enterprise Data Warehouses**

**Alexander A. Afanasiev, Vitaly A. Kudinov**

Kursk State University  
33 Radishcheva St, Kursk, Kursk region, 305000, Russia  
E-mail: afanasyevkursk@gmail.com, kudinovva@yandex.ru

**Abstract.** An important role in the processes of data mining, building corporate data warehouses and developing information and analytical systems is played by data quality. The article is devoted to the formulation and substantiation of the relevance of the problem of extracting expectations for the quality of corporate storage data. Despite the numerous publications devoted to the problem of extracting expert knowledge, in the field of data quality (Data quality) and data governance in general (Data governance), this problem is underdeveloped. The aim of the study is to determine the possibility of expanding the existing problem of extracting expert knowledge by the problem of extracting expectations for the quality of data of corporate storages, proposed for further research, as well as developing an ontological model for extracting expectations for the quality of data of corporate storages. The study was carried out on the basis of the ontological approach. As a result of the study, the need for a centralized collection of requirements for the

quality of data of an organization was substantiated, the problem of extracting expectations for the quality of data of corporate storages was posed by expanding the existing problem of extracting expert knowledge, and an ontological model for extracting expectations for the quality of data of corporate storages was developed.

**Keywords:** enterprise ontology, ontological modeling, expert systems, knowledge extraction problem, data quality

**For citation:** Afanasiev A.A., Kudinov V.A. 2022. Using an Ontological Approach to Extract Expectations About the Data Quality of Enterprise Data Warehouses. Economics. Information technologies, 49(3): 566–574 (in Russian). DOI 10.52575/2687-0932-2022-49-3-566-574

## Введение

На сегодняшний день одним из активно развивающихся направлений деятельности предприятий является цифровизация. Цифровизация сфер деятельности предприятий ведет к увеличению количества используемых информационных систем, их развитию и усложнению, что способствует увеличению массивов хранимой информации. Этот факт ставит перед предприятиями задачу по управлению качеством данных в корпоративных хранилищах, для выполнения которой необходимо сформировать понимание о том, какие данные следует считать качественными.

Целью исследования является определение возможности расширения существующей проблемы извлечения экспертных знаний предлагаемой для дальнейших исследований проблемой извлечения ожиданий к качеству данных корпоративных хранилищ, а также разработка онтологической модели извлечения ожиданий к качеству данных корпоративных хранилищ.

Для достижения цели исследования необходимо выполнить следующие задачи:

1. Установить место и роль качества данных в аналитических методологиях и процессах интеллектуального анализа данных;
2. Установить место и роль качества данных при построении корпоративных хранилищ данных и разработке информационно-аналитических систем;
3. Обосновать необходимость централизованного сбора требований к качеству данных организации;
4. Определить возможность расширения существующей проблемы извлечения экспертных знаний предлагаемой для дальнейших исследований проблемой извлечения ожиданий к качеству данных корпоративных хранилищ;
5. Разработать онтологическую модель извлечения ожиданий к качеству данных корпоративных хранилищ.

Новизна проведенного исследования состоит в обосновании актуальности проблемы извлечения ожиданий к качеству данных корпоративных хранилищ и ее постановке с помощью расширения существующей проблемы извлечения экспертных знаний.

Теоретическое применение полученных результатов состоит в установлении места и роли качества данных в аналитических методологиях и процессах интеллектуального анализа данных, а также при построении корпоративных хранилищ данных и разработке информационно-аналитических систем, и обоснованию акцента именно на извлечении знаний как процессе выявления знаний, когда источником информации являются эксперты в проблемной области, а в качестве методов извлечения знаний рассматривать в первую очередь активные индивидуальные и групповые методы.

Практическое применение полученных результатов состоит в возможности использования разработанной онтологической модели извлечения ожиданий к качеству данных корпоративных хранилищ для построения онтологической модели данных, с помощью которой проектируются базы знаний и базы данных.



