

УДК 004.82

DOI: 10.18413/2518-1092-2026-11-1-0-6

Пимешков В.К.<sup>1,2</sup>

**ПОСТРОЕНИЕ МУЛЬТИДОМЕННЫХ СИСТЕМ ЗНАНИЙ  
НА ОСНОВЕ МНОГОСЛОЙНЫХ ГРАФОВ ЗНАНИЙ**

<sup>1)</sup> Институт информатики и математического моделирования им. В.А. Путилова ФИЦ КНЦ РАН,  
ул. Ферсмана, 24а, Мурманская область, г. Апатиты, 184209, Россия

<sup>2)</sup> филиал ФГАОУ ВО «Мурманский арктический университет» в г. Апатиты,  
ул. Лесная, 29, Мурманская область, г. Апатиты, 184209, Россия

*e-mail: v.pimeshkov@ksc.ru*

**Аннотация**

Системы знаний обеспечивают представление и применение структурированной информации для решения самых различных проблем. При разработке и применении современных систем знаний неизбежно возникают условия разнородности обрабатываемых данных, множества заинтересованных сторон или поддержки нескольких целей. Для решения задач в подобных условиях становится актуальна разработка мультидоменных систем знаний, способных согласованно работать с несколькими доменами в рамках единой системы. В данной работе предлагается концепция мультидоменных систем знаний, описывающая системы знаний, работающие в подобных условиях. Рассматриваются различные примеры подходов к построению систем знаний, так или иначе проявляющие мультидоменное поведение. Рассматриваются тематическое моделирование, в частности динамическое тематическое моделирование, в качестве основы для выделения слоев в многослойных графах знаний. Предлагается концепция многослойного динамического графа знаний (МДГЗ), использующего для разбиения на слои дрейфующую динамическую тематическую модель, позволяющую также отслеживать появление новых и угасание старых тем. Предполагается, что такой МДГЗ может применяться для определения релевантности текстовых документов в рамках задачи мониторинга социальных медиа.

**Ключевые слова:** системы знаний; графы знаний; социальные медиа; релевантность

**Для цитирования:** Пимешков В.К. Построение мультидоменных систем знаний на основе многослойных графов знаний // Научный результат. Информационные технологии. – Т.11, №1, 2026. – С. 54-70. DOI: 10.18413/2518-1092-2026-11-1-0-6

Pimeshkov V.K.<sup>1,2</sup>

**CONSTRUCTING MULTI-DOMAIN KNOWLEDGE SYSTEMS  
BASED ON MULTI-LAYER KNOWLEDGE GRAPHS**

<sup>1)</sup> Putilov Institute for Informatics and Mathematical Modeling FRC KSC RAS,  
24a Fersman St., Murmansk region, Apatity, 184209, Russia

<sup>2)</sup> Apatity branch of Murmansk Arctic University,  
29 Lesnaya St., Murmansk region, Apatity, 184209, Russia

*e-mail: v.pimeshkov@ksc.ru*

**Abstract**

Knowledge systems enable the representation and application of structured information to solve a wide variety of problems. The development and application of modern knowledge systems inevitably encounter heterogeneous data, multiple stakeholders, or the support of multiple objectives. To address these challenges, the development of multidomain knowledge systems capable of coherently working with multiple domains within a single system becomes essential. This paper proposes the concept of multidomain knowledge systems, which describes knowledge systems operating in such conditions. Various examples of approaches to constructing knowledge systems that exhibit multidomain behavior in one way or another are considered. Topic modeling, in particular dynamic topic modeling, is considered as a basis for identifying layers in multilayer knowledge graphs. A multilayer dynamic knowledge graph (MDKG) concept is proposed, which

uses a drifting dynamic topic model for layering, also allowing for tracking the emergence of new topics and the extinction of old ones. It is assumed that such a MDKG can be used to determine the relevance of text documents in the context of social media monitoring.

**Keywords:** knowledge systems; knowledge graphs; social media; relevance

**For citation:** Pimeshkov V.K. Constructing Multi-Domain Knowledge Systems Based on Multi-Layer Knowledge Graphs // Research result. Information technologies. – Т.11, №1, 2026. – P. 54-70. DOI: 10.18413/2518-1092-2026-11-1-0-6

## ВЕДЕНИЕ

Говоря о системах знаний, мы неизбежно сталкиваемся с понятиями предметной области и домена, которые в литературе часто обозначают одно и то же, но все же имеют различные коннотации. Под **предметной областью** будем понимать сферу знаний или дисциплину в целом, задающую используемую в ее границах терминологию, соответствующие концепты и отношения между ними или, иначе говоря, знания. Такие знания как правило представляют собой структурированную информацию, которая целенаправленно используется в рамках конкретной прикладной задачи. Следовательно, понятие “знание” приобретает конкретный смысл только в контексте мета задачи, для решения которой производится обработка данных. А под **доменом** будем понимать точку зрения на знания, задающую конкретный способ их интерпретации и оперирования над ними. Таким образом, различные домены являются не просто подмножествами знаний, а несовместимыми точками зрения на них. Можно сказать, что домены отражают то, как люди рассуждают и организуют работу. Например, в области общественного здравоохранения эпидемиологический домен интерпретирует понятия «случай», «риск» и «вспышка» с точки зрения моделей популяции и статистических данных, в то время как в клиническом домене эти же понятия интерпретируются с точки зрения отдельных пациентов и принятия медицинских решений.

Таким образом **система знаний** – это совокупность методов, моделей и средств, предназначенных для представления, хранения, обработки и использования знаний, как правило из одной предметной области. Они применяются практически везде, где людям или организациям приходится неоднократно принимать решения в условиях неопределенности, и где нельзя полагаться только на необработанные данные.

Но по мере усложнения решаемых задач и вовлечения все более разнородных источников данных становится очевидным, что представления знаний в виде единой предметной области становится избыточно сложным, а знаний одного домена недостаточно для решения комплексных задач. Таким образом, когда система знаний работает с разнородными данными, поддерживает несколько целей или обслуживает множество заинтересованных сторон, она неизбежно вводит внутренние различия, функционально эквивалентные доменам. Даже если такие домены не были выделены явно на этапе проектирования, они проявляются в виде модулей, контекстов или областей применимости правил и начинают играть роль самостоятельных точек зрения. Во многих практических сценариях система должна одновременно учитывать несколько различных точек зрения на одни и те же знания, которые могут различаться по целям использования, способам интерпретации и даже по используемой терминологии. Такие ситуации характерны, в частности, для анализа больших массивов текстовых данных, мониторинга социальных медиа, поддержки принятия решений и других задач, где данные зашумлены, динамичны и отражают эволюционирующие предметные области.

Для решения подобных задач требуется класс систем знаний, способных согласованно работать с несколькими доменами в рамках единого представления знаний. Такие системы будем называть **мультидоменными системами знаний (МСЗ)**, под которыми понимается системы, обеспечивающие представление и использование знаний как минимум из двух различных доменов, которые различаются по своему назначению или структуре. При этом ни одного из этих доменов недостаточно самого по себе для решения поставленных перед системой задач. Тем самым МСЗ выступает посредником между человеческими намерениями и сложными, неоднородными и изменяющимися данными реального мира.

Одна из прикладных задач, для которых применение МСЗ является особенно актуальным, это мониторинг социальных медиа, в рамках которого, в частности, производится автоматическая обработка больших массивов пользовательских текстовых данных с целью выявления публикаций, релевантных заданным информационным потребностям, таким как определенные темы, события, личности или общественные настроения. При этом существенной особенностью социальных медиа является их семантическая неоднородность, высокая степень зашумленности, быстрое изменение лексики и контекста, а также наличие множества пересекающихся точек зрения на одни и те же события и явления. Определение релевантности публикаций в данном контексте не является тривиальной задачей, так как релевантность напрямую зависит от текущей точки зрения задаваемой удовлетворяемой информационной потребностью. Одно и то же сообщение может быть релевантным с точки зрения общественного дискурса, мониторинга социально-политической обстановки или маркетингового анализа и при этом нерелевантным в любом другом контексте. Таким образом, задача мониторинга социальных медиа естественным образом формулируется как мультидоменная, требующая одновременного учета нескольких, потенциально несовместимых, точек зрения на наблюдаемые данные.

