

МРНТИ 67.01.33

З.Е. Кабжан¹ – основной автор, ©
А.Ю. Шахнович², Н.Т. Шогелова³,
Е.Д. Глызно⁴, С.В. Исаенко⁵

^{1,3,4,5}Магистр, ²Канд. техн. наук

ORCID

¹<https://orcid.org/0009-0004-9957-0034> ²<https://orcid.org/0009-0009-5128-6970>³<https://orcid.org/0000-0002-5220-1459> ⁴<https://orcid.org/0000-0001-9914-0764>⁵<https://orcid.org/0009-0005-8336-4428>^{1,2,3,4}Казахский научно-исследовательский и проектный институт
строительства и архитектуры, г. Алматы, Казахстан⁵ТОО «QazCode», г. Астана, Казахстан³nazymshogelova@gmail.com<https://doi.org/10.55956/EVGB7641>

ФОРМИРОВАНИЕ ЦИФРОВОГО ТЕЗАУРУСА СТРОИТЕЛЬНОЙ ОТРАСЛИ: МЕТОДОЛОГИЯ И РЕАЛИЗАЦИЯ

Аннотация. Современное развитие строительной отрасли требует унификации и стандартизации терминологии, особенно в условиях цифровизации и внедрения технологий информационного моделирования. В статье рассматривается процесс формирования цифрового тезауруса, направленного на систематизацию понятий, используемых в нормативно-правовой и нормативно-технической документации. Проведён анализ 1566 нормативных документов, что позволило выделить 4210 существующих терминов и 10226 терминов-кандидатов. Для обработки данных применялись методы частотного анализа, лингвистической обработки и экспертной оценки. В качестве платформы управления тезаурусом была выбрана Tematres, обеспечивающая структурированное представление терминов, поддержку многоязычности и интеграцию с международными стандартами SKOS и RDF. Среди преимуществ платформы – открытый исходный код, гибкость настройки, автоматизированный поиск и поддержка эквивалентных терминов. Однако выявлены ограничения: отсутствие инструментов визуализации семантических связей и ограниченная поддержка онтологических моделей. Полученные результаты подтверждают эффективность предложенного подхода для автоматизации нормативного контроля и интеллектуального поиска. Разработанный тезаурус может быть интегрирован в цифровые системы проектирования и экспертизы, способствуя устранению терминологических коллизий и повышению точности нормативных требований.

Ключевые слова: цифровой тезаурус, строительная отрасль, стандартизация терминологии, нормативно-техническая документация, автоматизированный анализ данных, Tematres.



Кабжан, З.Е. Формирование цифрового тезауруса строительной отрасли: методология и реализация [Текст] / З.Е. Кабжан, А.Ю. Шахнович, Н.Т. Шогелова, Е.Д. Глызно, С.В. Исаенко // *Механика и технологии* / Научный журнал. – 2025. – №2(88). – С.284-299. <https://doi.org/10.55956/EVGB7641>

Введение. Цифровая трансформация строительной отрасли является ключевым направлением модернизации в условиях активного развития

284

информационных технологий. Одним из важных элементов этого процесса является создание цифровых инструментов, обеспечивающих структурирование и стандартизацию терминологической базы. Важность единого терминологического аппарата обусловлена необходимостью унификации нормативных требований, автоматизации процессов проектирования и экспертизы, а также интеграции технологий информационного моделирования строительных объектов (ТИМСО, BIM) [1]. В международной практике использование цифровых тезаурусов позволяет обеспечить единообразие в трактовке терминов и улучшить процессы автоматизированной обработки данных, что способствует повышению качества проектной документации и снижению рисков, связанных с человеческим фактором [2].

Формирование цифрового тезауруса строительной отрасли представляет собой сложную задачу, требующую системного подхода к отбору, классификации и интерпретации терминов, используемых в нормативно-правовых актах (НПА) и нормативно-технической документации (НТД). Внедрение цифрового тезауруса позволяет устранить терминологические коллизии, повысить точность правоприменительной практики, а также создать единую основу для автоматизированных проверок информационных моделей строительных объектов. Аналогичные проекты по созданию отраслевых тезаурусов в международной практике уже доказали свою эффективность в таких областях, как медицина, экология и строительство [3].

В статье Sunny 2017 года проведен систематический обзор литературы, целью которого является исследование применения и потенциальных функций тезаурусов в цифровых библиотеках [4]. В результате анализа 29 исследований из 466 публикаций, извлеченных из Library and Information Science Abstracts (LISA) и Scopus, были выявлены различные подходы к выбору и применению тезаурусов, а также их роль в улучшении поиска и навигации для пользователей цифровых библиотек. К достоинствам исследования можно отнести предоставление рекомендаций для администраторов цифровых библиотек, однако ограничение заключается в узкой выборке работ, что может повлиять на универсальность выводов.

Тезаурус документального языка, как средство стандартизации терминов, является важной разработкой для эффективного управления документами, особенно при использовании специализированных программных приложений, таких как TemaTres. TemaTres представляет собой сервер контролируемого словаря, позволяющий организовать и управлять терминологией. Это бесплатное веб-приложение предоставляет постоянный доступ к коллекциям документов, сгруппированных по темам, и поддерживает процессы анализа, исследований и принятия решений. TemaTres обеспечивает гибкую установку, функциональный интерфейс и широкие возможности визуализации, что делает его удобным инструментом для создания и управления контролируемыми словарями.

Gonzales-Aguilar в своей работе рассматривают значение тезауруса документального языка, как стандартизированного набора терминов, для управления документами с использованием программных приложений, таких как TemaTres, который служит сервером контролируемого словаря [5]. TemaTres, будучи бесплатным веб-приложением, обеспечивает постоянный доступ к тематическим наборам документов, поддерживая исследовательскую и аналитическую деятельность. Основным достоинством

является удобство установки и функциональность для визуального представления и управления словарями, однако ограничения могут возникать из-за узкой специализации приложения и необходимости точной настройки для разных отраслей.

