

КЛАСТЕРИЗАЦИЯ ДОКУМЕНТОВ ПО ТЕМЕ УПРАВЛЕНИЯ АРХИТЕКТУРОЙ ПРЕДПРИЯТИЯ

А.Р. Денисов, В. Миридаштаки

Артем Руфимович Денисов* (ORCID 0000-0002-3359-4103), Вахид Миридаштаки (ORCID 0000-0001-6619-1880)

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина), ул. Профессора Попова, 5 лит. Ф, Санкт-Петербург, 197376, Россия

E-mail: iptema@yandex.ru*, vahidmiri7@gmail.com

В условиях стремительного роста объёма научных публикаций в различных областях возникает проблема информационной избыточности, затрудняющая систематизацию и анализ существующих исследований. В данной статье представлен подход к автоматизированной кластеризации и тематической классификации 200 научных документов (статей, монографий, стандартов), посвящённых теме управления архитектурой предприятия. Для анализа текстов использовались современные инструменты обработки естественного языка, включая нейросетевые сервисы. Методология исследования включала несколько этапов: извлечение и унификацию ключевых слов с построением графа знаний, формирование TF-IDF матрицы для количественной оценки значимости понятий, а также кластеризацию документов с применением алгоритмов DBSCAN и EM. В результате были выделены 10 содержательных кластеров, отражающих основные направления исследований в области управления архитектурой предприятия. Среди них – теоретические и методологические основы данной темы (фреймворки TOGAF, Zachman), интеграция цифровых двойников, вопросы цифровой трансформации и внедрения технологий Индустрии 4.0, применение эталонных архитектур (RAMI 4.0, IIRA), а также практические аспекты внедрения управления архитектурой предприятия в различных отраслях. Для визуализации результатов использовались сжатые в двумерное пространство распределения публикаций, объединённых минимальным остовным деревом, а также тепловые карты. В результате был построен иерархический классификатор документов в виде дерева решений, что позволило наглядно представить структуру кластеров и взаимосвязи между темами. Разработанный алгоритм демонстрирует высокую эффективность для тематического структурирования больших массивов текстов и может быть применён в других предметных областях, сталкивающихся с проблемой информационной перегрузки.

Ключевые слова: архитектура предприятия, алгоритм анализа документов, обработка естественного языка, граф знаний, кластеризация документов, тематический классификатор.

DOCUMENT CLUSTERING ON THE TOPIC OF ENTERPRISE ARCHITECTURE MANAGEMENT

A.R. Denisov, V. Miridashtaki

Artem R. Denisov* (ORCID 0000-0002-3359-4103), Vahid Miridashtaki (ORCID 0000-0001-6619-1880)

Saint Petersburg Electrotechnical University "LETI", Popova str., 5 F, St. Petersburg, 197376, Russia,

E-mail: iptema@yandex.ru*, vahidmiri7@gmail.com

With the rapid growth of scientific publications in various fields, the problem of information redundancy arises, which manifests in difficulties in systematizing and analyzing existing research. This article presents an approach to automated clustering and thematic classification of 200 scientific documents (articles, monographs, standards) devoted to the topic of enterprise architecture management. Modern natural language processing tools, including neural network services, are used for text processing. The research methodology includes several stages: extraction and unification of keywords with the construction of a knowledge graph, the formation of a TF-IDF matrix for quantifying the significance of concepts, as well as

document clustering using DBSCAN and EM algorithms. As a result, 10 meaningful clusters have been identified, reflecting the main research directions in the field of enterprise architecture management. Among them are the theoretical and methodological foundations on this topic (TOGAF and Zachman frameworks), the integration of digital twins, issues of digital transformation and the introduction of Industry 4.0 technologies, the use of reference architectures (RAMI 4.0, IIRA), as well as practical aspects of implementing enterprise architecture management in various industries. To visualize the results, the distribution of publications compressed into a two-dimensional space, combined with a minimal spanning tree, as well as heat maps are used. As a result, a hierarchical classifier of documents has been built in the form of a decision tree, which makes it possible to visualize the structure of clusters and the relationship between topics. The developed algorithm demonstrates high efficiency for the thematic structuring of large arrays of texts and can be applied in other subject areas facing the problem of information overload.

Keywords: enterprise architecture, document analysis algorithm, natural language processing, knowledge graph, document clustering, thematic classifier.

Для цитирования:

Денисов А.Р., Миридаштаки В. Кластеризация документов по теме управления архитектурой предприятия. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2026. № 02(68). С. 138-149. DOI: 10.6060/ivecofin.2026682.788

For citation:

Denisov A.R., Miridashtaki V. Document clustering on the topic of enterprise architecture management. *Ivecofin*. 2026. N 02(68). P. 138-149. DOI: 10.6060/ivecofin.2026682.788 (in Russian)

ВВЕДЕНИЕ

Методология управления архитектурой предприятия (Enterprise Architecture Management, EA) [1] является важной практической дисциплиной, позволяющей предприятиям сформировать «дорожную карту» стратегического развития, объединяющую разрозненные точечные инновационные проекты в единый управляемый и стратегически выверенный процесс построения организации будущего. Ключевой ценностью EA является ее способность создавать единое и непротиворечивое видение, охватывающее все домены компании: бизнес-стратегию, данные, приложения и технологическую инфраструктуру. Таким образом, EA не просто описывает ИТ-ландшафт, а обеспечивает глубокую взаимосвязь бизнес-стратегии, операционных процессов, информационных активов и технологий. Для современных высокотехнологичных компаний такой целостный и управляемый подход становится не просто преимуществом, а условием сохранения конкурентоспособности и устойчивого развития в эпоху цифровой экономики.

Для практической реализации принципов EA разработаны такие стандарты как TOGAF, а также специализированные языки и инструменты моделирования, например, ArchiMate [2]. Они позволяют формализовать текущее состояние компании и ее целевой образ, выявить разрывы между ними и спланировать последовательность проектов по преобразованию. Это превращает EA в работающий механизм управления. Например, в критически важной области управления

производственным оборудованием и техническим обслуживанием применение EA обеспечивает стратегическое согласование инвестиций, интеграцию данных с датчиков, установленных на оборудовании, цифровых двойников (Digital Twins) и систем оперативного управления, что, например, позволяет перейти к предиктивному обслуживанию (predictive maintenance), радикально снижая простои и затраты [3].

Однако из-за популярности темы EA изучение и систематизация передовых наработок в этой области сталкивается с проблемой «избыточности» информации, когда в существующих базах индексирования представлено большое количество статей, посвященных различным аспектам ее применения. При этом в литературе представлен сравнительный анализ публикаций, но либо он описывает широко известные концепты теории и практики EA, например [1, 4], либо связан с описанием практических кейсов внедрения [2], либо содержит описание отдельных аспектов данной теории, например [5]. Такие публикации не позволяют получить комплексное представление о текущем состоянии данной методологии, что затрудняет последующие исследования по данной теме вследствие усложнения процедур поиска и анализа уже существующих решений. Так, авторами была сформирована коллекция текстов (<https://disk.yandex.ru/i/G8GtgLoqcn6RGg>), включающая в себя 200 различных статей, монографий, стандартов и учебных пособий с темами от практического построения распределенных корпоративных систем с использованием эталон-

ных архитектур, таких как RAMI4.0 [6, 7], до обсуждения теории и методики реализации ЕА. Для анализа такого объема информации требуются специализированные инструменты обработки естественного языка (Nature Language Processing, NLP) [8]. В рамках исследования использовались нейросетевые сервисы GigaChat, DeepSeek и Grok.

МАТЕРИАЛЫ И МЕТОДЫ

Реализация процедуры анализа публикаций осуществлялась по алгоритму, представленному на рис. 1. Часть операций алгоритма выполнялась программой на языке Python (<https://drive.google.com/file/d/1MUCaNOWqLepgLI141t4iaIzK-zESJt60/view?usp=sharing>).

