

СИСТЕМНЫЙ АНАЛИЗ И УПРАВЛЕНИЕ SYSTEM ANALYSIS AND MANAGEMENT

УДК 004.82
DOI 10.52575/2687-0932-2026-53-1-144-152
EDN LFDANJ

Использование моделей BERT и GPT как эффективное решение для автоматизации построения онтологий

Катышев А.М., Аникин А.В.

Волгоградский государственный технический университет
Россия, 400005, г. Волгоград, пр. Ленина, д. 28
anton@anikin.name

Аннотация. В статье рассматривается проблема автоматизации построения онтологических баз знаний из неструктурированных русскоязычных текстов. Основная цель исследования – разработка и оценка комбинированного подхода, повышающего качество извлечения знаний. Методология основана на синергии двух передовых моделей обработки естественного языка (Natural Language Processing): Bidirectional Encoder Representations from Transformers (BERT) для точной идентификации концептов и базовых отношений, и Generative Pre-trained Transformer (GPT) для генерации неявных связей и обогащения онтологии. Экспериментальная оценка на корпусе текстов по веб-разработке показала, что предложенный метод BERT+GPT достигает F1-меры 0.82, что значительно превосходит существующие решения, такие как Text2Onto (0.52) и FRED (0.62). Наиболее важным результатом является статистически значимое улучшение полноты извлечения и итоговой F1-меры, что доказывает эффективность гибридного подхода. Практическая значимость работы заключается в возможности автоматизировать создание баз знаний для русскоязычных информационных и образовательных систем.

Ключевые слова: онтологическая база знаний, граф знаний, извлечение онтологий, BERT, GPT

Для цитирования: Катышев А.М., Аникин А.В. 2026. Использование моделей BERT и GPT как эффективное решение для автоматизации построения онтологий. *Экономика. Информатика*, 53(1): 144–152. DOI 10.52575/2687-0932-2026-53-1-144-152. EDN LFDANJ

Effective BERT and GPT Integration for Ontology Development

Aleksandr M. Katyshev, Anton V. Anikin

Volgograd State Technical University
28 Lenin Ave., Volgograd 400005, Russia
anton@anikin.name

Abstract. This paper addresses the challenge of automated ontology construction, particularly for morphologically rich languages like Russian, where existing tools such as Text2Onto and FRED show significant limitations. We introduce a novel hybrid methodology that synergistically integrates two powerful transformer-based models to build comprehensive ontological knowledge bases from Russian text corpora. The primary objective is to overcome the trade-off between precision and recall inherent in single-model approaches. Our proposed framework operates in a two-stage process. Initially, a Russian-adapted Bidirectional Encoder Representations from Transformers (BERT) model is employed for high-precision

extraction of explicit knowledge. Leveraging its deep contextual understanding, BERT performs named entity recognition to identify candidate concepts and extracts a foundational set of semantic relationships through a sentence-pair classification approach. Subsequently, a fine-tuned Generative Pre-trained Transformer (GPT) model is utilized for knowledge enrichment and recall enhancement. GPT generates plausible hypotheses about unstated or implicit relationships between concepts, refines and verifies relations found by BERT, and resolves logical conflicts, thereby filling knowledge gaps. An empirical evaluation was conducted on a corpus of educational texts on web development to validate the method efficacy. The combined BERT+GPT approach demonstrated superior performance, achieving an F1-measure of 0.82, which significantly surpasses standalone BERT (0.80), FRED (0.62), and Text2Onto (0.52). This improvement is primarily attributed to a substantial increase in recall (0.81) while maintaining high precision (0.82). The practical application and utility of the generated ontologies are discussed in the context of their integration with knowledge management platforms like Stardog, enabling advanced semantic search, data enrichment, and logical inference capabilities.

