



МГУ имени М.В. Ломоносова

Научно-исследовательский вычислительный центр

Лаборатория анализа информационных ресурсов (404)

Лаборатория информационных систем математических наук (418)

Отчет за 2025 год

Тема:

Методы структурирования трудноформализуемых предметных областей на основе автоматизированного формирования больших графов знаний и онтологий по разнородным потокам текстовых данных

Этап 6. (2025 г.) Исследование методов применения больших нейросетевых языковых моделей для задач автоматизированного формирования графов знаний

Кадровый состав (по теме)

Добров Борис Викторович	завлаб, к.фил.н.	418
Ионова Нина Ростиславовна		404
Ковалев Григорий Павлович (совм.)		404
Кремнева Марина Дмитриевна		404
Лукашевич Наталья Валентиновна	внс, дтн	404
Мячина Анастасия Владимировна		404
Рожков Игорь Сергеевич (совм.)		404
Сидоров Алексей Валерьевич (совм.)		404
Студеникина Ксения Андреевна (до 08.2025)		415
Тихомиров Михаил Михайлович	снс, к.фил.н.	404
Чернышев Даниил Иванович (совм.)		404
Штернов Сергей Владимирович (совм.)		404
Ярошенко Полина Владимировна	к.фил.н.	404
Григорьев Денис Андреевич		404
Гращенко Павел Валерьевич (совм.) (до 08.2025)	завлаб, д.фил.н.	415
Паско Лада Игоревна (до 08.2025)		415
Герасимова Анастасия Алексеевна (до 01.2025)	к.фил.н.	415
Лютикова Екатерина Анатольевна (совм.) (до 08.2025)	внс, д.фил.н.	404

Планы на 2025 год

Исследование методов применения больших нейросетевых языковых моделей для задач автоматизированного формирования графов знаний

- исследовать методы улучшения качества результатов информационного поиска для решения задач RAG (подключения новых знаний) для больших языковых моделей;
- исследовать методы автоматического структурирования материалов при формировании аналитических статей;
- исследовать методы организации научно-технических знаний.

Практическая значимость ожидаемых результатов заключается в создании новых инструментов для формирования графов знаний в форме связанных текстовых статей.

Формы завершения: Отчет, WoS+Scopus=6, другие=2.

НИРы

➤ **Грант РНФ**

25-71-00206

25 декабря 2024 - 16 декабря 2026 Семантический поиск на основе нейросетевых моделей информационного-поиска для русского языка: создание бенчмарка и разработка новых моделей

Руководитель: Добров Б.В. Участники : Лукашевич Н.В., Тихомиров М.М. Ковалев Г.П.

➤ **Грант РНФ**

25-11-00191

27 мая 2025 - 31 декабря 2027 Методы адаптации больших языковых моделей на русский язык и конкретные предметные области

Руководитель: Тихомиров М.М. Ответственные исполнители: Добров Б.В., Ярошенко П.В., Чернышев Д.И.

➤ **+ 1 НИР**

Публикации 2025

	Тема	
	ЛАИР (404)	ЛАЛС (415)
Q1	1	
WoS, Scopus	14 (+3)	
Ядро РИНЦ	6 (+3)	1
Всего	24 (+3)	

Исследование методов улучшения качества результатов информационного поиска для решения задач RAG (подключения новых знаний) для больших языковых моделей

- Создан RusBEIR, комплексный бенчмарк, разработанный для оценки моделей информационного поиска на русском языке без предварительного обучения. Он включает 17 наборов данных из различных областей, интегрируя адаптированные, переведенные и вновь созданные наборы данных, что позволяет систематически сравнивать лексические и нейронные модели.
- Описаны методы адаптации мультязычных LLM для обработки русского языка. Представлена комплексная и вычислительно эффективная методология Ruadapt для языковой адаптации LLM с заменой токенизации. Приведено подробное эмпирическое исследование каждого шага адаптации с целью определения оптимальных гиперпараметров.