Первоначально необходимо было выделить из текстов статей ключевые слова, названия и аннотации на русском языке. Для этого использовался сервис GigaChat, способный обрабатывать ссылки с Яндекс-диска. Проблемой данного этапа является то, что документы написаны в авторском стиле, в том числе и при формулировании ключевых слов. По этой причине каждая статья связана

с практически уникальными ключевыми словами, слабо пересекающимися с другими текстами.

Для унификации ключевых слов был использован онтологический подход [9], предполагающий классификацию всех имеющихся понятий в виде иерархического графа знаний со структурой «Направление-Тема-Понятие». При этом все синонимы и близкие по смыслу ключевые слова объединялись в единое понятие. Для решения этой задачи использовался сервис DeepSeek. Далее с помощью этого же сервиса тексты документов были проанализированы на соответствие выделенным понятиям графа знаний. При этом было обнаружено, что сервис добавляет избыточные («лишние») ключевые слова, имеющие очень слабую связь с анализируемым текстом (например, простое упоминание в тексте). Такие «лишние» связи были выявлены через соответствующий запрос в новом чате для каждого текста. Данная операция выполнялась в автоматизированном режиме, когда сервис предлагал ключевые слова для исключения с объяснением причин этого, но итоговое решение об исключении принимал эксперт.

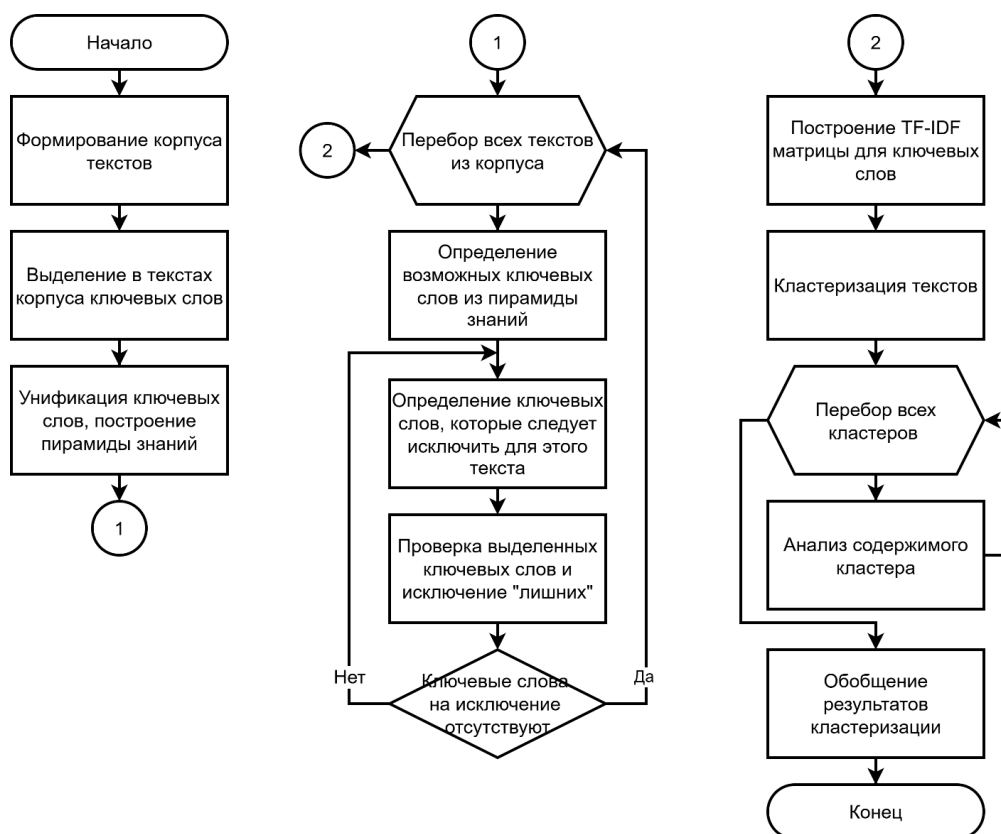


Рисунок 1. Алгоритм анализа документов в коллекции
Figure 1. Algorithm for analyzing texts in a document collection

Далее на основе выделенных унифицированных ключевых слов была сформирована TF-IDF матрица [10], которая была кластеризована с использованием косинусной меры близости. Документы сформированных кластеров были проанализированы на предмет совпадения ключевых слов, а

также через определение основного содержания статей кластеров. Для последней задачи использовался сервис Grok, не имеющий ограничений на размер анализируемых текстов. Полученные результаты были объединены в виде дендрограммы.

РЕЗУЛЬТАТЫ И ДИСКУССИЯ

На основе представленного на рис. 1 алгоритма был построен граф знаний по теме ЕА (рис. 2). В целом, она показывает основные направления исследований в области ЕА: Архитектура и стратегия (99,5% текстов), Данные и аналитика (73%), Цифровая трансформация (71%), Безопасность и надежность (45,5%) и Устойчивое развитие (25%). Причем теме стратегического управления, включая «Стратегическое планирование» и «Согласование целей бизнеса и ИТ» посвящено 137 статей корпуса (68,5%). Такое распределение отражает значимость понятий в методологии ЕА, однако оно не дает возможности понять специфику отдельных текстов.

Для выделения типовых тем в коллекции текстов была проведена их кластеризация, для чего использовались методы DBSCAN и EM (рис. 3 и 4) [11]. Кластеризация осуществлялась над TF-IDF матрицей с использованием косинусной меры близости. Результаты кластеризации выводились в двумерном пространстве, полученном методом многомерного шкалирования (Multidimensional Scaling, MDS) [12]. Поверх кластеризованных данных выведено минимальное остовное дерево (Minimum spanning tree) [13], показывающее ми-

нимальные расстояния между документами коллекции. Это позволяет увидеть, как связи внутри кластеров, так и между ними.

Кластеризация методом DBSCAN позволяет увидеть отдельные группы объектов, отстоящие на некотором расстоянии от других кластеров. В результате было выделено 3 кластера:

0 – документы, посвященные построению распределенных корпоративных информационных систем, в том числе цифровых двойников, на базе эталонных архитектур, в первую очередь RAMI4.0 (17%);

1 – документы, посвященные описанию методологии ЕА (47%);

2 – документы, посвященные внедрению цифровых технологий на основе принципов ЕА (2%).

Остальные тексты (34%) сильно отличаются от остальных, что не позволяет объединить их в кластеры методом DBSCAN.

При группировании с помощью EM-алгоритма документы распределялись по 10 кластерам, что позволяет качественно определить назначение каждого из них (в среднем 20 статей на кластер). Кластеры перечислены «слева-направо».



Рисунок 2. Граф знаний по теме ЕА
Figure 2. Knowledge Graph on the EA topic

Кластеры, выделенные EM-алгоритмом:

Кластер 6 (24 текста). Наиболее значимые ключевые слова: Industry 4.0, RAMI 4.0, Interoperability, System Integration, Cyber-Physical systems, Smart Manufacturing, IoT & IIoT. Тексты кластера представляют собой комплексный обзор концепций, российских и международных стандартов и практических подходов к переходу от традиционного производства к цифровому (Индустрия 4.0,

Промышленный интернет вещей (IoT/IIoT), Smart Manufacturing). Центральной темой кластера является описание и реализация референсных архитектур, таких как RAMI 4.0 (Reference Architectural Model Industrie 4.0), IIRA (Industrial Internet Reference Architecture) и связанные модели. В документах определяются, описываются и анализируются возможности интеграции киберфизических систем (CPS), Интернета вещей, циф-

ровых двойников, облачных технологий и много-агентных систем в производственные процессы. Основные темы включают интероперабельность (совместить) систем, вертикальную и горизонтальную интеграцию, стандартизацию (включая российские и международные стандарты), разработку MES (Manufacturing Execution Systems) и

MOM (Manufacturing Operations Management), а также адаптацию этих технологий для умных фабрик (Smart Factories) и сетевых предприятий. Значительное внимание уделяется цифровым двойникам и их интеграции в RAMI 4.0. Три наиболее репрезентативных текста: [6, 7, 14].