## **Место и роль качества данных в корпоративном управлении данными и их анализе**

Выявлению проблем с качеством данных уделяется внимание в методологиях интеллектуального анализа данных: KDD (шаг «Очистка данных и предварительная обработка») [Fayyad, Piatetsky-Shapiro, Smyth, 1996] и CRISP-DM (фаза «Понимание данных») [Chapman, Clinton, Kerber, Khabaza, Reinartz, Shearer, Wirth, 2000]. Также вопросы качества данных рассматриваются в таких методологиях, как CASP-DM (фаза «Понимание данных», задача «Проверить качество данных») [Martinez-Plumed, Ochando, Ferri, Flach, Hernandez-Orallo, Kull, Lachiche, Ramirez-Quintana, 2017], ASUM-DM (фаза «Анализ-Проектирование-Настройка и Сборка», активность «Подготовка данных», задача «Очистить данные») [IBM Analytics Solutions Unified Method (ASUM)], RAMSYS (фаза «Понимание данных», задача «Проверка качества данных») [Moyle, Jorge, 2001], которые усовершенствовали выделенные в CRISP-DM фазы интеллектуального анализа данных [Martinez-Plumed et al. 2021].

Исследователи обращают внимание на важность работ в области качества данных при построении корпоративных хранилищ данных [Килимова, 2022], разработке информационно-аналитических систем [Любицын, 2012]. А. Д. Килимова утверждает, что особую роль в мониторинге больших данных играет оценка и поддержание их качества, так как именно эта задача влияет на достоверность и корректность этих данных, следовательно, и на качество принимаемых на их основе управленческих решений [Килимова, 2022]. Так, В.Н. Любицын утверждал, что обеспечение требуемого качества данных, используемых в информационно-аналитической системе (ИАС) любого вида и назначения, почти всегда является одной из ключевых проблем создания подобной системы [Любицын, 2012].

В общем виде важность определения качества данных в корпоративном управлении данными и их анализе обуславливается принципом в информатике GIGO (garbage in – garbage out). Принцип «мусор на входе – мусор на выходе» основан на выводе, что хорошие входные данные обычно приводят к хорошим выходным результатам, и плохие входные данные приводят к плохим результатам на выходе. Хотя концепция «мусор на входе – мусор на выходе» была известна во времена Чарльза Беббиджа (1864) и даже раньше, первым этот термин ввел Джордж Фьючел, программист, использовавший GIGO как обучающий метод в конце 1950-х годов [Лидвелл, Холден, Батлер, 2021]. Этот термин чаще всего используется в контексте разработки алгоритмов и программного обеспечения, но на сегодняшний день он актуален и при корпоративном управлении данными (Data Governance) и их анализе (Data Mining, Business intelligence), так как в исследовании, сделанном KPMG в 2019 году, было отмечено, что 69 % опрошенных руководителей крупнейших российских компаний из ключевых отраслей экономики планируют внедрить в ближайшие два года технологии на основе анализа больших данных (big data) и предиктивной аналитики [Цифровые технологии в российских компаниях, 2019]. Так, использование некачественных данных при их анализе (garbage in) может привести к ошибочным результатам анализа (garbage out), что в дальнейшем станет причиной принятия неверных управленческих решений.

### **Обоснование необходимости централизованного сбора требований к качеству данных организации**

На сегодняшний день измерения (в других источниках критерии, параметры) качества данных довольно хорошо изучены. Одним из наиболее популярных источников методической информации в области управления и в частности качества данных является Data Management Body of Knowledge [DAMA-DMBOK, 2020] от международной организации DAMA International, которая изучает вопросы в области управления данными. Другим источником, в котором уделяется внимание к качеству данных промышленных предприятий, является Международные рекомендации по статистике промышленности ООН [Международные рекомендации по статистике промышленности, 2010]. В 2019 году Аналитический центр при Правительстве РФ презентовал собственную методику оценки и повышения каче-

ства данных [Методология оценки и повышения качества данных, 2019] в рамках национальной программы «Цифровая экономика». Также существует методическая документация отраслевых ассоциаций, изучающих организацию управления данными в конкретной отрасли экономики. Например, в нефтяной промышленности такой организацией является Professional Petroleum Data Management (PPDM) Association [Data Rules Program Strategy: International Petroleum Data Standards, 2018].