Комплексный стенд для оценки моделей информационного поиска на русском языке

Бенчмарк RusBEIR



rusBEIR
IR benchmark

Модели обучаются и тестируются на **англоязычных данных**



Как объективно оценить **качество модели** при **переносе на другие языки**?

Русскоязычный аналог бенчмарка BEIR

- **27 датасетов:**
 - **Переведенные** датасеты из BEIR
 - Русскоязычные части **мультиязычных** датасетов
 - Русскоязычные датасеты для информационного поиска
 - **Собственные новые** наборы данных
- Zero-shot оценка для русского языка

Kovalev G. et al. Building Russian Benchmark for Evaluation of Information Retrieval Models //arXiv preprint arXiv:2504.12879. – 2025.

Датасеты бенчмарка rusBEIR

Source (↓)	Task (↓)	Dataset (↓)	Origin (↓)	Relevancy	Train	Dev	Test	Corpus	Avg. Word Lengths (D/Q)
BEIR	Bio-Medical IR	rus-NFCorpus	Translation	Binary	2,590	324	323	3,633	216.6 / 3.5
BEIR	Argument Retrieval	rus-ArguAna	Translation	Binary	—	—	1,406	8,674	147.8 / 173.8
BEIR	Fact Checking	rus-SciFact	Translation	Binary	809	—	300	5,183	185.8 / 11.2
BEIR	Citation-Prediction	rus-SCIDOCs	Translation	Binary	—	—	1000	25,657	153.1 / 9.8
BEIR	Bio-Medical IR	rus-TREC-COVID	Translation	Binary	171,332	—	—	50	138.9 / 8.5
BEIR	Question Answering (QA)	rus-FiQa	Translation	Binary	57,638	500	5,500	648	122.0 / 9.9
BEIR	Duplicate Question Retrieval	rus-Quora	Translation	Binary	522,931	5,000	—	10,000	9.8 / 7.9
BEIR	Duplicate Question Retrieval	rus-CQADupstack	Translation	Binary	457,199	—	—	13,145	117.6 / 7.5
BEIR	Argument Retrieval (QA)	rus-Touche	Translation	Binary	382,545	—	—	49	252.5 / 6.8
BEIR	Information-Retrieval	rus-MMARCO	Part of multilingual	Binary	502,939	6980	—	8,841,823	49.6 / 5.95
Open-Source Dataset	Information-Retrieval	rus-MIRACL	Part of multilingual	Binary	4,683	1,252	—	9,543,918	43 / 6.2
Open-Source Dataset	Question Answering (QA)	rus-XQuAD	Part of multilingual	Binary	—	1,190	—	240	112.9 / 8.6
Open-Source Dataset	Question Answering (QA)	rus-XQuAD-sentences	Part of multilingual	Binary	—	1,190	—	1212	22.4 / 8.6
Open-Source Dataset	Question Answering (QA)	rus-TyDi QA	Part of multilingual	Binary	—	1,162	—	89,154	69.4 / 6.5
									44 100.4 / 8.7
									32 89.9 / 9.2
									6 28.1 / 23.9
									26 62.07 / 6.4
									44 155.2 / 8.8

Новые наборы данных в rusBEIR

- **8 новых** наборов данных
- Fact checking, Information Retrieval, RAG
- Одни и те же данные в **разных форматах**