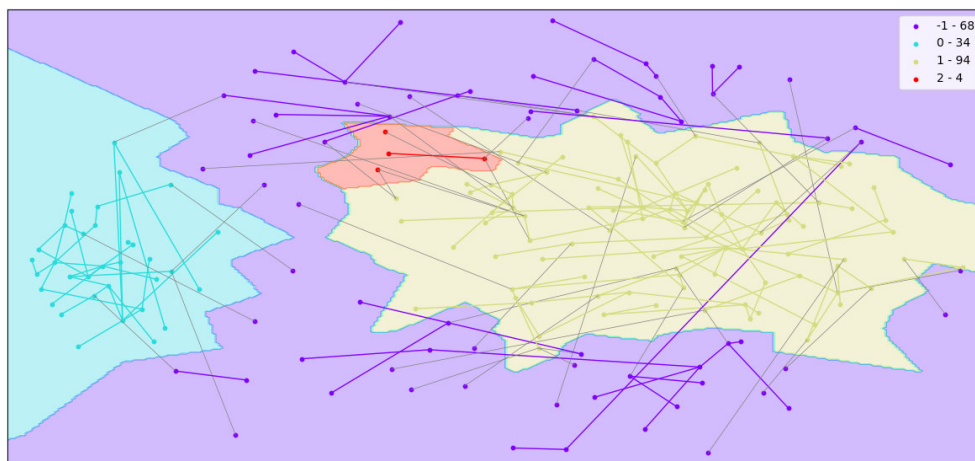


Рисунок 3. Результаты кластеризации, полученные методом DBSCAN
Figure 3. Clustering results obtained using the DBSCAN method

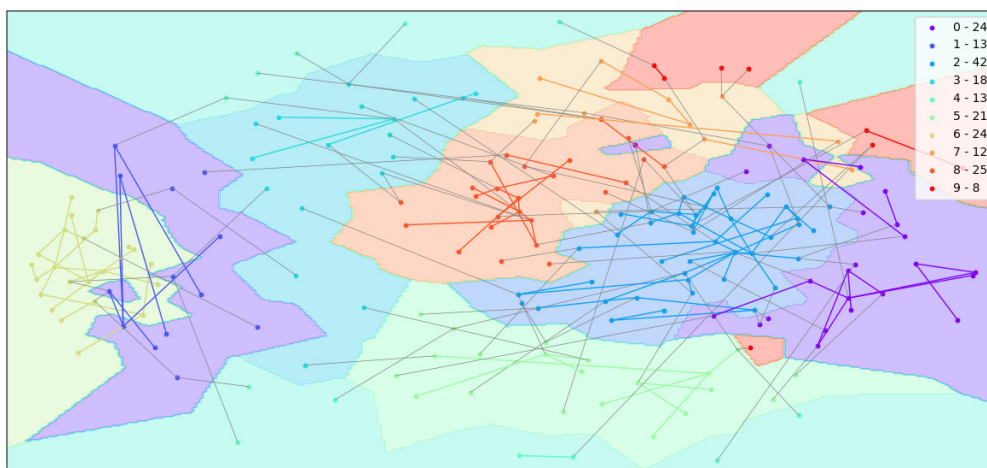


Рисунок 4. Результаты кластеризации, полученные с помощью EM-алгоритма
Figure 4. Clustering results obtained using the EM algorithm

Кластер 1 (13 текстов). Наиболее значимые ключевые слова: Data Architecture, System Integration, IoT & IIoT, Artificial Intelligence, Digital Twin, Predictive Analytics, Big Data, Data Models, Machine Learning. Документы кластера посвящены архитектурам, моделям и технологиям цифровизации производства в контексте Индустрии 4.0. Центральной темой являются цифровые двойники, их применение для мониторинга, прогнозирования, предиктивного обслуживания, интеграции физических и виртуальных моделей в различных отраслях, в первую очередь, энергетике и машиностроении. Референсные архитектуры (RAMI4.0, IIRA и др.) в этом кластере выполняют обеспечивающую функцию, реализуя вертикальную и горизонтальную интеграцию, цепочки со-

здания ценности, интероперабельность и переход к умному производству. Важным аспектом в данном кластере является переход к технологиям управления на основе данных, например, решающим задачи снижения простоев, повышения устойчивости человеко-машинных систем, что обеспечивается через широкое использование промышленного интернета вещей, искусственного интеллекта, технологий больших данных, облачных технологий и стандартов. Три наиболее репрезентативных текста: [15, 16, 17].

Кластер 4 (13 текстов). Наиболее значимые ключевые слова: Digital Transformation. Кластер сформирован по остаточному принципу. В него вошли документы, описывающие процесс внедрения разного рода технологий на основе

принципов ЕА, не попавшие в другие кластеры. По этой причине тексты кластера представляют собой разнообразный сборник научных статей, методических разработок и аналитических материалов, объединенных в рамках контекста цифровой трансформации, стратегического управления и интеграции ИТ-систем с бизнес-процессами. Общий акцент сделан на интеграции бизнес-стратегий с ИТ-инфраструктурой для повышения эффективности, устойчивости и адаптивности организаций в условиях изменяющейся внешней среды. Вследствие наличия специализированных под теоретические вопросы кластеров, большинство статей кластера имеют практическую направленность, в частности, они предлагают инструменты, модели оценки (качества ЕА, эффективности систем) и подходы к автоматизации, с примерами из реальных отраслей (инфокоммуникации, производство, ИТ-компания). Три наиболее репрезентативных текста: [18, 19, 20].

Кластер 3 (текстов). Наиболее значимые ключевые слова: Data Architecture, Digital Transformation, Risk Management, System Integration, Business Processes, Data Governance, Data Models, EA Models, Strategic Planning. Документы кластера ориентированы на интеграцию технологий цифрового производства и Индустрии 4.0 в бизнес-процессы промышленных организаций на основе методологии ЕА. Основные темы кластера: моделирование и диагностика бизнес-архитектуры, интеграция цифровых технологий (AI, микросервисы, облачные сервисы, цифровые двойники), стратегическое управление трансформацией, экономическое обоснование сетевых предприятий, внедрение бережливого производства, а также методы проектирования сложных систем для различных промышленных отраслей (энергетика, производство, горнодобыча). Общий акцент делается на переходе к дата-центричным и сервис-ориентированным системам. В частности, часть документов кластера ориентирована на совершенствование методов проектирования ЕА через создание динамических адаптивных моделей, учитывающих распределенную организацию и экосистемы компаний или цепочки создания ценности. Статьи сочетают концептуальные модели (онтологии, алгоритмы, матрицы) с практическими примерами, реализованными на основе стандартных референсных моделей, таких как RAMI4.0 или TOGAF. Три наиболее репрезентативных текста: [21, 22, 23].

Кластер 8 (25 документов). Наиболее значимые ключевые слова: Business-IT Alignment, Business Architecture, Data Architecture, Architectural Design, IT Governance, Architectural Principles, Data Governance, Digital Transformation, Risk Management, Business Processes, Strategic Planning, Technological Architecture, Archi-

tecture Governance, Change Management, Application Architecture. Тексты кластера посвящены ЕА как ключевому инструменту для цифровой трансформации бизнеса, обеспечивающему интеграцию инновационных технологий в бизнес-стратегии компании в различных отраслях экономики. Обсуждаются фреймворки и модели (например, TOGAF, FEA), преимущества внедрения ЕА (экономия затрат, повышение гибкости), вызовы (интеграция с цифровыми технологиями, оценка зрелости) и практические примеры реализации ЕА на различных предприятиях. Основная тема: как интеграция бизнес-стратегий с технологическими инновациями на основе принципов ЕА помогает компаниям адаптироваться к быстрым изменениям, повысить свою устойчивость и минимизировать риски, максимизировать ценность от ИТ-инвестиций. В документах определяется необходимость руководства процессами управления архитектурой (Architecture governance), которое позволит динамически управлять интегрированной системой, включающей в себя бизнес- и ИТ-архитектуру в соответствии со стратегией развития компании целей. Это позволяет организациям согласовывать процессы развития в условиях геополитической нестабильности, экономических трансформаций и технологических изменений. Три наиболее репрезентативных текста: [24, 25, 26].