В работе Rejowski рассматривается процесс создания Бразильского тезауруса туризма на основе анализа диссертаций по туризму, изданной литературы и мнения специалистов, собранного в 2017 году [6]. Тезаурус включает 17 категорий, в которых термины структурированы по концептуальным связям в области туризма, а для его онлайн-реализации был использован TemaTres. К достоинствам работы можно отнести междисциплинарный подход и многофункциональность тезауруса как исследовательского инструмента, однако сложность терминологической структуры может затруднять его применение в специфических исследованиях.

В статье Berg описан процесс создания цифрового архива SACDA, предназначенного для хранения и классификации материалов, связанных с южноазиатской диаспорой в Канаде, и разработка специализированного тезауруса для структурирования данных архива [7]. SACDA сотрудничает с различными учреждениями и организациями для оцифровки и каталогизации культурного наследия, заполняя существующие таксономические пробелы и обогащая текущие тезаурусы, такие как AIATSIS и Международный тезаурус терминологии по беженцам. Одним из основных достижений проекта стало создание тезауруса, который учитывает особенности южноазиатского наследия в канадском контексте, однако возможным ограничением является сложность интеграции с другими международными стандартами тезаурусов.

В рамках работы Autiero описывается создание «Тезауруса для оцифровки артефактов Гандхары», цифрового ресурса для описания искусства Гандхары и буддийского искусства в целом [8]. Проект DiGA реализовал семиступенчатый процесс, включающий сбор источников, оцифровку ключевых материалов, создание иерархии концепций для сюжетов и персонажей, а также сравнение с другими существующими тезаурусами. В результате создания тезауруса был сделан значительный шаг в изучении искусства Гандхары, открывший новые перспективы как в исследовательской, так и в цифровой среде. Однако одной из трудностей проекта стало обеспечение совместимости с международными стандартами и интеграции тезауруса в более широкие цифровые исследования.

В статье Zeng исследуется автоматизированное создание многоязычных тезаурусов на основе ресурсов цифровых библиотек, с акцентом на извлечение терминов из многоязычного параллельного корпуса [9]. В работе предложены методы автоматического извлечения и фильтрации терминов, построения взаимосвязей между ними, а также сопоставления терминов между китайским и другими языками. Основными достижениями исследования стали разработка инновационного подхода, основанного на сочетании лингвистического и статистического анализа, и улучшение качества автоматического сопоставления терминов по сравнению с традиционными методами. Ограничением исследования является тестирование на платформе Java, что может затруднить интеграцию метода в другие системы.

Создание цифрового тезауруса строительной отрасли обусловлено необходимостью стандартизации терминологии и унификации понятий в нормативно-правовой базе Республики Казахстан. В условиях активного

развития технологий информационного моделирования (ТИМСО, BIM) цифровизация нормативных требований играет ключевую роль в автоматизации процессов проектирования, экспертизы и управления строительными объектами. Формирование тезауруса направлено на обеспечение однозначной интерпретации терминов, устранение терминологических коллизий, а также оптимизацию процессов информационного поиска и верификации строительных моделей [10].

Целью данной статьи является разработка и обоснование методологии формирования цифрового тезауруса строительной отрасли на основе анализа НПА и НТД. В работе рассматриваются основные этапы создания цифрового тезауруса, включая систематизацию терминологии, устранение терминологических коллизий, унификацию определений и внедрение электронной базы данных для поддержки автоматизированных процессов проектирования и экспертизы. Кроме того, статья направлена на выявление ключевых преимуществ цифрового тезауруса в контексте цифровизации строительной отрасли, а также анализ возможных трудностей, возникающих при его разработке и применении.

Условия и методы исследований. Исходной базой для разработки цифрового тезауруса послужил корпус нормативных документов строительной отрасли, включающий 1566 единиц актуальных и устаревших НПА и НТД. В рамках исследования был проведен их комплексный анализ, включающий автоматизированные и экспертные методы обработки текстовых данных.

Анализ текстового содержимого осуществлялся с использованием методов статистического и лингвистического анализа, направленных на выявление отраслевых терминов. Выделенные термины были проверены на соответствие нормативным требованиям, после чего была проведена их классификация по тематическим группам и установлены семантические связи между ними. В ходе работы исключались дублирующиеся, устаревшие и нерелевантные термины, а для терминов, не имеющих четких определений, формировались уникальные дефиниции с привлечением экспертного сообщества.

Результатом исследования стало формирование унифицированной терминологической базы строительной отрасли, представленной в цифровом формате. Структурированный тезаурус включает термины, их определения, семантические связи и ссылки на соответствующие нормативные документы. Полученные результаты могут быть использованы для автоматизированной обработки проектной документации, интеллектуального поиска, а также для интеграции в системы цифровой экспертизы строительных объектов.

Результаты исследований. Формирование цифрового тезауруса строительной отрасли основывается на комплексном анализе существующей нормативно-правовой и нормативно-технической документации. В рамках исследования была реализована многоэтапная процедура идентификации, классификации и верификации терминологии с применением автоматизированных методов анализа текстов и экспертной оценки [11].

Анализ корпуса документов позволил выделить две основные группы терминов: существующие термины, зафиксированные в НПА и НТД (действующих или отмененных), и потенциальные термины, выявленные на основе статистической обработки текстов. Список существующих терминов был сформирован полуавтоматическим способом путем извлечения данных из разделов «Термины и определения» нормативных документов, что

позволило идентифицировать 19056 терминов. При этом предварительная выборка терминов выполнялась по заданным лингвистическим шаблонам, после чего осуществлялась ручная проверка и верификация. В процессе обработки дублирующих определений и терминов, представленных в нескольких нормативных источниках, окончательное количество существующих терминов было сокращено до 12080 единиц, из которых в основу тезауруса вошло 4210 терминов.