Source (↓)	Task (↓)	Dataset (↓)	Origin (↓)	Relevancy	Train	Dev	Test	Corpus	Avg. Word Lengths (D/Q)
rusBEIR	Information-Retrieval	wikifacts-articles	Originally Russian	3-level	—	5433	—	12,848	2,059.1 / 11.4
rusBEIR	Fact Checking	wikifacts-para	Originally Russian	3-level	—	5433	—	176,364	150.0 / 11.4
rusBEIR	Information-Retrieval	wikifacts-sents	Originally Russian	3-level	—	5433	—	1,500,778	17.6 / 11.4
rusBEIR	Fact Checking	wikifacts-window_2	Originally Russian	3-level	—	5433	—	1,500,776	35.2 / 11.4
rusBEIR	Fact Checking	wikifacts-window_3	Originally Russian	3-level	—	5433	—	1,500,774	52.8 / 11.4
rusBEIR	Fact Checking	wikifacts-window_4	Originally Russian	3-level	—	5433	—	1,500,772	70.5 / 11.4
rusBEIR	Fact Checking	wikifacts-window_5	Originally Russian	3-level	—	5433	—	1,500,770	88.1 / 11.4
rusBEIR	Fact Checking	wikifacts-window_6	Originally Russian	3-level	—	5433	—	1,500,768	105.8 / 11.4

Проект Ruadapt

Цель - перенос мультязычных LLM на русский язык

Основные шаги:

- Подготовка более подходящей для русского языка токенизации
- Адаптация базовой версии модели на новую токенизацию
- Отображение из базовой модели в инструктивную (LER)
- Дополнительное дообучение на инструкциях

На текущий момент этап адаптации базы происходит на датасете в 60-80GB:

- Википедия
- Новости
- Хабр
- Книги (либрусек)

RuadaptQwen2.5: результаты сравнения

1	Rank* (UB)	Model	Arena Elo	Votes	Organization	License
2	1	gpt-4o-2024-11-20	1114	749	OpenAI	Proprietary
3	1	Google: Gemini Pro 1.5	1111	1253	Google	Proprietary
4	1	DeepSeek R1	1109	509	DeepSeek	Open Source
5	1	claude-3-5-sonnet-20241022	1105	1006	Anthropic	Proprietary
6	1	claude-3-7-sonnet-20250219	1095	390	Anthropic	Proprietary
7	1	Google: Gemini Flash 2.0	1092	525	Google	Proprietary
8	1	claude-3-5-sonnet-20240620	1088	1563	Anthropic	Proprietary
9	1	DeepSeek V3	1087	710	DeepSeek	Open Source
10	3	gpt-4o-2024-05-13	1078	2032	OpenAI	Proprietary
11	6	gpt-4-turbo-2024-04-09	1059	2007	OpenAI	Proprietary
12	6	claude-3-5-haiku-20241022	1051	904	Anthropic	Proprietary
13	10	Google: Gemini Flash 1.5	1047	1280	Google	Proprietary
14	10	Llama 3.1 405B Instruct Turbo	1047	1719	Meta*	Proprietary
15	5	GigaChat 2 Max	1046	302	Sber	Proprietary
16	10	Llama 3.3 70B Instruct	1043	834	Meta*	Open Source
17	10	Qwen2.5 72B Instruct	1040	1156	Qwen	Open Source
18	10	GigaChat-Max-preview 4.0.26.20	1037	1309	Sber	Proprietary
19	10	gpt-4-0613	1036	1989	OpenAI	Proprietary
20	10	MiniMax: MiniMax-Text-01	1036	576	MiniMaxAI	Open Source
21	10	gpt-4o-mini-2024-07-18	1035	1795	OpenAI	Proprietary
22	11	Gemma 2 27B	1025	1673	Google	Proprietary
23	11	RefalMachine/RuadaptQwen2.5-32B-Pro-Beta	1019	539	RefalMachine	Open Source
24	14	Llama 3.1 70B Instruct Turbo	1014	1724	Meta*	Proprietary
25	11	YandexGPT 5 Pro	1013	575	Yandex	Proprietary
26	11	t-tech/T-pro-it-1.0	1011	387	t-bank-ai	Open Source
27	14	saiga llama3 70b	1008	676	Ilya Gusev	Open Source
28	14	Mistral Small 3	1008	552	Mistral	Open Source

Исследование методов автоматического структурирования материалов при формировании аналитических статей

- Разработаны методы автоматического формирования энциклопедических статей методами абстрактного реферирования с использованием больших нейросетевых языковых моделей на материалах ресурсов Wikipedia и РуВики
- Проведены исследования для лучшего понимания возможностей LLM в различных задачах базовых методов обработки текстов
- Исследованы методы определения тональности и выражения эмоций, что является важным для извлечения экспертных оценок о перспективности, достоинствах и недостатках рассматриваемых процессов для формирования информационно-аналитических материалов.