Кластер 5 (21 документ). Наиболее значимые ключевые слова: Business Architecture, Business Processes, Digital Transformation, Application Architecture. Тексты кластера описывают ЕА как инструмент развития бизнес-архитектуры в условиях цифровой трансформации. В статьях на основе стандартных фреймворков (TOGAF, ArchiMate, IDEF0, BPMN и др.) и их расширений предлагаются модели и подходы для проектирования, анализа и оптимизации бизнес-архитектуры в различных секторах экономики. Ключевые темы: интеграция ИТ с бизнес-процессами, обеспечение надежности, поддержка цифровых сервисов, переход от линейной к циклической экономике, использование технологий машинного обучения для обеспечения устойчивости компаний. Подчеркивается роль ЕА в повышении эффективности, адаптивности и экологичности предприятий, минимизации рисков. Также рассматриваются вопросы развития методологии ЕА через внедрение стратегических и мотивационных (заинтересованные лица, драйверы и т.п.) элементов, элементов обеспечения экологической устойчивости и т.п. Три наиболее репрезентативных текста: [27, 28, 29].

Кластер 0 (24 документа). Наиболее значимые ключевые слова: EA Models, Architectural Design, Architectural Principles, Business Architecture, TOGAF, Zachman Framework, Business Pro-

cesses, Strategic Planning, Business-IT Alignment. Теоретическое ядро. Содержит комплексный обзор развития концепции ЕА от исторических корней до современных подходов к цифровой трансформации. Тексты анализируют теоретические основы ЕА, включая ключевые теории и методологии моделирования (TOGAF, Zachman, FEAF, GERAM) и их интеграцию с бизнес-стратегией, ИТ-инфраструктурой и процессами управления. Особое внимание уделяется преимуществам ЕА, таким как повышение эффективности бизнеса, повышение качества процессов стратегического планирования, повышение уровня согласованности между бизнес и ИТ целями, улучшение процессов принятия решений, адаптивность к технологическим изменениям, снижение рисков инвестиций. Три наиболее репрезентативных текста: [1, 30, 31].

Кластер 2 (42 документа). Наиболее значимые ключевые слова: Business Architecture, Technological Architecture, Application Architecture, Data Architecture, Business-IT Alignment, TOGAF, Strategic Planning, Business Processes, EA Models, Architectural Design, Zachman Framework, Architecture Governance, IT Governance, Architectural Principles. Методическое ядро. Общая тематика кластера фокусируется на переходе от описания ИТ-инфраструктуры к комплексному проектированию ЕА, способствующему повышению конкурентоспособности, эффективности и адаптивности компаний в динамичной экономической среде. В основе большинства текстов лежит идея, что ЕА – это методология, позволяющая организациям адаптироваться к вызовам цифровой экономики через согласование бизнес и ИТ целей. Документы кластера содержат обзоры и сравнительный анализ фреймворков (в первую очередь, TOGAF, Zachman Framework, Gartner), кейсы применения ЕА (в основном с использованием TOGAF ADM) на промышленных предприятиях и в государственных структурах. Часть работ фокусируется на визуализации результатов проектирования ЕА, подчеркивая, что ЕА модели позволяют абстрагировать сложные системы для анализа, планирования и реализации изменений, способствуя лучшему пониманию бизнес-стратегий и технологических нужд. Три наиболее репрезентативных текста: [4, 5, 32].

Кластер 7 (12 документов). Наиболее значимые ключевые слова: Business-IT Alignment, Architecture Governance, Strategic Planning, Digital Transformation, IT Governance, TOGAF. Документы кластера описывают ЕА как стратегический инструмент управления сложностью организаций в эпоху цифровой трансформации, цифровизации бизнес-моделей и интеграции ИТ с бизнес-процессами. Основные темы: ценность, выгоды и

критические факторы успеха внедрения ЕА, инвестиционные модели и оценка вложений в ИТ-проекты и ЕА-решения. Часть статей исследует «мифы» про ЕА: почему многие утверждения о выгоде не подкреплены практическими результатами. В ряде статей описана практика применения и кейсы в различных отраслях и странах. Общий вывод: ЕА – не статичный набор диаграмм, а динамичная дисциплина, требующая доказательных подходов, зрелой практики и интеграции с бизнес-стратегией. При правильном использовании в условиях Индустрии 4.0 ЕА помогает организациям сохранять конкурентоспособность через лучшее управление изменениями, инвестициями и экосистемами. Однако успех зависит от преодоления мифов, эмпирического подтверждения выгод и адаптации под конкретный контекст (отрасль, масштаб, культура). Три наиболее репрезентативных текста: [33, 34, 35].

Кластер 9 (8 документов). Наиболее значимые ключевые слова: Business-IT Alignment, Strategic Planning, IT Governance, Business Processes. Документы кластера посвящены практике внедрения ЕА в крупных организациях, а также инфраструктурной поддержке данных процессов. Большинство работ – описание кейсов с акцентом на крупные организации, госсектор и телеком. Цель – помочь руководству и специалистам компании (CIO, ЕА-архитекторам) и исследователям понять, как на основе ЕА обеспечить стратегическое развитие бизнеса, в т.ч. в условиях регуляторного давления, как обеспечить успешное внедрение ЕА, как связать стратегию с архитектурой и как адаптировать подходы к цифровой трансформации. Для этого предлагаются различные инструменты, такие как системы поддержки принятия решений и системы стратегического планирования. Три наиболее репрезентативных текста: [2, 36, 37].

Проверка кластеров методом силуэта [38] (рис. 5) показала, что часть из них (2, 6, 8) характеризуются хорошей изолированностью. Это связано со специфичностью тем данных кластеров, что, в частности, видно по результатам кластеризации методом DBSCAN, где также были выделены схожие темы. Часть кластеров (1, 3, 4, 7 и 9) характеризуются низким качеством, что связано либо с «размытостью» темы, как, например, у 7 и 9 кластеров, либо с формированием этих кластеров по остаточному принципу. При оценке результатов методом силуэта также нужно учитывать, что он разрабатывался для метода К-средних [38], что снижает результаты оценки при кластеризации другими методами, в частности, с помощью EM-алгоритма.

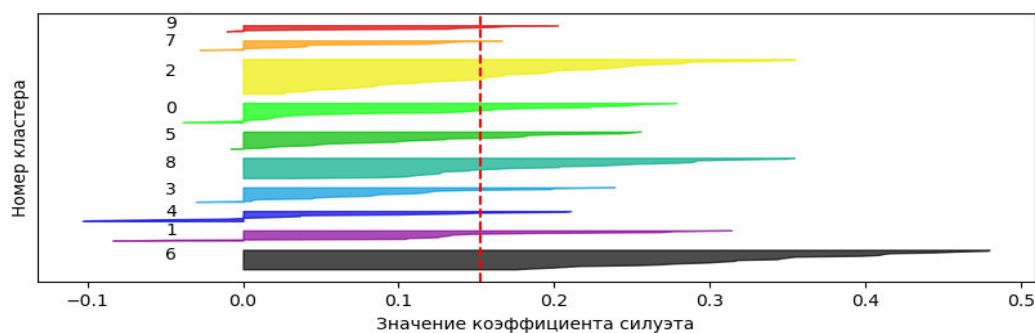


Рисунок 5. Силуэты сформированных кластеров
Figure 5. Silhouettes of the formed clusters

