

Научная статья | Original paper

УДК 004.85:004.932

Обзор методов автоматизированного создания онтологий предметных областей науки

Г.В. Филиппов¹ ✉, А.В. Рыкалин^{1,2}, А.А. Хорошилов^{1,2}, С.Э. Чурилов¹

¹ Московский авиационный институт (Национальный исследовательский университет)
Москва, Российская Федерация

² Всероссийский институт научной и технической информации
Российской академии наук, Москва, Российская Федерация

✉ german.filippov2000@yandex.ru

Резюме

Контекст и актуальность. Современные интеллектуальные системы обработки и анализа смыслового содержания отраслевых текстов работают с текстовой информацией, содержащей тысячи классов объектов и неограниченное число смысловых отношений. Все эти знания о высокотехнологичных отраслях могут быть реализованы в онтологических справочниках. Необходимость автоматизированного формирования таких инструментов является актуальной задачей, в том числе для обеспечения доступа исследователей к научным ресурсам. Тем не менее, ручное создание онтологий для высокотехнологичных отраслей чрезвычайно трудоемко, а применение «сырых» больших языковых моделей (LLM, Large Language Models) порождает проблемы «галлюцинаций» и неоднородности данных. Так возникает потребность использования гибридных методов автоматизированного построения отраслевых онтологий, сочетающих точность лингвистического анализа с масштабируемостью LLM. **Цель.** Исследовать существующие методы создания онтологий и методов их сопоставления, рассмотреть возникающие при этом проблемы и предложить новые методы автоматизированного создания отраслевых онтологий предметных областей науки. **Гипотеза.** Интеграция лингвостатистических методов, метода концептуального фразеологического анализа и специализированных LLM в едином цикле позволит повысить полноту и точность автоматического извлечения понятий и родовидовых отношений между ними из научно-технической документации, снижая при этом уровень «галлюцинаций» и неоднородность результирующих онтологий. **Методы и материалы.** Исследование базируется на анализе четырех групп методов: лексико-синтаксических шаблонов (LSP, Lexical Syntax Patterns),



статистических методов частотного анализа, методов глубокого обучения (включая LLM) и гибридных подходов. Для верификации предложенного подхода собран массив из 3000 полнотекстовых документов по авиационной тематике. **Результаты.** В результате анализа существующих методов построения онтологий выявлены ключевые их недостатки. Проведенные предварительные исследования показали, что предложенный метод автоматизированного создания отраслевых онтологий работоспособен и обеспечит возможность решить поставленную задачу с минимальными трудозатратами. Разработан алгоритм создания отраслевых онтологий, в котором LLM используются не изолированно, а в связке с правилами лингвистического анализа. **Выводы.** Показано, что ни один из существующих классов методов автоматического построения онтологий не является универсальным: лингвистические шаблоны требуют ручной адаптации, статистика ограничена по точности, а нейросетевые модели страдают от «галлюцинаций». Разработанный гибридный подход, в отличие от известных решений, позволяет применять LLM не как «черный ящик», а в связке с лингвистическими правилами, обеспечивая замкнутый цикл итеративного уточнения онтологии. Перспективы дальнейшей работы включают масштабирование подхода на другие области науки и интеграцию с существующими базами знаний. Результаты работы будут внедрены в технологический процесс ВИНТИ РАН для решения задач анализа текстов различных предметных областей.

Ключевые слова: онтологии, машинное обучение, автоматизированная обработка текстов, большие языковые модели, семантико-синтаксический анализ, смысловая структура текста

Финансирование. Исследование выполнено при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках государственного задания № FFFU-2025-0010.

Для цитирования: Филиппов, Г.В., Рыкалин, А.В., Хорошилов, А.А., Чурилов, С.Э. (2026). Обзор методов автоматизированного создания онтологий предметных областей науки. *Моделирование и анализ данных*, 16(2), 97–111. <https://doi.org/10.17759/mda.2026160205>

Overview of methods for automated creation of ontologies for scientific domains

G.V. Filippov¹ ✉, A.V. Rykalin^{1,2}, A.A. Khoroshilov^{1,2}, S.E. Churilov¹

¹ Moscow Aviation Institute (National Research University), Moscow, Russian Federation

² All-Russian Institute of Scientific and Technical Information of the Russian Academy of Sciences, Moscow, Russian Federation