Для представления, интеграции и применения знаний применимо к такой задаче целесообразно использовать графы знаний, поскольку они позволяют в формальном виде на основе наблюдаемых данных моделировать концепты и отношения между ними. Вместе с этим они обеспечивают гибкость при работе с неструктурированными или слабо структурированными данными, а также допускают расширение и модификацию структуры по мере появления новых знаний. В условиях мультидоменной интерпретации данных естественным развитием графов знаний являются многослойные графы знаний, в которых различные слои соответствуют различным доменам или аспектам интерпретации. Каждый слой задает собственную семантику связей между сущностями и отражает отдельную точку зрения на решаемую задачу, при этом слои могут опираться на общее множество сущностей или данных. Такая организация позволяет согласованно представлять и анализировать знания из нескольких доменов в рамках единой структуры, не сводя их к единой интерпретации.

В контексте анализа социальных медиа одним из подходов к формированию слоев в многослойных графах знаний можно рассматривать тематическое моделирование, позволяющее автоматизировано выявлять латентные тематические структуры в больших коллекциях текстов. При этом оно позволяет формировать темы на основе данных, без предварительного задания жестких семантических границ, а выявленные таким образом темы могут интерпретироваться как домены отражающие социальные стереотипы наблюдаемых сообществ. Но поскольку социальные медиа характеризуются высокой динамикой, важным расширением классического тематического моделирования является динамическое тематическое моделирование, учитывающее изменение тематической структуры и состава во времени. Такой подход позволяет отслеживать эволюцию, появление новых тем и изменения существующих. В контексте МСЗ это позволяет обеспечить возможность адаптивного обновления слоев графа знаний в ответ на изменения в наблюдаемых данных.

Дальнейшая работа организована следующим образом. В начале рассматриваются примеры существующих подходов к построению систем знаний проявляющих мультидоменное поведение. Затем рассматриваются методы тематического моделирования как средство выявления доменов и формирования многослойных графов знаний в контексте задачи мониторинга социальных медиа. Предлагается формализация мультидоменного динамического графа знаний, использующего динамическое тематическое моделирование для выделения слоев для задачи определения релевантности документов в рамках мониторинга социальных медиа. В заключении подводятся итоги работы и обсуждаются направления дальнейших исследований.

## **МУЛЬТИДОМЕННЫЕ СИСТЕМ ЗНАНИЙ**

Хотя термин «мультидоменная система знаний» исторически не использовался для описания устоявшихся подходов к представлению знаний, можно заметить, что широкий спектр систем демонстрирует мультидоменное поведение, несмотря на то что они были разработаны с использованием различных терминологий, технологий и по разным мотивам. В контексте данного обзора будем считать, что система обладает мультидоменными характеристиками если она удовлетворяет одному или нескольким из следующих условий:

- требует явной или неявной координации между разнородными источниками знаний или моделями знаний;
- поддерживает множество интерпретационных контекстов, перспектив или режимов рассуждений в рамках решаемой задачи;
- поддерживает ограниченную или условную видимость элементов знаний или правил вывода.

Рассмотрим некоторые из подходов к построению систем знаний, с точки зрения проявления мультидоменного поведения, выявляя механизмы, посредством которых они реализуют его, или то, как оно проявляется неявно.

**Системы знаний, построенные по принципу «Доски объявлений\Классной доски» (Blackboard system).** В данном подходе системы знаний разрабатываются как средство координации разнородных источников знаний посредством общей рабочей памяти, для решения сложных, плохо определенных проблем, где решение является суммой своих частей. Каждый источник знаний работает на основе собственных предположений, представлений и стратегий вывода, предоставляя частичные решения, которые постепенно объединяются в общей памяти.

Например, можно выделить новаторскую для своего времени систему распознавания речи Hearsay-II [27] разработанную для обработки неструктурированной, естественной разговорной речи. Она восстанавливает намерение (ту мысль что он хотел передать через речь) говорящего на основе гипотетических интерпретаций, сформулированных на различных уровнях абстракции. При этом ключевые функции генерации, комбинирования и оценки гипотетических интерпретаций выполняются различными и независимыми программами, называемыми в рамках системы источниками знаний (ИЗ). Каждый ИЗ представляет собой независимый модуль, работающий по принципу «условие-действие». Они взаимодействуют через глобальную базу данных, называемую «доской». Любой ИЗ может как сгенерировать гипотезу (записать ее на доске), так и модифицировать существующую. Эти действия, в свою очередь, могут создавать структуры, удовлетворяющие условиям применимости других ИЗ. Таким образом доска в данной системе выполняет две основные роли: предоставляет промежуточные состояния деятельности по решению проблем, и принимает гипотезы от одних ИЗ для активации других ИЗ.

Хотя эта архитектура не рассматривается в терминах предметных областей или доменов, она естественным образом поддерживает множество сосуществующих режимов интерпретации в общем проблемном пространстве с механизмами управления определяющими, когда и как каждый режим становится активным. С точки зрения данного обзора, такие системы можно рассматривать как МСЗ, в которых домены представлены независимыми ИЗ, а их взаимодействие, направленное на решения общей задачи, реализуется при помощи общей памяти.

**Мультидоменные экспертные системы (Multi-Domain Expert Systems).** Классические экспертные системы (ЭС) имитируют процесс принятия решения экспертом узкой предметной области используя базу знаний и механизм логического вывода. Изначально создание таких систем имело предметно-ориентированный характер, и в каждой предметной области и конкретной прикладной задаче требовалось создание своей, адаптированной к их требованиям, экспертной системы. И с нарастанием их популярности и вездесущим внедрением, стал очевиден ряд проблем, которые привели к созданию так называемых мультидоменных ЭС [24]. Можно выделить три такие основные проблемы:

1. непреднамеренная передача ответственности за выбор подходящих ЭС от машины к пользователю;

2. сложность поиска решения сложных задач, требующих более чем одного источника знаний, горизонтально или вертикально;

3. общая разобщенность в разработке ЭС, которая приводит к сложностям в агрегировании всех знаний необходимых для решения поставленной задачи.

Определение понятие домена в случае таких систем отличается от определения домена, данного в данной работе. Доменом в рамках мультидоменных ЭС считается «область интереса, для которой разрабатывается экспертная система». Данные области интереса имеют иерархическую древовидную структуру, где внутри каждой области интереса существуют подобласти интереса, которые являются основной областью других областей более низкого уровня. Таким образом мультидоменные ЭС являются расширением классических ЭС, вводя явное разделение областей ответственности и механизмы активации соответствующих правил или контекстов, для управления растущими базами правил и контекстно-зависимыми рассуждениями. Что позволяет эффективно ограничивать применимость и видимость знаний в зависимости от ситуационных предпочтений в рамках решаемых задач.

Например, в [15] предлагается гибридная экспертная система, использующая искусственные нейронные сети и классические правила для решения задачи прогнозирования токсикологии (границы безопасности химических веществ) соединений. На вход данная система получает молекулярные дескрипторы (числовые характеристики, описывающие структуру и свойства молекулы) а на выходе дает значение токсичности включая его объяснение. В начале модуль классификации, основанный на искусственной нейронной сети, присваивает соединению химический класс, а затем активируются соответствующие количественные модели и базы знаний, позволяющие получить объяснение результата. Полученная ЭС расценивается как гибридная, где необходимые знания активируются только при необходимости.

Таким образом мультидоменные ЭС хоть и описываются с точки зрения стратегий управления или модульности, они демонстрируют мультидоменное поведение, поскольку поддерживают несколько частично несовместимых точек зрения на рассуждения, которые активируются избирательно в соответствии с решаемой в данный момент задачей. Выбор подходящей ЭС или координация глобального решения при этом достигается некоторой мета-системой управления.