Генерация статей типа Wikipedia по тематике физики

Prompt Type	Contents
Basic +RAG +forced lang.	Твоя задача - ответить на запрос пользователя. [Пиши на русском языке.] Релевантная информация: <RAG-context>] Расскажи о следующем термине: <target term>
Detailed +RAG	Ты - большой эксперт в области физики. Ты должен предоставить основную информацию о запрашиваемом термине. Структурируй ответ по разделам: Определение, Отличительные особенности, Дополнительно. Следуй указанному порядку разделов. Раздели ответ на разделы. Если считаешь, что на какой-то раздел не стоит делить его. Главный критерий качества твоего ответа - полнота информации. Пропускай то, в чем не уверен, не лги и не используй русский язык. [При ответе помимо собственных знаний используй релевантную информацию. Если считаешь, что информация не подходит, не используй ее. Релевантная информация: <RAG-context>] Расскажи о следующем термине: <target term>
Encyclopedic +target term	Построй текст энциклопедической статьи о следующем термине: <target term>]. Статья должна быть структурирована по разделам. Предложения: <RAG-context> Ответ:

Калибровочные бозоны – это частицы, переносящие фундаментальные взаимодействия в Стандартной модели физики элементарных частиц. Они являются квантами соответствующих калибровочных полей, описывающих эти взаимодействия.

К калибровочным бозонам относятся:

1. Фотон – переносчик электромагнитного взаимодействия.
2. W^+ , W^- , Z^0 бозоны – переносчики слабого ядерного взаимодействия.
3. 8 глюонов – переносчики сильного ядерного (цветового) взаимодействия между кварками.
4. Предполагаемый гравитон – гипотетическая частица-переносчик гравитационного взаимодействия (пока не обнаружен экспериментально).

Основные свойства калибровочных бозонов:

- Являются безмассовыми (кроме W^+ , W^- , Z^0) векторными частицами с целочисленным спином.
- Описываются калибровочными теориями, основанными на принципе калибровочной инвариантности.
- Взаимодействуют только с частицами, несущими соответствующий заряд (электрический, цветовой и т.д.)
- Играют фундаментальную роль, будучи переносчиками всех известных фундаментальных взаимодействий (кроме гравитации).

Существование калибровочных бозонов является одним из столпов Стандартной модели и подтверждено многочисленными экспериментами в физике высоких энергий.

Claude-3 + BasicRAG

Тестирование качества сформированных статей с привлечением ученых-физиков

Вопрос 2 - калибровочные бозоны *

Калибровочные бозоны - это частицы, которые переносят элементарные силы в природе. Они являются промежуточными частицами, которые передают силы между другими частицами. Существует три типа калибровочных бозонов: фотоны, векторные бозоны и гравитоны. Фотоны переносят электромагнитные силы, векторные бозоны - слабые силы, а гравитоны - гравитационные силы. Калибровочные бозоны были открыты в ходе экспериментов на ускорителях частиц. Они играют важную роль в теории Великого объединения, которая стремится объединить все известные силы в единую теорию.	Определение: Калибровочные бозоны - это частицы, которые являются переносчиками фундаментальных взаимодействий в квантовой теории поля. Они взаимодействуют с другими частицами, перенося силу взаимодействия между ними. Структура: Калибровочные бозоны могут быть как продольными, так и поперечными компонентами. Продольные компоненты калибровочных бозонов могут возникать как голдстоуновские бозоны. Использование: Калибровочные бозоны используются в квантовой теории поля для описания фундаментальных взаимодействий между частицами. Они могут взаимодействовать с другими частицами, такими как кварки и лептоны, перенося силу взаимодействия между ними. Отличительные особенности: Калибровочные бозоны отличаются от других бозонов тем, что они являются переносчиками фундаментальных взаимодействий. Они могут взаимодействовать с другими частицами, перенося силу взаимодействия между ними. Дополнительные сведения: Калибровочные бозоны могут нетривиально взаимодействовать с хиггсовским бозоном, а также с кварками и лептонами (посредством юкавского взаимодействия). Это позволяет описать сложные взаимодействия между частицами в квантовой теории поля.
---	--