Выявление потенциальных терминов осуществлялось на основе анализа текстового содержимого нормативных документов с использованием методов частотного анализа и морфологического разложения. В результате предварительного автоматизированного анализа было выделено 570901 единица, включающих как термины, так и словосочетания, не обладающие точным концептуальным закреплением. Для повышения точности отбора был разработан метод ранжирования терминов по критериям значимости, включающим отсутствие или наличие определения, принадлежность к общему словарю русского языка, а также пороговые значения частотного распределения. В ходе ранжирования терминов и их последующей экспертной оценки количество потенциальных терминов было сокращено до 527363, из которых 10226 терминов-кандидатов были включены в дальнейшую классификационную обработку.

Дополнительно был проведен анализ «Пилотного набора» из 36 нормативных документов, что позволило идентифицировать 7272 уникальных термина с четкими определениями, а также 10226 терминов-кандидатов, требующих дальнейшей экспертной проработки. Из этого числа 813 терминов ранее не встречались в нормативной базе, что подтверждает необходимость продолжения исследования в направлении детального анализа их контекстного использования.

Одной из ключевых задач формирования цифрового тезауруса являлась тематическая классификация терминологии по предметным областям. В рамках исследования были выделены категории знаний, включающие фундаментальные дисциплины, архитектуру и проектирование, инженерные сети, транспортные инфраструктуры, строительные материалы и механизмы, что позволило структурировать терминологию в логически взаимосвязанную систему. Для автоматизации классификации были разработаны списки якорных терминов и частотные модели, позволившие предварительно отнести 33187 потенциальных терминов к определенной предметной области.

В результате проведенной работы сформирован структурированный цифровой тезаурус строительной отрасли, включающий унифицированные термины, их определения, тематическую классификацию и ссылки на соответствующие нормативные документы. Разработанная методология обеспечивает высокую точность выделения и обработки терминов, что делает возможным использование тезауруса в системах цифровой экспертизы, автоматизированного проектирования и нормативного контроля.

На текущем этапе остается ряд задач, связанных с уточнением терминологической базы, выявлением новых терминов по мере актуализации нормативных документов, а также интеграцией тезауруса в цифровые платформы строительной отрасли. Полученные результаты подтверждают эффективность выбранного подхода и создают основу для дальнейшего развития цифровых инструментов автоматизированного анализа нормативных требований в строительстве.

В рамках дальнейшей работы над тезаурусом строительной отрасли была реализована задача систематизации терминологии путем их классификации по тематическим группам. В отличие от традиционной алфавитной организации словарей, была разработана функциональная система классификации, ориентированная на практическое применение в условиях автоматизации процессов проектирования, экспертизы и цифрового управления нормативными документами.

С целью повышения функциональности цифрового тезауруса была проведена систематическая группировка терминов на основе структур, регламентированных Системой проектной документации для строительства (СПДС). Данный подход позволил сформировать логически структурированную терминологическую базу, где термины были распределены по соответствующим областям знаний. При этом дополнительно выделена группа терминов общего характера, охватывающих многопрофильные области и базовые инженерные дисциплины.

После первичной классификации терминов была проведена работа с экспертным сообществом, в ходе которой осуществлялось распределение терминов с учетом требований ЕСККС (Единая система классификации и кодирования строительных ресурсов) [12]. Использование данной системы предусмотрено нормативной базой и Концепцией внедрения технологий информационного моделирования строительных объектов (ТИМСО) [13], что делает классификацию терминов цифрового тезауруса согласованной с международными требованиями. Однако адаптация структуры ЕСККС под задачи формирования тезауруса потребовала дополнительной методологической проработки и анализа.

В строительной отрасли терминология должна соответствовать особенностям инженерного проектирования и содержать взаимосвязи между компонентами строительных объектов и специализированными областями знаний. Например, при проектировании колонны требуется владение знаниями в таких областях, как строительная механика, железобетонные конструкции, строительные материалы, нагрузки и воздействия. В свою очередь, при проектировании технической системы, такой как водонапорная башня, необходимо привлечение специалистов по гидравлике, металлическим конструкциям, нагрузкам и фундаментам. Таким образом, для каждого элемента или системы возможно определить перечень знаний, необходимых для проектирования и эксплуатации [14].

На основании этого анализа вся терминология была условно разделена на две основные группы:

- Политехнические и описательные термины, являющиеся единообразными и понятными для всех участников строительного процесса;
- Прикладные термины, требующие детального отнесения к специализированным областям знаний.

Применение данного принципа можно проиллюстрировать на примере термина «колонна». Для всех специалистов в строительстве он представляет собой вертикальную несущую опору. Однако с точки зрения инженера-конструктора, использующего методы расчета по железобетонным конструкциям и строительной механике, колонна определяется как внецентренно сжатый элемент, что уточняет инженерное поведение. Включение такой дифференциации терминов требует участия экспертного сообщества и может быть реализовано в несколько этапов.

В результате проведенной аналитической работы была сформирована базовая классификация терминов строительной отрасли, обеспечивающая логическое структурирование информации. Термины в тезаурусе представляются преимущественно в виде имени существительного в именительном падеже либо в форме определительного словосочетания. В зависимости от структуры словосочетания были выделены следующие типы терминов:

- Согласованные словосочетания (A + N) – прилагательное стоит перед существительным (например: главный фасад, ландшафтный дизайн, астрономический азимут).

- Несогласованные словосочетания (N + N) – определение выражено существительным в косвенном падеже (система водоотведения потребителя, рециркуляция сточных вод).