На рисунке 1 демонстрируется, что в изученных методологиях находятся общие измерения качества данных. Тем не менее, несмотря на разработанность измерений качества данных, актуальны вопросы реализации проверок на соответствие конкретных данных измерениям: какие данные требуют проверки, как сформулировать задачу проверки для инженера-разработчика, как отображать результаты проверок качества данных потребителю данных, заинтересованному в их качестве.

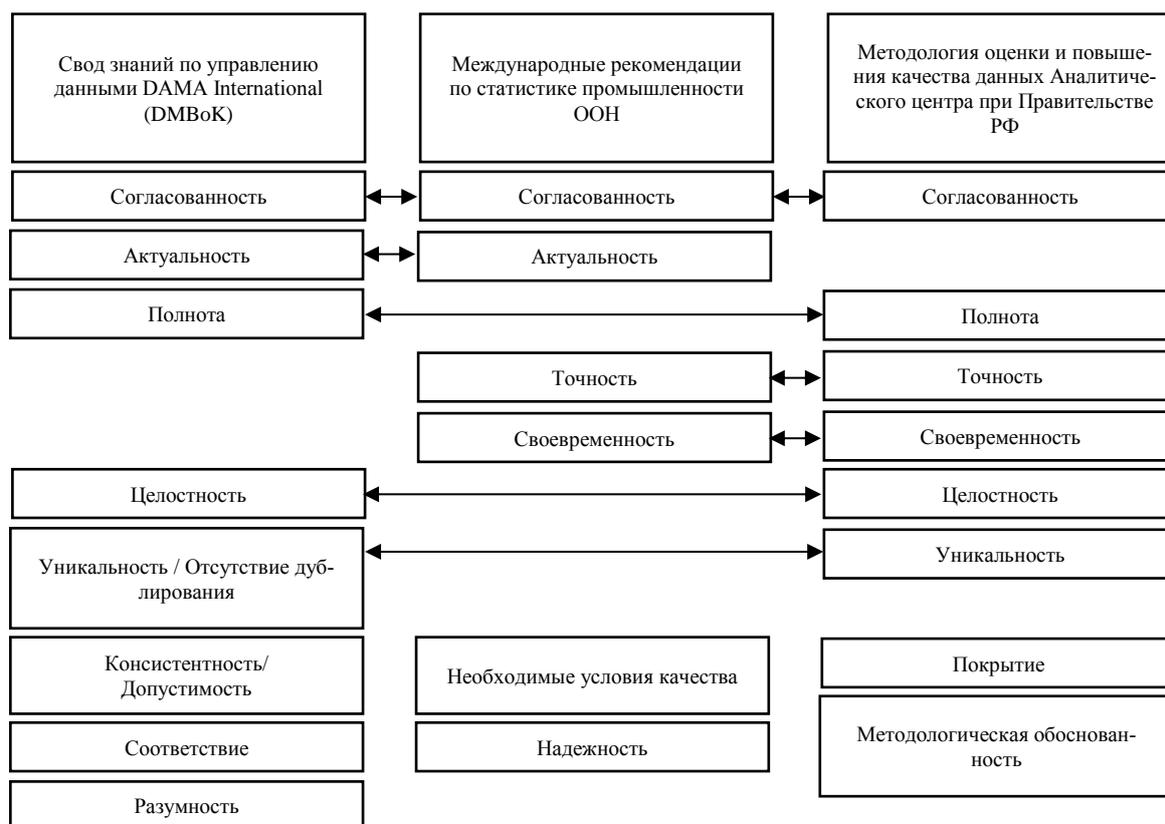


Рис. 1. Измерения (критерии, параметры) качества данных в стандартах и методологических источниках

Fig. 1. Measurements (criteria, parameters) of data quality in standards and methodological sources

В.Н. Любицын [Любицын, 2012] отмечает, что набор факторов, влияющих на качество данных, весьма разнообразен и может в процессе эксплуатации информационно-аналитических систем меняться. Поэтому, по его мнению, формированию и систематической модификации методики оценки качества, используемой для анализа данных, необходимо уделять большое внимание, поскольку именно она является основной для выбора места и технологии доведения качества данных до требуемого в конкретной ситуации уровня.