Keywords: Ontological Knowledge Base, Knowledge Graph, Ontology Learning, BERT, GPT

For citation: Katyshev A.M., Anikin A.V. 2026. Effective BERT and GPT Integration for Ontology Development. *Economics. Information technologies*, 53(1): 144–152 (in Russian). DOI 10.52575/2687-0932-2026-53-1-144-152. EDN LFDANJ

Введение

Онтологические базы знаний (ОБЗ) и графы знаний все чаще используются для организации и применения накопленной информации в различных областях науки и техники [Hogan et al., 2021]. С развитием интернет-технологий объем неструктурированных текстовых данных, в том числе на русском языке, стремительно растет [Anikin et al., 2014; Katyshev et al., 2021]. Автоматическое извлечение знаний из этих текстовых корпусов и последующее формирование ОБЗ позволяют преобразовать разрозненные тексты в структурированное представление знаний, пригодное для семантического поиска, анализа данных и логического вывода. Задача создания онтологических баз знаний из текстов является крайне актуальной в связи с необходимостью организации знаний и поддержки принятия решений в условиях больших данных, а также развитием семантического веба.

Несмотря на значительный прогресс в области обработки естественного языка (NLP) и машинного обучения, задача автоматического построения онтологий остается сложной [Petroni et al., 2019]. Требуются модели, способные понимать контекст русскоязычных текстов, извлекать ключевые концепты и их взаимосвязи, а также формировать формальное представление – онтологию.

Современное состояние вопроса. В области извлечения онтологий из текстовых данных разработан широкий спектр подходов. Систематические обзоры показывают, что использование больших языковых моделей для инженерии онтологий и построения графов знаний является одним из наиболее активно развивающихся направлений [Al-Aswadi et al., 2020; Pan et al., 2023; Viemann, 2005]. Ранние исследования опирались на сочетание лингвистического анализа и статистических методов. Например, Text2Onto является одной из ранних систем для извлечения онтологий из текста, использующей вероятностные методы для выявления терминов, концептов и отношений [Cimiano, Völker, 2005]. Text2Onto представляет собой дальнейшее развитие Text-To-Onto, использующего моделирование знаний с помощью примитивов (концептов, отношений, аксиом), извлекаемых из корпуса текстов посредством анализа частей речи и идентификации совместных упоминаний на основе частоты. Однако последующие исследования выявили ограничения Text2Onto, включая сложность настройки и установки инструмента, зависимость качества от объема обучающих данных и трудности масштабирования на большие текстовые корпуса. В частности, отмечалась необходимость ручного улучшения качества получаемых онтологий.

Другим подходом является инструмент OntoGen, который реализует принцип полуавтоматического построения онтологий [Fortuna et al., 2007]. OntoGen сочетает алгоритмы

кластеризации текстов и интерактивный пользовательский интерфейс для вовлечения эксперта в процесс генерации онтологии. Основные возможности OntoGen включают автоматическое предложение потенциальных концептов на основе тематического моделирования, упрощение их именования и визуализацию онтологии для пользователя. Ограничением OntoGen является то, что он требует активного участия эксперта и не является полностью автоматическим, а качество онтологии в значительной степени зависит от действий пользователя. Кроме того, его применение к русскоязычным текстам требует дополнительных этапов предварительной обработки (например, морфологического анализа) и может приводить к снижению качества кластеризации из-за богатства русской морфологии.

Современные методы все чаще используют достижения в области семантического анализа и лингвистического парсинга. Например, FRED (Friendly Efficient Deductor) – это инструмент машинного чтения для семантического веба, который автоматически преобразует неструктурированные тексты в формальные онтологии (RDF/OWL) [Gangemi et al., 2017]. FRED опирается на комбинаторную категориальную грамматику и фреймовую семантику для глубокой обработки текста: выполняется синтаксический анализ предложений, разрешение анафоры, распознавание именованных сущностей и снятие омонимии в словаре. FRED может обогащать полученную онтологию ссылками на внешние базы знаний (например, DBpedia) и лексические базы (WordNet). Однако у этого инструмента есть ограничения: он в первую очередь ориентирован на английский язык, а визуализация полученных сложных графов знаний часто затруднена из-за их объема. Использование FRED в реальном времени ограничено производительностью глубокой обработки текста.

Помимо перечисленных, существуют и другие подходы, такие как шаблонные методы на основе лексико-синтаксических паттернов (например, шаблоны типа «X есть Y») и системы извлечения открытых отношений (OpenIE). Однако большинство традиционных систем либо требуют трудоемкой настройки, либо не в полной мере используют контекст, что особенно критично для русского языка. Это мотивирует нас обратиться к современным контекстным языковым моделям для повышения качества извлечения знаний.