- Для систематизации структуры терминов была предложена буквенная кодировка:

- N – существительное (например, технология, виброизолятор).

- AN – согласованное словосочетание (например, водопроводная труба).

- NN – несогласованное словосочетание (например, проверка приборов учета).

- AAN, ANN, NNN, NAN – сложные словосочетания с более глубокой структурой (нормальная скорость распространения пламени, инженерные системы жизнеобеспечения здания).

При анализе нормативных документов было выявлено, что термин, состоящий из одного слова (существительного), чаще всего обозначает общее понятие или строительное изделие (например, труба, линия, система). Однако для уточнения значения термина и его принадлежности к конкретной области знаний требуется добавление дополнительных слов (водопроводная труба, линия электропередачи, система газоснабжения).

Определительная часть словосочетания может отражать различные семантические связи между терминами:

- «имеет» (многопустотная плита – железобетонная плита с пустотами).

- «состоит из материала» (деревянная колонна – колонна, выполненная из дерева).

- «монтируется по технологии» (железобетонная свая – свая, которая устанавливается методом вибропогружения или ударной забивки).

- «воспринимает нагрузки» (опорная плита, несущая стена).

Дополнительно был проведен анализ составных терминов (знак безопасности с внешней подсветкой, эвакуация людей на пожаре), которые представляют собой комбинацию двух или более понятий. Для таких терминов требуется отдельное решение о включении в тезаурус в исходном виде или разбиении на более простые элементы.

С учетом анализа структуры терминов был предложен стандарт представления данных в цифровом тезаурусе:

- Термины подаются в нижнем регистре.

- Используется единственное число, за исключением случаев, когда множественное число является нормативным (сточные воды, строительные конструкции).

– Рекомендуемая длина термина – не более трех слов, если не требуется более детализированное определение.

После структурирования терминов в цифровом тезаурусе строительной отрасли был проведен комплексный процесс поиска и верификации их определений, направленный на исключение терминологических коллизий и обеспечение единообразия понятийного аппарата.

В ходе работы была разработана иерархия приоритетности источников нормативных определений:

– В случае наличия нескольких определений к одному термину из разных нормативных документов, приоритет отдавался определению из документа более высокого уровня иерархии. Например, если термин присутствовал в СП РК и РДС РК, то в тезаурус включалось определение, зафиксированное в РДС РК.

– Если термин имел разные определения в равнозначных нормативных источниках, окончательный вариант определения утверждался экспертной группой, после чего проходил дополнительную верификацию рецензентами-экспертами строительной отрасли.

– Если нормативные документы Республики Казахстан не содержали определения к термину, поиск осуществлялся в специализированных технических, строительных и политехнических словарях, справочниках и учебных пособиях.

– В случае отсутствия определения даже в специализированных источниках, термин проходил этап экспертной разработки, в ходе которого формировалась дефиниция при участии специалистов в области строительства, а окончательная формулировка утверждалась экспертной группой.

В качестве примера экспертного подхода к формированию определений можно рассмотреть термин «зонное водоснабжение», определение которого было выведено на основе уже существующего термина «зонная система водопровода». Первоначальное определение «зонная система водопровода» – «система водопровода, разделенная на части по предельно-допустимым гидростатическим напорам» – стало основой для новой дефиниции, которая была адаптирована с учетом контекста использования термина: «использование для водоснабжения здания нескольких водопроводных систем, разделенных на части по предельно-допустимым гидростатическим напорам». Такой методологический подход позволил создать логически выверенные и нормативно корректные определения терминов.

В процессе формирования тезауруса были выявлены случаи дублирования определений, вызванные разделением синонимичных терминов на две отдельные единицы. Например, термин «зиговое соединение (зиг)», определяемый как «разновидность фальцевого соединения, выполненного из листового металлического материала в виде буквы «З»» [15], в тезаурусе был представлен в виде двух отдельных терминов: «зиговое соединение» и «зиг», имеющих идентичные определения.

Для оптимизации структуры тезауруса такие случаи были проанализированы с точки зрения семантических взаимосвязей. Было принято решение обозначать более полный термин (например, «зиговое соединение») как дескриптор, то есть основной термин, рекомендованный к употреблению в нормативно-правовых актах и нормативно-технической документации. В то же время сокращенные или разговорные версии терминов (например, «зиг») обозначались как аскрипторы, то есть термины-

синонимы, не рекомендованные для использования в нормативных источниках, но учитываемые в цифровом тезаурусе для поиска и анализа взаимосвязей между понятиями.

Для обеспечения объективной, квалифицированной и независимой оценки терминологической базы была сформирована экспертная группа, в состав которой вошли высококвалифицированные специалисты строительной отрасли. Экспертная группа осуществляла рецензирование тезауруса, а также консультативную и методологическую поддержку в вопросах верификации и согласования терминологии.

Экспертам был предоставлен доступ к актуальной редакции перечня терминов, распределенных по предметным областям, соответствующим их профессиональной компетенции. В круг задач экспертов входили:

- рассмотрение и согласование классификации предметных областей в строительстве;
- анализ сформированного перечня терминов, выявление возможных ошибок и дублирующихся элементов;
- консультирование по вопросам терминологии и нормативных определений;
- рекомендации по дополнению перечня терминами, ранее не включенными в классификацию;
- рекомендации по исключению или перемещению терминов в более релевантные предметные области;
- согласование и утверждение корректных определений терминов, а также формулирование новых определений для терминов, не имеющих нормативного закрепления.

Процесс экспертного рецензирования носил итеративный характер, то есть проводился в несколько последовательных этапов. По мере уточнения терминологической базы количество итераций обсуждения тезауруса увеличивалось, что позволило добиться высокого качества содержания и минимизировать терминологические коллизии.