Давид Лошин [Loshin, 2010] утверждал, что изучение существующей документации дает только статическую картину того, что может быть (а может и не быть) верным в отношении состояния среды данных. Более полную картину можно составить, собрав то, что может считаться «веским доказательством», от ключевых лиц, связанных с бизнес-процессами, использующими данные. Поэтому на следующем этапе предложенного им процесса анализа требований к данным проводится интервью с ранее определенными ключевыми заинтересо-



ванными сторонами, отмечаются их критические области, вызывающие обеспокоенность, и обобщаются эти опасения, чтобы выявить пробелы, которые необходимо заполнить в форме требований к данным.

Данные мнения в том числе подтверждают и сложность формирования требований к качеству данных от потребителей данных, использующих их в рамках своих направлений профессиональной деятельности, так и необходимость использовать активные индивидуальные и групповые методы извлечения знаний, например, интервьюирование.

Таким образом, одной из первых актуальных задач процесса управления качеством данных является сбор требований к качеству данных организации. Согласно ГОСТ Р ИСО 8000-2-2019, под требованием к качеству данных понимается потребность или ожидание, которое установлено, предполагается или является обязательным, а под качеством данных понимается степень соответствия совокупности присущих характеристик объекта требованиям [ГОСТ Р ИСО 8000-2-2019, 2019]. Выполнение этой задачи позволит определить ожидания сотрудников организации, которым должны соответствовать данные.

Подобное мнение высказывает А. Д. Килимова: «Необходимо централизовать процесс управления качеством данных и создать главный его документ – реестр требований к данным (так называемый мастер данных), поскольку эта информация становится нужной всем системам» [Килимова, 2022].

Для выполнения этой задачи необходимо разработать инструмент извлечения знаний в области качества данных в корпоративных хранилищах. Организации, использующие для работы с данными корпоративные хранилища, как правило, характеризуются большими объемами данных и множеством сотрудников, заинтересованных в их качестве. Инструментом сбора, хранения и анализа ожиданий к качеству данных может стать экспертная информационная система, адаптированная для этой области.

### Результаты и их обсуждение

Проблема извлечения экспертных знаний была поставлена в конце 1970-х годов в процессе разработки первых экспертных систем. Так, Э.А. Фейгенбаум утверждал: «Поскольку сила экспертных систем заключается в их базах знаний, успешное применение ИИ требует, чтобы знания переместились из голов экспертов в программы.» [Feigenbaum, 1980]. Среди отечественных исследователей проблеме посвящали свои труды О.И. Ларичев и В.К. Моргоев, давшие анализ проблем приобретения и извлечения экспертных знаний [Ларичев, Моргоев, 1991], Т.А. Гаврилова и К.Р. Червинская, которые классифицировали методы извлечения знаний [Гаврилова, Червинская, 1992].

В рамках исследования предлагается расширить существующую проблему извлечения экспертных знаний проблемой извлечения ожиданий к качеству данных корпоративных хранилищ.

При этом планируется сделать акцент именно на извлечении знаний как процессе выявления знаний, когда источником информации являются эксперты в проблемной области [Ларичев, Моргоев, 1991], а в качестве методов извлечения знаний рассматривать в первую очередь активные индивидуальные и групповые методы [Гаврилова, Червинская, 1992]. Это обусловлено следующими причинами:

1. Согласно ГОСТ Р ИСО 8000-2-2019, под требованием к качеству данных понимается потребность или ожидание, которое установлено, предполагается или является обязательным, а под качеством данных понимается степень соответствия совокупности присущих характеристик объекта требованиям [ГОСТ Р ИСО 8000-2-2019, 2019];

2. Согласно DAMA-DMBOK данные можно считать высококачественными в той мере, в которой они соответствуют потребностям и ожиданиям потребителей. То есть данные обладают высоким или низким качеством, если они, соответственно, пригодны или непригодны к использованию по назначению. Следовательно, качество данных зависит от контекста и потребностей потребителей данных [DAMA-DMBOK, 2020];

3. Согласно DAMA-DMBOK для того, чтобы данные были надежными и достоверными, профессионалам в области управления данными нужно сделать всё возможное для наилучшего понимания требований клиентов к качеству данных и способов измерения степени соответствия данных этим требованиям. И делаться это должно в режиме постоянного обсуждения, поскольку требования к данным и качеству данных меняются не менее динамично, чем потребности и приоритеты бизнеса, зависящие, в свою очередь, от не менее переменчивых внешних сил и условий [DAMA-DMBOK, 2020].