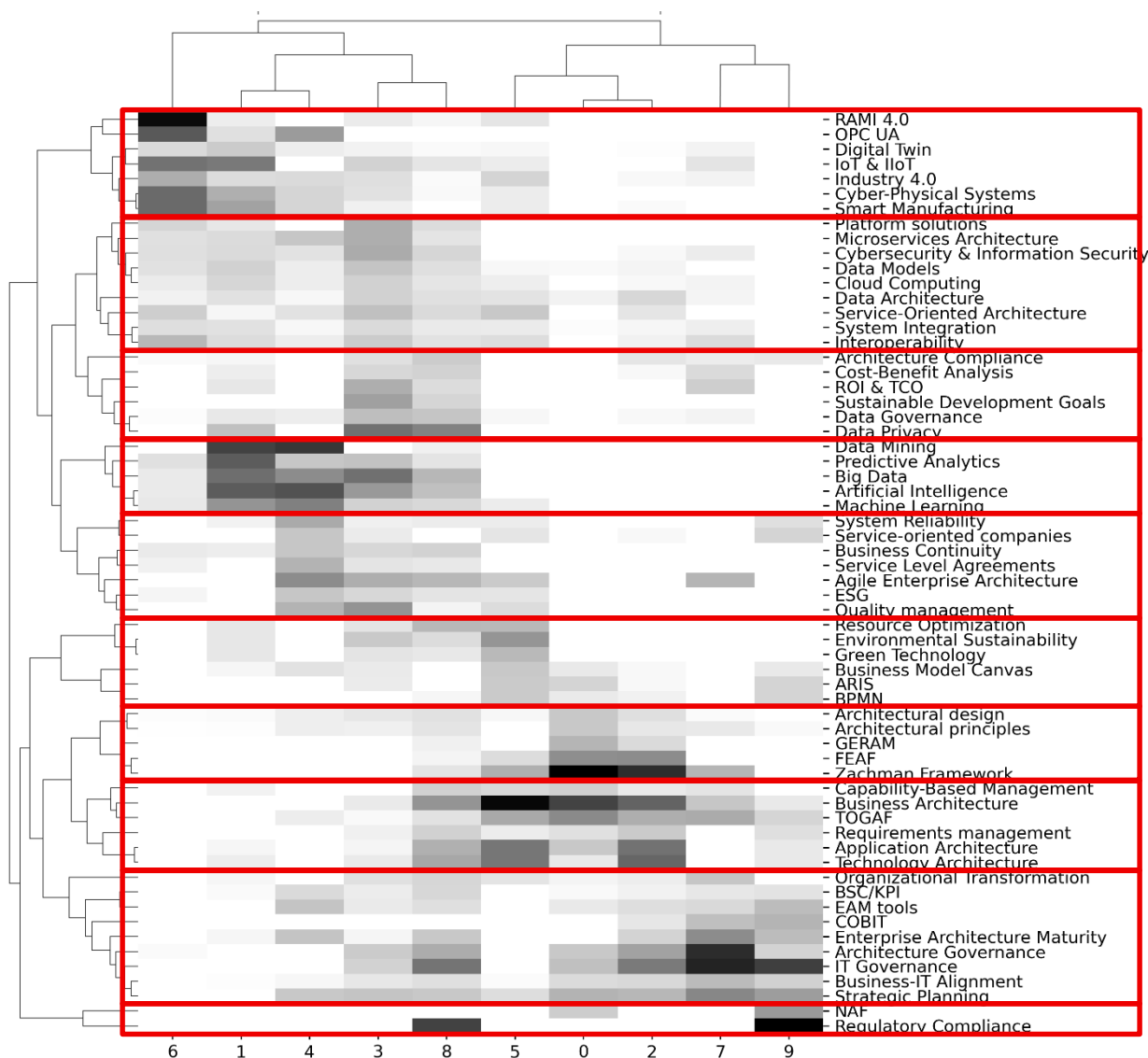


Рисунок 6. Тепловая карта распределения ключевых слов по кластерам
Figure 6. Heat map of keyword distribution by cluster

Проверка кластеров методом силуэта [38] (рис. 5) показала, что часть из них (2, 6, 8) характеризуются хорошей изолированностью. Это связано со специфичностью тем данных кластеров, что, в частности, видно по результатам кластеризации методом DBSCAN, где также были выделены схожие темы. Часть кластеров (1, 3, 4,

7 и 9) характеризуются низким качеством, что связано либо с «размытостью» темы, как, например, у 7 и 9 кластеров, либо с формированием этих кластеров по остаточному принципу. При оценке результатов методом силуэта также нужно учитывать, что он разрабатывался для метода К-средних [38], что снижает результаты оценки

при кластеризации другими методами, в частности, с помощью ЕМ-алгоритма.

Для понимания содержания каждого из кластеров была построена тепловая карта (cluster map) [38] кластеров на основе TF-IDF матрицы. Ее особенностью является иерархическая кластеризация объектов и признаков, используемых при построении карты (рис. 6), что позволяет выделить группы понятий, которые схожим образом используются в коллекции текстов. Эти группы определяют ключевые темы, которым посвящены статьи (рис. 7). Таким образом можно отследить распространенность тем в сформированных кластерах. Кроме этого, на основе иерархического группирования самих кластеров можно построить классификатор тем статей в виде дерева решений (рис. 8), что и являлось задачей исследования.

Для понимания содержания каждого из кластеров была построена тепловая карта (cluster map) [38] кластеров на основе TF-IDF матрицы. Ее особенностью является иерархическая кластеризация объектов и признаков, используемых при построении карты (рис. 6), что позволяет выделить группы понятий, которые схожим образом используются в коллекции текстов. Эти группы определяют ключевые темы, которым посвящены статьи (рис. 7). Таким образом, можно отследить распространенность тем в сформированных кластерах. Кроме этого, на основе иерархического группирования самих кластеров можно построить классификатор тем статей в виде дерева решений (рис. 8), что и являлось задачей исследования.

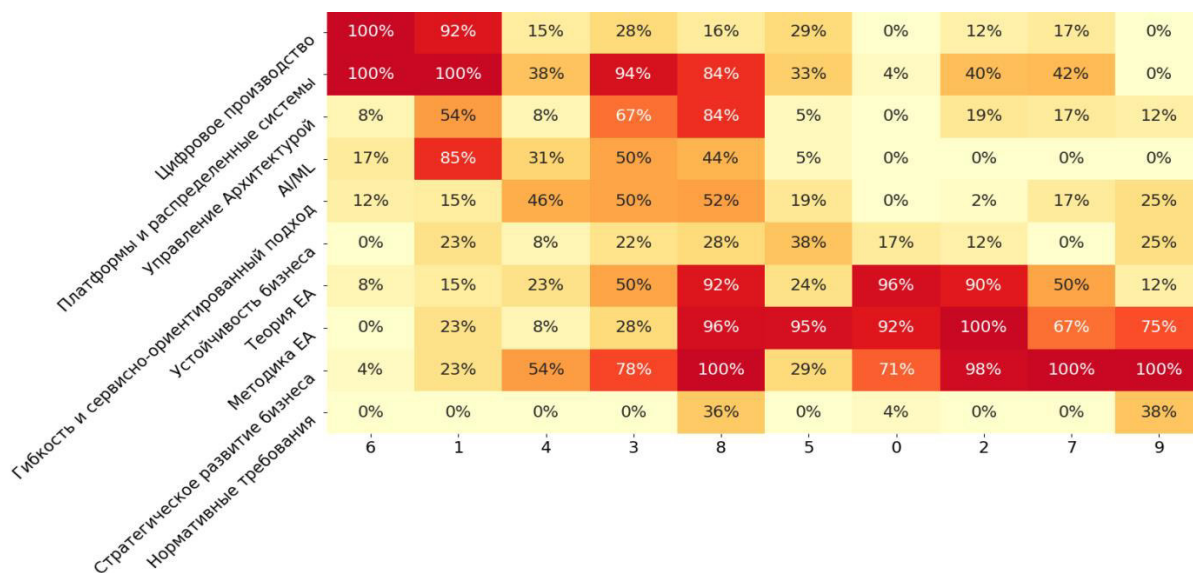


Рисунок 7. Тепловая карта распределения ключевых тем по кластерам

Figure 7. Heat map of the distribution of key topics by clusters

ЗАКЛЮЧЕНИЕ

Проведенное исследование направлено на решение проблемы информационной «избыточности» в научных публикациях, которая существенно затрудняет системный анализ и синтез научных знаний. В рамках работы был разработан и апробирован комплексный подход к автоматизированной кластеризации и тематической классификации корпуса научных текстов на примере коллекции из 200 документов, посвященных управлению архитектурой предприятия. Предложенная методология, объединяющая современные инструменты обработки естественного языка (GigaChat, DeepSeek, Grok), онтологический подход для унификации понятийного аппарата и алгоритмы машинного обучения (DBSCAN, EM), позволила не просто сгруппировать документы, но и выявить смысловую структуру анализируемой предметной области.