Разработанная методология и итерационный подход к созданию тезауруса позволили не только сформировать структурированную терминологическую систему, но и создать инструмент, адаптированный для цифрового применения в процессах нормативного контроля, автоматизированной экспертизы и информационного поиска в строительной отрасли.

Обсуждение научных результатов. Современные технологические решения для управления тезаурусами представляют собой широкий спектр инструментов, ориентированных на автоматизацию процессов разработки, поддержки и актуализации терминологических данных. В ходе исследования проведен анализ наиболее распространенных программных решений, предназначенных для управления цифровыми тезаурусами, их функциональных возможностей, преимуществ и ограничений. Данный анализ позволил определить их применимость в рамках разработки и внедрения тезауруса строительной отрасли, а также выбрать оптимальное программное обеспечение для его реализации [16].

Выбор платформы Tematres в качестве инструмента для хранения и управления цифровым тезаурусом строительной отрасли обоснован совокупностью факторов, включающих функциональные характеристики, экономическую эффективность и адаптивность платформы к требованиям

проекта [5]. Использование Tematres обусловлено следующими преимуществами:

- Экономическая доступность. Платформа представляет собой программное обеспечение с открытым исходным кодом, что исключает необходимость приобретения лицензий и позволяет минимизировать финансовые издержки при внедрении, особенно на начальном этапе проекта.

- Соответствие международным стандартам. Tematres поддерживает стандарты SKOS (Simple Knowledge Organization System) и RDF (Resource Description Framework), что обеспечивает интеграцию тезауруса в международные информационные системы и облегчает обмен данными с другими платформами.

- Многоязычная поддержка. Платформа предоставляет возможности для работы с несколькими языками, включая казахский и русский, что особенно актуально для реализации проекта в условиях многоязычной среды Казахстана.

- Интуитивно понятный интерфейс. Простота освоения и удобство использования системы позволяют минимизировать временные затраты на обучение пользователей и обеспечивают быструю интеграцию в рабочие процессы.

- Гибкость настройки. Открытый исходный код платформы позволяет адаптировать систему под конкретные требования проекта, включая структуру тезауруса, параметры интерфейса и функции экспорта данных.

Однако при выборе платформы Tematres необходимо учитывать ряд функциональных ограничений, которые могут повлиять на эффективность использования в проектах с особыми требованиями. Во-первых, ограниченные аналитические возможности платформы не позволяют выполнять сложный анализ данных и визуализировать семантические связи между терминами. Это может создавать определённые затруднения при её применении в проектах, где требуется детальный семантический анализ или углубленная работа с взаимосвязями терминов. Во-вторых, Tematres не поддерживает сложные онтологические модели, поскольку изначально ориентирована на управление тезаурусами. В частности, платформа не предусматривает использование OWL (Web Ontology Language) – международного стандарта, необходимого для построения полноценных онтологических систем. Это ограничение может быть критичным в тех случаях, когда требуется моделирование сложных концептуальных структур. В-третьих, технические требования к серверной инфраструктуре также являются важным фактором при внедрении Tematres. Для полноценного развертывания и эксплуатации платформы требуется базовый сервер, что может потребовать привлечения технических специалистов, особенно при отсутствии внутренних ресурсов для поддержки и администрирования системы.

Несмотря на указанные ограничения, Tematres остается гибким инструментом, который благодаря открытому исходному коду может адаптироваться к текущим потребностям проекта. Кроме того, платформа может рассматриваться как промежуточное решение, позволяющее на ранних этапах стандартизации и структурирования терминологической базы заложить основу для дальнейшего перехода к более сложным и функциональным инструментам управления знаниями [17].

Для хранения и управления тезаурусом был внедрен специализированный электронный ресурс Tematres, который предоставляет

функциональные возможности для структурированного поиска, классификации и отображения терминов. Итоговый перечень терминов с их определениями размещен в открытом доступе на веб-платформе (рис. 1): <http://tematres.kazniisa.kz>.

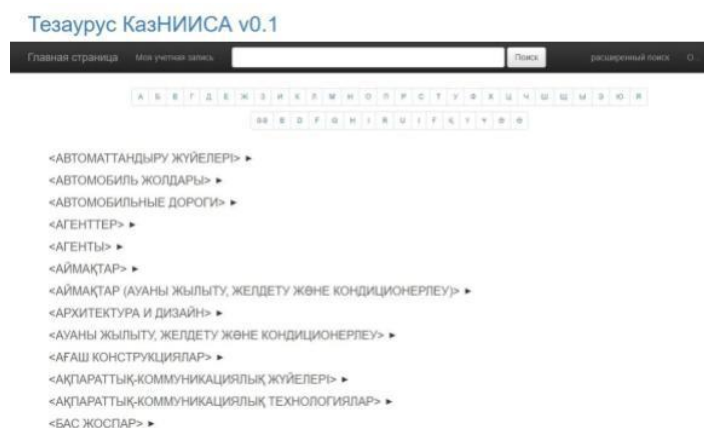


Рис. 1. Главная страница электронного ресурса Tematres

Платформа позволяет осуществлять:

- поиск терминов по ключевым словам;
- просмотр терминов по предметным областям и в алфавитном порядке;
- отображение эквивалентных терминов на втором языке тезауруса (казахский/русский);
- переход на страницу термина, где доступны его определение, источник нормативного закрепления и тематическая область.

Разработанный цифровой тезаурус строительной отрасли, реализованный на базе Tematres, обеспечивает систематизацию и унификацию строительной терминологии, создавая основу для интеграции в процессы цифрового проектирования, экспертизы и нормативного контроля.