Постановка проблемы. Целью данного исследования является разработка подхода для извлечения знаний из русскоязычных текстовых корпусов и генерации на их основе онтологической базы знаний. Для достижения этой цели решаются следующие задачи:

- 1) анализ существующих методов и инструментов для извлечения онтологических знаний из текстов;
- 2) разработка комбинированной модели на основе современных языковых моделей BERT и GPT для повышения качества извлечения концептов и отношений;
- 3) экспериментальная проверка эффективности предложенного подхода;
- 4) демонстрация работы модели на конкретном примере;
- 5) обсуждение возможностей интеграции полученных онтологий с платформами управления знаниями.

Объект и методы исследования

Объект исследования. Объектом исследования являются русскоязычные текстовые корпуса из различных предметных областей. В качестве основного примера в данной работе используется корпус учебных материалов по теме «Разработка веб-сайтов на PHP», включающий разделы, описывающие язык PHP, принципы веб-разработки, взаимодействие PHP с базами данных, HTML-разметку и сопутствующие технологии.

Методы исследования. Мы предлагаем комбинированный подход, который объединяет мощь двух современных типов моделей на основе трансформеров, BERT и GPT, для автоматического построения онтологической базы знаний.

Этап 1: извлечение знаний с помощью BERT. На первом этапе используется BERT (Bidirectional Encoder Representations from Transformers) – двунаправленная модель-трансформер, предварительно обученная на большом текстовом корпусе [Devlin et al., 2019]. Мы применяем

русскоязычную версию BERT, адаптированную для учета специфики морфологии и лексики русского языка [Kuratov, Arkhipov, 2019]. BERT выполняет следующие шаги:

1. Распознавание именованных сущностей и ключевых терминов. Модель BERT, дообученная для задачи извлечения сущностей (NER), проходит по текстам и размечает последовательности слов, соответствующие потенциальным концептам. Благодаря использованию контекстного представления распознаются общие сущности независимо от их формальной вариативности. Например, «PHP» и «язык PHP» будут распознаны как связанный концепт PHP (язык программирования). Ключевое преимущество BERT заключается в его способности создавать контекстуальные представления (эмбединги) для каждого слова. Это позволяет модели понимать, что «PHP» и «язык PHP» относятся к одному и тому же концепту, даже если они выражены по-разному. BERT учитывает окружающие слова, чтобы точно определить значение и категорию сущности. Например, если в тексте встречается "PHP", BERT может определить, что это язык программирования, а не случайное сочетание букв, исходя из контекста предложения или всего документа. Результатом этого шага является набор размеченных сущностей из входного текста, каждая из которых ассоциирована с определенной категорией. Например: [(PHP, ЯзыкПрограммирования), (MySQL, СУБД), (HTML, ЯзыкРазметки)].

2. Классификация и кластеризация сущностей. Используя латентные представления (эмбединги) извлеченных терминов из BERT, система вычисляет семантическое сходство между ними. Термины автоматически группируются по значению, что помогает выявить общие концепты более высокого уровня. Например, термины «MySQL» и «PostgreSQL» могут иметь схожие эмбединги и быть классифицированы как реляционные СУБД, в то время как «HTML» и «Markdown» – как язык разметки. Таким образом строится первичная таксономия (иерархия «класс-подкласс»). Результатом является структурированная иерархия концептов, где общие категории объединяют более специфические сущности.

3. Извлечение семантических отношений. BERT также используется для обнаружения отношений между концептами с помощью контекстных представлений. Мы применяем подход классификации пар предложений: для каждой пары кандидатов формируется вход для BERT, состоящий из фразы, в которой упоминаются обе сущности. Модель, дообученная на размеченных данных, определяет тип отношения между X и Y или указывает на его отсутствие. В результате формируется каркас онтологии в виде набора триплетов (субъект, отношение, объект), содержащий отношения «класс-подкласс», «элемент множества», «функциональная связь» и т. д. Главное преимущество здесь – высокая точность (precision) благодаря глубокому контекстуальному пониманию, которое обеспечивает BERT. Модель способна выявлять даже неявно выраженные, но подразумеваемые связи, минимизируя количество ложных срабатываний. BERT в основном ограничен извлечением тех отношений, которые явно или достаточно очевидно выражены в тексте. Это создает потребность в этапе обогащения с помощью GPT.