☐ Левая

☒ Правая

☐ Ничья

☐ Обе плохие

Пример параллельного сравнения для термина "калибровочные бозоны"

В последние два десятилетия интерес к квантовым биениям резко возрос в связи с наблюдением биений в квантово-размерных полупроводниковых наноструктурах, рассматриваемых сейчас как наиболее перспективные материалы для электронной и вычислительной техники. Это позволяет более отчетливо регистрировать квантовые биения. Простейшим примером, демонстрирующим квантовые биения в радикальной паре, может служить пара радикалов, не взаимодействующих друг с другом и отличающихся лишь значениями изотропных g-факторов. Они показаны для демонстрации квантовых биений волновой функции. В литературе, однако, описаны лишь единичные примеры, когда при лазерном облучении удавалось наблюдать квантовые биения в рекомбинации пар. Эти динамические переходы (квантовые биения) аналогичны квантовым биениям в других объектах, что позволяет рассматривать спиновую химию в одном ряду с другой новой научной областью - фемтохимией и с некоторыми спектроскопическими методами исследования, в которых также проявляется квантовая интерференция. И в спиновой химии, и в фемтохимии квантовые биения в заселенностях состояний реагентов модулируют выход продуктов реакции.

Фактологическая точность *



Смысловая согласованность *



Соответствие формату энциклопедической статьи *



Пример обособленного оценивания для термина "квантовые биения"

Исследование методов организации научно-технических знаний

- Участие в деятельности Национальной платформы «Ковчег знаний МГУ
- Исследование потенциала LLM для задачи автоматического пополнения таксономий на материале русского языка
- Участие в кооперации с другими научными коллективами по организации научных соревнований по решению задач обработки и анализа научно-технических текстов, в частности участие в деятельности международного коллектива по формированию наборов данных BioASQ для поддержки решения задач семантического индексирования биомедицинских данных и автоматического формирования ответов на вопросы
- Предложен многоэтапный подход к автоматическому порождению ключевых слов для русскоязычных научных статей.
- Для решения проблемы сложности с формированием представительных коллекций научно-технических текстов, рассматривались методы распознавания речи, как источника информации

Планы на 2026 год

Разработка макета корпоративной базы знаний в виде Вики-ресурса

Исследование методов:

- дообучения адаптированных больших языковых моделей общей направленности;
- обучения больших языковых моделей специализированных для некоторых задач;
- оценки актуализации знаний при применении больших нейросетевых моделей.

Практическая значимость ожидаемых результатов заключается в создании новых инструментов для формирования графов знаний в форме связанных текстовых статей.

Формы завершения: Отчет, WoS+Scopus=6, другие=2.