Главная страница электронного ресурса (рис. 1) предоставляет пользователям следующие инструменты навигации и поиска:

- поисковая строка – для быстрого поиска терминов по ключевым словам;
- кнопка перехода на главную страницу – обеспечивает возврат к стартовой странице тезауруса;
- кнопка «Моя учетная запись» – предоставляет пользователям доступ к персонализированным функциям платформы;
- кнопка расширенного поиска – позволяет проводить поиск с учетом различных параметров (по категориям, частичному совпадению и т. д.);
- кнопка перехода к общей информации о тезаурусе – содержит сведения о целях, методах формирования и использовании тезауруса;
- кнопки алфавитного указателя статей тезауруса – обеспечивают доступ к терминам в алфавитном порядке;
- список предметных областей – структурированное отображение терминов в соответствии с их тематической принадлежностью.

При выборе пользователем определенной предметной области система отображает перечень терминов, относящихся к данной категории,

представленный в алфавитном порядке. Для каждого термина доступен переход на его индивидуальную страницу с детализированной информацией.

На странице конкретного термина отображаются следующие элементы (рис. 2):

- наименование термина – основной элемент страницы;
- кнопки «Термин» и «Metadata» – обеспечивают доступ к метаданным, связанным с данным термином;
- кнопка «Definition note» – раскрывает определение термина при нажатии;
- кнопка «Source document» – предоставляет информацию о нормативном или справочном источнике, из которого заимствовано определение;
- предметная область – категория, к которой относится термин (например, область «Агенты»);
- эквивалентный термин на втором языке тезауруса – для русскоязычных терминов отображается их казахский эквивалент, а для казахоязычных терминов – русскоязычный аналог.

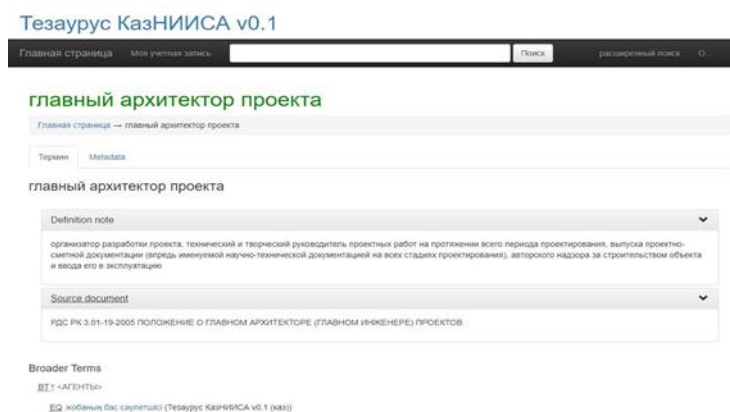


Рис. 2. Страница термина в Тезаурусе

На главной странице платформы реализован алфавитный указатель, обеспечивающий навигацию по терминологической базе. Выбор буквы алфавита позволяет отобразить перечень терминов, начинающихся на соответствующую букву, с указанием их эквивалентов на втором языке тезауруса. При выборе конкретного термина осуществляется переход на его страницу, где представлена полная информация о термине, его определении, источнике и связях с другими терминами.

Использование платформы Tematres позволяет систематизировать терминологическую базу строительной отрасли, обеспечивая доступность, структурированность и удобство поиска информации. Электронно-цифровая форма представления тезауруса способствует его интеграции в информационные системы строительного проектирования, нормативного контроля и автоматизированного анализа, тем самым поддерживая процессы цифровизации отрасли.

Закключение. Формирование цифрового тезауруса строительной отрасли представляет собой значимый этап в развитии системы терминологической стандартизации, обеспечивающей унификацию понятийного аппарата и повышение эффективности нормативного

регулирования. В ходе проведённой работы был реализован комплекс мероприятий по систематизации терминов, включающих анализ нормативно-правовой и нормативно-технической документации, статистическую обработку текстов, выявление и классификацию терминологии, а также экспертную оценку и верификацию данных.

В рамках проекта была проведена статистическая обработка 1566 нормативных документов, что позволило выделить 4210 существующих терминов с определениями, включённых в цифровой тезаурус. Дополнительно были идентифицированы 7870 проблемных терминов, которые не были включены в тезаурус по причине неопределённости их трактовки или несоответствия отраслевой классификации. Также был сформирован перечень 10226 терминов-кандидатов, которые прошли процедуру экспертной оценки и частично были включены в тезаурус.

Дополнительно были обработаны 36 нормативных документов, в результате чего было выявлено 813 терминов, внесённых в тезаурус. В результате экспертной оценки и анализа текстов НПА и НТД было отобрано 2249 терминов, что в совокупности позволило сформировать единый цифровой тезаурус, включающий 7272 термина, распределённых по 81 предметной области.

Разработанный тезаурус является не только инструментом унификации строительной терминологии, но и основой для автоматизированной обработки нормативной документации, интеллектуального поиска и интеграции в цифровые платформы проектирования и экспертизы. Внедрение электронного ресурса Tematres в качестве базы данных обеспечило удобный доступ к терминологической информации и структурированное представление.

Реализация цифрового тезауруса строительной отрасли создаёт основу для дальнейшей цифровизации нормативно-технического регулирования, повышения точности правоприменительных практик и поддержки процессов автоматизированного анализа строительной документации.