4. Практический опыт профессиональной деятельности в сфере поддержки качества данных показывает, что часто через определение качества данных корпоративных хранилищ в организациях стремятся определять качество выполнения бизнес-процессов, что повышает роль экспертов, владеющих знаниями об этих бизнес-процессах, как источника ожиданий к качеству данных.

С целью проработки предложенной проблемы исследования была разработана онтологическая модель извлечения ожиданий к качеству данных корпоративных хранилищ (рис. 2).

Онтологические модели используются для обеспечения поддержки деятельности по накоплению, совместному использованию и повторному использованию знаний [Загоруйко, Загоруйко, 2016].

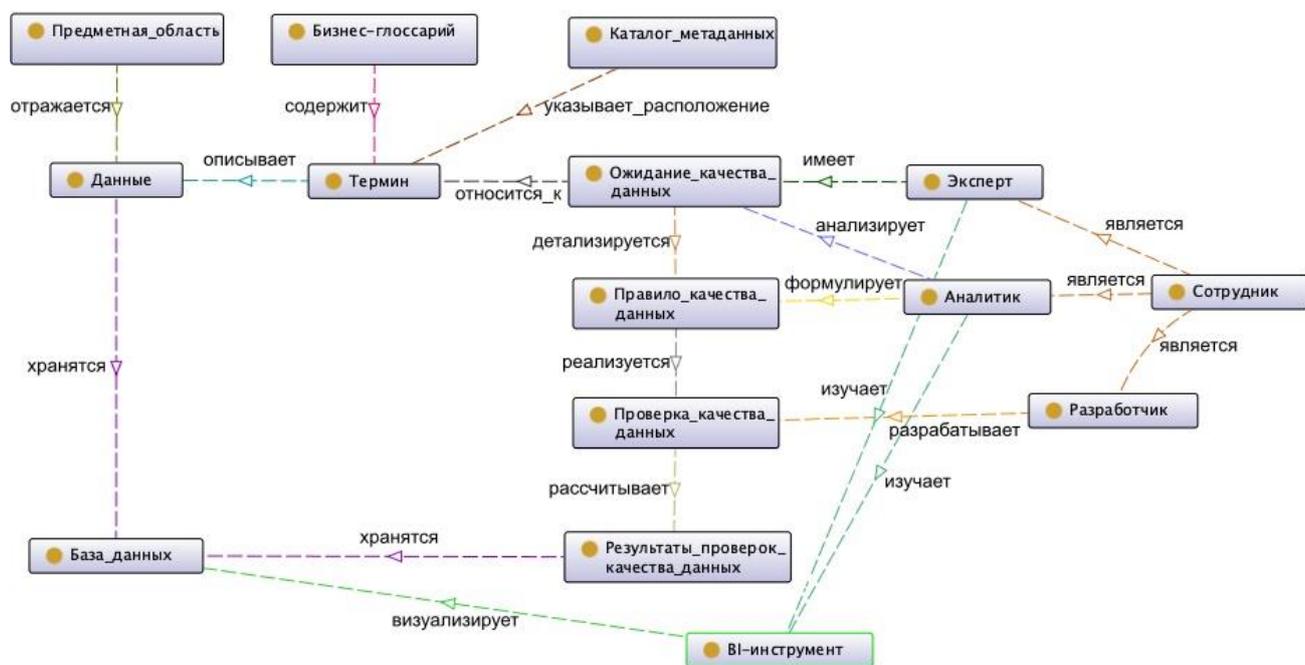


Рис. 2. Онтология извлечения ожиданий к качеству данных корпоративных хранилищ  
Fig. 2. Ontology of extraction of expectations about data quality of enterprise warehouses

Представленная онтологическая модель описывает процесс извлечения (триплеты <Эксперт, имеет, Ожидание качества данных> и <Аналитик, анализирует, Ожидание качества данных>) и последующей обработки экспертных знаний – ожиданий к качеству данных.