✉ german.filippov2000@yandex.ru



Abstract

Context and relevance. Modern intelligent systems for processing and analyzing the semantic content of industry texts work with textual information containing thousands of classes of objects and an unlimited number of semantic relationships. All this knowledge about high-tech industries can be implemented in ontological reference books. The need for automated generation of such tools is an urgent task, including to ensure researchers' access to scientific resources. Nevertheless, the manual creation of ontologies for high-tech industries is extremely time-consuming, and the use of «raw» Large Language Models (LLM) creates problems of «hallucinations» and data heterogeneity. Thus, there is a need to use hybrid methods for automated construction of industry ontologies that combine the accuracy of linguistic analysis with the scalability of LLM. **Objective.** To investigate the existing methods of creating ontologies and methods of their comparison, to consider the problems that arise in this case, and to propose new methods for the automated creation of industry-specific ontologies of subject areas of science. **Hypothesis.** The integration of linguistic statistical methods, the method of conceptual phraseological analysis and specialized LLM in a single cycle will increase the completeness and accuracy of automatic extraction of concepts and generic relationships between them from scientific and technical documentation, while reducing the level of «hallucinations» and the heterogeneity of the resulting ontologies. **Methods and materials.** The research is based on the analysis of four groups of methods: Lexical Syntax Patterns (LSP), statistical methods of frequency analysis, Deep Learning methods (including LLM) and hybrid approaches. To verify the proposed approach, an array of 3,000 full-text documents on aviation topics has been collected. **Results.** As a result of the analysis of existing methods for constructing ontologies, their key disadvantages have been identified. Preliminary studies have shown that the proposed method of automated creation of industry-specific ontological reference books is workable and will provide an opportunity to solve the task with minimal effort. An algorithm for creating industry-specific ontologies has been developed, in which LLMs are used not in isolation, but in conjunction with the rules of linguistic analysis. **Conclusions.** It is shown that none of the existing classes of methods for automatic ontology construction is universal: linguistic patterns require manual adaptation, statistics are limited in accuracy, and neural network models suffer from «hallucinations». The developed hybrid approach, unlike well-known solutions, allows LLM to be applied not as a «black box», but in conjunction with linguistic rules, providing a closed loop of iterative ontology refinement. The prospects for further work include scaling the approach to other scientific fields and integration with existing knowledge bases. The results of the work will be implemented in the technological process of VINITI RAS to solve the problems of text analysis in various subject areas.

Keywords: ontologies, machine learning, automated text processing, large language models, semantic and syntactic analysis, semantic structure of the text

Funding. The study was supported by the Ministry of Science and Higher Education of the Russian Federation, the state task number FFFU-2025-0010.



For citation: Filippov, G.V., Rykalin, A.V., Khoroshilov, A.A., Churilov, S.E. (2026). Overview of methods for automated creation of ontologies for scientific domains. *Modelling and Data Analysis*, 16(2), 97—111. (In Russ.). <https://doi.org/10.17759/mda.2026160205>

Введение

В настоящее время встает ряд задач обеспечения доступа исследователям и исследовательским коллективам к национальным и международным ресурсам, результатам научных исследований на территории страны. Для их решения все большую популярность приобретают технологии семантического полнотекстового поиска, базирующиеся на смысловом анализе информации (Захарова, 2009). Одним из наиболее эффективных инструментов, позволяющих проводить данную модель поиска, являются онтологии. В соответствии с концепцией Грубера, онтологией называется «явная спецификация концептуализации» (Gruber, 1995). Концептуализация — это абстрактная модель понятийного состава отрасли. Явная спецификация — это формальное, машиночитаемое представление понятийной модели отрасли. Таким образом, понятие онтологии можно представить как машиночитаемое представление упрощенной модели знания, необходимое для точного и эффективного обмена знаниями в семантической паутине той или иной отрасли.

Помимо этого, понятийная система высокотехнологичных отраслей может включать сотни тысяч классов объектов, которые участвуют в неограниченном числе отношений и ситуаций. Недостаточное использование лингвистических и онтологических знаний (формализованных представлений терминологии предметной области), используемых в приложениях информационного поиска и автоматической обработки текстов (АОТ), часто не обеспечивают требуемое качество результатов обработки и анализа содержания текстов по причине высокой трудоемкости ручной обработки текстов и создания на их основе онтологий. Существенно повысить качество интеллектуальной обработки можно только путем внедрения в технологии автоматической обработки текстов дополнительных знаний о современном техническом отраслевом языке и знаний об отраслевых технологических процессах.

Для решения этих проблем необходимо получить и описать эти знания, а также зафиксировать их в специальных онтологических справочниках: глоссариях, семантических и концептуальных словарях, тезаурусах, онтологиях. Именно упомянутые справочники должны содержать описания сотен тысяч отраслевых объектов и их смысловых отношений.

Разработчики информационных систем обычно применяют современные методы обработки и анализа текстов, учитывающие как привычные лингвистические, так и нейросетевые методы обработки и анализа смыслового содержания текстов, в частности, бурно развивающиеся большие языковые модели. Тем не менее, их внедрение добавляет ряд технических проблем, таких как эпистемическая неопределенность (EU, Epistemic Uncertainty), «галлюцинации» LLM, неоднородность терминологии, низкий уровень сопоставления онтологий. Все это также необходимо учитывать при составлении онтологий.