Основное преимущество BERT – высокая точность (precision) извлечения фактов за счет учета контекста, что снижает количество ложных срабатываний. Однако его использование ограничено извлечением только тех отношений, которые явно выражены в тексте.

Этап 2: обогащение знаний с помощью GPT. На втором этапе используется GPT (Generative Pre-trained Transformer) – авторегрессионная языковая модель [Brown et al., 2020], которая в нашей системе выступает в роли генератора и уточнителя отношений. Мы использовали модель, дополнительно дообученную на текстовом корпусе в области программирования, чтобы максимально приблизить ее к нужному контексту [Karpukhin et al., 2021]. GPT выполняет следующие задачи:

1. Генерация дополнительных отношений. После того как BERT сформировал базовый граф знаний, GPT проверяет возможное наличие семантической связи для каждой пары концептов, не связанных явно. Модели подается описание каждого концепта и вопрос в свободной форме, например: «Как связаны концепты X и Y?». GPT генерирует вероятное

утверждение. Если в сгенерированном тексте явно присутствует отношение (например, «PHP взаимодействует с MySQL для доступа к базе данных»), из него извлекается новый триплет. Это помогает улучшить полноту (recall) онтологии, добавляя триплеты, которые не были явно выражены в исходном корпусе. Если в сгенерированном GPT тексте явно присутствует новое отношение (например, «взаимодействует_с» между «PHP» и «MySQL»), оно извлекается и добавляется в онтологию как новый триплет (Субъект, Отношение, Объект). Это позволяет «наполнить» граф знаний связями, которые не были очевидны из прямого текстового анализа BERT. Этот процесс значительно улучшает полноту онтологии, добавляя ценные связи, которые не были явно выражены в исходном корпусе текстов. GPT восполняет пробелы, используя свои «знания мира» и способность к логическому выводу.

2. Уточнение и верификация существующих отношений. Для каждого отношения, найденного BERT, GPT может сгенерировать уточняющее предложение или краткое описание этого отношения. Например, для триплета (PHP, используется_для, Веб-разработка) GPT может сгенерировать описание: «PHP используется для серверной веб-разработки». Это служит как для проверки правдоподобности отношения (если GPT «согласен» с фактом и может его связно описать), так и для обогащения онтологии аннотациями, что ценно для конечных пользователей. Этот процесс служит двойной цели: проверка правдоподобия – если GPT «согласен» с фактом и может его связно и логично описать, это служит косвенным подтверждением правдоподобия отношения. Неспособность GPT сгенерировать адекватное описание может указывать на потенциальную ошибку или неточность в извлеченном триплете, требующую дальнейшей проверки; обогащение онтологии аннотациями – сгенерированные описания могут быть добавлены к отношениям в онтологии в качестве аннотаций или комментариев. Это делает онтологию более «читаемой» и полезной для конечных пользователей, которым не всегда очевидно значение формальных связей.

3. Разрешение конфликтов и заполнение пробелов. Если в графе знаний обнаруживаются противоречия, GPT используется для их разрешения. Модели на основе GPT способны учитывать огромное количество мировых знаний, поэтому они могут предположить, например, что один из конфликтующих суперклассов является более общим понятием [Bosselut et al., 2019; Petroni et al., 2019]. Современные подходы также включают методы промпт-тюнинга с использованием онтологических знаний, что повышает эффективность при малом количестве примеров [Ye et al., 2022]. Аналогично, GPT может предлагать новые обобщающие концепты для нескольких несвязанных сущностей с общими чертами. GPT также может быть использован для выявления «несвязанных» сущностей, которые, тем не менее, имеют общие черты. Модель может предложить новый, более общий концепт, который объединит эти сущности. Например, если в онтологии присутствуют «PHP», «Python» и «Java» как отдельные языки программирования, но нет общего класса «ЯзыкПрограммирования», GPT может предложить создать такой класс, основываясь на их общих характеристиках и функциях. Это помогает в дальнейшей иерархической организации знаний.