Поскольку качество данных является частью Платформы управления данными (по DAMA-DMBOK2) [DAMA-DMBOK:..., 2020], при разработке онтологической модели были учтены такие сущности управления данными, как Бизнес-гlossарий и Каталог метаданных. Ожидание качества данных относится к терминам, содержащимся в Бизнес-гlossарии с указанием их расположения в базах данных (согласно Каталогу метаданных).

Данную онтологическую модель можно использовать при построении инфологической модели данных, с помощью которой проектируются базы знаний и базы данных.



## Заключение

Таким образом, проведенное исследование указывает на существующее внимание к вопросам качества данных в методологиях интеллектуального анализа данных, а также при построении корпоративных хранилищ данных, разработке информационно-аналитических систем. В процессе исследования проведено обоснование необходимости централизованного сбора требований к качеству данных организации, так как существуют сложности при формировании требований к качеству данных от потребителей данных и необходимость использования активных индивидуальных и групповых методов извлечения знаний, например, интервьюирования.

В процессе исследования поставлена проблема извлечения ожиданий к качеству данных корпоративных хранилищ путем расширения существующей проблемы извлечения экспертных знаний, а также разработана онтологическая модель извлечения ожиданий к качеству данных корпоративных хранилищ.

## Список источников

- Гаврилова Т.А., Червинская К.Р. 1992. Извлечение и структурирование знаний для экспертных систем / Т. А. Гаврилова, К. Р. Червинская. М.: Радио и связь, 1992. 199 с.: ил.; 22 см.
- ГОСТ Р ИСО 8000-2-2019: Национальный стандарт Российской Федерации. Качество данных. Часть 2. Словарь [Электронный ресурс]. Стандартинформ. – Электронные данные. – Электронный текст документа подготовлен АО «Кодекс» и сверен по: М.: Стандартинформ, 2019. – режим доступа: <https://docs.cntd.ru/document/1200169126> (дата обращения: 15.05.2022).
- Загоруйко Ю.А., Загоруйко Г.Б. 2016. Инженерия знаний: учебное пособие. [Электронный ресурс] / Ю.А. Загоруйко, Г.Б. Загоруйко; Новосиб. гос. ун-т. – Новосибирск: РИЦ НГУ, 2016. 93 с. Режим доступа: <http://e-lib.nsu.ru/dsweb/Get/Resource-1052/page001.pdf>
- Лидвелл У., Холден К., Батлер Дж. 2021. Универсальные принципы дизайна. Пер. А. Мороз. СПб.: Питер, 2021. 272 с.: ил.
- Международные рекомендации по статистике промышленности. 2010. [Электронный ресурс] Организация Объединенных Наций. Серия «Статистические документы». – Электронные данные. – Нью-Йорк: Издательство Организации Объединенных Наций, 2010. Режим доступа: <https://www.un-ilibrary.org/content/books/9789210563826/read> (дата обращения: 05.11.2021).
- Методология оценки и повышения качества данных. 2019. [Электронный ресурс] Аналитический центр при Правительстве Российской Федерации. – Электронные данные. – Режим доступа: [https://digital.ac.gov.ru/upload/iblock/215/%D0%9A%D0%B0%D1%87%D0%B5%D1%81%D1%82%D0%B2%D0%BE%20%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85%20\\_f.pdf](https://digital.ac.gov.ru/upload/iblock/215/%D0%9A%D0%B0%D1%87%D0%B5%D1%81%D1%82%D0%B2%D0%BE%20%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85%20_f.pdf) (дата обращения: 05.11.2021).
- Цифровые технологии в российских компаниях. 2019. [Электронный ресурс] Результаты исследования КПМГ. Режим доступа: <https://assets.kpmg/content/dam/kpmg/ru/pdf/2019/01/ru-ru-digital-technologies-in-russian-companies.pdf> (Дата обращения: 21.04.2022)
- DAMA-DMBOK: Свод знаний по управлению данными. Второе издание. 2020. Dama International [пер. с англ. Г. Агафонова]. М.: Олимп-Бизнес, 2020. 828 с.: ил.
- Data Rules Program Strategy: International Petroleum Data Standards. 2018. [Электронный ресурс] The Professional Petroleum Data Management (PPDM) Association. – Электронные данные. – Калгари: Ассоциация профессионального управления нефтяными данными (PPDM), 2018. – Режим доступа: <https://dl.ppdm.org/dl/2263> (дата обращения: 07.11.2021).
- IBM Analytics Solutions Unified Method (ASUM) [Электронный ресурс] / IBM Analytics. Режим доступа: [http://gforge.icesi.edu.co/ASUM-DM\\_External/index.htm#cognos.external.asum-DM\\_Teaser/tasks/sps\\_clean\\_data\\_F9A96B23.html?proc=\\_0eKIHlt6EeW\\_y7k3h2HTng&path=\\_0eKIHlt6EeW\\_y7k3h2HTng,\\_0eHEyVt6EeW\\_y7k3h2HTng,\\_0eIS8Vt6EeW\\_y7k3h2HTng,\\_0eIS51t6EeW\\_y7k3h2HTng](http://gforge.icesi.edu.co/ASUM-DM_External/index.htm#cognos.external.asum-DM_Teaser/tasks/sps_clean_data_F9A96B23.html?proc=_0eKIHlt6EeW_y7k3h2HTng&path=_0eKIHlt6EeW_y7k3h2HTng,_0eHEyVt6EeW_y7k3h2HTng,_0eIS8Vt6EeW_y7k3h2HTng,_0eIS51t6EeW_y7k3h2HTng) (Дата обращения: 22.04.2022)
- Loshin David. 2010. The Practitioner's Guide to Data Quality Improvement (1st. ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