Классификация методов автоматизированного создания онтологий предметных областей

Одной из современных парадигм интеллектуальной обработки текстовой информации, описывающих знания о мире и предметных областях, являются онтологические справочники и формальные онтологии. Многие исследователи в этой сфере стремятся разработать сложные формальные подходы и практически аксиоматизированные теории. Однако при таком подходе автоматическая обработка неструктурированных текстов на естественном языке затруднительна ввиду их неоднозначности и неточности (Лукашевич, 2016).

В настоящее время разработано значительное число методов, подходов и технологий, решающих задачи автоматизированного формирования онтологий предметных областей. Существующие методы условно разделяются на следующие группы в зависимости от использования основного подхода:

- 1) методы, основанные на лексико-синтаксических шаблонах;
- 2) методы глубокого и машинного обучения, в частности на основе использования LLM;
- 3) статистические методы;
- 4) гибридные методы.

Методы на основе лексико-синтаксических шаблонов

Такие классические методы, основанные на правилах, предполагают использование лексико-синтаксических шаблонов, которые интегрируют лексические, синтаксические и семантические представления для автоматического извлечения релевантных слов и фраз из текста. Согласно исследованию (Sadimekova et al., 2024), существует возможность автоматического создания LSP на основе шаблонов проектирования онтологий (ODP, *Ontology Design Patterns*) (Gangemi, Presutti, 2009). Анализ эффективности показывает, что методы на основе лексико-синтаксических шаблонов демонстрируют высокую результативность лишь при работе с текстами узкоспециализированных предметных областей с унифицированной терминологией, при этом обладая высоким уровнем прозрачности и не требуя обучающих выборок. Однако использование таких методов требует точечной настройки и доработки шаблонов под конкретную предметную отрасль, что включает в себя необходимость добавления специфических понятий и свойств для каждой конкретной задачи. Кроме того, методы на основе LSP показывают низкие показатели полноты при вариативности естественного языка.

Подход на основе статистических методов

Подход, базирующийся на использовании статистических методов, относится к методам статистического анализа данных на естественном языке (Хорошилов, Никитин, Будзко, 2014). Для реализации этих методов требуются большие объемы текстовой информации — репрезентативные корпуса текстов.

На первом этапе в каждой коллекции документов выделяют термины-кандидаты (как правило, существительные и именные группы), определяется их частота



встречаемости. Использование частотных критериев позволяет значительно сократить число предполагаемых классов понятий.

На втором этапе проводится ранжирование терминов по частоте встречаемости и другим статистическим критериям, оценивается пересечение различных коллекций по используемым терминам.

Этап выделения отношений между классами является наиболее сложным. В рамках этого подхода предполагается, что первоначально строится не произвольная онтология, а тезаурус — таксономия с терминами. В качестве базовых отношений, действующих между терминами, определяются отношения «is-a» и «synonym-of». Для выделения отношения «is-a» обычно используют количественный подход к информации. Он позволяет выделить только базовые отношения, необходимые для построения таксономии. Однако возможно его расширение для выделения и других отношений (Хорошилов, Кан, Филиппов, 2024).

Основной недостаток такого подхода — необходимость наличия огромных массивов текстовой информации, а также выполнения требований к репрезентативности этих наборов текстов. Помимо этого, подход на основе статистических методов обладает недостаточной семантической точностью, особенно при работе с редкими терминами, а также присутствием значительного числа гиперонимов, гипонимов.

Методы машинного обучения и нейросетевые подходы

В настоящее время методы создания онтологий предметных отраслей науки на основе машинного и глубокого обучения набирают все большую популярность.

В работе (Sivaramakrishnan, Suchithra, 2019) авторами предложено применение методов интеллектуального анализа данных посредством использования сверточных нейросетей на основе семантических и тематических графов. Предлагаемый метод устраняет ограничение на поиск ключевых слов, которым обладают большинство классических подходов к рассматриваемой задаче, путем применения метода преобразования запросов. В его основе — формирование внутренней пары «запрос — значение» и последующее выведение списка предыдущих запросов пользователей, что в итоге сокращает время и затраты на вычисления и повышает производительность. Тем не менее, стоит учитывать неинтерпретируемость процесса обучения модели.

Авторами работы (Сидорова, Иванов, Овчинникова, 2025) предложено использование нейросетевых моделей и больших языковых моделей для извлечения из текстов терминов и семантических связей между ними. В данной работе представлены динамически формирующиеся инструкции для LLM на основе результатов предыдущих этапов анализа. Общая схема принципа работы LLM представлена на рис. 1. Результаты экспериментов показали хорошие значения F1-меры для задач извлечения терминов и связей между ними.

Однако при использовании методов и моделей искусственного интеллекта необходимо учитывать проблему неоднородности создаваемых онтологий, а также высокую вероятность проявления «галлюцинаций» у LLM. В целях снижения данных показателей необходимо подготавливать целостные онтологии, которые обладают



внутренним единством, логической связностью и достаточной полнотой для последующей передачи данных нейросетевой модели. Далее будут рассмотрены алгоритмы, используемые для решения указанных проблем.