Публикации по теме

1. 2025 Gousyatskaya Polina, Loukachevitch Natalia, Word Sense Disambiguation in Russian: A Generative LLM Approach //Communications in Computer and Information Science, место издания Springer Nature Switzerland Cham, том 2671, с. 145-156. – DOI 10.1007/978-3-032-04958-2_11 (**Scopus**)
2. 2025 Kirillovich Alik, Karpov Ilya, Ilvovsky Dmitry, Kulaev Maxim, Loukachevitch Natalia, RuSemCor: a Word Sense Disambiguation corpus for Russian //ACM International Conference on Information and Knowledge Management (**Scopus**)
3. 2025 Kovalev G., Tikhomirov M., Kozhevnikov E., Kornilov M., Loukachevitch N., Building Russian Benchmark for Evaluation of Information Retrieval Models //In Proceedings of the International Conference "Dialogue-2025" (**Scopus**)
4. 2025 Lee G., Loukachevitch N., Khokhlov A., Generating Encyclopedic Articles Based on a Collection of Scientific Publications //Proceedings of the International Conference "Dialogue-2025" (**Scopus**)
5. 2025 Loukachevitch N., Tkachenko N., Lapanitsyna A., Tikhomirov M., Rusnachenko N., RuOpinionNE-2024: Extraction of Opinion Tuples from Russian News Texts //Proceedings of the International Conference "Dialogue-2025" (**Scopus**)
6. 2025 Nentidis A., Katsimpras G., Krithara A., Krallinger M., Ortega M.R., Loukachevitch N., BioASQ at CLEF2025: The Thirteenth Edition of the Large-Scale Biomedical Semantic Indexing and Question Answering Challenge //Proceedings of European Conference on Information Retrieval 2025, Springer (New York), с. 407-415 (**Scopus**)
7. 2025 Nentidis Anastasios, Katsimpras Georgios, Krithara Anastasia, Krallinger Martin, Rodriguez-Ortega Miguel, Rodriguez-Lopez Eduard, Loukachevitch Natalia, Sakhovskiy Andrey, Tutubalina Elena, Dimitriadis Dimitris, Tsoumakas Grigorios, Giannakoulas George, Bekiaridou Alexandra, Samaras Athanasios, Nunzio Giorgio Maria Di, Ferro Nicola, Marchesin Stefano, Martinelli Marco, Silvello Gianmaria, Paliouras Georgios, Overview of BioASQ 2025: The Thirteenth BioASQ Challenge on Large-Scale Biomedical Semantic Indexing and Question Answering //Lecture Notes in Computer Science, Springer Nature (Switzerland), с. 173-198. – DOI: 10.1007/978-3-032-04354-2_12 (**Scopus**)
8. 2025 Polevoi Anton, Kragin Alexander, Loukachevitch Natalia, Ground Truth-Free WER Prediction for ASR via Audio Quality and Model Confidence Features //International Conference on Speech and Computer, Springer Nature (Switzerland), с. 10-1007. – DOI: 10.1007/978-3-032-07959-6_3 (**Scopus**)
9. 2025 Rozhkov I., Loukachevitch N., Methods for Recognizing Nested Terms //Proceedings of the International Conference "Dialogue-2025" (**Scopus**)
10. 2025 Sadkovskii F., Loukachevitch N., Grishin I. Fine-tuning large language models for hypernym discovery task: Sister terms do their part // Записки научных семинаров Санкт-Петербургского отделения математического института им. В.А. Стеклова РАН. — 2025. — Vol. 546. — P. 146–
11. 2025 Vatolin A., Gerasimenko N., Loukachevitch N., Ianina A., Vorontsov K., ruSciFact: Open Benchmark for Verifying Scientific Facts in Russian //In Proceedings of the International Conference "Dialogue-2025" (**Scopus**)
12. 2025 Volkov E.V., Dobrov B.V., Automated Transcription of Unfamiliar Words to Improve Recognition of Terms in the Subject Domain //Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications, Pleiades Publishing, Ltd (Road Town, United Kingdom), том 35, № 3, с. 245-254. – DOI: 10.1134/S1054661825700154 (**Scopus**, **Ядро РИНЦ**, **RSCI**)

Публикации по теме (2)