### Список литературы

- Килимова А.Д. 2022. Потоки данных в легкой промышленности. Компетентность. № 3. [Электронный ресурс] URL: <https://cyberleninka.ru/article/n/potoki-dannyh-v-legkoy-promyshlennosti> (дата обращения: 11.05.2022).
- Ларичев О.И., Моргоев В.К. 1991. Проблемы, методы и системы извлечения экспертных знаний. Автомат. и телемех., выпуск 6, 3–27. [Электронный ресурс]. Режим доступа: <http://www.mathnet.ru/links/15765250a80252d52d362d9091115564/at4189.pdf> (дата обращения: 04.07.2022).
- Любицын В.Н. 2012. Повышение качества данных в контексте современных аналитических технологий. Вестник ЮУрГУ. Серия: Компьютерные технологии, управление, радиоэлектроника. № 23. [Электронный ресурс] URL: <https://cyberleninka.ru/article/n/povyshenie-kachestva-dannyh-v-kontekste-sovremennyh-analiticheskikh-tehnologiy> (дата обращения: 11.05.2022).
- Chapman P., Clinton J., Kerber R., Khabaza T., Reinartz T., Shearer C., Wirth R. 2000. CRISP-DM 1.0 step-by-step data mining guide. SPSS, 2000, 78 p.
- Fayyad U., Piatetsky-Shapiro G., Smyth P. 1996. "The kdd process for extracting useful knowledge from volumes of data," Commun. ACM, Nov. 1996. 39(11): 27–34.
- Feigenbaum E. 1980. Knowledge engineering: the applied side of artificial intelligence. Computer Science Department of Stanford University. Stanford, [Электронный ресурс] – Режим доступа: <https://stacks.stanford.edu/file/druid:cn981xh0967/cn981xh0967.pdf> (дата обращения: 04.07.2022).
- Martínez-Plumed F. et al. 2021. "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories," in IEEE Transactions on Knowledge and Data Engineering, 33(8): 3048-3061, 1 Aug. 2021, doi: 10.1109/TKDE.2019.2962680.
- Martinez-Plumed F., Ochando L.C., Ferri C., Flach P. A., Hernandez-Orallo J., Kull M., Lachiche N., Ramirez-Quintana M. J. 2017. "CASP-DM: context aware standard process for data mining," CoRR, vol. abs/1709.09003, 2017. [Online]. Available: <http://arxiv.org/abs/1709.09003>
- Moyle S., Jorge A. 2001. "Ramsys-a methodology for supporting rapid remote collaborative data mining projects," in ECML/PKDD 2001 Workshop on Integrating Aspects of Data Mining, Decision Support and Meta-Learning: Internal SolEuNet Session, 2001, 20–31.