Этап 3: устранение предвзятости. Для решения проблем предвзятости были проведены специальные работы, что является важным аспектом, так как предвзятость и справедливость в больших языковых моделях – активно исследуемая проблема [Haque et al., 2025]. Применялись техники дебиасинга, такие как перевзвешивание и методы проецирования; использовались промпты и механизмы фильтрации для направления модели к более нейтральным и справедливым ответам; формулировались запросы, стимулирующие модель рассматривать различные точки зрения; а также применялись инструкции, поощряющие генерацию оригинальных решений и дополнительное обучение на редких, но успешных примерах.

Метрики оценки. Для оценки качества извлечения информации мы использовали стандартные метрики: полнота (Recall), точность (Precision) и их гармоническое среднее – F1-мера. Эталонные онтологии для оценки были подготовлены вручную экспертами.

Результаты и их обсуждение

Для оценки эффективности разработанного метода были проведены эксперименты на корпусах русскоязычных текстов. В качестве сравниваемых решений были выбраны классический инструмент Text2Onto, современный метод FRED, а также версии нашего подхода, использующие только BERT или только GPT. Поскольку некоторые из сравниваемых инструментов не поддерживают русский язык напрямую, для них тексты предварительно переводились на английский язык. Результаты для корпуса по веб-разработке представлены в табл. 1.

Таблица 1
Table 1

Результаты извлечения онтологии для корпуса по веб-разработке (метрики усреднены по 10 запускам)
Ontology extraction results for the web development corpus (metrics averaged over 10 runs)

Метод	Точность	Полнота	F1-мера
Предложенный BERT+GPT	0,82	0,81	0,82
BERT	0,85	0,75	0,80
GPT	0,78	0,80	0,79
Text2Onto	0,60	0,45	0,52
FRED	0,70	0,55	0,62
OntoGen	0,65	0,50	0,57

Обсуждение результатов. Анализ результатов выявляет явное преимущество предложенного комбинированного подхода. Метод BERT+GPT достиг наивысшей полноты (в некоторых прогонах до 88 %), значительно превосходя классические системы. Увеличение полноты объясняется способностью GPT генерировать скрытые отношения, которые не были явно представлены в тексте. Точность нашего метода (82 %) незначительно уступает чистому подходу на основе BERT (85 %), что является ожидаемым, поскольку генерация новых фактов может приводить к небольшому количеству ошибок. Несмотря на это, итоговая F1-мера (0.82) является самой высокой среди всех сравниваемых методов. Статистически значимое улучшение (на 5–20 процентных пунктов по F1-мере) демонстрирует эффективность синергии BERT и GPT.

Качественный анализ извлеченных онтологий показал, что наш метод обнаруживает как все основные концепты, так и ряд дополнительных, которые были лишь неявно связаны с темой. Инструменты, такие как Text2Onto, часто ограничивались узким набором сущностей, упуская, например, обобщающие категории. FRED, в свою очередь, генерировал избыточный граф со множеством нерелевантных для данной предметной области узлов (шум). Наш метод эффективно фильтрует нерелевантные сущности на этапе BERT и добавляет недостающие важные связи с помощью GPT, в то время как другие модели могут испытывать трудности с обработкой сложных табличных данных [Zhao et al., 2023].

Пример практической реализации. В результате обработки корпуса по теме «Разработка веб-сайтов на PHP» была автоматически построена онтология предметной области. Она включает основные концепты: PHP, Веб-сайт, веб-разработка, база данных, HTML и др., а также отношения между ними: например, PHP «является» языком программирования, PHP «используется в» веб-разработке, PHP «взаимодействует с» MySQL, MySQL «является» базой данных и т. д.

Обсуждение интеграции. Практическая ценность автоматически построенных онтологий раскрывается при их интеграции в существующие инструменты управления знаниями, такие как OntoDog [Zheng et al., 2014] и StarDog.

OntoDog может использоваться для визуализации, редактирования и создания упрощенных «представлений для сообщества» на основе сгенерированной онтологии [Zheng

et al., 2014]. Это позволяет экспертам: (а) скрывать части онтологии, нерелевантные для конкретного приложения; (б) добавлять пользовательские описания или метки; (в) генерировать упрощенную версию для образовательных целей, доводя автоматически сгенерированную структуру до удобного для пользователя состояния.