Полученная в ходе анализа дендрограмма отражает направления дискурса по теме EA. Во-первых, классификатор разделяет тексты на два крупных блока: «внедрение цифровых технологий» (46,5%) и «управление архитектурой предприятия» (53,5%), что соответствует фундаментальному разделению дисциплины на практику построения корпоративных информационных систем и стратегическое управление. Во-вторых, полученная классификация позволяет выделить специализированные направления: от теоретических основ (фреймворки TOGAF и Zachman) до прикладных аспектов, таких как интеграция цифровых двойников, применение эталонных архитектур (RAMI 4.0, IIRA) и отраслевая специфика внедрения. Такая структура не только облегчает навигацию по существующему массиву знаний, но и позволяет исследователям и практикам быстро идентифицировать наиболее релевантные источники для своих задач, включая возможность использования автоматизированных инструмен-

тов поиска и анализа. Дополнительно необходимо указать, что приведенный в статье алгоритм классификации текстов может быть использован при

анализе документов других направлений, где также стоит проблема «избыточности» существующих публикаций.

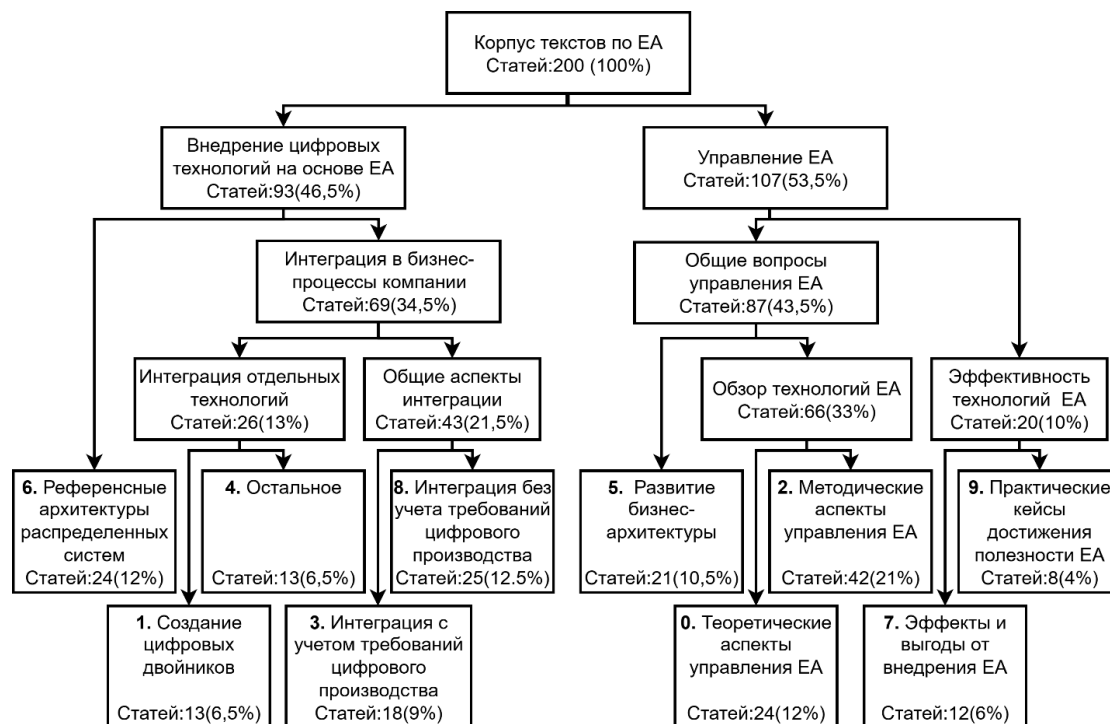


Рисунок 8. Классификация документов в области EA

Figure 8. Classification of EA documents

Ключевым направлением дальнейшего развития предложенного подхода является разработка механизмов оценки результатов классификации документов. Важно правильно оценивать работу на каждом шаге алгоритма: выделение унифицированных ключевых слов, кластеризация корпуса текстов, определение темы для каждого из кластеров. Следующим этапом станет автоматизация полного цикла анализа, что поз-

волит обрабатывать растущие коллекции документов в режиме реального времени. В дальнейшем полученный иерархический классификатор может стать основой для разработки интеллектуальной системы поддержки принятия решений для исследователей по теме EA.

Авторы заявляют об отсутствии конфликта интересов.

The authors declare no conflict of interest.

ЛИТЕРАТУРА

1. Kotusev S., Kurnia Sh. The theoretical basis of enterprise architecture: A critical review and taxonomy of relevant theories. *Journal of Information Technology*. 2021. Vol. 36(3). P. 275-315. DOI: 10.1177/0268396220977873.
2. Abunadi I. Enterprise architecture best practices in large corporations. *Information*. 2019. Vol. 10. 293. DOI: 10.3390/info10100293.
3. Невиницын В.Ю., Грименицкий П.Н., Ваняйки И.К., Отлягузов Д.С. Разработка цифровых двойников типовых технологических процессов. *Современные наукоемкие технологии. Региональное приложение*. 2025. № 1(81). С.111-117. DOI: 10.6060/snt.20258101.00010.
4. Штейнгарт Е.А., Бурмистров А.Н. Обзор и сравнительная характеристика методологий разработки архитектуры предприятий. *Научно-технические ведомости СПбГПУ. Серия «Экономические науки»*. 2016. № 3(245). С. 111-129. DOI: 10.5862/JE.245.11.
5. Zhou Zh., Zhi Q., Morisaki Sh., Yamamoto Sh. A systematic literature review on enterprise architecture visualization methodologies. *IEEE Access*. 2020. Vol. 8. 96405.
6. Schweichhart K. Reference Architectural Model Industrie 4.0 (RAMI 4.0): An introduction. Plattform Industrie 4.0. 2016. <https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/rami40-an-introduction.html>.

REFERENCES

1. Kotusev S., Kurnia Sh. The theoretical basis of enterprise architecture: A critical review and taxonomy of relevant theories. *Journal of Information Technology*. 2021. Vol. 36(3). P. 275-315. DOI: 10.1177/0268396220977873.
2. Abunadi I. Enterprise architecture best practices in large corporations. *Information*. 2019. Vol. 10. 293. DOI: 10.3390/info10100293.
3. Nevinityn V.Yu., Grimenitsky P.N., Vanyaikin I.K., Otlyaguzov D.S. Digital twins design of typical technological processes. *Modern high technologies. Regional application*. 2025. N 1(81). P.111-117. DOI: 10.6060/snt.20258101.00010. (in Russian).
4. Shteingart E.A., Burmistrov A.N. Review and comparative characteristics of methodologies for the development of enterprise architecture. *π-Economy*. 2016. N 3(245). P. 111-129. DOI: 10.5862/JE.245.11. (in Russian).
5. Zhou Zh., Zhi Q., Morisaki Sh., Yamamoto Sh. A systematic literature review on enterprise architecture visualization methodologies. *IEEE Access*. 2020. Vol. 8. 96405.
6. Schweichhart K. Reference Architectural Model Industrie 4.0 (RAMI 4.0): An introduction. Plattform Industrie 4.0. 2016. <https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/rami40-an-introduction.html>.