**Мультиконтекстные системы (Multi-Context Systems, MCS).** Возникли как ответ на ограничения монолитного представления знаний и вывода, особенно в условиях, когда знания из разнородных источников не могли быть последовательно объединены. Они обеспечивают явную формализацию рассуждений в различных контекстах, каждый из которых задается своим собственным словарем и аксиомами, через введение так называемых правил-мостов, без необходимости в глобальной согласованности. Под контекстом при том понимаются «люди, агенты, базы данных, модули или базы знаний» [9]. При этом такие контексты могут быть разнородными по своей природе, и использовать различные логические языки и системы вывода. Правила-мосты же обеспечивают передачу информации из одного контекста в другой без изменений, обеспечивая в том числе средства для выбора подходящей абстракции и разрешения конфликтов между контекстами. Таким образом они позволяют только добавлять информацию к контекстам, но не удалять или изменять ее. Чтобы сделать такую формализацию применимой к большему числу задач и ввести недостающие типы операций над потоками данных, в [20] была предложена концепция так называемых регулируемых мультиконтекстных систем (Managed Multi-Context Systems, mMCS), которые являются обобщением позволяющим свободно определять произвольные операции над контекстами, например, операторы удаления или изменения.

В такой трактовке понятия MCS и mMCS значительно перекликается с предложенной концепцией МСЗ, поскольку все они рассматривают сосуществование множества потенциально несовместимых способов рассуждения в рамках одной системы знаний. Ключевое различие заключается в уровне абстракции и фокусе на различных деталях. MCS и mMCS предписывают явный механизм координации, основанный на формально определенных контекстах и правилах взаимодействия, регулирующих передачу информации между ними. В отличие от них, предлагаемая концепция МСЗ является более общей и не обязывает к конкретному механизму

согласования знаний между доменами, что позволяет применить предлагаемую концепцию для описания более широкого класса систем знаний, включая те в которых домены являются неявными.

Научные работы в области MCS и mMCS [11, 26, 32, 34, 36] концентрируются преимущественно на теоретических аспектах взаимодействия в различных контекстах, не углубляясь в решение практических задач.

**Сверхбольшие онтологии.** Обычные онтологии задают формальное описание системы понятий некоторой области знаний, включающее их свойства и взаимосвязи, через набор сущностей, содержащий понятия, аксиомы, отношения и экземпляры. Обычные онтологии, как правило, разрабатываются для поддержки решения конкретной задачи, приложения или узко определенной предметной области и, следовательно, оптимизированы для концептуальной ясности и управляемого рассуждения в рамках четко ограниченной области знаний. Сверхбольшие онтологии же представляют собой всеобъемлющие, формально структурированные представления знаний для целой предметной области, а не для узкоспециализированной задачи. Они включают сотни тысяч понятий, организованных в глубокие иерархии и реляционные базы данных. Вследствие такого масштаба и широты их применение сопряжено с повышенной сложностью логического вывода, требующей модульности, аппроксимации или стратегий, ограничивающих логический вывод. Хотя модульность часто представляется как вопрос проектирования программного обеспечения или удобства сопровождения, модули онтологий обычно кодируют различные точки зрения на общие сущности, с ограниченным и контролируемым соответствием между ними.

Например, медицинская онтология SNOMEDCT [10] предоставляющая основную общую терминологию для электронных медицинских карт, различных систем поддержки принятия врачебных решений, каталогов клинических услуг, ведения лабораторной отчетности и т.д. Включает термины отражающие понятия таких категорий медицины и здравоохранения как симптоматика заболеваний, медицинские процедуры и вмешательства, химические вещества, фармацевтические продукты. С точки зрения устройства, она основана на логике описаний, в которой уникально идентифицированные независимые от языка концепты организованы в родовидовую таксономию отношений IS-A. Каждый концепт как правило относится к одной иерархии понятий, за исключением комбинированных понятий. Такие иерархии понятий можно расценивать как своеобразные домены, задающие интерпретации понятий.

Или, например онтология OpenCyc [37] охватывающая основные понятия и правила функционирования мира как такового. Создавалась с целью обеспечения решения программами сложных задач из области искусственного интеллекта. Основана на логике первого порядка и разделена на так называемые микротеоии (являющиеся объектами первого класса), обеспечивающие совокупность утверждений, которые считаются истинными при определенном наборе допущений, таких как область исследования, временной период, уровень абстракции или точка зрения. Вместо того чтобы требовать глобальной согласованности всех знаний, OpenCyc разделяет знания на множество микротеоий, каждая из которых имеет свои аксиомы. Микротеоии можно расценивать как домены, задающие контекст интерпретации знаний применимо к предметной области или решаемой задаче.

Таким образом подобные сверхбольшие онтологии можно рассматривать как МСЗ, в которых отдельные домены выделяются, например, неявно через иерархии понятий или явно в виде отдельных модулей знаний. При этом при неявном выделении доменов сохраняется общая логика размышлений на данных и соответственно согласованность, возможно ограниченная лишь какими-то контекстными правилами. А при явном выделении логика размышлений становится частью выделенного модуля, отделяясь от общего контекста.

**Многослойные графы знаний.** В общем смысле под графом знаний (ГЗ) обычно понимают какую-либо графовую структуру данных, внутри которой некоторые сущности (вершины) связаны между собой как правило семантическими связями (ребра). На практике лучше подходит определение с саммита *Ontology 2020* [23]: «граф знаний – это не просто статичное представление фактов, а некоторая активная архитектура, включающая в себя помимо способа представления

знаний операции по генерации, хранению и применению этих знаний». Данное определение отражает как структурный, так и функциональный аспект ГЗ, подчеркивая их роль в качестве динамических систем, в которых представление знаний неотделимо от механизмов их генерации, интерпретации и применения.

Классические ГЗ обычно строятся с использованием взаимодополняющих подходов, основанных на знаниях и данных. В подходе, основанном на знаниях, структура графа выводится из базовой онтологии, которая определяет соответствующие понятия, отношения и ограничения предметной области, а наполнение ГЗ экземплярами происходит на основе анализа наблюдаемых данных. В отличие от этого, подходы, основанные на данных, строят ГЗ путем извлечения сущностей и отношений непосредственно из данных, таких как текстовые корпуса (обработанная коллекция текстовых документов) или базы данных. В этом случае структура графа возникает из самих данных и может быть лишь слабо ограничена формальными моделями. На практике многие крупномасштабные ГЗ сочетают оба подхода, сопоставляя извлеченные данные с онтологическими структурами для повышения согласованности и интерпретируемости.

Многослойные ГЗ же интегрируют различные измерения знаний в рамках одной единой системы. Потребность в них возникла вместе с вовлечением в процессы обработки все более сложных и неоднородных данных. Такие графы разделены на несколько слоев, каждый из которых задает свою семантику отношений (например, социальные, профессиональные или контекстуальные) или разные точки зрения на предметную область (структура понятий и связей на каждом слое различна).

Известным примером многослойного ГЗ служит KnowWhereGraph [35], разработанный для поддержки размышлений с пространственными данными, и включающий более 30 слоев (пространственные, демографические, инфраструктурные, сельскохозяйственные данные и др.). Базовая структура KnowWhereGraph основана на кортежной модели многоуровневого графа, где каждое ребро принадлежит одному или нескольким измерениям: (субъект, предикат, объект, слой, временная метка, достоверность, геокоординаты). При этом географический и временной контекст являются обязательными. Граф хранится как набор матриц смежности для каждого слоя с межслойными указателями, обеспечивающими навигацию по измерениям. Он пополняется информацией через гибридный конвейер, сочетающий автоматическое извлечение и экспертное курирование, а соответствие между слоями поддерживается посредством межслойного связывания сущностей.

Таким образом многослойные ГЗ можно рассматривать как МСЗ, в которых роль доменов играют отдельные слои, задающие конкретную семантику отношений, а согласованность знаний достигается общей схемой знаний.

