

Научная статья

УДК 811'111

DOI: <https://doi.org/10.18721/JHSS.16203>EDN: <https://elibrary/DCDYJE>

ОНТОЛОГИЯ КАК ИНСТРУМЕНТ ТЕРМИНОЛОГИЧЕСКОЙ ОПТИМИЗАЦИИ МАШИННОГО ПЕРЕВОДА ЛЕКСИКИ НЕФТЕГАЗОВОЙ ОТРАСЛИ

С.В. Калинина ✉, **Л.В. Коцюбинская**Ленинградский государственный университет имени А.С. Пушкина,
Санкт-Петербург, г. Пушкин, Российская Федерация✉ movable@yandex.ru

Аннотация. Актуальность исследования обусловлена проблемой терминологической неточности машинного перевода узкоспециализированных текстов нефтегазовой тематики. Владение терминологией предметной области является важнейшим условием обеспечения эффективной межязыковой коммуникации в глобальных бизнес-проектах. Основанные на архитектуре трансформеров современные большие языковые модели демонстрируют ограничения в обработке лексики языка для специальных целей, что требует обращения к компьютерной лексикографии. В эпоху искусственного интеллекта онтология выступает ключевым лексикографическим инструментом, обеспечивая семантическую точность в определенных предметных областях, а лингвистическая компетенция специалиста по обработке текстов на естественных языках выступает основой для эффективного управления техническими знаниями, обеспечивая адаптацию трансформеров к узкоспециализированным контекстам. Цель статьи — обосновать целесообразность применения онтологического подхода к машинной обработке узкоспециализированных текстов на естественных языках. Для достижения цели необходимо решить ряд задач: аргументировать использование онтологического подхода к лексике языка для специальных целей на современном этапе развития прикладной лингвистики; описать существенные характеристики актуальной английской нефтегазовой терминологии; создать фрагмент онтологии нефтегазового оборудования с использованием редактора Protégé; обработать технический текст в паре EN_RU при помощи модели-трансформера DeepSeek R1 с учетом фрагмента онтологии; представить результаты в виде сравнительной таблицы. Материалом исследования является актуальная терминологическая лексика, функционирующая в англоязычной научно-технической литературе и периодике нефтегазовой предметной области. Методологическую базу статьи составляют методика фреймового анализа терминологии, метод построения онтологии. Результаты исследования имеют практическое значение для переводческих нефтегазовых проектов: использование онтологии в трансформерах обеспечивает повышенную корректность перевода терминов по сравнению с базовой моделью, позволяет адаптировать большие языковые модели к требованиям отраслевой терминологии. Таким образом, онтологический подход трансформирует переводческие практики, сочетая мощь технологических инноваций с лингвистической экспертизой для глубинного понимания отраслевой лексики при передаче технического знания.

Ключевые слова: термины в нефтегазовой отрасли, модель-трансформер, автоматизация лексикографии, онтология, терминологический менеджмент.

Для цитирования: Калинина С.В., Коцюбинская Л.В. Онтология как инструмент терминологической оптимизации машинного перевода лексики нефтегазовой отрасли // Terra Linguistica. 2025. Т. 16. № 2. С. 41–55. DOI: 10.18721/JHSS.16203



ONTOLOGY AS A TOOL FOR MASHINE TRANSLATION TERMINOLOGICAL OPTIMIZATION OF OIL AND GAS INDUSTRY LEXIS

S.V. Kalinina  , L.V. Kotsyubinskaya

Pushkin Leningrad State University,
St. Petersburg, Pushkin, Russian Federation

 movable@yandex.ru

Abstract. The study relevance is stipulated by terminological inaccuracy in machine translation of highly specialized oil and gas industry texts. Competence in domain-specific terminology is the essential prerequisite for effective cross-linguistic communication in global business projects. The modern large language models (LLMs) based on the transformer architecture feature the limitations processing language for special purposes (LSP); thereby appealing to computational lexicography becomes necessary. In the AI era, ontology is a key lexicographic tool providing semantic precision within the certain domains. The linguistic competence of a specialist engaged in natural language text processing is the base for technical knowledge effective management, ensuring transformer adaptation to highly specialized contexts. The article is aimed to substantiate the reasonability of ontological approach application to machine processing of highly specialized texts in natural language. To achieve that aim several tasks shall be addressed: justification of ontological approach to the LSP at the present stage of the applied linguistics progress; description of essential properties of actual English oil and gas terminology; creation of oil and gas equipment ontology fragment using Protégé editor; processing of EN_RU technical text using the DeepSeek R1 transformer model considering the ontology fragment; presentation of the results in a comparative table format. The research material comprises the current terminology lexis functioning in English scientific and technical literature and oil and gas industry periodicals. Methodological basis includes terminology frame analysis procedure, ontology building method. The research results are practically significant for oil and gas translation projects: the ontology application in transformers ensures higher terminological accuracy in comparison with a basic model, providing LLMs adaptation to the technical terminology requirements. Thus, the ontological approach transforms translation procedures through the combination of technology innovation and linguistic expertise for deeper understanding of the technical lexis in specialized knowledge spreading.

Keywords: oil and gas industry terminology, transformer model, lexicographic automation, ontology, terminology management.

Citation: Kalinina S.V., Kotsyubinskaya L.V., Ontology as a tool for mashine translation terminological optimization of oil and gas industry lexis, *Terra Linguistica*, 16 (2) (2025) 41–55. DOI: 10.18721/JHSS.16203

Введение: постановка проблемы

Достижения научно-технического прогресса сопровождаются необходимостью именования появляющихся понятий, что объясняет причины возрастания объемов специальной лексики – «номинативных единиц языка для специальных целей, обозначающих научные, технические, экономические, общественно-политические, шире – специальные понятия» [1, с. 76]. Так, лексический состав естественных языков является средством организации, хранения и передачи накопленной человечеством информации [2]. Значимость адекватной передачи знания не вызывает сомнения, при этом особая важность придается передаче специальной лексики той или иной предметной области, поскольку адекватная и согласованная терминология является



залогом перевода должного качества¹ и, как следствие, результативности профессиональной коммуникации².

В свете эволюции машинной обработки текстов на естественных языках особенно актуально понимание теоретических и прикладных основ языкознания. Технологии искусственного интеллекта прочно внедряются во все сферы деятельности человека. Переводческая деятельность и сопряженная с ней терминологическая работа не являются исключением. Отечественные и зарубежные исследователи говорят о современных системах машинного перевода [3, 4] и больших языковых моделях как об удобном инструменте оперативного выполнения переводческой задачи [5]. Вместе с тем задачи семантического переосмысления вторичной номинации терминологической лексики предметной области, анализа ошибок и разработки способов их исправления пока не могут решаться полностью автоматически.

Терминологическое редактирование переводов специальных текстов заключается в достижении «правильного использования терминологических единиц с позиции их роли в конкретном тексте и места среди других единиц данной отрасли знаний и/или деятельности внутри терминосистемы» [1, с. 196–203]. Учитывая тот факт, что используемые в современном машинном переводе нейронные сети разработаны на огромных массивах информации, которые не включают в себя ориентированные на терминологию областей знания тексты, возрастает необходимость лингвистической экспертизы сгенерированных большими языковыми моделями материалов [6, 7].

Статья фокусируется на выявлении способов адекватной передачи терминологической лексики при обработке узкоспециализированных текстов на естественных языках при помощи трансформеров – терминологической адаптации больших языковых моделей через создание онтологии.