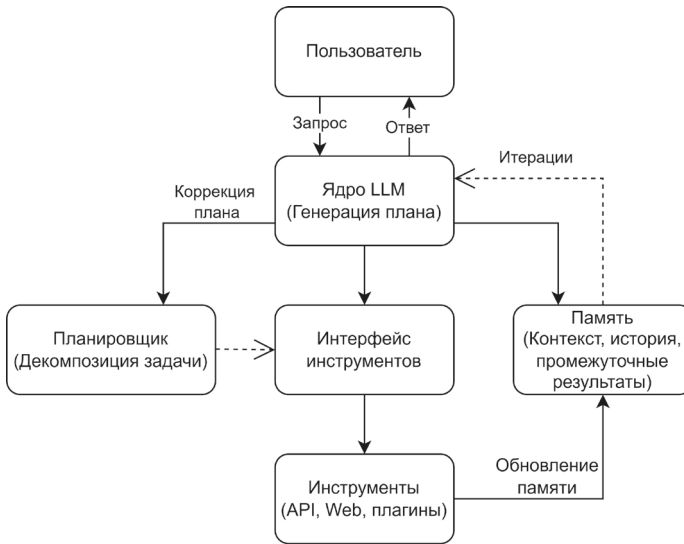


Рис. 1. Схема принципа работы LLM

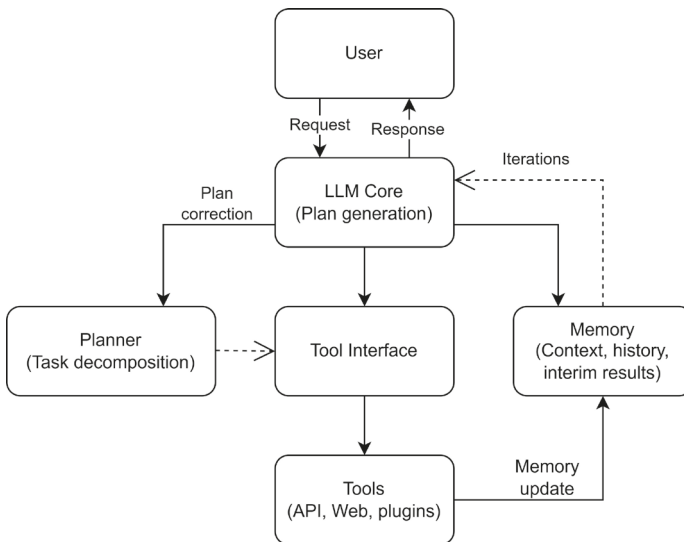


Fig. 1. LLM operating principle diagram



Гибридные методы

Современные исследования в области онтологического инжиниринга демонстрируют устойчивый тренд к гибридизации различных подходов (Ghidalia, 2023).

В работе (Чуднов, 2025) рассматривается система динамического управления знаниями (СДУЗ). Она состоит из модуля извлечения знаний, основанного на дообученной модели ruBERT, и логического интерпретатора на основе Prolog.

СДУЗ позволяет решать задачи распознавание именованных сущностей (NER, Named Entity Recognition) и извлечения отношений (RE, Relation Extraction). Разработчики предложили механизм обратной связи, позволяющий обнаружить противоречия между предсказаниями нейросети и логическими правилами, проводить экспертную коррекцию и дообучать трансформер на скорректированных данных. Однако сложность настройки правил требует большего вовлечения экспертов в работу с системой.

Иной подход реализован в работе (Kukreja, 2025), где предложен end-to-end пайплайн изучения онтологий, использующий LLM (GPT-4o) для извлечения концептов и отношений в сочетании с гибридным интерпретатором принятия решений на основе multi-factor similarity score. Система включает human-in-the-loop интерфейс для валидации и обеспечивает хранение результатов в графовой базе данных Neo4j.

Для преодоления недостатков отдельных методов создания онтологий в настоящей работе предлагается гибридный подход, включающий лингвостатистические методы, метод концептуального фразеологического анализа (Хорошилов, Кан, Филиппов, 2024) и использование LLM.

Данный метод должен позволять построить замкнутый цикл автоматизированного создания онтологий на основе текстов различных предметных областей, в частности, автоматического установления родовидовых отношений между понятиями.