**Большие языковые модели (Large language models, LLMs).** Хотя в своем классическом представлении LLMs не являются системами знаний, так как знания в них представлены в неявном виде, они заслуживают отдельного упоминания в рамках данного обзора, в силу их близкого взаимодействия с системами знаний в целом. LLMs представляют собой модели на основе глубоких нейронных сетей, обученные на огромных объемах текстовых данных, которые неявно усваивают знания закодированные в языке. Знания в них кодируются неявно на этапе обучения напрямую в модель посредством распределенных числовых параметров, полученных из больших текстовых корпусов. В результате семантические закономерности кодируются в многомерных векторных пространствах, а не в явных символах, правилах или отношениях. Таким образом отдельные факты, понятия или ограничения не могут быть непосредственно идентифицированы, проверены или обработаны. Стоит также заметить, что современные LLMs обучаются в том числе на различных мультимодальных данных, и выходят за рамки кодирования только семантических знаний.

Несмотря на это LLMs применяются для поддержки инженерии знаний. Например, для извлечения знаний с целью разработки экспертных систем [13; 33] или поддержки построения онтологий и графов знаний [17; 19]. Или они могут применяться способами функционально схожими с классическими системами знаний. Например, как вопросно-ответные системы общего или специализированного характера [29]. А такие архитектуры, как Mixture-of-Experts (MoE), еще

больше усиливают эту аналогию, вводя модульность и условную активацию, где разные подсети или отдельные модели, называемые «экспертами», специализируются в разных областях предметной области или пространстве знаний. Концептуально такая модульная специализация перекликается с предлагаемой концепцией МСЗ, в которых различные области или контексты сосуществуют и совместно способствуют решению единой проблемы.

Например, в [21] предлагается модель для идентификации фейковых новостей, полагающаяся на архитектуру МоЕ и механизм управления, так называемый «domain gate». Модель начинает работу с преобразования обрабатываемого текста в эмбединги предложений и домена (выделенная авторами тематическая область), при помощи модели BERT. Затем полученные эмбединги обрабатываются механизмом выбора для определения релевантных сетей-экспертов. Каждая такая сеть обрабатывает эмбединги и выдает набор признаков, которые подаются классификатору для финального решения.

Вместе с этим, все чаще LLMs применяются как компоненты комбинированных систем знаний. Например, через методы RAG (Retrieval-Augmented Generation) и RIG (Retrieval-Integrated Generation), представляющие собой современную форму классической парадигмы «база знаний + механизм вывода», где механизмы поиска обеспечивают явный доступ к знаниям, а LLMs выполняют функцию рассуждений и генерацию ответов на естественном языке, принимая во внимание реальные знания. Например, можно выделить вычислительную систему знаний Wolfram|Alpha [12], работающую на основе методов RAG.

Таким образом хотя LLMs не являются системами знаний в классическом понимании, они все чаще тесно интегрируются с таковыми. А модульная архитектура комбинированных систем знаний, включающих LLMs как отдельные модули или инструменты вывода может рассматриваться как близкая к предлагаемой концепции МСЗ.

Рассмотренные подходы к построению систем знаний различаются по уровню абстракции и зависимости от конкретных способов представления знаний. Некоторые тесно связаны с конкретными типами систем, в то время как другие работают на архитектурном или концептуальном уровне и могут быть реализованы для широкого спектра систем знаний. С точки зрения используемых для построения систем знаний данных, можно разделить подходы на управляемые знаниями (экспертные системы, онтологии), данными (ГЗ, LLMs) или с комбинированным управлением. Такое разделение существенно поскольку не для каждой ситуации могут быть доступны достаточные объемы экспертных знаний или наоборот наблюдаемых данных. Вместе с этим с развитием систем представления знаний в целом, характерен постепенный переход от систем на основе экспертных знаний к системам на основе данных и впоследствии к комбинированным системам.

Также следует отметить, что мультидоменность не является универсальным или обязательным свойством систем знаний. Существуют узкоспециализированные системы знаний, ориентированные на решение четко определенных задач в рамках одной предметной области, которые демонстрируют высокую эффективность без необходимости явного или неявного разделения на домены. Например, база знаний UniProtKB [38], посвященная белковым соединениям и их описаниям. Хотя она и представлена двумя секциями, одна для экспертно-размеченных данных, вторая для сгенерированных аннотаций, но данное разделение скорее техническое и отражает не столько разные точки зрения, сколько технические характеристики знаний. В этом смысле мультидоменность можно рассматривать как архитектурный ответ на требования обработки разнородных данных, поддержки множества целей или обслуживания различных заинтересованных сторон.

### **ТЕМАТИЧЕСКИЕ МОДЕЛИ КАК СПОСОБ ВЫДЕЛЕНИЯ ДОМЕНОВ ДЛЯ МНОГОСЛОЙНЫХ ГРАФОВ ЗНАНИЙ**

Тематическое моделирование представляет собой, как правило статистические и вероятностные методы, предназначенные для выявления абстрактных «тем», характеризующих

коллекцию документов. Тематические модели изучают текстовые корпуса, выявляя скрытые тематические закономерности, объясняющие наблюдаемые случаи совместного употребления слов или терминов. Таким образом задачу тематического моделирования можно рассматривать как задачу одновременной нечеткой кластеризации документов и слов по одному и тому же множеству кластеров (темам) [1]. При этом как правило количество тем выделенное в документах меньше числа слов во всем наборе, что позволяет представить документ в виде вектора в пространстве тем, который имеет меньшее число компонент по отношению к пространству всех слов. Следовательно, тематическое моделирование может рассматриваться в качестве инструмента уменьшения размерности данных, а выделенные темы могут применяться для семантического анализа текстов. Тематические модели используются в очень широком кругу задач, например в рамках задач информационного поиска для семантического поиска или расширения запросов [39], в различных рекомендательных системах [4, 5], в задачах анализа социального дискурса [6, 25], в различных задачах составления кратких пересказов текстов [3, 31] и т.д. Можно сказать, что тематическое моделирование прекрасно справляется с выявлением скрытых тематических структур в текстовых данных, что делает его бесценным для организации, обобщения, исследования и осмысления огромных объемов информации в самых разных предметных областях.

Различные методы тематического моделирования представляют собой не столько альтернативные способы упрощенного представления текста, сколько последовательные разработки призванные решить проблемы, обнаруженные в прошлых методах. Наиболее популярными классическими методами тематического моделирования на сегодняшний день являются LDA [8] и LSA [18]. LDA в контексте тематического моделирования можно определить как генеративную вероятностную модель, которая генерирует темы классифицируя слова и документы по этим темам в соответствии с их вероятностными распределениями. Основная идея модели заключается в том, что документы представлены случайным распределением скрытых тем, где каждая тема характеризуется распределением по словам. LSA же предполагает, что слова, семантически близкие друг к другу, как правило, встречаются в схожих контекстах (документах).

Так же с развитием технологий и появлением больших языковых моделей (LLM), они тоже стали применяться для задач тематического моделирования. Применяются они как правило в двух основных форматах: для генерации эмбеддингов или в промптовом режиме. В первом случае они применяются на начальном этапе для генерации семантически-богатых эмбеддингов текста, которые затем применяются для кластеризации другими методами [14]. Во втором случае они применяются для генерации тем в промптовом режиме, те через создание текстовых запросов к модели [28]. Подобный подход к тематическому моделированию может отличаться от классического, так как в зависимости от выбранного подхода в результате не обязательно будут получены тематические распределения для терминов и документов. Результатом может являться множество терминов обозначающие выделенные темы, и размеченные темами документы, без их тематических распределений.

Формально классическое тематическое моделирование можно описать следующим образом. Пусть имеется корпус документов  $D$  и словарь терминов данного корпуса документов  $Vc(D)$ :

$$D = \{d_1, \dots, d_N\},$$

где каждый документ  $d \in D$  представлен множеством предложений  $d = \{snt_1, \dots, snt_K\}$ , где каждое предложение  $snt_k$  состоит из последовательности токенов  $snt = \{tkn_{k1}, \dots, tkn_{kl}\}$ , где каждый токен  $tkn \in T \mid T = W \cup P$  принадлежит множеству слов  $W$  или знаков препинания  $P$ .

$$Vc(D) = \{term_1, \dots, term_M\},$$

где каждый термин  $term_m \in Vc(D)$ ,  $term_m = \{tkn_1, \dots, tkn_j\}$ ,  $j \geq 1$  состоит хотя бы из одного токена.