7. ГОСТ Р 59799-2021. Умное производство. Модель эталонной архитектуры Индустрии 4.0 (RAMI 4.0). <https://files.stroyinf.ru/Data/778/77860.pdf>.
8. Герасимов А.С., Бобков С.П. Разработка самообучающейся системы обработки естественного языка с динамической реорганизацией знаний. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2025. № 3 (65). С. 77-84. DOI: 10.6060/ivecofin.2025653.734.
9. Косовская Т.М. Сравнение различных представлений знаний для сложных структурированных объектов при решении задач ИИ. *Компьютерные инструменты в образовании*. 2021. № 2. С. 41–57. DOI: 10.32603/2071-2340-2021-2-41-57.
10. Шерстнев П.А. Вычисление вектора документа с использованием меры TF-IDF. В сб.: «Актуальные проблемы авиации и космонавтики». Матер. VII Межд. н.-пр. конф., посвященной Дню космонавтики. Красноярск: СибГУНиТ им.Решетнева. 2021. Т. 2. С. 213-215.
11. Дмитриев Н.Д., Тюлькова А.В., Сорокожердьев В.В. Кластеризация российских компаний технологического сектора на основе рыночных и финансовых показателей. *Вестник ЧелГУ*. 2024. № 10 (492). С. 116-127. DOI: 10.47475/1994-2796-2024-492-10-116-127.
12. Искандаров Д.З., Бороздина С.М. Метод многомерного шкалирования для оценки эффективности взаимодействия предприятий инвестиционно-строительной сферы и особых экономических зон в аспекте совершенствования финансового и инновационного потенциала строительной отрасли. *Вестник МГСУ*. 2022. № 9. С. 1250-1263. DOI: 10.22227/1997-0935.2022.9.1250-1263.
13. Кисляков А.Н. Алгоритм минимального связующего дерева для оценки точек роста в системе стратегического планирования территорий. *Развитие территорий*. 2024. № 2 (36). С. 33-40. DOI: 10.32324/2412-8945-2024-2-33-40.
14. Jaskó S., Skropa A., Holczinger T., Chováň T., Abonyi J. Development of manufacturing execution systems in accordance with Industry 4.0 requirements: A review of standard- and ontology-based methodologies and tools. *Computers in Industry*. 2020. N 123. 103300. DOI: 10.1016/j.compind.2020.103300.
15. Helmann A., Neto A.A., Deschamps F., de Freitas Rocha Loures E. Industrial maintenance and the digital twin – An architectural assessment. *FAIM 2022, LNME*. 2023. P. 696-703. DOI: 10.1007/978-3-031-17629-6_73.
16. Денисов А.Р., Никулин А.Е. Организация процессов проактивного технического обслуживания на предприятии топливно-энергетического сектора. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2025. № 4 (66). С. 105-114. DOI: 10.6060/ivecofin.2025664.750.
17. Кабанов А.А., Федоров И.А. Архитектура системы управления цифровыми двойниками производств как основа интеграции различных моделей их представлений. *Вестник РГРТУ*. 2022. № 82. С. 162-176.
18. Gampfer F., Jürgens A., Müller M., Buchkremer R. Past, current and future trends in enterprise architecture – A view beyond the horizon. *Computers in Industry*. 2018. N 100. P. 70-84. DOI: 10.1016/j.compind.2018.03.006.
19. Mirsalar S.R., Ranjbarfard M. A model for evaluation of enterprise architecture quality. *Evaluation and Program Planning*. 202. N 83. 101853. DOI: 10.1016/j.eval-progplan.2020.101853.
20. Yenuganti G. Evolving the modern enterprise stack: From monoliths to adaptive cognitive architectures. *JCSTS*. 2025. N 7(5). P. 356-361. DOI: 10.32996/jcsts.2025.7.5.45.
21. Korhonen J.J., Halén M. Enterprise architecture for digital transformation. *IEEE 19th Conference on Business Informatics*. 2017. P. 350-358. DOI: 10.1109/CBI.2017.45.
22. Deryabin S., Temkin I., Rzazade U., Kondratev E. Models and methods of designing data-centric microservice architectures of digital enterprises. *Informatics*. 2023. N 10. 4. DOI: 10.3390/informatics10010004.
7. GOST R 59799-2021. Smart manufacturing. Reference architecture model for Industry 4.0 (RAMI 4.0). <https://files.stroyinf.ru/Data/778/77860.pdf>. (in Russian).
8. Gerasimov A.S., Bobkov S.P. Development of a self-learning system for natural language processing with dynamic knowledge reorganization. *Ivecofin*. 2025. N 3 (65). P. 77-84. DOI: 10.6060/ivecofin.2025653.734. (in Russian).
9. Kosovskaya T.M. Comparison of different knowledge representations for complex structured objects in solving AI problems. *Computer tools in education*. 2021. N 2. P. 41–57. DOI: 10.32603/2071-2340-2021-2-41-57. (in Russian).
10. Sherstnev P.A. Calculating a document vector using the TF-IDF measure. «Actual problems of aviation and cosmonautics». *Materials of the VII International Scientific and Practical Conference dedicated to Cosmonautics Day*. Krasnoyarsk: SibGUNiT named after Reshetneva. 2021. Vol. 2. P.213-215. (in Russian).
11. Dmitriev N.D., Tyulkova A.D., Sorokozherdiev V.V. Clustering of Russian companies in the technology sector based on market and financial indicators. *Bulletin of Chelyabinsk State University*. 2024. N 10 (492). P. 116-127. DOI: 10.47475/1994-2796-2024-492-10-116-127. (in Russian).
12. Iskandarov D.Z., Borozdina S.M. Method of multidimensional scaling for evaluating the effectiveness of interaction between investment and construction sector enterprises and special economic zones in the aspect of improving the financial and innovative potential of construction industry. *Vestnik MGSU*. 2022. N 9. P. 1250-1263. DOI: 10.22227/1997-0935.2022.9.1250-1263. (in Russian).
13. Kislyakov A.N. Minimum spanning tree algorithm for evaluating growth points in territories strategic planning system. *Territory Development*. 2024. N 2 (36). P. 33-40. DOI: 10.32324/2412-8945-2024-2-33-40. (in Russian).
14. Jaskó S., Skropa A., Holczinger T., Chováň T., Abonyi J. Development of manufacturing execution systems in accordance with Industry 4.0 requirements: A review of standard- and ontology-based methodologies and tools. *Computers in Industry*. 2020. N 123. 103300. DOI: 10.1016/j.compind.2020.103300.
15. Helmann A., Neto A.A., Deschamps F., de Freitas Rocha Loures E. Industrial maintenance and the digital twin – An architectural assessment. *FAIM 2022, LNME*. 2023. P. 696-703. DOI: 10.1007/978-3-031-17629-6_73.
16. Denisov A.R., Nikulin A.E. Organization of proactive maintenance processes at an enterprise in the energy industry. *Ivecofin*. 2025. N 4 (66). P. 105-114. DOI: 10.6060/ivecofin.2025664.750 (in Russian)
17. Kabanov A.A., Fedorov I.A. Architecture of management system for digital twins of productions as a basis for integration of different models of their representations. *Vestnik of RSREU*. 2022. N 82. P. 162-176. (in Russian).
18. Gampfer F., Jürgens A., Müller M., Buchkremer R. Past, current and future trends in enterprise architecture – A view beyond the horizon. *Computers in Industry*. 2018. N 100. P. 70-84. DOI: 10.1016/j.compind.2018.03.006.
19. Mirsalar S.R., Ranjbarfard M. A model for evaluation of enterprise architecture quality. *Evaluation and Program Planning*. 202. N 83. 101853. DOI: 10.1016/j.eval-progplan.2020.101853.
20. Yenuganti G. Evolving the modern enterprise stack: From monoliths to adaptive cognitive architectures. *JCSTS*. 2025. N 7(5). P. 356-361. DOI: 10.32996/jcsts.2025.7.5.45.
21. Korhonen J.J., Halén M. Enterprise architecture for digital transformation. *IEEE 19th Conference on Business Informatics*. 2017. P. 350-358. DOI: 10.1109/CBI.2017.45.
22. Deryabin S., Temkin I., Rzazade U., Kondratev E. Models and methods of designing data-centric microservice architectures of digital enterprises. *Informatics*. 2023. N 10. 4. DOI: 10.3390/informatics10010004.