### References

- Kilimova A.D. 2022. Data flows in light industry. Competence. №3 (in Russian). [Electronic resource]. URL: <https://cyberleninka.ru/article/n/potoki-dannyh-v-legkoy-promyshlennosti> (date of access: 11.05.2022).
- Larichev O. I., Morgoev V. K. 1991. Problems, methods and systems for extracting expert knowledge. Avtomat. i telemekh., issue 6, 3–27 (in Russian). [Electronic resource]. URL: <http://www.mathnet.ru/links/15765250a80252d52d362d9091115564/at4189.pdf> (accessed 04.07.2022).
- Lyubitsyn V.N. 2012. Improving the quality of data in the context of modern analytical technologies. Vestnik SUSU. Series: Computer technologies, control, radio electronics. No. 23 (in Russian). [Electronic resource]. URL: <https://cyberleninka.ru/article/n/povyshenie-kachestva-dannyh-v-kontekste-sovremennyh-analiticheskikh-tehnologiy> (date of access: 11.05.2022).
- Chapman P., Clinton J., Kerber R., Khabaza T., Reinartz T., Shearer C., Wirth R. 2000. CRISP-DM 1.0 step-by-step data mining guide. SPSS, 2000, 78 p.
- Fayyad U., Piatetsky-Shapiro G., Smyth P. 1996. "The kdd process for extracting useful knowledge from volumes of data," Commun. ACM, Nov. 1996. 39(11): 27–34.
- Feigenbaum E. 1980. Knowledge engineering: the applied side of artificial intelligence. Computer Science Department of Stanford University. Stanford, [Electronic resource]. URL: <https://stacks.stanford.edu/file/druid:cn981xh0967/cn981xh0967.pdf> (accessed: 04.07.2022).
- Martínez-Plumed F. et al. 2021. "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories," in IEEE Transactions on Knowledge and Data Engineering, 33(8): 3048-3061, 1 Aug. 2021, doi: 10.1109/TKDE.2019.2962680.
- Martinez-Plumed F., Ochando L.C., Ferri C., Flach P. A., Hernandez-Orallo J., Kull M., Lachiche N., Ramirez-Quintana M. J. 2017. "CASP-DM: context aware standard process for data mining," CoRR, vol. abs/1709.09003, 2017. [Electronic resource]. URL: <http://arxiv.org/abs/1709.09003>



Moyle S., Jorge A. 2001. “Ramsys-a methodology for supporting rapid remote collaborative data mining projects,” in ECML/PKDD 2001 Workshop on Integrating Aspects of Data Mining, Decision Support and Meta-Learning: Internal SolEuNet Session, 2001, 20–31.

**Конфликт интересов:** о потенциальном конфликте интересов не сообщалось.

**Conflict of interest:** no potential conflict of interest related to this article was reported.

#### ИНФОРМАЦИЯ ОБ АВТОРАХ

**Афанасьев Александр Александрович**, аспирант кафедры программного обеспечения и администрирования информационных систем, Курский государственный университет, г. Курск, Россия

**Кудинов Виталий Алексеевич**, доктор педагогических наук, профессор, профессор кафедры программного обеспечения и администрирования информационных систем, Курский государственный университет, г. Курск, Россия

#### INFORMATION ABOUT THE AUTHORS

**Alexander A. Afanasiev**, Postgraduate Student, Department of Software and Administration of Information Systems, Kursk State University, Kursk, Russia

**Vitaly A. Kudinov**, Doctor of Pedagogy, Professor, Professor of the Department of Software and Information Systems Administration, Kursk State University, Kursk, Russia