Тогда тематическое моделирование можно формально описать как выделение тематических распределений с заданным количеством тем  $tn$ :

$$TM(D, Vc, tn) \rightarrow \{\Theta, \Phi\},$$

где  $\Theta$  – матрица распределения документов по темам;

$\Phi$  – матрица распределения терминов по темам.

Но подобные классические модели могут быть применены только к задачам, для которых справедливо что последовательность документов и слов в них не имеет значения, а также количество тем известно заранее и не меняется со временем. Соответственно для решения задач, где эти условия не выполнимы, разрабатывались специфичные тематические модели. Наибольший интерес в рамках данной работы представляют модели, учитывающие временной фактор, или так называемые динамические тематические модели [7]. Стандартные динамические тематические модели были созданы для анализа временной эволюции тем в больших коллекциях документов. Они вводят зависимость по времени между тематическими моделями, обученными на данных, разбитых на последовательные временные отрезки. Данная зависимость основана на предположении о том, что тематическое распределение на текущем временном отрезке  $t$  зависит от распределения на предыдущем отрезке  $t - 1$ , что предотвращает слишком резкое изменение тем и открывает возможность для отслеживания динамики их изменения. В результате такие тематические модели могут демонстрировать как изменяется словарный запас, связанный с темой.

Формально в таком случае наблюдаемый корпус документов должен быть разбит на подмножества в соответствии с некоторым выбранным временным окном  $\Delta t$ :

$$D_{\Delta t} = \{D_{t_1}, \dots, D_{t_p}\},$$

Следовательно, множества терминов для каждого отрывка времени будет своими и в общем случае будут различны  $V_{c_t} \neq V_{c_{t'}}$ . В таком случае динамическое тематическое моделирование можно формально определить как такое, где распределение терминов по темам зависит от предыдущего отрезка времени:

$$DTM(D_t, V_{c_t}, t_n) \rightarrow \{\theta_t, \Phi_t\} \mid \Phi_t \sim f(\Phi_{t-1})$$

Например, в [22] предлагается динамическая тематическая модель реального времени, основанная на LDA, учитывающая изменение тем в различных временных промежутках, подразумевая что для каких-то тем словарный запас может оставаться постоянным на протяжении многих лет, а для каких-то изменяться за дни или недели. Учет различных временных промежутков достигается путем априори заданной временной шкалы для разбиения временных данных на такие промежутки. Динамичность модели достигается за счет предлагаемого авторами алгоритма обновления модели, который обновляет текущую модель и мультимасштабную модель предыдущей эпохи на основе новых полученных документов (для текущей эпохи).

Подобные модели позволяют улавливать динамику изменения тем, что особенно актуально в задаче мониторинга социальных медиа, когда темы в обрабатываемых данных могут быстро эволюционировать. Но такие модели все еще не обладают способностью к идентификации рождения новых и угасания старых тем, иначе говоря, количество тем в модели остается неизменным на протяжении ее работы. А при работе с данными социальных медиа появление новых и угасание старых тем является неизбежным процессом. Для решения этой проблемы были разработаны различные модели модифицирующие стандартные динамические тематические модели, возможностью изменения количества тем. Назовем такие модели дрейфующими динамическими тематическими моделями и формально определим их следующим образом:

$$DTM^+(D_t, V_{c_t}) \rightarrow \{\theta_t, \Phi_t, Z_t\} \mid \Phi_t \sim f(\Phi_{t-1}),$$

где  $Z_t$  множество тем, актуальных во временном окне  $t$ . Стоит сразу отметить, что как правило в самом начале для первой итерации количество тем может как быть задано вручную, так и определено каким-либо методом подбора оптимального количества тем. А в дальнейшем тематические распределения терминов и соответствующие им темы будут зависеть от полученного на прошлом шагу тематического разбиения. Добавление новых или исключение старых тем будет зависеть от выбранного подхода и решаемой задачи.

Например, в [30] предлагается модель, модифицирующая стандартный LDA для учета тематического дрейфа, под которым подразумевается изменение в тематических распределениях, полученных для некоторого нового набор документов по отношению к начальному набору документов. Гипотеза, лежащая в основе модели, заключается в том, что если небольшой дрейф является нормой, то значительный дрейф указывает на появление новой темы. Для определения момента рождения новых тем при анализе дрейфов выделяются аномальные значения или выбросы,

которые указывают на появление новой темы. Для закрытия угасающих тем производится оценка количества токенов тем в новых данных относительно других тем присутствующих в модели, иначе говоря тема должна сохранять некоторый заданный априори минимальный процент токенов из числа всех, иначе считается угасшей и закрывается.

Таким образом в контексте мониторинга социальных медиа такие дрейфующие динамические тематические модели позволяют выявлять темы, которые можно интерпретировать как временно устойчивые домены, задающие специфическую терминологию и контекст интерпретации данных. Подобные домены в свою очередь позволяют определять релевантность принимая во внимание актуальную семантическую интерпретацию данных. Появление новых тем соответствует формированию новых доменов, а угасание отражает снижение их значимости вплоть до исключения из рассмотрения в рамках текущей модели знаний. Таким образом, дрейфующее динамическое тематическое моделирование предоставляет естественный механизм выявления и отслеживания эволюции доменной структуры в МСЗ.

Но несмотря на то, что динамические тематические модели позволяют выявлять и отслеживать эволюцию доменов, получаемые в их результате темы сами по себе не являются знаниями в строгом смысле этого слова. Темы как правило представляют собой статистические распределения терминов и документов, отражающие латентную структуру данных. В них отсутствуют явно заданные концепты, отношения между ними, а также формализованные утверждения, которые могли бы быть использованы для интерпретации, сопоставления или логического вывода. Вследствие чего возникает необходимость в переходе от тематических моделей к более выразительным формам представления знаний, позволяющих связывать различные понятия и их характеристики с позиции различных точек зрения. Многослойные графы знаний выступают как естественное продолжение данного подхода, позволяя интерпретировать выявленные домены в качестве слоев, внутри которых формируются явно заданные концепты и отношения между ними. Такой переход обеспечивает интеграцию статистически выявленных доменных структур с формализованным, пригодным для анализа и повторного использования представлением знаний.

### **МУЛЬТИДОМЕННЫЙ ДИНАМИЧЕСКИЙ ГРАФ ЗНАНИЙ**

Переход от тематического моделирования к многослойному графу знаний может быть осуществлен путем интерпретации выявленных динамических тем в качестве доменов, задающих структуру слоев графа, а тематические распределения могут быть использованы для выделения конкретных понятий и отношений в рамках этих слоев. Общая согласованность знаний достигается за счет единства словаря терминов наблюдаемых данных. Выбор конкретного типа извлекаемых отношений и соответствующего метода при этом будет зависеть от решаемой задачи. Подходящим вариантом в рамках задач мониторинга социальных медиа, не требующим задания явной схемы отношений, могут быть взвешенные ассоциативные отношения, получаемые на основе статистики совместного употребления терминов с учетом их тематического распределения [2].

Таким образом для построения многослойного графа знаний мы располагаем наблюдаемым потоком документов  $D_{\Delta t}$ , соответствующим ему словарем  $V_{C_{\Delta t}}$  и результатами их динамического тематического моделирования  $\{\theta_t, \phi_t, Z_t\}$ . Назовем подобный многослойный граф знаний, построенный с учетом результатов дрейфующего динамического тематического моделирования – мультидоменным динамическим графом знаний (МДГЗ). Формально определим его следующим образом:

$$MDKG(D_{\Delta t}, V_{C_{\Delta t}}, \{\theta_t, \phi_t, Z_t\}) := (Vr, E, init, fin, l, w),$$

где  $Vr \neq \emptyset$  – множество вершин, представленное терминами из словаря  $Vr \subset V_{C_{\Delta t}}$ ;

$E$  – множество ребер, представленное отношениями между терминами;

$init(\theta_t, \phi_t): E \rightarrow V$  – функция, присваивающая каждому ребру начальную вершину;

$fin(\theta_t, \phi_t): E \rightarrow V$  – функция, присваивающая каждому ребру конечную вершину;

$l: E \rightarrow Z_t$  – функция, присваивающая каждому ребру метку домена (слоя), соответствующую теме из тематической модели;

$w(\theta_t, \phi_t): E \rightarrow R$  – функция, присваивающая вес каждому ребру,  $R$  – множество вещественных чисел.