Stardog, промышленная платформа графов знаний, открывает возможности для выполнения сложных семантических запросов (SPARQL) к данным, извлеченным из текстов. Например, можно найти все технологии, которые «взаимодействуют с базами данных», даже если эти связи были разбросаны по разным документам. Stardog также позволяет обогащать онтологию путем связывания ее узлов с внешними источниками данных (CSV, SQL, API) и поддерживает логический вывод для автоматического получения новых фактов. Например, если определить отношение «взаимодействует_с» как симметричное, система сможет автоматически вывести, что если А взаимодействует с В, то и В взаимодействует с А, повышая полноту знаний.

Перед интеграцией в промышленные репозитории рекомендуется верификация и очистка онтологии, для чего наш подход может предоставлять уровни уверенности для каждого извлеченного факта.

Заключение

В данной работе был предложен и исследован новый комбинированный подход к автоматическому построению онтологических баз знаний из русскоязычных текстов, основанный на совместном использовании моделей BERT и GPT. BERT используется для точного извлечения концептов и базовых отношений, а GPT – для генерации дополнительных связей и обогащения онтологии, что повышает ее полноту.

Экспериментальная оценка продемонстрировала превосходство предложенного метода над существующими решениями со значительным улучшением F1-меры и полноты извлечения онтологии. На практическом примере из области веб-разработки было показано, что метод способен конструировать осмысленную онтологию, отражающую ключевые концепты предметной области и их взаимосвязи.

Научная значимость результатов заключается в развитии методов онтологического инжиниринга, применимых к русскоязычным текстам, а также в демонстрации эффективности гибридного использования предобученных языковых моделей для задач семантического анализа. Практическая ценность работы заключается в возможности автоматизировать создание баз знаний для русскоязычных информационных ресурсов, которые могут быть использованы в системах электронного обучения, интеллектуальных справочных системах и семантическом поиске.

Будущие исследования будут направлены на дальнейшее совершенствование подхода. Планируется инкорпорация более современных моделей (например, GPT-4) [Bhatt et al., 2024], разработка механизмов контроля качества генерируемых онтологий с использованием логических правил и обратной связи от экспертов, а также адаптация метода для многоязычных корпусов, что является важным в специализированных областях, например, в юриспруденции [Zheng et al., 2021]. Конечной целью является интеграция решения в полнофункциональный программный прототип с веб-интерфейсом для генерации онтологий «на лету».

References

- Al-Aswadi F.N., Chan H.Y., Gan K.H. 2020. Automatic ontology construction from text: a review from shallow to deep learning trend. *Artificial Intelligence Review*, 53: 3901–3928. DOI: 10.1007/s10462-019-09782-9.
- Anikin A., Kultsova M., Irina Z., Sadovnikova N., Litovkin D. 2014. Knowledge based models and software tools for learning management in open learning network. In: *Communications in Computer and Information Science*. Vol. 466. Springer, 156-171. DOI: 10.1007/978-3-319-11854-3_15.
- Bhatt A., Vaghela N., Dudhia K. 2024. Generating knowledge graphs from large language models: A comparative study of GPT-4, LLAMA 2, and BERT. arXiv preprint arXiv:2401.07412.