Методология исследования

Современные научные исследования основываются на интеграции различных наук: «...междисциплинарность отчетливо проявляется в лингвистическом пространстве, где многообразие направлений, новые оригинальные подходы к анализу языка выдвигаются перед исследователями в большом количестве разноплановых взглядов и теорий» [8, с. 11].

Методология статьи включает в себя комбинирование метода лингвистического анализа терминологии, метода ее стандартизации как практической сферы терминоведения, метода создания лингвистической онтологии с учетом архитектуры модели-трансформера DeepSeek R1 [9].

Лексические единицы современной англоязычной терминосистемы нефтегазовой отрасли проанализированы в диссертационном исследовании³ при помощи следующих методов: наблюдения, описания, анализа по непосредственным составляющим, дефиниционного, этимологического, компонентного, контекстологического видов анализа, структурно-функционального метода, метода языкового поля, категориального анализа языковых единиц, метода моделирования с использованием инструментария теории когнитивной метафоры и теории идеализированных когнитивных моделей, методики фреймового анализа.

Цифровая трансформация и междисциплинарный характер словарной науки расширили перечень подходов к исследованию терминологии: информация должна быть представлена в удобном для компьютерной обработки виде [10]. Использование метода стандартизации терминологической лексики предполагает следование рекомендациям определенных руководств

¹ Герд А.С. Специальный текст как предмет прикладного языкознания // Прикладное языкознание: учебник / отв. редактор А.С. Герд. СПб.: Издательство С.-Петербургского университета, 1996. С. 68–90.

² Cabre M.T. Terminology and translation // Handbook of Translation Studies. Vol. 1. / ed. by Y. Gambier, L. van Doorslaer. Amsterdam; Philadelphia: John Benjamins Publishing Company, 2010. P. 356–365. URL: <https://opencourses.ionio.gr/modules/document/file.php/DFLT1112/Handbook%20of%20Translation%20Studies%20Vol.%201.pdf> (дата обращения: 20.01.2025).

³ Калинина С.В. Структурно-семантические и функциональные особенности англоязычно терминологии нефтегазовой сферы: дисс. ... канд. филол. наук; специальность: 5.9.6. СПб., 2024. 271 с.



и стандартов при разработке лингвистических информационных ресурсов. Для цели статьи применимы следующие:

ISO 1087: 2019 (E)⁴, определяющий понятия «язык для специальных целей» (3.1.9), «формальный язык» (3.1.10), «терминологическая статья» (3.6.2), «терминологический словарь» (3.7.4); ISO 704:2022 (E)⁵ определяющий понятия «терминологическая статья» (3.4), «терминолог» (3.6, 3.7), регулирующий отношения между термином и понятием, иерархические отношения «часть – целое»; ISO 24156-1:2014⁶ определяющий понятие «формальный язык» (3.6), например Web Ontology Language (OWL)⁷; ISO 10241-1:2011, характеризующий относящиеся к термину данные: вербальные обозначения (3.4.1.1.1), сокращенные и полные формы (3.4.1.2.3, 3.4.1.2.4), буквенные символы и варианты (3.4.1.1.4, 3.4.1.2.2), омографы и антонимы (3.4.1.4, 3.4.1.5), нормативные статус лексики, примечания к терминологической статье (4.4.3.2.2); ГОСТ 7.25-2001⁸ и ISO 25964-1:2011⁹ (2.23), поясняющие применяемые в информационно-поисковых тезаурусах иерархические отношения.

Для автоматической обработки текстов на естественных языках и разработки приложений информационного поиска используются и онтологические ресурсы, в частности лингвистические онтологии [11]. Метод создания онтологий должен учитывать особенности архитектуры применимой модели-трансформера. В статье используется трансформер DeepSeek R1 – языковая модель для обработки естественного языка с акцентом на контекстное понимание и работу со сложной терминологией [9].

Теоретические предпосылки исследования

При обработке текстовой информации в современных информационно-поисковых и информационно-аналитических системах используются минимальные знания о языке и мире – приоритет отдается учету частотности встречаемости слов, в то время как для выявления основного содержания текста человеком требуется большой объем знаний. «Недостаток лингвистических (о языке) и онтологических (о мире) знаний, используемых в приложениях информационного поиска и автоматической обработки текстов, приводит к разнообразным проблемам» [11, с. 47].

В этой связи одной из значимых задач прикладной лингвистики на современном этапе ее развития является преодоление ограничений больших языковых моделей в обеспечении адекватной передачи специальной лексики. Задача решается «внедрением в современные методы автоматической обработки текстов дополнительных объёмов знаний о языке и мире» [11, с. 48]. Организация управления терминологией включает в себя моделирование профессиональных терминосистем¹⁰; создание в зависимости от выбранной системы автоматизации перевода терминологических баз / автоматических словарей / онтологий [13–15]; интеграцию терминов в системы искусственного интеллекта для корректной вербализации ими соответствующих понятий [16–18].

В настоящее время распространены базы знаний онтологического типа. Онтологии представляют собой «компьютерные ресурсы, содержащие формализованное описание фрагмента знаний о мире» [11, с. 49]. Онтология является способом документирования знаний о предметной области, ее семантической основой – определенной системе понятий соответствуют

⁴ ISO 1087: 2019 (E). Terminology work and terminology science – Vocabulary. ISO, 2019. URL: <https://www.iso.org/standard/62330.html> (дата обращения: 20.06.2025).

⁵ ISO 704:2022 (E). Terminology work – Principles and methods. 4th ed. ISO, 2022. URL: <https://www.iso.org/standard/79077.html> (дата обращения: 04.06.2025).

⁶ ISO 24156-1:2014. Graphic notations for concept modelling in terminology work and its relationship with UML – Part 1: Guidelines for using UML notation in terminology work. 1st ed. ISO, 2014. URL: <https://www.iso.org/standard/57420.html> (дата обращения: 20.06.2025).

⁷ OWL Web Ontology Language Reference. W3C. URL: <https://www.w3.org/TR/owl-ref> (дата обращения: 03.04.2025).

⁸ ГОСТ 7.25.-2001. Тезаурус информационно-поисковый одноязычный: Правила разработки: структура, состав и форма представления // Система стандартов по информации, библиотечному и издательскому делу. Минск: Межгосударственный совет по стандартизации, метрологии и сертификации, 2001.

⁹ ISO 25964-1:2011. Information and documentation – Thesauri and interoperability with other vocabularies. 1st ed. ISO, 2011. URL: <https://www.iso.org/standard/53657.html> (дата обращения: 20.06.2025).

¹⁰ Калинина С.В. Структурно-семантические и функциональные особенности англоязычно терминологии нефтегазовой сферы: дисс. ... канд. филол. наук; специальность: 5.9.6. СПб., 2024. 271 с.



наборы языковых выражений [19]. В связи с тем, что онтологии применяют в автоматической обработке текстов, появляется понятие *лингвистической онтологии* — особого вида лексической базы данных, «понятия которой в значительной мере связаны со значениями языковых единиц, терминов предметной области» [11, с. 49]. Таким образом, компонентом онтологической модели становится лексика естественного языка, называющая понятия и отношения онтологии.