Алгоритм состоит из следующих этапов:

1. сбор массива научно-технической документации (НТД) по интересующей тематике;
2. проведение обработки текста посредством разработанного программного комплекса обработки текстов, включающего средства токенизации и морфологического анализа (МА);
3. определение системы объектов в собранном массиве на основе использования семантико-синтаксического (ССА) и статистического анализа;
4. проведение вторичной обработки на основе применения методов концептуального анализа (КА);
5. преобразование полученных результатов в онтологическую структуру посредством генерации RDF-триплетов и импорт в редактор Protégé через OWL-интерфейс;
6. дообучение нескольких LLM (в частности, моделей семейств Qwen и Deepseek) на массивах текстов и результатах предыдущих этапов. Каждая модель должна решать свою задачу в процессе:
 - a) Модель А — токенизация и морфологический анализ;
 - b) Модель В — статистический анализ;
 - c) Модель С — семантико-синтаксический анализ (на основе результатов первой модели),
 - d) Модель D — концептуальный анализа (на основе результатов модели ССА),



- е) Модель Е — формирование онтологической модели, преобразование результатов концептуального анализа и ССА в онтологические структуры, в частности, в онтологии / тезаурусы / таксономии в виде RDF / OWL-файлов;
7. проведение тестирования дообученных моделей. В качестве метрик качества использовать точность, полноту и F1-меру. В первых итерациях планируется достичь значения F1-меры не ниже 0,8 на тестовых данных.
- Подробная схема алгоритма представлена на рис. 2.

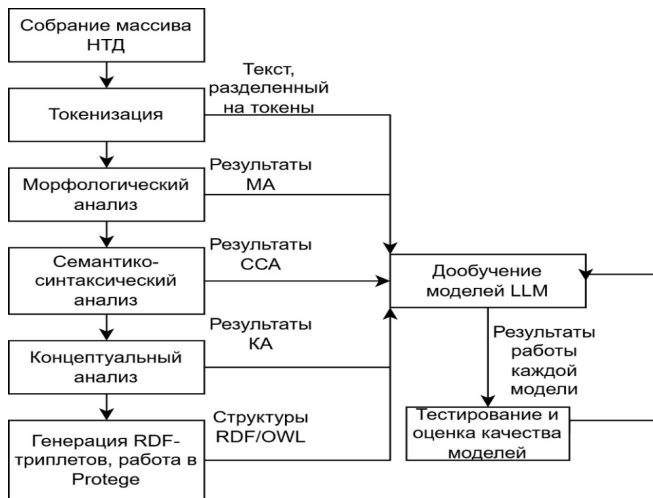


Рис. 2. Схема алгоритма гибридного подхода, основанного на лингвостатистических методах и LLM

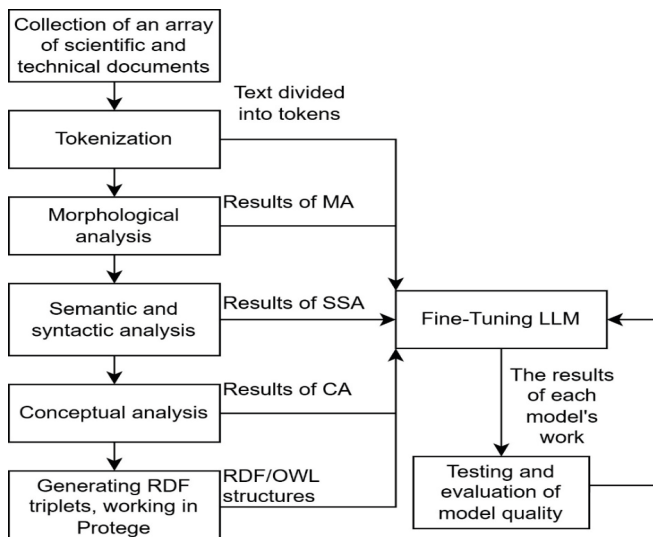


Fig. 2. Algorithm diagram of a hybrid approach based on linguo-statistical methods and LLMs



На данном этапе работы собран массив из 3000 полнотекстовых документов из открытых источников по авиационной тематике, реализована обработка текстов при помощи средств токенизации и морфологического анализа.

Классификация методов сопоставления онтологий предметных областей

В настоящее время среди основных методов сопоставления онтологий предметных областей выделяют традиционные и эволюционные методы (рис. 3).