Обновление МДГЗ осуществляется вместе с обновлением тематической модели в соответствии с изменением ее параметров. Изменение параметров отдельных тем  $(\theta_t, \phi_t)$  приводит к модификации структуры соответствующих слоев графа. Изменения в структуре будут зависеть от выбранного подхода к извлечению отношений. Появление новых тем  $(Z_t^+ = Z_t \setminus Z_{t-1})$  инициирует построение новых слоев, а исчезновение тем  $(Z_t^- = Z_{t-1} \setminus Z_t)$  – их деактивацию. Таким образом, динамика тематической модели непосредственно определяет динамику структуры многослойного графа знаний.

Что бы формально описать как именно полученный МДГЗ может быть применен в задаче определения релевантности публикаций, сначала формально опишем задачу мониторинга социальных медиа. В рамках мониторинга система наблюдает поток документов, поступающих со временем  $DF(t) = \{d_1, \dots, d_k\}$  из некоторого множества источников  $DS$ . Аналогично документам, на основе которых строится тематическая модель, они состоят из предложений, которые состоят из токенов. Документами в рамках социальных медиа, как правило считаются публикации и комментарии к ним.

Цель мониторинга – выделить из наблюдаемого потока документов  $DF(t)$  упорядоченное подмножество документов релевантных запросу  $q$ :

$$(DF(t)', \leq) = \{d \in DF(t) | f_{rel}(d, q) > th\},$$

$$f_{rel} : DF \times \{q\} \rightarrow C,$$

$$d_i \leq_q d_j \Leftrightarrow f_{rel}(d_i, q) \leq f_{rel}(d_j, q)$$

где  $th$  – некоторое пороговое значение релевантности документа;

$C$  – множество оценок (например, множество вещественных чисел);

$q \in Q | q = \{kw_1, \dots, kw_n | kw \in C^{IA}\}$  – запрос, представленный набором ключевых слов ( $n$ -грамм), где  $Q$  – множество всех возможных запросов пользователя, ограниченное областью его интересов  $IA = \{C^{IA}, Rel^{IA}\}$  задающей ожидания пользователя, где  $C^{IA}$  – множество концептов области интересов (мыслительных единиц),  $Rel^{IA}$  – некоторые отношения на множестве концептов области интересов. Выделение подобной области интересов рассматривается в информационном поиске в рамках задач моделирования пользователей;

$f_{rel}$  – функция оценки релевантности документа  $d$  запросу  $q$ .

Благодаря тематической модели  $DTM^+$  открывается возможность получить тематическое распределение запроса  $\theta_t(q)$ , как для любого другого документа, и на его основе выбрать релевантное запросу подмножество доменов МДГЗ  $Z_t^q$ :

$$Z_t^q = \{z \in Z_t | \theta_t(q, z) > dth\},$$

где  $dth$  – пороговое значение релевантности домена.

Что дает возможность дополнения функции релевантности информацией из соответствующих релевантных слоев графа:

$$f_{rel}^*: DF \times \{q\} \times Z_t^q \times MDKG_t^q \rightarrow C$$

В качестве такой информации может выступать структура затронутых терминов, важность соответствующих вершин и ребер, наличие поддерживающих или противоречащих связей, информация о пересечении терминов запроса. В результате открывается возможность, например расширения запроса, то есть модификации исходного поискового запроса, или иногда целевых документов, терминами релевантными (в смысле социальных стереотипов) запросу. Или сама функция определения релевантности может быть реализована на основе структуры графа, например, как мера расстояния на графе от терминов запроса до терминов публикации, или при помощи эмбедингов графов на подобии `node2vec` [15], через сравнение эмбедингов запроса и публикаций. Но следует отметить, что в предлагаемой постановке релевантность публикаций интерпретируется не как абсолютная характеристика соответствия документа запросу, а как

контекстно обусловленная величина, определяемая релевантными доменами МДГЗ. Поскольку домены выявляются на основе тематической структуры наблюдаемого потока данных, они отражают устойчивые способы интерпретации информации, характерные для различных сообществ или актуальных событий. В этом смысле оценка релевантности может рассматриваться как согласованность публикации этим интерпретациям, зафиксированными в структуре графа знаний. Таким образом, расширенная функция релевантности фактически реализует оценку соответствия публикации не только запросу пользователя, но и доменно-обусловленному контексту, сформированному в результате анализа дискурса в наблюдаемых источниках.

### **ЗАКЛЮЧЕНИЕ**

В статье проведен обзор различных подходов к построению систем знаний проявляющих мультидоменное поведение. Несмотря на то, что рассмотренные подходы к построению систем знаний различаются по уровню абстракции и зависимости от конкретных способов представления знаний, все они обеспечивают построение таковых в условиях множества интерпретационных контекстов или разнородности обрабатываемых данных. Предложенная концепция мультидоменных систем знаний позволяет описать широкий круг систем знаний создаваемых и работающих в таких условиях, без строгой привязки к такой-либо архитектуре или способу представления знаний. Рассмотрено тематическое моделирование, в частности динамическое тематическое моделирование, как инструмент выделения доменов применимо к ГЗ. Предложена формализация МДГЗ, полагающегося на такую дрейфующую динамическую тематическую модель для выделения тематических слоев графа. Подобный граф знаний предполагается использовать для определения релевантности публикаций в рамках задач мониторинга социальных медиа.

В продолжении данного исследования планируется рассмотрение возможности дополнения концепции МДГЗ предметным доменом, отражающим информационную потребность пользователя, например при помощи онтологии или гибридного подхода. Дополнение системы таким стабильным доменом должно обеспечить однозначность применимо к предметным знаниям, а именно в используемой терминологии, и отношениях между предметными понятиями. Вместе с этим подобное дополнение должно обеспечивать стабильное определение релевантности, например в условиях, когда запрос не находит достаточного отражения в наблюдаемых данных.

### **Список литературы**

1. Антон К. Тематическое моделирование текстов на естественном языке / К. Антон, Г. Андрей // Труды Института системного программирования РАН. – 2012. – Т. 23. – С. 215-244.
2. Формирование многослойных графов знаний на базе тематического моделирования текстов / В.К. Пимешков [и др.] // Труды Кольского Научного Центра РАН. Серия: Технические Науки. – 2024. – Т. 15. – № 3. – С. 50-60.
3. A topic modeling-based bibliometric exploration of automatic summarization research / X. Chen [et al.] // Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. – 2024. – Т. 14. – № 5. – С. e1540.
4. Bagul D.V. A novel content-based recommendation approach based on LDA topic modeling for literature recommendation / D.V. Bagul, S. Barve // 2021 6th International Conference on Inventive Computation Technologies (ICICT) 2021 6th International Conference on Inventive Computation Technologies (ICICT). – 2021. – С. 954-961.
5. Bansal S. Topic Modeling Driven Content Based Jobs Recommendation Engine for Recruitment Industry: 5th International Conference on Information Technology and Quantitative Management, ITQM 2017 / S. Bansal, A. Srivastava, A. Arora // Procedia Computer Science. – 2017. – Т. 122. – С. 865-872.
6. Benites-Lazaro L.L. Topic modeling method for analyzing social actor discourses on climate change, energy and food security: Special Issue on the Problems of Methods in Climate and Energy Research / L.L. Benites-Lazaro, L. Giatti, A. Giarolla // Energy Research & Social Science. – 2018. – Т. 45. – С. 318-330.
7. Blei D.M. Dynamic topic models / D.M. Blei, J.D. Lafferty // Proceedings of the 23rd international conference on Machine learning: ICML '06. – New York, NY, USA: Association for Computing Machinery, 2006. – С. 113-120.
8. Blei D.M. Latent Dirichlet allocation / D.M. Blei, A.Y. Ng, M.I. Jordan // Journal of machine Learning research. – 2003. – Т. 3. – С. 993-1022.