Несмотря на вариативность онтологий, исследователи едины в определении наборов основных компонентов онтологии: классов (понятий), атрибутов (свойств), экземпляров (отдельных индивидов), отношений между классами и экземплярами, которые обеспечивают значения для репрезентативного словаря и формальные ограничения на его согласованное использование [20, 21]. Для создания онтологий используются машиночитаемые языки, например Web Ontology Language (OWL)¹¹. Современными исследованиями подтверждается практическая польза использования онтологий: в обучении англоязычной технической терминологии [22], на базе терминологии брендинга [23], в биомедицинской предметной области [24], с использованием возможностей редактора Concept Maker [26].

Поскольку лингвистические онтологии при построении опираются на значения реально существующих языковых единиц естественного языка, значимым является вопрос выбора лексики. В отсутствие единого исследовательского подхода к отбору терминологической лексики для улучшения нейронного перевода принципы и критерии отбора зависят от профессиональной компетентности лингвиста-терминолога, задач того или переводческого проекта, наличия ограничений в архитектуре выбранного программного обеспечения автоматизации перевода.

В статье используется актуальная англоязычная терминология нефтегазовой сферы, проанализированная в диссертационном исследовании¹², отобранная из языка для специальных целей — современной отраслевой периодики¹³ и узкоспециализированной технической литературы. Язык для специальных целей представляет собой функциональную разновидность языка для общих целей, служащую для профессионального общения: «номинативные единицы языка для специальных целей обозначают специальные понятия»; в этой «прикладной терминологии репрезентируются понятия, в которых концептуализированы результаты научно-производственной деятельности человека по изменению окружающей действительности» [1, с. 76; 26, с. 202]. Лексическая единица языка для специальных целей становится термином, если начинает выполнять терминологические функции и приобретает признаки термина [1, с. 79–80]. Для цели статьи понятия «специальная лексика», «терминологическая лексика», «терминология» являются идентичными.

Принцип отбора указанных языковых единиц базируется на сущностных характеристиках термина: соответствие отражаемому термином понятию и нормам языка, полисемантность, однозначность, краткость, способность к деривации, внедренность в круг профессионального общения, инвариантность, мотивированность, воспроизводимость, интернациональность, благозвучность, эзотеричность. Учитывается временной промежуток появления лексики (за последние 25 лет), новизна содержания как факт вербализации появившегося понятия, отсутствие в современных отраслевых лексикографических источниках¹⁴ / многоязычных терминологических базах¹⁵, полисемантность как проявление антропоцентрической природы лексической единицы.

¹¹ OWL Web Ontology Language Reference. W3C. URL: <https://www.w3.org/TR/owl-ref> (дата обращения: 03.04.2025).

¹² Калинина С.В. Структурно-семантические и функциональные особенности англоязычно терминологии нефтегазовой сферы: дисс. ... канд. филол. наук; специальность: 5.9.6. СПб., 2024. 271 с.

¹³ Oil and Gas Journal. URL: <https://digital.ogj.com/ogjournal/library/> (дата обращения: ежемесячная 2019–2021 гг. по подписке).

¹⁴ Булатов, А.И., Пальчиков, В.В. Англо-русский словарь по нефти и газу. М.: РУССО, 2001. 400 с.; Кедринский В.В. Англо-русский словарь по химии и переработке нефти. М.: РУССО, 2004. 768 с.; Морозов Н.В. Англо-русский и русско-английский словарь по нефти и газу: Компактное издание. М.: Живой язык, 2010. 512 с.

¹⁵ UNTERM. URL: <https://unterm.un.org/unterm2/en/> (дата обращения: 03.04.2025); WIPO Pearl. URL: <https://www.wipo.int/en/web/wipo-pearl> (дата обращения: 02.04.2025).



Проведенный в указанном диссертационном исследовании анализ показывает, что на современном этапе знание нефтегазовой отрасли представлено многословными терминами-словосочетаниями и инициальными аббревиатурами; преобладают семантические способы образования лексических единиц, что увеличивает количество мотивированных терминов – актуальная нефтегазовая терминосистема высоко антропоцентрична и многозначна; номинирующие отраслевые понятия термины встраиваются в тематические группы, соотносимые со следующими понятийными областями технического знания: «Аппараты и Оборудование», «Вещества и Продукты», «Свойства», «Технологические процессы», «Методы и Приемы».

Выявленная в исследовании специфика отраслевой терминологии, равно как и практика переводческой деятельности авторов статьи, позволяет говорить о не всегда корректной передаче лексики при обработке узкоспециализированных текстов при помощи больших языковых моделей, что подчеркивает необходимость использования предметно-ориентированных информационных ресурсов для улучшения нейронного перевода.

Архитектура и обучение модели-трансформера Deepseek R1 оптимизированы для перевода насыщенных терминологией материалов. Модель обладает рядом преимуществ для технического перевода: способна обрабатывать длинные тексты, сохранять связность терминологии на всем протяжении текста; корректно интерпретирует многозначные термины; хорошо справляется с таблицами и формулами; модель можно дообучать на своих глоссариях, в общении с моделью достаточно системных промптов [9].

Обсуждение результатов исследования

В цифровом архиве Ассоциации Компьютерной Лингвистики [27] не обнаружено работ по использованию онтологии при машинной обработке текстов нефтегазовой тематики. Вместе с тем, в силу появления новых языковых единиц, терминология этой предметной области представляет неоспоримый исследовательский интерес. «Неологизация – закономерная реакция отраслевой лексики на развитие нефтегазовой отрасли и необходимость называния новых понятий. Инновационные процессы особенно интенсивно происходят в областях, связанных с усовершенствованием продукции, механизмов и *оборудования*, именно эти области описываются особыми языковыми единицами – терминологической лексикой, которая в отношении обновления является наиболее подвижной и динамичной» [28, с. 47].

Значительное количество новых терминологических единиц нефтегазовой сферы образовано семантическим способом: инструментом построения нового профессионального знания выступает эвристический потенциал вторичной номинации. При этом каждое новое понятие пишется поверх предыдущего, как палимпсест: «Метафорическое моделирование по аналогии возможно при условии, что в воспринимающем сознании уже существует, с той или иной мерой полноты, представление, требующее знакового выражения» [29, с. 176]. В исследовании актуальной отраслевой лексики выявлено, что «превалирующий концепт „*оборудование*“ (а без его наличия процесс добычи/переработки/транспортировки нефти невозможен) выражается в терминах человека и его деятельности, а также в терминах животного мира и мира растений» [30, с. 221]. В этой связи интересен результат машинной обработки насыщенных обновленной терминологией технических текстов.

Редактор онтологии Protégé от других отличается многоплатформенностью – запускается на операционной системе как Windows, так и Linux, написан на языке программирования общего назначения Java и использует язык семантической разметки OWL [31]. Терминологические единицы отраслевой терминосистемы структурированы¹⁶ исходя из их тематической принадлежности к определенной понятийной области: представлена фреймовая схема (рис. 1).

¹⁶ Калинина С.В. Структурно-семантические и функциональные особенности англоязычно терминологии нефтегазовой сферы: дисс. ... канд. филол. наук; специальность: 5.9.6. СПб., 2024. 271 с.