- Biemann C. 2005. Ontology Learning from Text: A Survey of Methods. In: Proceedings of the LDV-Forum, Band 20(2): 75-93.
- Bosselut A., Rashkin H., Sap M., Malaviya C., Celikyilmaz A., Choi Y. 2019. Comet: Commonsense transformers for automatic knowledge graph construction. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy, Association for Computational Linguistics, 1530-1540. DOI: 10.18653/v1/P19-1146.
- Brown T., Mann B., Ryder N., et al. 2020. Language models are few-shot learners. In: Advances in Neural Information Processing Systems 33 (NeurIPS 2020), 1877-1901.
- Cimiano P., Völker J. 2005. A framework for ontology learning and data-driven change discovery. In: Natural Language Processing and Information Systems. Alicante, Spain, Springer, 227-238. DOI: 10.1007/11428817_22.
- Devlin J., Chang M.-W., Lee K., Toutanova K. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. Minneapolis, Minnesota, Association for Computational Linguistics, 4171-4186. DOI: 10.18653/v1/N19-1423.
- Fortuna B., Grobelnik M., Mladenic D. 2007. OntoGen: Semi-automatic Ontology Editor. In: Knowledge Discovery in Databases: PKDD 2007. Warsaw, Poland, Springer, 65-76. DOI: 10.1007/978-3-540-74976-9_9.
- Gangemi A., Presutti V., Reforgiato Recupero D., et al. 2017. Semantic web machine reading with FRED. *Semantic Web*, 8(6): 873-893. DOI: 10.3233/SW-160240.
- Haque F., Xu D., Niu X. 2025. A Comprehensive Survey on Bias and Fairness in Large Language Models. In: Trends and Applications in Knowledge Discovery and Data Mining. Springer, 83-101. DOI: 10.1007/978-981-96-8197-6_7.
- Hogan A., Blomqvist E., Cochez M., et al. 2021. Knowledge graphs. *ACM Computing Surveys (CSUR)*, 54(4): 1-37. DOI: 10.1145/3447790.
- Karpukhin V., Baranchukov A., Burtsev M., Tsetlin Y., Gusev G. 2021. RuGPT-3: Large-scale russian language models with few-shot learning capabilities. arXiv preprint arXiv:2109.04351.
- Katyshev A., Anikin A., Denisov M., Petrova T. 2021. Intelligent Approaches for the Automated Domain Ontology Extraction. In: Advanced Network Technologies and Intelligent Computing. Springer, 81-91. DOI: 10.1007/978-981-96-8197-6_7.
- Kuratov Y., Arkhipov M. 2019. Adaptation of deep bidirectional multilingual transformers for russian language. arXiv preprint arXiv:1905.07213.
- Pan S., Luo L., Wang Y., et al. 2023. Unifying Large Language Models and Knowledge Graphs: A Roadmap. arXiv preprint arXiv:2306.08302.
- Petroni F., Rocktäschel T., Lewis P., et al. 2019. Language models as knowledge bases? In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China, Association for Computational Linguistics, 2763-2773. DOI: 10.18653/v1/D19-1282.
- Ye H., Zhang N., Deng S., et al. 2022. Ontology-enhanced Prompt-tuning for Few-shot Learning. arXiv preprint arXiv:2201.11332.
- Zhao B., Ji C., Zhang Y., et al. 2023. Large language models are complex table parsers. arXiv preprint arXiv:2312.11521.
- Zheng J., Xiang Z., Stoeckert Jr C.J., He Y. 2014. Ontodog: a web-based ontology community view generation tool. *Bioinformatics*, 30(9): 1340-1342. DOI: 10.1093/bioinformatics/btt761.
- Zheng L., Guha N., Anderson B.R., Henderson P., Ho D.E. 2021. When does pre-training help? assessing self-supervised learning for law and the casehold dataset of 53,000+ legal holdings. In: Proceedings of the 18th International Conference on Artificial Intelligence and Law (ICAIL 2021). São Paulo, Brazil, ACM, 159-168. DOI: 10.1145/3462757.3462772.

Конфликт интересов: о потенциальном конфликте интересов не сообщалось.

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

Поступила в редакцию 25.08.2025

Received August 25, 2025

Поступила после рецензирования 27.11.2025

Revised November 27, 2025

Принята к публикации 02.12.2025

Accepted December 02, 2025



ИНФОРМАЦИЯ ОБ АВТОРАХ

Катышев Александр Михайлович, преподаватель кафедры программного обеспечения автоматизированных систем, Волгоградский государственный технический университет, г. Волгоград, Россия

Аникин Антон Викторович, кандидат технических наук, доцент кафедры программного обеспечения автоматизированных систем, Волгоградский государственный технический университет, г. Волгоград, Россия

INFORMATION ABOUT THE AUTHORS

Aleksandr M. Katyshev, Lecturer of the Department of Software for Automated Systems, Volgograd State Technical University, Volgograd, Russia

Anton V. Anikin, Candidate of Technical Sciences, Associate Professor of the Department of Software for Automated Systems, Volgograd State Technical University, Volgograd, Russia