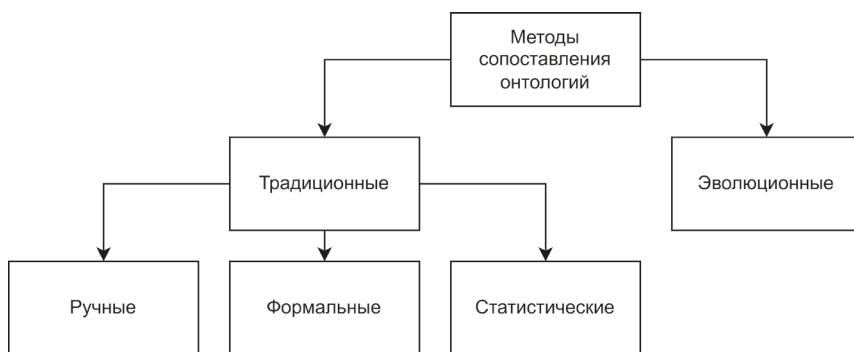


Рис. 3. Классификация методов сопоставления онтологий

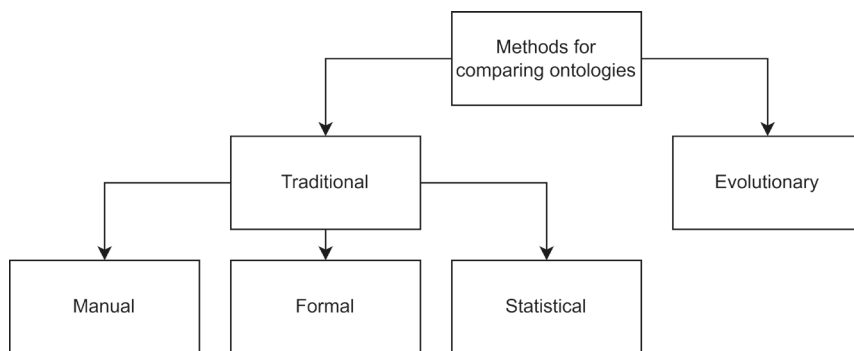


Fig. 3. Classification of ontology comparison methods

Традиционные методы сопоставления

Традиционные методы включают ручные, формальные и статистические подходы. Первые включают лексикографическое сравнение, сравнение с учетом значений и создание «онтологий верхнего уровня» (Maedche, Staab, 2002). Применение таких методов обычно сопровождается финансовыми трудностями, однако позволяет решить задачу с относительно высокой точностью.



Формальные методы включают в себя структурный анализ (оценка сходства связей между сущностями), анализ сходства по перекрестным связям (оценка сходства через похожие классы и типы связей в онтологиях) и логический вывод (анализ свойств классов и экземпляров классов онтологий, с помощью которых из имеющихся соответствий логически выводятся другие соответствия) (Рогошина, 2017). Такой подход обладает высокой вычислительной сложностью и дает качественные результаты при работе лишь с узкоспециализированной терминологией.

Статистический подход к сопоставлению онтологий учитывает частоту встречаемости терминов в корпусах текстов и применяет меры сходства, количественно оценивая, насколько выявленные сущности подобны. Основные недостатки таких методов описаны выше.

Эволюционные алгоритмы сопоставления

Эволюционная оптимизация (мета-сопоставление) основывается на поиске оптимальных весовых коэффициентов для комбинации различных мер сходства (лексических, структурных). Однако применение эволюционных алгоритмов (ЕА) в задачах сопоставления онтологий сопряжено с рядом ограничений, среди которых преждевременная сходимость (конвергенция), приводящая к субоптимальным решениям, а также необходимость априорного задания эталонных соответствий для оценки качества найденных решений.

Для устранения данных недостатков авторами работы (Lv, Jiang, Li, 2021) предложен усовершенствованный подход к сопоставлению онтологий на основе ЕА, в котором для оценки качества решения представлены два приблизительных оценочных показателя — псевдовоспоминание и псевдоточность. В свою очередь, для преодоления преждевременной сходимости используется адаптивное давление отбора, которое регулируется в зависимости от стадии эволюции.

В работе (Lv, Jiang, Li, 2021) выделяют применение ЕА на основе входных данных без эталонного выравнивания (NRA, No Reference Alignment) и с ним (RA, Reference Alignment). Разработанный авторами алгоритм Improved EA+NRA отмечается высокой точностью обработки, но низкой производительностью по сравнению с ЕА+РА.

Примером применения эволюционного алгоритма можно указать работу (Кравченко Д., Кравченко Ю., Марков, 2020), где авторы для решения задач сопоставления крупных онтологий предложили в перспективу биоинспирированные алгоритмы. Гибридизация бактериального поиска и алгоритма кукушки позволила им достичь экономии времени до 13% по сравнению с традиционными методами сопоставления при обработке онтологий с 500 000 вершин.

Заключение

Проведенный анализ существующих подходов к решению задачи автоматизированного построения онтологий предметных областей науки показал, что ни один из существующих классов методов не является универсальным:



- лингвистические шаблоны требуют трудоемкой адаптации;
- статистические методы ограничены недостаточной семантической точностью;
- нейросетевые подходы имеют проблемы с нехваткой размеченных данных и «галлюцинациями».

В связи с этим, был предложен гибридный подход автоматизированного создания тематических онтологий предметных областей науки, основывающийся на интеграции лингвостатистических методов и LLM. Разработан алгоритм данного подхода, который позволяет построить замкнутый цикл автоматизированного создания онтологий.

В отличие от известных решений, алгоритм позволяет:

- применять LLM не как «черный ящик», а в связке с правилами лингвистического анализа;
- обеспечить замкнутый цикл: онтология — обучение модели — уточнение онтологии.

В дальнейшем планируется масштабирование подхода на другие области науки и интеграция с базами знаний.