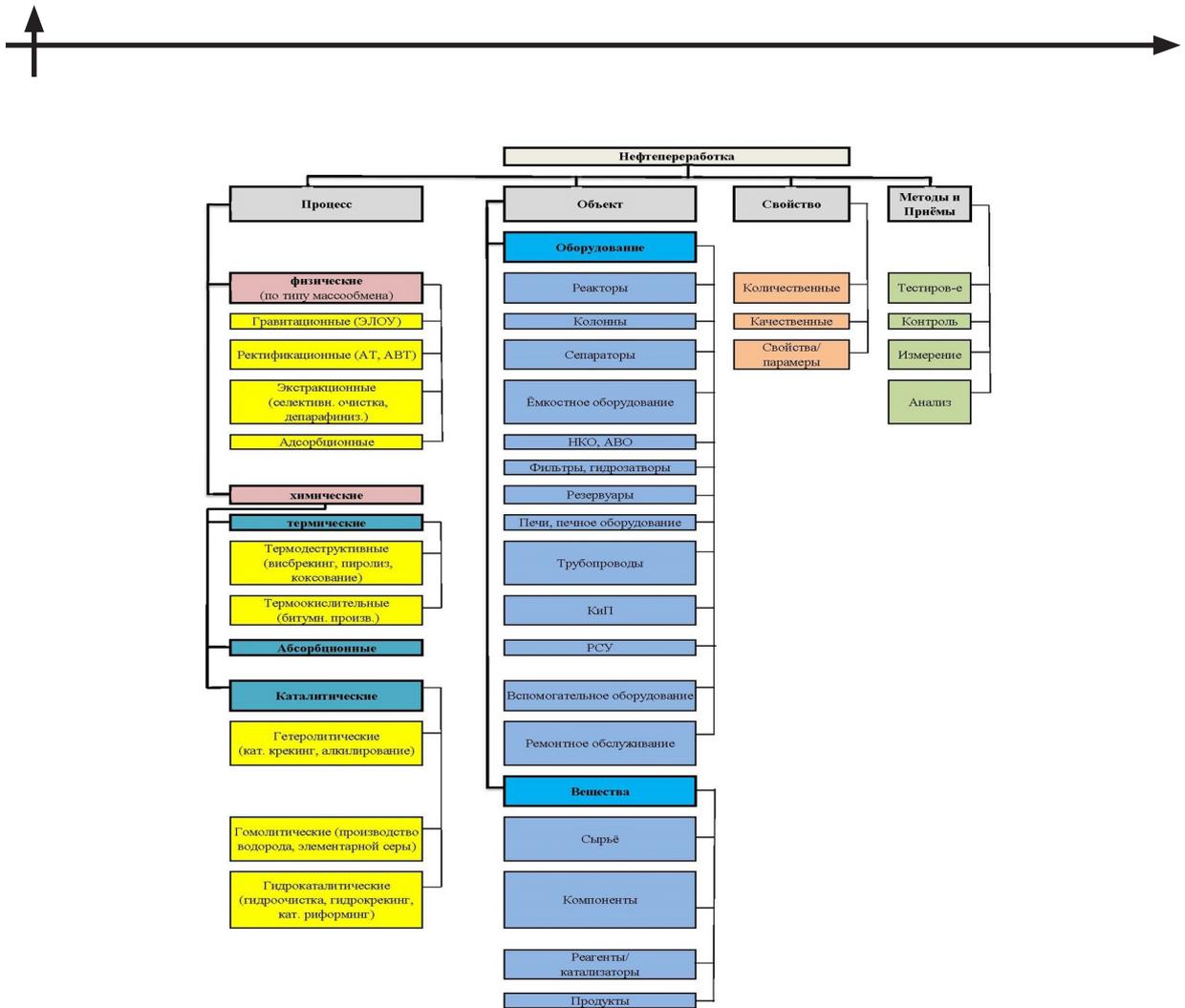


Рис. 1. Фреймовая схема терминосистемы «Нефтегазопереработка»
 Fig. 1. Oil and gas processing term system frame-based chart

В Protégé на основании фреймовой схемы терминосистемы «Нефтегазопереработка» создан фрагмент онтологии нефтегазового *оборудования* “Oil&gas equipment_fragment.rdf”. Следует отметить, что для улучшения визуализации (отображения онтологии в виде графа) в Protégé возможно подключение плагина OntoGraf, однако для целей статьи не существенно.

Обозначающие оборудование термины структурированы ручным вводом: резервуары, насосы, компрессоры, контрольно-измерительные приборы и автоматика (КИПиА), печи, емкости, вспомогательное, ремонтное оборудование (рис. 2). В соответствии с ISO 10241-1:2011¹⁷ каждому вводимому понятию соответствует однозначное имя. Вариативность терминов фиксируется сразу при вводе понятия – посредством указания предметной области снижается полисемия терминов: например, coil – змеевик трубопровода (механика) и катушка индуктивности (электрика).

По ГОСТу 7.25-2001¹⁸ в информационно-поисковых системах для повышения эффективности поиска, разрешения многозначности языковых единиц используются иерархические отношения: родовидовые (класс – подкласс), часть – целое (характерные свойства, роль в процессе).

¹⁷ ISO 10241-1:2011. Terminology entries in standards. Part 1: General requirements and examples of presentation. 1st ed. ISO, 2011. URL: <https://www.iso.org/standard/40362.html> (дата обращения: 20.06.2025).

¹⁸ ГОСТ 7.25-2001. Тезаурус информационно-поисковый одноязычный: Правила разработки: структура, состав и форма представления // Система стандартов по информации, библиотечному и издательскому делу. Минск: Межгосударственный совет по стандартизации, метрологии и сертификации, 2001.