13. 2025 Григорьев Д.А., Чернышев Д.И., RUWIKIBENCH: оценка больших языковых моделей посредством воспроизведения энциклопедических статей //Доклады Российской академии наук. Математика, информатика, процессы управления, Российская академия наук (Москва), том 527, № S, с. 171-181. – DOI: 10.7868/S2686954325070148 (**Ядро РИНЦ, переводная версия Web of Science, Scopus**)
14. 2025 Зубарева Е.В., Лукашевич Н.В., Пичугов А.А., Семенов А.Л., Национальная платформа «Ковчег знаний МГУ»: онтологический компас цифрового суверенитета и трансформации образования //Информатизация образования и методика электронного обучения : цифровые технологии в образовании: материалы IX Междунар. науч. конф. Красноярск, 23–26 сентября 2025 г.: в 4 ч. Ч. 1, место издания Красноярский государственный педагогический университет им. В.П. Астафьева г.Красноярск, том 1, с. 215-219 (РИНЦ)
15. 2025 Лемаев В.И., Лукашевич Н.В., Выражение эмоций в разных языках: возможности межъязыкового переноса моделей распознавания //Вестник Московского университета. Серия 9: Филология, Изд-во Моск. ун-та (М.), № 4, с. 9-20 (Ядро РИНЦ, RSCI)
16. 2025 Лемаев В.И., Лукашевич Н.В., Межъязыковой перенос моделей автоматического распознавания эмоций в речи на русский язык: данные и тенденции //Научно-техническая информация. Серия 2: Информационные процессы и системы, ВИНТИ (М.), № 5, с. 28-38 (Ядро РИНЦ, RSCI)
17. 2025 Рогов А.А., Лукашевич Н.В., Сравнение подходов к интерпретации языковых моделей: анализ метода на основе маскированного языкового моделирования и традиционных методов //Труды XII национальной конференции по искусственному интеллекту с международным участием, КИИ-2025, с. 287-297 (РИНЦ)
18. 2025 Садковский Ф.А., Лукашевич Н.В., Гришин И.Ю., Автоматическое пополнение таксономий на русском языке с помощью больших языковых моделей //Труды XII национальной конференции по искусственному интеллекту с международным участием, том 3, с. 152-163 (РИНЦ)
19. 2025 Студеникина К.А., Лютикова Е.А., Герасимова А.А., Оценка лингвистической компетенции больших языковых моделей на материале корпуса согласовательной вариативности //Компьютерная лингвистика и вычислительные онтологии, издательство Университет ИТМО (Санкт-Петербург) (РИНЦ)
20. 2025 Тихомиров М.М., Чернышев Д.И., Ruadapt: вычислительно эффективная языковая адаптация больших языковых моделей //Доклады Российской академии наук. Математика, информатика, процессы управления, Российская академия наук (Москва), том 527, с. 290-300. – DOI: 10.7868/S2686954325070252 (**Ядро РИНЦ, переводная версия Web of Science, Scopus**)
21. 2025 Уразов С.О., Лукашевич Н.В., Задачи анализа тональности в новостных текстах //Труды XII НАЦИОНАЛЬНОЙ КОНФЕРЕНЦИИ ПО ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ С МЕЖДУНАРОДНЫМ УЧАСТИЕМ, КИИ-2025, с. 319-329 (РИНЦ)
22. 2025 Шамаева Е.Д., Лукашевич Н.В., Применимость методов оценки качества синтаксического анализа к результатам синтаксического анализатора на основе большой языковой модели //Труды XII НАЦИОНАЛЬНОЙ КОНФЕРЕНЦИИ ПО ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ С МЕЖДУНАРОДНЫМ УЧАСТИЕМ, том 1, с. 1-330 (РИНЦ)
23. 2025 Ярошенко Полина Владимировна, Лукашевич Наталья Валентиновна, RusEmoLex: эмоциональный лексикон для русского языка //Russian Journal of Linguistics, Российский университет дружбы народов (РУДН) (Москва), том 29, № 3, с. 659-687 (Scopus Q1). – DOI: 10.22363/2687-0088-44439
24. 2025 Яушев К.Ш., Лукашевич Н.В., Автоматическая генерация ключевых слов для русскоязычных научных статей с использованием псевдоразметки и контрастивного обучения //Труды XII национальной конференции по искусственному интеллекту с международным участием, том 1, с. 341-352 (РИНЦ)