Список литературы

1. Яскевич, В. Анализ национального опыта внедрения BIM [Текст] / В. Яскевич, Б. Куспангалиев, А. Ходжиков // Проект Байкал Учредители: Григорьева Елена Ивановна. – 2023. – Т. 20. – № 75. – С. 105-111.
2. Доржеева, О.А. Тезаурусное моделирование терминологии строительства [Текст] / О.А. Доржеева // Вестник Челябинского государственного университета. – 2013. – № 31 (322). – С. 85-87.
3. Абдурахманова, А.З. Терминология строительства как объект тезаурусного моделирования [Текст] / А.З. Абдурахманова // Ученые записки Крымского федерального университета имени В.И. Вернадского. Филологические науки. – 2016. – № 2 (68) (2-1). – С. 17-23.
4. Sunny S.K., Angadi M. Applications of thesaurus in digital libraries // DESIDOC Journal of Library and Information Technology. Defence Scientific Information and Documentation Centre. – 2017.
5. Gonzales-Aguilar A., Ramirez-Posada M., Ferreyra D. TemaTres: software para gestionar tesauros // Profesional de La Informacion. – 2012. – No. 21(3). – P. 319-325.
6. Rejowski M., Barbanti C.H. Construção de um Tesouro Brasileiro de Turismo // Revista Turismo Em Análise. – 2019. – No. 29 (2). – P. 182-195.
7. Berg M., Bains S., Suri S. South Asian Canadian Digital Archive Thesaurus // KULA: Knowledge Creation, Dissemination, and Preservation Studies. – 2022. – No. 6 (3). – P. 1-7.

8. Autiero S., Elwert F., Moscatelli C., Pons J. The Seven Steps: Building the DiGA Thesaurus // Journal of Open Humanities Data. – 2023. – No. 9.
9. Zeng W. Exploration and study of multilingual thesauri automation construction for digital libraries in China // Electronic Library. – 2012. – No. 30 (2). – P. 233-247.
10. Григорьев, Д.С. Современные взгляды на роль цифровых технологий в обеспечении экономической безопасности строительной отрасли [Текст] / Д.С. Григорьев // Инновации и инвестиции. – 2023. – № 8. – С. 339-344.
11. Zhang J., El-Gohary N. Automated Information Transformation for Automated Regulatory Compliance Checking in Construction // J. Comput. Civ. Eng. – 2015. – No. 29.
12. Национальный научно-исследовательский и проектный институт строительства и архитектуры. Единая система классификации и кодирования строительных ресурсов (ЕСККС) [Электронный ресурс]. – Режим доступа: <https://kazniisa.kz/eskks/>. Дата обращения: 17.02.2025.
13. Zakon.kz. Нормативно-правовой документ [Электронный ресурс]. – Режим доступа: https://online.zakon.kz/Document/?doc_id=39585910. Дата обращения: 17.02.2025
14. Kessler R., Béchet N., Berio G. Extraction of Terminology in the Field of Construction // First International Conference on Digital Data Processing (DDP). – 2019. – No. 22-26.
15. Комитет по делам строительства и жилищно-коммунального хозяйства Министерства регионального развития Республики Казахстан. (2013). СП РК 4.01-102-2013. Внутренние санитарно-технические системы (Введен впервые, соответствует СП РК 4.01-05-2013). Утвержден приказом КДСиЖКХ МРР РК от 28.12.2013 № 407-НҚ (утратившим силу).
16. Martínez-González M., Alvite-Díez M. On the evaluation of thesaurus tools compatible with the Semantic Web // Journal of Information Science. – 2014. – No. 40. – P. 711-722.
17. Mandal S. Integration of External Repositories with TemaTres for Easy Access of Related Terms // Library Philosophy and Practice. – 2021.

Данное исследование выполнено в соответствии с Договором №225-2020-05 от 04.02.2020 года в рамках выполнения государственного задания на оцифровку АГСК 1.1. «Нормативные правовые и нормативные технические документы», реализуемого в рамках республиканской бюджетной программы 225 «Реализация мероприятий по совершенствованию архитектурной, градостроительной и строительной деятельности».

Материал поступил в редакцию 17.02.25, принят 12.05.25.

З.Е. Кабжан¹, А.Ю. Шахнович¹, Н.Т. Шогелова¹, Е.Д. Глызно¹, С.В. Исаенко²

¹Қазақ құрылыс және сәулет ғылыми-зерттеу және жобалау институты,
Алматы қ., Қазақстан

²«QazCode» ЖШС, Астана қ., Қазақстан

ҚҰРЫЛЫС САЛАСЫНЫҢ ЦИФРЛЫҚ ТЕЗАУРУСЫН ҚАЛЫПТАСТЫРУ: ӘДІСТЕМЕСІ ЖӘНЕ ІСКЕ АСЫРУЫ

Аңдатпа. Құрылыс саласының қазіргі даму кезеңі терминологияны біріздендіру мен стандарттауды талап етеді, әсіресе цифрландыру және ақпараттық модельдеу технологияларын енгізу жағдайында. Бұл мақалада нормативтік-құқықтық және нормативтік-техникалық құжаттамада қолданылатын ұғымдарды жүйелеуге бағытталған құрылыс саласының цифрлық тезаурусын қалыптастыру

процесі қарастырылады. 1566 нормативтік құжатқа талдау жүргізіліп, оның нәтижесінде 4210 қолданыстағы термин мен 10226 термин-кандидат анықталды. Деректерді өңдеу үшін жиілік талдауы, лингвистикалық өңдеу және сараптамалық бағалау әдістері қолданылды. Тезаурусты басқару платформасы ретінде Tematres таңдалды, ол терминдерді құрылымдық түрде көрсету, көптілділікті қолдау және SKOS пен RDF халықаралық стандарттарына интеграциялау мүмкіндігін қамтамасыз етеді. Платформаның негізгі артықшылықтарына ашық бастапқы код, икемді баптау параметрлері, автоматтандырылған іздеу және эквивалентті терминдерді қолдау жатады. Алайда, кейбір шектеулер анықталды, соның ішінде семантикалық байланыстарды визуализациялау құралдарының болмауы және онтологиялық модельдерді қолдаудың шектеулігі. Зерттеу нәтижелері ұсынылған әдістің нормативтік бақылауды автоматтандыру және интеллектуалды іздеуді жетілдіру үшін тиімділігін растайды. Қалыптастырылған тезаурус сандық жобалау және сараптау жүйелеріне интеграцияланып, терминологиялық қайшылықтарды жоюға және нормативтік талаптардың дәлдігін арттыруға ықпал етеді.