Результаты работы будут внедрены в технологический процесс ВИНТИ РАН для решения задач анализа текстов различных предметных областей.

Список источников / References

1. Захарова, И.В. (2009). *Математическая модель семантического поиска с использованием онтологического подхода: Автореф. дис. канд. физ.-мат. наук*. Челябинский государственный университет. Челябинск.
Zakharova, I.V. (2009). *Mathematical model of semantic search using an ontological approach: Extended abstr. Diss. Cand. Sci. (Ph.-Mat.)*. Chelyabinsk State University. Chelyabinsk. (In Russ.).
2. Кравченко, Д.Ю., Кравченко, Ю.А., Марков, В.В. Гибридный биоинспирированный алгоритм отображения онтологий в задачах извлечения и управления знаниями. — *Известия ЮФУ. Технические науки*. 2020. — Ростов-на-Дону: ЮФУ, 2020. — с. 16—28.
Kravchenko, D.Yu., Kravchenko, Yu.A., Markov, V.V. Hybrid bioinspired algorithm for mapping ontologies in knowledge extraction and management tasks. — *Izvestiya YuFU. Tekhnicheskie nauki*. 2020. — Rostov-na-Donu: YuFU, 2020. — pp. 16—28. (In Russ.).
3. Лукашевич, В.Н. (2016). *Модели и методы автоматического анализа текстов на основе лингвистических онтологий в информационно-аналитических системах: Автореф. дис. докт. тех. наук*. Нижегородский государственный технический университет им. Р.Е. Алексеева. Нижний Новгород.
Lukashevich, V.N. (2016). *Models and methods of automatic text analysis based on linguistic ontologies in information and analytical systems: Extended abstr. Diss. Dr. Sci. (Techn.)*. Nizhny Novgorod State Technical University named after R.E. Alekseev. Nizhny Novgorod. (In Russ.).
4. Рогушина, Ю.В. Сопоставление семантических информационных ресурсов Web на основе онтологического анализа. *International Journal «Information Technologies & Knowledge»*. — Vol. 11, N. 1. — 2017. — с. 49—71.



- Rogushina, Yu.V. Comparison of semantic Web information resources based on ontological analysis. *International Journal «Information Technologies & Knowledge»*. — Vol. 11, N. 1. — 2017. — pp. 49—71. (In Russ.).
5. Сидорова, Е.А. Иванов, А.И., Овчинникова, К.А. Извлечение информации из текстов на основе онтологии и больших языковых моделей. *Онтология проектирования*. — № 1, Т. 15. — 2025. — с. 114—129.
Sidorova, E.A., Ivanov, A.I., Ovchinnikova, K.A. Information extraction from texts based on ontology and large language models. *Ontologiya proektirovaniya*. — № 1, Т. 15. — 2025. — pp. 114—129. (In Russ.).
6. Хорошилов, А.А., Никитин, Ю.В., Будзко, В.И. Автоматическое создание формализованного представления смыслового содержания неструктурированных текстовых сообщений СМИ и социальных сетей. *Системы высокой доступности*. — 2014. — Т. 10, № 3. — С. 52—69.
Khoroshilov, A.A., Nikitin, Yu. V., Budzko, V.I. Automatic creation of a formalized representation of the semantic content of unstructured text messages in the media and social networks. *Sistemy vysokoi dostupnosti*. — 2014. — Т. 10, № 3. — P. 52—69. (In Russ.).
7. Хорошилов, А.А., Кан, А.В., Филиппов, Г.В. Методы автоматического формирования комплекса онтологических справочников по авиакосмической отрасли. *Научно-технический вестник Поволжья*. № 5. — 2024. — Казань: ООО «Рашин Сайнс», 2024. — с. 150—154.
Khoroshilov, A.A., Kan, A.V., Filippov, G.V. Methods of automatic formation of a complex of ontological reference books on the aerospace industry. *Nauchno-tehnicheskii vestnik Povolzh'ya*. № 5. — 2024. — Kazan': ООО «Rashin Sains», 2024. — pp. 150—154. (In Russ.).
8. Чуднов, И.И. Нейро-символические гибридные системы для динамического управления знаниями. *КОД науки — исследования молодых ученых*. № 5 (73). — 2025. — Москва: МГПУ, 2025.
Chudnov, I.I. Neuro-symbolic hybrid systems for dynamic knowledge management. *KOD nauki — issledovaniya molodykh uchenykh*. № 5 (73). — 2025. — Moskva: MGPU, 2025. (In Russ.).
9. Gangemi, A., Presutti, V. Ontology Design Patterns. *Handbook on Ontologies*. — 2009. — pp. 221—243.
10. Ghidalia, S. *Combining Machine Learning and Ontology: A Systematic Literature Review*. — 2023. URL: <https://ar5iv.labs.arxiv.org/html/2401.07744> (дата обращения: 20.12.2025).
11. Gruber, T.R. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human — Computer Studies*. — 1995. — Т. 43. — № 5—6. — pp. 907—928.
12. Kukreja, S. *End-to-End Ontology Learning Pipeline for Technical Documents*. — 2025. URL: <https://github.com/itsssnehin/Ontology-Learning> (дата обращения: 20.12.2025).
13. Lv, Q., Jiang, C., Li, H. *Solving Ontology Meta — Matching Problem Through an Evolutionary Algorithm with Approximate Evaluation Indicators and Adaptive Selection Pressure*. *IEEE Access*. — Vol. 9, 2021. — pp. 3046—3064. DOI: 10.1109/ACCESS.2020.3047875
14. Maedche, A., Staab, S. Measuring Similarity between Ontologies Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web. *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management, EKAW02*. — Vol. 2473 of LNAI, Berlin, Springer Verlag. — 2002.