23. **Liao M.-H., Wang C.-T.** Using enterprise architecture to integrate lean manufacturing, digitalization, and sustainability: A lean enterprise case study in the chemical industry. *Sustainability*. 2021. N 13. 4851. DOI: 10.3390/su13094851.
24. **Афанасьева С.И., Чернышева К.В.** Цифровая трансформация с использованием систем управления взаимоотношений с клиентами. *Современные наукоемкие технологии. Региональное приложение*. 2024. №1(77). С.60-97. DOI: 10.6060/snt.20247701.0008.
25. **Vandevenne N., Van Riel J., Poels G.** Green enterprise architecture (GREAN) – Leveraging EA for environmentally sustainable digital transformation. *Sustainability*. 2023. N 15. 14342. DOI: 10.3390/su151914342EA.
26. **Анисифоров А.Б.** Особенности проектов цифровой трансформации промышленных предприятий и некоторые аспекты подготовки к их реализации. *Научный журнал НИУ ИТМО. Серия «Экономика и экологический менеджмент»*. 2022. № 2. С. 32-41. DOI: 10.17586/2310-1172-2022-16-2-32-41.
27. **Oberlea M., Yesilyurta O., Schleretha A., Rislinga M., Schela D.** Enterprise IT architecture greenfield design combining IEC 62264 and TOGAF by example of battery manufacturing. *Procedia Computer Science*. 2023. N 217. P. 136-146. DOI: 10.1016/j.procs.2022.12.209.
28. **Ильин И.В., Лёвина А.И., Дубгорн А.С.** Цифровая трансформация как фактор формирования архитектуры и IT-архитектуры предприятия. *Научный журнал НИУ ИТМО. Серия «Экономика и экологический менеджмент»*. 2019. № 3. С. 50-55. DOI: 10.17586/2310-1172-2019-12-3-50-55.
29. **Laumann F., Tambo T.** Enterprise architecture for a facilitated transformation from a linear to a circular economy. *Sustainability*. 2018. N 10. 3882. DOI:10.3390/su10113882.
30. **Tamm T., Seddon P.B., Shanks G.** How enterprise architecture leads to organisational benefits. *International Journal of Information Management*. 2022. N 67. 102554. DOI: 10.1016/j.ijinfomgt.2022.102554.
31. **Денисов А.Р., Миридаштаки В.** Референсная модель подсистемы стратегического развития компании на основе принципов архитектуры предприятия. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2025. № 3 (65). С. 85-93. DOI: 10.6060/ivecofin.2025653.735.
32. **Saffranto R.** Enterprise architecture planning with TOGAF ADM: A case study of a heavy equipment rental service company. *Applied Information System and Management*. 2025. Vol. 8. N 2. P. 183-194. DOI: 10.15408/aism.v8i2.46335.
33. **Gonga Y., Janssen M.** The value of and myths about enterprise architecture. *International Journal of Information Management*. 2019. N 46. P. 1-9. DOI: 10.1016/j.ijinfomgt.2018.11.006.
34. **Van den Berg M., Slot R., van Steenbergen M., Faasse P., van Vliet H.** How enterprise architecture improves the quality of IT investment decisions. *The Journal of Systems and Software*. 2019. N 152. P. 134-150. DOI: 10.1016/j.jss.2019.02.053.
35. **Niemi E., Pekkola S.** The benefits of enterprise architecture in organizational transformation. *Bus Inf Syst Eng*. 2020. N 62(6). P. 585-597. DOI: 10.1007/s12599-019-00605-3.
36. **Kotusev S., Kurnia Sh., Dillnutt R., Taylor P.** Can enterprise architecture be based on the business strategy? *Proceeding of the 53rd Hawaii International Conference on System Sciences*. 2020. P. 5613-5622.
37. **Dang D., Pekkola S.** Institutional perspectives on the process of enterprise architecture adoption. *Information Systems Frontiers*. 2020. N 22. P. 1433-1445.
38. **Репина С.И.** Проверка качества кластеров с помощью силуэтного анализа. *Экономика и социум*. 2024. № 9(124). С. 958-975. DOI: 10.5281/zenodo.13918451.
23. **Liao M.-H., Wang C.-T.** Using enterprise architecture to integrate lean manufacturing, digitalization, and sustainability: A lean enterprise case study in the chemical industry. *Sustainability*. 2021. N 13. 4851. DOI: 10.3390/su13094851.
24. **Afanasyeva S.I., Chernysheva K.V.** Digital transformation using customer relationship management systems. *Modern high technologies. Regional application*. 2024. N 1(77). P. 60-97. DOI: 10.6060/snt.20247701.0008. (in Russian).
25. **Vandevenne N., Van Riel J., Poels G.** Green enterprise architecture (GREAN) – Leveraging EA for environmentally sustainable digital transformation. *Sustainability*. 2023. N 15. 14342. DOI: 10.3390/su151914342EA.
26. **Anisiforov A.B.** Features of digital transformation projects of industrial enterprises and some aspects of preparation for their implementation. *Scientific journal NRU ITMO. Series «Economics and Environmental Management»*. 2022. N 2. P. 32-41. DOI: 10.17586/2310-1172-2022-16-2-32-41. (in Russian).
27. **Oberlea M., Yesilyurta O., Schleretha A., Rislinga M., Schela D.** Enterprise IT architecture greenfield design combining IEC 62264 and TOGAF by example of battery manufacturing. *Procedia Computer Science*. 2023. N 217. P. 136-146. DOI: 10.1016/j.procs.2022.12.209.
28. **Ilin I.V., Levina A.I., Dubgorn A.S.** Digital transformation as a factor of the enterprise architecture and IT architecture development. *Scientific journal NRU ITMO. Series «Economics and Environmental Management»*. 2019. N 3. P. 50-55. DOI: 10.17586/2310-1172-2019-12-3-50-55. (in Russian).
29. **Laumann F., Tambo T.** Enterprise architecture for a facilitated transformation from a linear to a circular economy. *Sustainability*. 2018. N 10. 3882. DOI:10.3390/su10113882.
30. **Tamm T., Seddon P.B., Shanks G.** How enterprise architecture leads to organisational benefits. *International Journal of Information Management*. 2022. N 67. 102554. DOI: 10.1016/j.ijinfomgt.2022.102554.
31. **Denisov A.R., Miridashtaki V.** Reference model for a company's strategic development subsystem based on the principles of enterprise architecture. *Ivecofin*. 2025. N 3 (65). P. 85-93. DOI: 10.6060/ivecofin.2025653.735. (in Russian).
32. **Saffranto R.** Enterprise architecture planning with TOGAF ADM: A case study of a heavy equipment rental service company. *Applied Information System and Management*. 2025. Vol. 8. N 2. P. 183-194. DOI: 10.15408/aism.v8i2.46335.
33. **Gonga Y., Janssen M.** The value of and myths about enterprise architecture. *International Journal of Information Management*. 2019. N 46. P. 1-9. DOI: 10.1016/j.ijinfomgt.2018.11.006.
34. **Van den Berg M., Slot R., van Steenbergen M., Faasse P., van Vliet H.** How enterprise architecture improves the quality of IT investment decisions. *The Journal of Systems and Software*. 2019. N 152. P. 134-150. DOI: 10.1016/j.jss.2019.02.053.
35. **Niemi E., Pekkola S.** The benefits of enterprise architecture in organizational transformation. *Bus Inf Syst Eng*. 2020. N 62(6). P. 585-597. DOI: 10.1007/s12599-019-00605-3.
36. **Kotusev S., Kurnia Sh., Dillnutt R., Taylor P.** Can enterprise architecture be based on the business strategy? *Proceeding of the 53rd Hawaii International Conference on System Sciences*. 2020. P. 5613-5622.
37. **Dang D., Pekkola S.** Institutional perspectives on the process of enterprise architecture adoption. *Information Systems Frontiers*. 2020. N 22. P. 1433-1445.
38. **Repina S.I.** Verification of the quality of clusters using silhouette analysis. *Economy and society*. 2024. N 9(124). P. 958-975. DOI: 10.5281/zenodo.13918451. (in Russian).

Поступила в редакцию 10.02.2026
Принята к опубликованию 24.02.2026

Received 10.02.2026
Accepted 24.02.2026