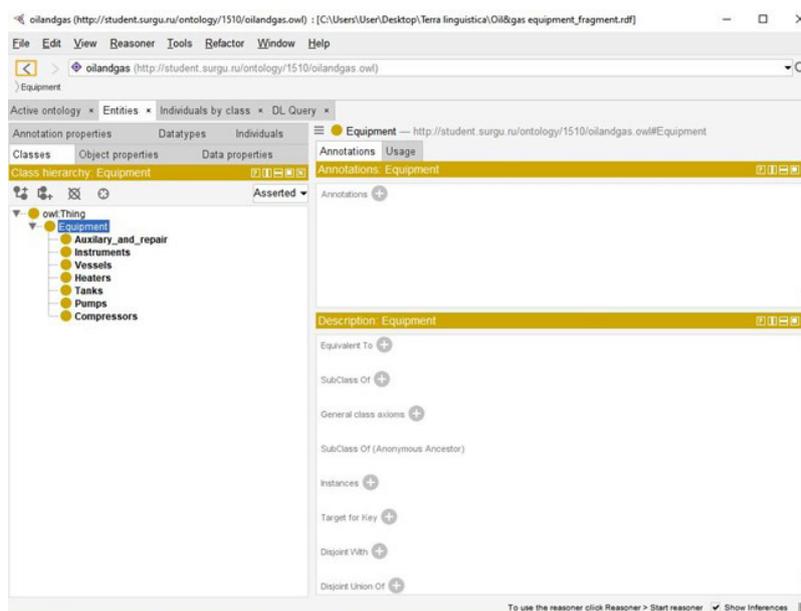


Рис. 2. Фрагмент заполнения классов

Fig. 2. Classes' input fragment

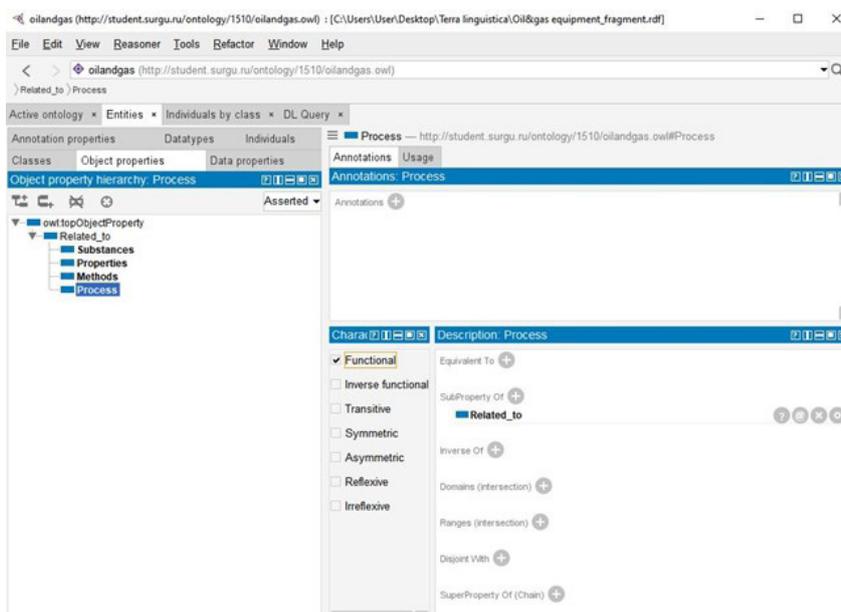


Рис. 3. Фрагмент заполнения связей

Fig. 3. Functional relations' input fragment

Описание иерархических отношений не зависит от контекста их упоминания [11, с. 57–65]. Данное требование обосновано в ISO 25964-1:2011¹⁹: контекст, используемый для подтверждения отношений не всегда доступен в автоматическом режиме. Обозначены функциональные связи (рис. 3), корректность заполнения различных функций отношений лингвистической онтологии проверена Reasoner – логическим выводом в редакторе онтологии Protégé.

¹⁹ ISO 25964-1:2011. Information and documentation – Thesauri and interoperability with other vocabularies. 1st ed. ISO, 2011. URL: <https://www.iso.org/standard/53657.html> (дата обращения: 20.06.2025).

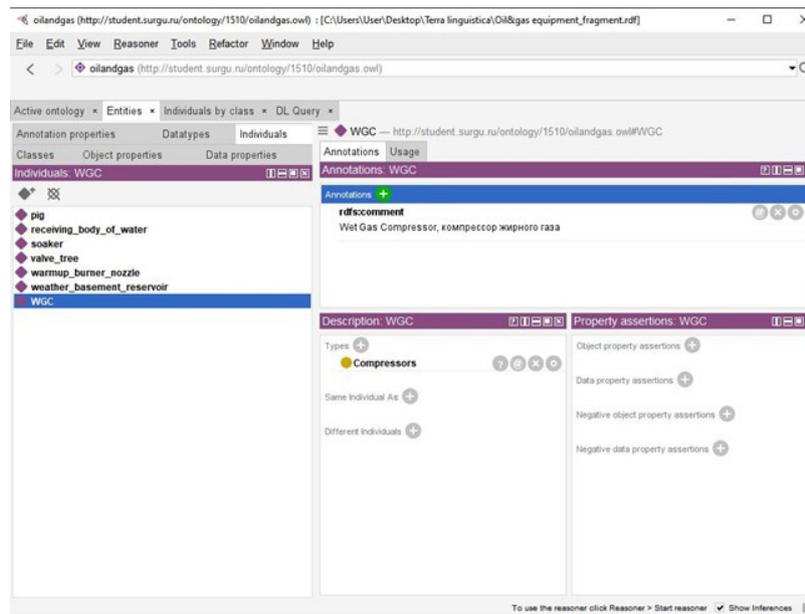


Рис. 4. Фрагмент заполнения индивидов

Fig. 4. Individuals' annotations and description input fragment



Рис. 5. Инструкция трансформеру

Fig. 5. Prompt to Deepseek

Для индивидов внесены метки на русском и английском языке – дефиниции специальной лексики (рис. 4).

В модель-трансформер Deepseek [12] выгружены два файла: созданный фрагмент онтологии и англоязычный текст из отраслевой технической литературы, добавлен системный промпт по использованию загруженного фрагмента онтологии при переводе терминов (рис. 5). Представленная как часть промпта онтология воспринимается моделью как инструкция к переводу. Архитектура модели сопоставляет термины из текста с узлами онтологии, использует связи онтологии для точного перевода лексики в контексте.

Выдача большой языковой модели безусловно нуждается в постредактировании для устранения присущих машинному переводу ошибок [32], однако запрашиваемая терминологическая точность достигнута (табл. 1). На основании системного промпта моделью-трансформером DeepSeek R1 выданы комментарии для экспертной обратной связи (рис. 6).



Таблица 1. Пример работы большой языковой модели (Фрагмент)
Table 1 Large language model's performance results (Fragment)

Оригинал	Перевод трансформера DeepSeek R1 без онтологии	Перевод трансформера DeepSeek R1 с онтологией
<p>“Field development is enabled by more efficient reservoir drainage through new vertical valve trees, and long distance, electrically heated flow lines to avoid hydrates in gas pipelines, the operator said, adding that the technology improves heat efficiency and enables longer tie-backs”²⁰</p>	<p>«Освоение месторождения стало возможным благодаря более эффективному дренированию пласта за счет новых вертикальных клапанных установок (вертикальных фонтанных арматур) и протяженных электроподогреваемых потоковых линий для предотвращения образования гидратов в газопроводах, — заявил оператор, добавив, что данная технология повышает тепловую эффективность и позволяет осуществлять более длинное присоединение удаленных месторождений».</p>	<p>«Разработка месторождения обеспечивается за счет более эффективного дренирования пласта с использованием новых вертикальных ***регулирующих арматур***, а также протяженных электроподогреваемых трубопроводов для предотвращения образования гидратов в газопроводах. По словам оператора, данная технология повышает тепловую эффективность и позволяет осуществлять более длительные подключения».</p>
<p>“Check status of the wet gas compressor (WGC) — on some units the WGC auto trips on blower shutdown; the reactor pressure must be higher than the fractionator and regenerator to prevent flow reversal”²¹.</p>	<p>«Проверьте статус компрессора влажного газа (КВГ) — на некоторых установках КВГ автоматически отключается при остановке нагнетателя; давление в реакторе должно быть выше, чем в фракционирующей колонне и регенераторе, чтобы предотвратить реверс потока».</p>	<p>«Проверьте статус ***компрессора жирного газа*** (WGC) — на некоторых установках WGC автоматически отключается при остановке нагнетателя; давление в реакторе должно превышать давление во фракционирующей колонне и регенераторе для предотвращения обратного потока».</p>
<p>“Rolvnes-Goddo, south of Edvard Grieg field, is targeting the same porous, weathered basement reservoir currently being produced from Edwad Grieg’s Tellus area”²².</p>	<p>«Рольвнес-Годдо, к югу от месторождения Эдвард Григ, нацелен на тот же пористый коллектор выветренного фундамента, который в настоящее время эксплуатируется в районе Теллус месторождения Эдвард Григ».</p>	<p>«Месторождение Рольвнес-Годдо, расположенное южнее месторождения Эдвард Григ, нацелено на те же пористые ***резервуары на фундаменте со стоком для дождевой воды***, которые в настоящее время эксплуатируются в районе Теллус месторождения Эдвард Григ».</p>
<p>“Once ignition temperatures are verified, fuel supply to the warmup burner nozzles is removed, with fuel then supplied to the Duplex nozzles to initiate flame transition”²³.</p>	<p>«После подтверждения достижения температур воспламенения подача топлива к форсункам подогревательной горелки прекращается, и топливо подается к Duplex-форсункам для инициации перехода пламени».</p>	<p>«После подтверждения температур воспламенения подача топлива к ***соплам горелок нагрева*** прекращается, и топливо начинает подаваться к Duplex-соплам для иницирования перехода пламени».</p>
<p>“Visbreaking units are classified in those without soaker (Figure 4.1), and those with soaker (Figure 4.2)”²⁴.</p>	<p>«Установки висбрекинга классифицируются на установки без камеры догрева (Рисунок 4.1) и установки с камерой догрева (Рисунок 4.2)».</p>	<p>«Установки висбрекинга классифицируются на установки без ***реакционной камеры*** (Рисунок 4.1) и с реакционной камерой (Рисунок 4.2)».</p>
<p>“Water quality criteria and standards vary from site to site, depending on the use classification of the receiving body of water”²⁵.</p>	<p>«Критерии и стандарты качества воды варьируются от объекта к объекту в зависимости от целевого назначения водного объекта-реципиента».</p>	<p>«Критерии и стандарты качества воды варьируются в зависимости от объекта, исходя из классификации использования ***водоприемника***».</p>
<p>“Pigging problems related to pig velocity and pipeline geometry”²⁶.</p>	<p>«Проблемы очистки скребками (пигинга), связанные со скоростью движения скребка и геометрией трубопровода».</p>	<p>«Проблемы очистки трубопроводов скребками связаны со скоростью движения ***скребка*** и геометрией трубопровода».</p>