15. Sadirmekova, Z. et al. Approach to automating the construction and completion of ontologies in a scientific subject field. *International Journal of Electrical and Computer Engineering*. Vol. 14, No. 3. — 2024. — p. 3064—3072.
16. Sivaramakrishnan, R.G., Suchithra, R. Automatic Relationship Construction in Domain Ontology Engineering using Semantic and Thematic Graph Generation Process and Convolution Neural Network. *International Journal of Recent Technology and Engineering*. — Vol. 8. — I. 3. — 2019. — pp. 4602—4610.

Информация об авторах

Герман Владиславович Филиппов, аспирант, институт «Компьютерные науки и прикладная математика», Московский авиационный институт (МАИ), Москва, Российская Федерация, ORCID: <https://orcid.org/0009-0009-6669-6518>, e-mail: german.filippov2000@yandex.ru

Андрей Владимирович Рыкалин, аспирант, ассистент кафедры математической кибернетики, институт «Компьютерные науки и прикладная математика», Московский авиационный институт (МАИ), Москва, Российская Федерация, младший научный сотрудник, Всероссийский институт научной и технической информации Российской академии наук, Москва, Российская Федерация, ORCID: <https://orcid.org/0009-0002-2791-8789>, e-mail: rykalin.2000@mail.ru

Александр Алексеевич Хорошилов, доктор технических наук, ведущий научный сотрудник, Всероссийский институт научной и технической информации Российской академии наук, Москва, Российская Федерация, профессор кафедры математической кибернетики, институт «Компьютерные науки и прикладная математика», Московский авиационный институт (МАИ), Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0001-6641-3105>, e-mail: khoroshilov@mail.ru

Сергей Эдуардович Чурилов, магистрант, институт «Компьютерные науки и прикладная математика», Московский авиационный институт (МАИ), Москва, Российская Федерация, e-mail: churilov.ser1204@gmail.com

Information about the authors

German V. Filippov, Graduate student, Institute of Computer Science and Applied Mathematics, Moscow Aviation Institute, Moscow, Russian Federation, ORCID: <https://orcid.org/0009-0009-6669-6518>, e-mail: german.filippov2000@yandex.ru

Andrey V. Rykalin, Graduate student, Assistant of the Department of Mathematical Cybernetics, Institute of Computer Science and Applied Mathematics, Moscow Aviation Institute (MAI), Moscow, Russian Federation, Junior Researcher, All-Russian Institute of Scientific and Technical Information of the Russian Academy of Sciences, Moscow, Russian Federation, ORCID: <https://orcid.org/0009-0002-2791-8789>, e-mail: rykalin.2000@mail.ru

Alexander A. Khoroshilov, Doctor of Technical Sciences, Senior Researcher, All-Russian Institute of Scientific and Technical Information of the Russian Academy of Sciences, Moscow, Russian Federation, Professor of the Department of Mathematical Cybernetics, Institute of Computer Science and Applied Mathematics, Moscow Aviation Institute (MAI), Moscow, Russian Federation, ORCID: <https://orcid.org/0000-0001-6641-3105>, e-mail: khoroshilov@mail.ru



Sergey E. Churilov, Master's student, Institute of Computer Science and Applied Mathematics, Moscow Aviation Institute (MAI), Moscow, Russian Federation, e-mail: churilov.ser1204@gmail.com

Вклад авторов

Все авторы внесли равный вклад в концепцию, проведение исследования, анализ данных и подготовку рукописи.

Все авторы приняли участие в обсуждении результатов и согласовали окончательный текст рукописи.

Contribution of the authors

The authors contributed equally to the research, data analysis, and preparation of this manuscript.

All authors participated in the discussion of the results and approved the final text of the manuscript.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Conflict of interest

The authors declare no conflict of interest.

Поступила в редакцию 16.02.2026

Поступила после рецензирования 02.03.2026

Принята к публикации 06.03.2026

Опубликована 30.06.2026

Received 2026.02.16

Revised 2026.03.02

Accepted 2026.03.06

Published 2026.06.30