Тірек сөздер: цифрлық тезаурус, құрылыс саласы, терминологияны стандарттау, нормативтік-техникалық құжаттама, автоматтандырылған деректерді талдау, Tematres.

Z. Kabzhan¹, A. Shakhnovich¹, N. Shogelova¹, Y. Glyzno¹, S. Issayenko²

¹Kazakh Research and Design Institute of Construction and Architecture,
Almaty, Kazakhstan

²LPP «QazCode», Astana, Kazakhstan

FORMATION OF A DIGITAL THESAURUS FOR THE CONSTRUCTION INDUSTRY: METHODOLOGY AND IMPLEMENTATION

Abstract. The modern development of the construction industry requires the unification and standardization of terminology, especially in the context of digitalization and the implementation of Building Information Modeling (BIM) technologies. This article examines the process of forming a digital thesaurus for the construction industry, aimed at systematizing concepts used in regulatory and technical documentation. An analysis of 1566 regulatory documents was conducted, resulting in the identification of 4210 existing terms and 10226 candidate terms. Data processing methods included frequency analysis, linguistic processing, and expert evaluation. The Tematres platform was chosen as the thesaurus management system, providing structured term representation, multilingual support, and integration with international standards SKOS and RDF. The platform's key advantages include open-source accessibility, flexible configuration, automated search, and equivalent term support. However, several limitations were identified, such as the lack of tools for semantic relationship visualization and limited support for ontological models. The obtained results confirm the efficiency of the proposed approach for automating regulatory control and enhancing intelligent search. The developed thesaurus can be integrated into digital design and expert evaluation systems, facilitating the elimination of terminological inconsistencies and improving the accuracy of regulatory requirements.

Keywords: digital thesaurus, construction industry, terminology standardization, regulatory and technical documentation, automated data analysis, Tematres.

References

1. Yaskevich V., Kuspangaliyev B., Khodzhikov A. Analiz natsional'nogo opyta vnedreniya BIM [Analysis of national experience in BIM implementation] // Project Baikal Founders: Elena Ivanovna Grigorieva. – 2023. – Vol. 20. – No. 75. – P. 105-111. [in Russian].

2. Dorzheyeva O.A. Tezaurusnoye modelirovaniye terminologii stroitel'stva [Thesaurus modeling of construction terminology] // Bulletin of the Chelyabinsk State University. – 2013. – No. 31 (322). – P. 85-87. [in Russian].
3. Abdurakhmanova A.Z. Terminologiya stroitel'stva kak ob'yekt tezaurusnogo modelirovaniya [Construction terminology as an object of thesaurus modeling] // Scientific notes of the Crimean Federal University named after V.I. Vernadsky. Philological sciences. – 2016. – No. 2 (68) (2-1). – P. 17-23. [in Russian].
4. Sunny S.K., Angadi M. Applications of thesaurus in digital libraries // DESIDOC Journal of Library and Information Technology. Defence Scientific Information and Documentation Centre. – 2017.
5. Gonzales-Aguilar A., Ramírez-Posada M., Ferreyra D. TemaTres: software para gestionar tesauros // Profesional de La Informacion. – 2012. – No. 21(3). – P. 319-325.
6. Rejowski M., Barbanti C.H. Construção de um Tesouro Brasileiro de Turismo // Revista Turismo Em Análise. – 2019. – No. 29 (2). – P. 182-195.
7. Berg M., Bains S., Suri S. South Asian Canadian Digital Archive Thesaurus // KULA: Knowledge Creation, Dissemination, and Preservation Studies. – 2022. – No. 6 (3). – P. 1-7.
8. Autiero S., Elwert F., Moscatelli C., Pons J. The Seven Steps: Building the DiGA Thesaurus // Journal of Open Humanities Data. – 2023. – No. 9.
9. Zeng W. Exploration and study of multilingual thesauri automation construction for digital libraries in China // Electronic Library. – 2012. – No. 30 (2). – P. 233-247.
10. Grigor'yev D.S. Sovremennyye vzglyady na rol' tsifrovyykh tekhnologiy v obespechenii ekonomicheskoy bezopasnosti stroitel'noy otrasli [Modern views on the role of digital technologies in ensuring economic security of the construction industry] // Innovations and Investments. – 2023. – No. 8. – P. 339-344. [in Russian].
11. Zhang J., El-Gohary N. Automated Information Transformation for Automated Regulatory Compliance Checking in Construction // J. Comput. Civ. Eng. – 2015. – No. 29.
12. National Research and Design Institute of Construction and Architecture. Unified System of Classification and Coding of Construction Resources (ESKKS) [Electronic resource]. - Access mode: <https://kazniisa.kz/eskks/>. Date of access: 17.02.2025. [in Russian].
13. Zakon.kz. Regulatory document [Electronic resource]. - Access mode: https://online.zakon.kz/Document/?doc_id=39585910. Date of access: 17.02.2025 [in Russian].
14. Kessler R., Béchet N., Berio G. Extraction of Terminology in the Field of Construction // First International Conference on Digital Data Processing (DDP). – 2019. – No. 22-26.
15. Committee for Construction and Housing and Communal Services of the Ministry of Regional Development of the Republic of Kazakhstan. (2013). SP RK 4.01-102-2013. Internal Sanitary and Technical Systems (Introduced for the first time; corresponds to SP RK 4.01-05-2013). Approved by Order No. 407-NK dated 28.12.2013 (now void). [in Russian].
16. Martínez-González M., Alvite-Díez M. On the evaluation of thesaurus tools compatible with the Semantic Web // Journal of Information Science. – 2014. – No. 40. – P. 711-722.
17. Mandal S. Integration of External Repositories with TemaTres for Easy Access of Related Terms // Library Philosophy and Practice. – 2021