²⁰ Oil and Gas Journal. 2020. Vol. 118, Iss. 2. P. 11. URL: <https://digital.ogj.com/ogjournal/library/> (дата обращения: ежемесячная 2019–2021 гг. по подписке).

²¹ Oil and Gas Journal. 2019. Vol. 117, Iss. 11. P. 44. URL: <https://digital.ogj.com/ogjournal/library/> (дата обращения: ежемесячная 2019–2021 гг. по подписке).

²² Oil and Gas Journal. 2019. Vol. 117, Iss. 7. P. 9. URL: <https://digital.ogj.com/ogjournal/library/> (дата обращения: ежемесячная 2019–2021 гг. по подписке).

²³ Oil and Gas Journal. 2019. Vol. 117, Iss. 4. P. 55. URL: <https://digital.ogj.com/ogjournal/library/> (дата обращения: ежемесячная 2019–2021 гг. по подписке).

²⁴ Raseev, S. Thermal and Catalytic Processes in Petroleum Refining. New York: Marcel Dekker, Inc., 2003. P. 149.

²⁵ Ibid. P. 105.

²⁶ Oil and Gas Journal. 2019. Vol. 117, Iss. 5. P. 56. URL: <https://digital.ogj.com/ogjournal/library/> (дата обращения: ежемесячная 2019–2021 гг. по подписке).



Примечания:

1. Термины из онтологии:

- "valve trees" → "регулирующие арматуры" (класс **Instruments**, комментарий: "регулирующая арматура").
- "wet gas compressor (WGC)" → "компрессор жирного газа" (класс **Compressors**, комментарий: "Wet Gas Compressor").
- "weathered basement reservoir" → "резервуары на фундаменте со стоком для дождевой воды" (класс **Tanks**, комментарий: "резервуар на фундаменте...").
- "warmup burner nozzles" → "сопла горелок нагрева" (класс **Heaters**, комментарий: "сопло горелки нагрева").
- "soaker" → "реакционная камера" (класс **Vessels**, комментарий: "реакционная камера").
- "receiving body of water" → "водоприёмник" (класс **Tanks**, комментарий: "водоприёмник").
- "pig" → "скребок" (класс **Auxiliary_and_repair**, комментарий: "скребок для очистки труб").

2. Контекстная адаптация:

- Для термина "weathered basement reservoir" добавлено множественное число ("резервуары") в соответствии с контекстом геологического образования.
- Сохранены технические детали (например, "Duplex-сопла") и ссылки на источники.

Рис. 6. Примечания модели-трансформера

Fig. 6. DeepSeek R1 comments

Выводы и заключение

Современная технологическая цепочка обработки текстов на естественных языках предполагает обращение к генеративным системам машинного перевода, использующим большие языковые модели и передовые технологии искусственного интеллекта. При этом проблема управления специальной лексикой требует вовлечения лингвиста-терминолога. Владение согласованной терминологией предметной области является основой адекватного перевода научно-технической литературы, способствует результативности профессиональной межъязыковой бизнес-коммуникации. Онтология выступает основой языкового значения представления данных той или иной предметной области, настраивает большую языковую модель на определенную терминологию. В ходе исследования доказана целесообразность применения онтологического подхода к машинной обработке узкоспециализированных текстов на естественных языках. Перспективы исследования видятся в финализации создания нефтегазовой онтологии, ее расширении на смежные отрасли (нефтехимия, энергетика и пр.), настройке алгоритма динамического обновления через обратную связь с экспертом, что обеспечит терминологическую точность перевода технических текстов, наставничество новых переводчиков на проекте, обучение языку для специальных целей.

СПИСОК ИСТОЧНИКОВ

1. **Лейчик В.М.** Терминоведение: Предмет, методы, структура. М.: ЛЕНАНД, 2022. 248 с.
2. **Никитина С.Е.** Семантический анализ языка науки: На материале лингвистики. М.: Либроком, 2014. 146 с.
3. **Никишова О.А., Потапова В.Ю.** Использование систем машинного перевода при переводе текстов научно-технического содержания // Вестник Российского нового университета. Серия: Человек в современном мире. 2023. № 4. С. 133–140. DOI: 10.18137/RNU.V925X.23.04.P.133
4. **Yue R., Ortega J., Church K.** On Translating Technical Terminology: A Translation Workflow for Machine-Translated Acronyms // Proceedings of the 16th Conference of the Association for Machine



Translation in the Americas (Volume 1: Research Track). Chicago, 2024. P. 48–54. URL: <https://aclanthology.org/2024.amta-research.6/> (дата обращения: 03.04.2025).

5. **He Z., Liang T., Jiao W., Zhang Z., Yang Y., Wang R., Tu Z., Shi S., Wang X.** Exploring Human-Like Translation Strategy with Large Language Models. *Transactions of the Association for Computational Linguistics* 2024. Vol. 12. P. 229–246. URL: http://direct.mit.edu/tacl/article-pdf/doi/10.1162/tacl_a_00642/2346100/tacl_a_00642.pdf (дата обращения: 04.04.2025).

6. **Беляева Л.Н.** Машинный перевод в работе переводчика: практический аспект // *Вестник Пермского национального исследовательского политехнического университета. Проблемы языкознания и педагогики*. 2019. № 2. С. 8–20. DOI: 10.15593/2224-9389/2019.2.1

7. **Беляева Л.Н.** Машинный перевод в современной технологии процесса перевода // *Вестник Российского государственного педагогического университета им. А.И. Герцена*. 2022. № 203. С. 22–30. DOI: 10.33910/1992-6464-2022-203-22-30

8. **Табанакова В.Д.** Междисциплинарный синтез лингвистического научного знания // *Вестник Пермского национального исследовательского политехнического университета. Проблемы языкознания и педагогики*. 2023. № 2. С. 8–24. DOI: 10.15593/2224-9389/2023.2.1

9. DeepSeek R1: Инновационный инструмент для будущего // DeepSeek. URL: <https://deepseek-ai.ru/deepseek-r1/> (дата обращения: 04.06.2025).