9. Brewka G. Equilibria in heterogeneous nonmonotonic multi-context systems / G. Brewka, T. Eiter // Proceedings of the 22nd national conference on Artificial intelligence - Volume 1: AAAI'07. – Vancouver, British Columbia, Canada: AAAI Press, 2007. – С. 385-390.
10. Chang E. The use of SNOMED CT, 2013-2020: a literature review / E. Chang, J. Mostafa // Journal of the American Medical Informatics Association: JAMIA. – 2021. – Т. 28. – The use of SNOMED CT, 2013-2020. – № 9. – С. 2017-2026.
11. Dao-Tran M. Streaming Multi-Context Systems / M. Dao-Tran, T. Eiter // Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. – 2017. – С. 1000-1007.
12. Davis E. Testing GPT-4 with Wolfram Alpha and Code Interpreter plug-ins on math and science problems / E. Davis, S. Aaronson arXiv:2308.05713 [cs]. – arXiv, 2025.
13. Garrido-Merchán E.C. GOFAI meets Generative AI: Development of Expert Systems by means of Large Language Models. GOFAI meets Generative AI / E.C. Garrido-Merchán, C. Puente arXiv:2507.13550 [cs]. – arXiv, 2025.
14. Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. BERTopic / M. Grootendorst arXiv:2203.05794 [cs]. – arXiv, 2022.
15. Grover A. node2vec: Scalable Feature Learning for Networks. node2vec / A. Grover, J. Leskovec arXiv:1607.00653 [cs]. – arXiv, 2016.
16. Hybrid toxicology expert system: architecture and implementation of a multi-domain hybrid expert system for toxicology / G. Gini [et al.] // Chemometrics and Intelligent Laboratory Systems. – 1998. – Т. 43. – Hybrid toxicology expert system. – № 1. – С. 135-145.
17. Kommineni V.K. From human experts to machines: An LLM supported approach to ontology and knowledge graph construction. From human experts to machines / V.K. Kommineni, B. König-Ries, S. Samuel arXiv:2403.08345 [cs]. – arXiv, 2024.
18. Landauer T.K. An introduction to latent semantic analysis: Discourse Processes / Landauer T.K., Foltz P.W., Laham D. – 1998. – Т. 25. – № 2-3. – С. 259-284.
19. LLM-assisted Knowledge Graph Engineering: Experiments with ChatGPT / L.-P. Meyer [et al.] // arXiv:2307.06917 [cs]. – 2024. – LLM-assisted Knowledge Graph Engineering. – С. 103-115.
20. Managed Multi-Context Systems. / G. Brewka [et al.] journalAbbreviation: IJCAI International Joint Conference on Artificial Intelligence. – 2011. – 786 c.
21. MDFEND: Multi-domain Fake News Detection / Q. Nan [et al.] // Proceedings of the 30th ACM International Conference on Information & Knowledge Management / arXiv:2201.00987 [cs]. – 2021. – MDFEND. – С. 3343-3347.
22. Online multiscale dynamic topic models / T. Iwata [et al.] // Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining: KDD '10. – New York, NY, USA: Association for Computing Machinery, 2010. – С. 663-672.
23. Ontology summit 2020 communiqué: Knowledge graphs / K. Baclawski [et al.] // Applied Ontology. – 2021. – Т. 16. – Ontology summit 2020 communiqué. – № 2. – С. 229-247.
24. Poo D. Multi-domain expert systems / D. Poo, H. Lu // Expert Systems. – 2007. – Т. 8. – С. 67-73.
25. Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter / Jia Xue [et al.]. – 2020. – Т. 15. – № 9. – С. e0239441.
26. Reactive multi-context systems: Heterogeneous reasoning in dynamic environments / G. Brewka [et al.] // Artificial Intelligence. – 2018. – Т. 256. – Reactive multi-context systems. – С. 68-104.
27. The Hearsay-II Speech-Understanding System: Integrating Knowledge to Resolve Uncertainty / L.D. Erman [et al.] // ACM Comput. Surv. – 1980. – Т. 12. – The Hearsay-II Speech-Understanding System. – № 2. – С. 213-253.
28. TopicGPT: A Prompt-based Topic Modeling Framework. TopicGPT / C.M. Pham [et al.] arXiv:2311.01449 [cs]. – arXiv, 2024.
29. Towards leveraging LLMs for Conditional QA / S.-A. Hussain [et al.] arXiv:2312.01143 [cs]. – arXiv, 2023.
30. Wilson A. Tracking topic birth and death in LDA / A. Wilson, D. Robinson. – 2011.
31. Belwal R.C. Extractive text summarization using clustering-based topic modeling / R.C. Belwal, S. Rai, A. Gupta // Soft Computing. – 2023. – Vol. 27. – № 7. – P. 3965-3982.
32. Bikakis A. Alternative Strategies for Conflict Resolution in Multi-Context Systems / A. Bikakis, G. Antoniou, P. Hassapis // Artificial Intelligence Applications and Innovations III / eds. Iliadis [et al.]. – Boston, MA: Springer US, 2009. – P. 31-40.

33. From Text to Structure: Using Large Language Models to Support the Development of Legal Expert Systems / S. Janatian [et al.] // Legal Knowledge and Information Systems. – IOS Press, 2023. – From Text to Structure. – P. 167-176.
34. Gonçalves R. Evolving Bridge Rules in Evolving Multi-Context Systems / R. Gonçalves, M. Knorr, J. Leite // Computational Logic in Multi-Agent Systems / eds. N. Bulling [et al.]. – Cham: Springer International Publishing, 2014. – P. 52-69.
35. Know, Know Where, Knowwheregraph: A Densely Connected, Cross-Domain Knowledge Graph and Geo-Enrichment Service Stack for Applications in Environmental Intelligence / K. Janowicz [et al.] // AI Magazine. – 2022. – Vol. 43. – Know, Know Where, Knowwheregraph. – № 1. – P. 30-39.
36. Le T. Multi-context Systems with Preferences / T. Le, T.C. Son, E. Pontelli // PRIMA 2015: Principles and Practice of Multi-Agent Systems / eds. Q. Chen [et al.]. – Cham: Springer International Publishing, 2015. – P. 449-466.
37. Smywiński-Pohl A. Classifying the Wikipedia articles into the OpenCyc taxonomy / A. Smywiński-Pohl. – 2012.
38. UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023 / UniProt Consortium // Nucleic Acids Research. – 2023. – T. 51. – UniProt. – № D1. – C. D523-D531.
39. Yi X. A Comparative Study of Utilizing Topic Models for Information Retrieval / X. Yi, J. Allan // Advances in Information Retrieval / eds. M. Boughanem [et al.]. – Berlin, Heidelberg: Springer, 2009. – P. 29-41.