10. **Карпова О.М.** Изменение современной лексикографической парадигмы в цифровую эпоху (на материале словарей английского языка) // *Вестник Воронежского государственного университета. Серия: Лингвистика и межкультурная коммуникация*. 2022. № 2. С. 6–12. DOI: 10.17308/lic.2022.2/9284

11. **Лукашевич Н.В., Добров Б.В.** Проектирование лингвистических онтологий для информационных систем в широких предметных областях // *Онтология проектирования*. 2015. Т. 5, № 1 (15). С. 47–69.

12. DeepSeek. URL: <https://chat.deepseek.com/> (дата обращения: 04.04.2025).

13. **Окунь Д.Б.** Терминологическая база фармацевтических терминов для медицинских интеллектуальных систем // *Материалы XVII международной научной конференции «Системный анализ в медицине» (САМ 2023)*. Благовещенск, 2023. С. 79–83.

14. **Сидорова Е.А., Иванов А.И., Овчинникова К.А.** Извлечение информации из текстов на основе онтологии и больших языковых моделей // *Онтология проектирования*. 2025. Т. 15, № 1 (55). С. 114–129. DOI: 10.18287/2223-9537-2025-15-1-114-129

15. **Gruber T.R.** A translation approach to portable ontologies // *Knowledge Acquisition*. 1993. Vol. 5, No. 2. P. 199–220. DOI: 10.1006/knac.1993.1008

16. **Kim S., Sung M., Lee J., L., Perez J.G.** Efficient Terminology Integration for LLM-based Translation in Specialized Domains // *Proceedings of the Ninth Conference on Machine Translation*. Miami, 2024. P. 636–642. DOI: 10.18653/v1/2024.wmt-1.51

17. **Sowa J.F.** Building, Sharing and Merging Ontologies . URL: <http://www.jfsowa.com/ontology/ontoshar.htm> (дата обращения: 10.06.2025).

18. **Stefaniak K.** Machine Translation and Terminology: The Experience of the European Commission // *Proceedings of the New Trends in Translation and Technology Conference – NeTTT 2022*. Rhodes Island, Greece, 2022. P. 134–141. URL: https://www.researchgate.net/publication/371045198_Machine_Translation_and_Terminology_The_Experience_of_the_European_Commission (дата обращения: 02.04.2025).

19. **Buitelaar P., Huang C., Calzolari N., Gangemi A., Lenci A., Oltramari A., Prevot L.** Ontology-Based Semantic Lexicons: Mapping Discourse to Concepts // *Ontology and the Lexicon: A Natural Language Processing Perspective*. Cambridge: Cambridge University Press, 2010. P. 212–223. DOI: 10.1017/CBO9780511676536.013

20. **Arp R., Smith B., Spear A.** Building Ontologies with Basic Formal Ontology. MIT Press, 2015. 248 p. DOI: 10.7551/mitpress/9780262527811.001.0001

21. **Guarino N.** Formal Ontology in Information Systems // *Proceedings of FOIS'98*. Trento, 1998. P. 3–15. URL: <https://cidoc-crm.org/sites/default/files/ontologies%20Guarino.pdf> (дата обращения: 02.04.2025).

22. **Андрианова С.В.** Онтология в обучении англоязычной технической терминологии // *Вестник Университета Российской академии образования*. 2013. № 5. С. 100–104.

23. **Зенько М.И.** Терминология брендинга в онтологическом аспекте // *Русский лингвистический бюллетень*. 2023. № 7 (43). DOI: 10.18454/RULB.2023.43.32



24. Schulz S., Stenzhorn H., Boeker M., Smith B. Strengths and limitations of formal ontologies in the biomedical domain // *Electronic Journal of Communication, Information and Innovation in Health (Special Issue on Ontologies, Semantic Web and Health)*. 2009. Vol. 3, Iss. 1. P. 31–45. DOI: 10.3395/Reciis.V3I1.241En
25. Антонов И.В. Автоматизированное построение онтологии предметной области с использованием редактора онтологии CONCEPT MAKER // *Современные инновации в технике и производстве: сборник материалов I Международной научно-практической конференции*, Псков, 2020. С. 135–138.
26. Акаева Х.А., Алимуратов О.А. О некоторых системных критериях разграничения фундаментальных и прикладных терминологий // *Российский гуманитарный журнал*. 2016. Т. 5, № 2. С. 200–211. DOI: 10.15643/libartrus-2016.2.9
27. ACL Anthology. URL: <https://aclanthology.org/> (дата обращения: 02.04.2025).
28. Коцюбинская Л.В., Калинина С.В. К вопросу классификации неонимов в английской терминосистеме нефтепереработки // *Вестник Ивановского государственного университета. Серия: Гуманитарные науки*. 2020. № 4. С. 46–52. DOI: 10.46726/И.2020.4.7
29. Чернявская В.Е. Визуальная метафора в персуазивной коммуникации // *Известия Российского государственного педагогического университета им. А.И. Герцена*. 2021. № 202. С. 175–182. DOI: 10.33910/1992-6464-2021-202-175-182
30. Калинина С.В., Коцюбинская Л.В. Метафорическое моделирование термина нефтяной отрасли (на материале английского языка) // *Филологические науки. Вопросы теории и практики*. 2019. Т. 12, № 10. С. 218–222. DOI: 10.30853/filnauki.2019.10.48
31. Protégé. URL: <https://protege.stanford.edu/> (дата обращения: 02.04.2025).
32. Камшилова О.Н. Машинный перевод в эпоху цифровизации: новые практики, процедуры и ресурсы // *Terra Linguistica*. 2023. Т. 14, № 1. С. 41–56. DOI: 10.18721/JHSS.14105

REFERENCES

- [1] Leychik V.M., Terminovedeniye: Predmet, metody, struktura [Terminology studies: Subject, methods, structure], LENAND, Moscow, 2022.
- [2] Nikitina S.E., Semanticheskiy analiz yazyka nauki: Na materiale lingvistiki [Semantic analysis of the language of science: Based on linguistics], Librokom, Moscow, 2014.
- [3] Nikishova O.A., Potapova V.Yu., Use of Machine Translation Systems in the Translation of Texts of Scientific and Technical Content, *Vestnik of the Russian New University. Series: Man in the Modern World*, 4 (2023) 133–140. DOI: 10.18137/RNU.V925X.23.04.P.133
- [4] Yue R., Ortega J., Church K., On Translating Technical Terminology: A Translation Workflow for Machine-Translated Acronyms, In *Proceedings of the 16th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, Chicago, 2024, pp. 48–54. Available at: <https://aclanthology.org/2024.amta-research.6/> (accessed 03.04.2025).
- [5] He Z., Liang T., Jiao W., Zhang Z., Yang Y., Wang R., Tu Z., Shi S., Wang X., Exploring Human-Like Translation Strategy with Large Language Models, *Transactions of the Association for Computational Linguistics*, 12 (2024) 229–246. Available at: http://direct.mit.edu/tacl/article-pdf/doi/10.1162/tacl_a_00642/2346100/tacl_a_00642.pdf (accessed 04.04.2025).
- [6] Belyaeva L.N., Machine Translations in a Translator Workflow: Practical View, *PNRPU Linguistics and Pedagogy Bulletin*, 2 (2019) 8–20. DOI: 10.15593/2224-9389/2019.2.1
- [7] Belyaeva L.N., Machine Translation and Modern Translation Technology, *Izvestia: Herzen University Journal of Humanities & Sciences*, 203 (2022) 22–30. DOI: 10.33910/1992-6464-2022-203-22-30
- [8] Tabanakova V.D., Interdisciplinarity Synthesis of Linguistic Knowledge, *PNRPU Linguistics and Pedagogy Bulletin*, 2 (2023) 8–24. DOI: 10.15593/2224-9389/2023.2.1
- [9] DeepSeek R1: Innovatsionnyy instrument dlya budushchego [DeepSeek R1: An innovative tool for the future], DeepSeek. Available at: <https://deepseek-ai.ru/deepseek-r1/> (accessed 04.06.2025).
- [10] Karpova O.M., Changing of Modern Lexicographic Paradigm in Digital Era (with Special Reference to English Dictionaries), *Proceedings of Voronezh State University. Series: Linguistics and intercultural communication*, 2 (2022) 6–12. DOI: 10.17308/lic.2022.2/9284



[11] **Lukashevich N.V., Dobrov B.V.**, Developing Linguistic Ontologies in Broad Domains, *Ontology of Designing*, 5 (1 (15)) (2015) 47–69.

[12] DeepSeek. Available at: <https://chat.deepseek.com/> (accessed 04.04.2025).

[13] **Okun' D.B.**, Terminological Database of Pharmaceutical Terms for Medical Intelligent Systems, *Materialy XVII mezhdunarodnoy nauchnoy konferentsii "Sistemnyy analiz v meditsine" (SAM 2023)* [Materials of the XVII International Scientific Conference "System Analysis in Medicine" (SAM 2023)], Blagoveshchensk, 2023, pp. 79–83.

[14] **Sidorova Ye.A., Ivanov A.I., Ovchinnikova K.A.**, Information extraction from texts based on ontology and large language models, *Ontology of Designing*, 15 (1 (55)) (2025) 114–129. DOI 10.18287/2223-9537-2025-15-1-114-129

[15] **Gruber T.R.**, A translation approach to portable ontologies, *Knowledge Acquisition*, 5 (2) (1993) 199–220. DOI: 10.1006/knac.1993.1008

[16] **Kim S., Sung M., Lee J., L., Perez J.G.**, Efficient Terminology Integration for LLM-based Translation in Specialized Domains, *Proceedings of the Ninth Conference on Machine Translation*, Miami, 2024, pp. 636–642. DOI: 10.18653/v1/2024.wmt-1.51

[17] **Sowa J.F.**, Building, Sharing and Merging Ontologies. Available at: <http://www.jfsowa.com/ontology/ontoshar.htm> (accessed 10.06.2025).

[18] **Stefaniak K.**, Machine Translation and Terminology: The Experience of the European Commission, *Proceedings of the New Trends in Translation and Technology Conference – NeTTT 2022*. Rhodes Island, Greece, 2022, pp. 134–141. Available at: https://www.researchgate.net/publication/371045198_Machine_Translation_and_Terminology_The_Experience_of_the_European_Commission (accessed 02.04.2025).

[19] **Buitelaar P., Huang C., Calzolari N., Gangemi A., Lenci A., Oltramari A., Prevot L.**, *Ontology-Based Semantic Lexicons: Mapping Discourse to Concepts, Ontology and the Lexicon: A Natural Language Processing Perspective*, Cambridge University Press, Cambridge, 2010, pp. 212–223. DOI: 10.1017/CBO9780511676536.013

[20] **Arp R., Smith B., Spear A.**, *Building Ontologies with Basic Formal Ontology*, MIT Press, 2015. DOI: 10.7551/mitpress/9780262527811.001.0001

[21] **Guarino N.**, Formal Ontology in Information Systems, *Proceedings of FOIS'98*, Trento, 1998, pp. 3–15. Available at: <https://cidoc-crm.org/sites/default/files/ontologies%20Guarino.pdf> (accessed 02.04.2025).

[22] **Andrianova S.V.**, Ontologiya v obuchenii angloyazychnoy tekhnicheskoy terminologii [Ontology in teaching English-language technical terminology], *Herald of the University of the Russian Academy of Education*, 5 (2013) 100–104.

[23] **Zen'ko M.I.**, Terminology of Branding from an Ontological Perspective, *Russian Linguistic Bulletin*, 7 (43) (2023). DOI: 10.18454/RULB.2023.43.32

[24] **Schulz S., Stenzhorn H., Boeker M., Smith B.**, Strengths and limitations of formal ontologies in the biomedical domain, *Electronic Journal of Communication, Information and Innovation in Health (Special Issue on Ontologies, Semantic Web and Health)*, 3 (1) (2009) 31–45. DOI: 10.3395/Reciis.V3I1.241En

[25] **Antonov I.V.**, Avtomatizirovannoye postroyeniye ontologii predmetnoy oblasti s ispolzovaniyem redaktora ontologii CONCEPT MAKER [Automated construction of subject area ontology using the ontology editor CONCEPT MAKER], *Sovremennyye innovatsii v tekhnike i proizvodstve: sbornik materialov I Mezhdunarodnoy nauchno-prakticheskoy konferentsii* [Modern innovations in technology and production: collection of materials of the 1st International Scientific and Practical Conference], Pskov, 2020, pp. 135–138.

[26] **Akayeva Kh.A., Alimuradov O.A.**, Some systemic criteria of the differentiation between fundamental and applied terminologies, *Liberal Arts in Russia*, 5 (2) (2016) 200–211. DOI: 10.15643/libartus-2016.2.9

[27] ACL Anthology. Available at: <https://aclanthology.org/> (accessed 02.04.2025).

[28] **Kotsyubinskaya L.V., Kalinina S.V.**, To the question of neonyms' classification in the english terminology of oil refining, *Ivanovo State University Bulletin. Series: Humanities*. 4 (2020) 46–52. DOI: 10.46726/H.2020.4.7

[29] **Chernyavskaya V.Ye.**, Visual Metaphor in Persuasive Communication, *Izvestia: Herzen University Journal of Humanities & Sciences*. 202 (2021) 175–182. DOI: 10.33910/1992-6464-2021-202-175-182



[30] **Kalinina S.V., Kotsyubinskaya L.V.**, Metaphorical Modelling of Oil Industry Terms (By the Material of the English Language), *Philology. Theory & Practice*, 12 (10) (2019) 218–222. DOI: 10.30853/filnauki.2019.10.48

[31] Protégé. Available at: <https://protege.stanford.edu/> (accessed 02.04.2025).

[32] **Kamshilova O.N.**, Machine translation in the age of digitalization: new practices, procedures and resources, *Terra Linguistica*, 14 (1) (2023) 41–56. DOI: <https://doi.org/10.18721/JHSS.14105>

СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS

Калинина Светлана Валентиновна

Svetlana V. Kalinina

E-mail: movable@yandex.ru

Коцюбинская Любовь Вячеславовна

Lyubov' V. Kotsyubinskaya

E-mail: l.kocubinskaya@lengu.ru

Поступила: 18.04.2025; Одобрена: 18.06.2025; Принята: 30.06.2025.

Submitted: 18.04.2025; Approved: 18.06.2025; Accepted: 30.06.2025.