### References

1. Anton K. Topic modeling of natural language texts / K. Anton, G. Andrey // Proceedings of the Institute for System Programming of the Russian Academy of Sciences. - 2012. - Vol. 23. – P. 215-244.
2. Formation of multilayer knowledge graphs based on thematic texts modeling / Pimeshkov V. K. [et al.] // Transactions of the Kola Science Centre of RAS. Series: Engineering Sciences. 2024. Vol. 15, No. 3. P. 50–60.
3. A topic modeling-based bibliometric exploration of automatic summarization research / X. Chen [et al.] // Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. – 2024. – T. 14. – № 5. – P. e1540.
4. Bagul D.V. A novel content-based recommendation approach based on LDA topic modeling for literature recommendation / D.V. Bagul, S. Barve // 2021 6th International Conference on Inventive Computation Technologies (ICICT) 2021 6th International Conference on Inventive Computation Technologies (ICICT). – 2021. – P. 954-961.
5. Bansal S. Topic Modeling Driven Content Based Jobs Recommendation Engine for Recruitment Industry: 5th International Conference on Information Technology and Quantitative Management, ITQM 2017 / S. Bansal, A. Srivastava, A. Arora // Procedia Computer Science. – 2017. – T. 122. – P. 865-872.
6. Benites-Lazaro L.L. Topic modeling method for analyzing social actor discourses on climate change, energy and food security: Special Issue on the Problems of Methods in Climate and Energy Research / L.L. Benites-Lazaro, L. Giatti, A. Giarolla // Energy Research & Social Science. – 2018. – T. 45. – C. 318-330.
7. Blei D.M. Dynamic topic models / D.M. Blei, J.D. Lafferty // Proceedings of the 23rd international conference on Machine learning: ICML '06. – New York, NY, USA: Association for Computing Machinery, 2006. – P. 113-120.
8. Blei D.M. Latent Dirichlet allocation / D.M. Blei, A.Y. Ng, M.I. Jordan // Journal of machine Learning research. – 2003. – T. 3. – P. 993-1022.
9. Brewka G. Equilibria in heterogeneous nonmonotonic multi-context systems / G. Brewka, T. Eiter // Proceedings of the 22nd national conference on Artificial intelligence - Volume 1: AAAI'07. – Vancouver, British Columbia, Canada: AAAI Press, 2007. – P. 385-390.
10. Chang E. The use of SNOMED CT, 2013-2020: a literature review / E. Chang, J. Mostafa // Journal of the American Medical Informatics Association: JAMIA. – 2021. – T. 28. – The use of SNOMED CT, 2013-2020. – № 9. – P. 2017-2026.
11. Dao-Tran M. Streaming Multi-Context Systems / M. Dao-Tran, T. Eiter // Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. – 2017. – P. 1000-1007.
12. Davis E. Testing GPT-4 with Wolfram Alpha and Code Interpreter plug-ins on math and science problems / E. Davis, S. Aaronson arXiv:2308.05713 [cs]. – arXiv, 2025.
13. Garrido-Merchán E.C. GOF AI meets Generative AI: Development of Expert Systems by means of Large Language Models. GOF AI meets Generative AI / E.C. Garrido-Merchán, C. Puente arXiv:2507.13550 [cs]. – arXiv, 2025.
14. Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. BERTopic / M. Grootendorst arXiv:2203.05794 [cs]. – arXiv, 2022.

15. Grover A. node2vec: Scalable Feature Learning for Networks. node2vec / A. Grover, J. Leskovec arXiv:1607.00653 [cs]. – arXiv, 2016.
16. Hybrid toxicology expert system: architecture and implementation of a multi-domain hybrid expert system for toxicology / G. Gini [et al.] // Chemometrics and Intelligent Laboratory Systems. – 1998. – Т. 43. – Hybrid toxicology expert system. – № 1. – P. 135-145.
17. Kommineni V.K. From human experts to machines: An LLM supported approach to ontology and knowledge graph construction. From human experts to machines / V.K. Kommineni, B. König-Ries, S. Samuel arXiv:2403.08345 [cs]. – arXiv, 2024.
18. Landauer T.K. An introduction to latent semantic analysis: Discourse Processes / Landauer T.K., Foltz P.W., Laham D. – 1998. – Т. 25. – № 2-3. – P. 259-284.
19. LLM-assisted Knowledge Graph Engineering: Experiments with ChatGPT / L.-P. Meyer [et al.] // arXiv:2307.06917 [cs]. – 2024. – LLM-assisted Knowledge Graph Engineering. – P. 103-115.
20. Managed Multi-Context Systems. / G. Brewka [et al.] journalAbbreviation: IJCAI International Joint Conference on Artificial Intelligence. – 2011. – 786 p.
21. MDFEND: Multi-domain Fake News Detection / Q. Nan [et al.] // Proceedings of the 30th ACM International Conference on Information & Knowledge Management / arXiv:2201.00987 [cs]. – 2021. – MDFEND. – P. 3343-3347.
22. Online multiscale dynamic topic models / T. Iwata [et al.] // Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining: KDD '10. – New York, NY, USA: Association for Computing Machinery, 2010. – P. 663-672.
23. Ontology summit 2020 communiqué: Knowledge graphs / K. Baclawski [et al.] // Applied Ontology. – 2021. – Т. 16. – Ontology summit 2020 communiqué. – № 2. – P. 229-247.
24. Poo D. Multi-domain expert systems / D. Poo, H. Lu // Expert Systems. – 2007. – Т. 8. – P. 67-73.
25. Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter / Jia Xue [et al.]. – 2020. – Т. 15. – № 9. – P. e0239441.
26. Reactive multi-context systems: Heterogeneous reasoning in dynamic environments / G. Brewka [et al.] // Artificial Intelligence. – 2018. – Т. 256. – Reactive multi-context systems. – P. 68-104.
27. The Hearsay-II Speech-Understanding System: Integrating Knowledge to Resolve Uncertainty / L.D. Erman [et al.] // ACM Comput. Surv. – 1980. – Т. 12. – The Hearsay-II Speech-Understanding System. – № 2. – P. 213-253.
28. TopicGPT: A Prompt-based Topic Modeling Framework. TopicGPT / C.M. Pham [et al.] arXiv:2311.01449 [cs]. – arXiv, 2024.
29. Towards leveraging LLMs for Conditional QA / S.-A. Hussain [et al.] arXiv:2312.01143 [cs]. – arXiv, 2023.
30. Wilson A. Tracking topic birth and death in LDA / A. Wilson, D. Robinson. – 2011.
31. Belwal R.C. Extractive text summarization using clustering-based topic modeling / R.C. Belwal, S. Rai, A. Gupta // Soft Computing. – 2023. – Vol. 27. – № 7. – P. 3965-3982.
32. Bikakis A. Alternative Strategies for Conflict Resolution in Multi-Context Systems / A. Bikakis, G. Antoniou, P. Hassapis // Artificial Intelligence Applications and Innovations III / eds. Iliadis [et al.]. – Boston, MA: Springer US, 2009. – P. 31-40.
33. From Text to Structure: Using Large Language Models to Support the Development of Legal Expert Systems / S. Janatian [et al.] // Legal Knowledge and Information Systems. – IOS Press, 2023. – From Text to Structure. – P. 167-176.
34. Gonçalves R. Evolving Bridge Rules in Evolving Multi-Context Systems / R. Gonçalves, M. Knorr, J. Leite // Computational Logic in Multi-Agent Systems / eds. N. Bulling [et al.]. – Cham: Springer International Publishing, 2014. – P. 52-69.
35. Know, Know Where, Knowwheregraph: A Densely Connected, Cross-Domain Knowledge Graph and Geo-Enrichment Service Stack for Applications in Environmental Intelligence / K. Janowicz [et al.] // AI Magazine. – 2022. – Vol. 43. – Know, Know Where, Knowwheregraph. – № 1. – P. 30-39.
36. Le T. Multi-context Systems with Preferences / T. Le, T.C. Son, E. Pontelli // PRIMA 2015: Principles and Practice of Multi-Agent Systems / eds. Q. Chen [et al.]. – Cham: Springer International Publishing, 2015. – P. 449-466.
37. Smywiński-Pohl A. Classifying the Wikipedia articles into the OpenCyc taxonomy / A. Smywiński-Pohl. – 2012.
38. UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023 / UniProt Consortium // Nucleic Acids Research. – 2023. – Т. 51. – UniProt. – № D1. – P. D523-D531.

39. Yi X. A Comparative Study of Utilizing Topic Models for Information Retrieval / X. Yi, J. Allan // Advances in Information Retrieval / eds. M. Boughanem [et al.]. – Berlin, Heidelberg: Springer, 2009. – P. 29-41.

**Пимешков Вадим Константинович**, Стажер-исследователь, Лаборатория информационных технологий управления региональным развитием, Институт информатики и математического моделирования им. В.А. Путилова ФИЦ КНЦ РАН, Мурманская область, г. Апатиты, Россия; Ассистент, Кафедра информатики и вычислительной техники, филиал ФГАОУ ВО «Мурманский арктический университет» в г. Апатиты, Мурманская область, г. Апатиты, Россия

**Pimeshkov Vadim Konstantinovich**, Research Intern, Laboratory of Information Technologies for Regional Development Management, Putilov Institute for Informatics and Mathematical Modeling FRC KSC RAS, Murmansk region, Apatity, Russia; Assistant, Informatics and Computing Engineering Department, Apatity branch of Murmansk Arctic University, Murmansk region, Apatity, Russia