

УДК 004.89

doi: ... заполняется издательством

Автоматическое пополнение таксономий на русском языке с помощью больших языковых моделей¹

Ф.А. Садковский (*sadkovsky@iling-ran.ru*)^{A,B}

Н.В. Лукашевич (*louk_nat@mail.ru*)^A

И.Ю. Гришин (*grishin@sev.msu.ru*)^A

^A Московский государственный университет им. М.В.

Ломоносова, Москва

^B Институт языкознания РАН, Москва

В настоящей работе впервые исследуется потенциал больших языковых моделей для задачи автоматического пополнения таксономий на материале русского языка. Авторы адаптировали передовую методику TaxoLLaMA, которая ранее показала высокую эффективность для английского языка, используя для этого данные русскоязычного тезауруса RuWordNet. В рамках исследования было проведено сравнение производительности мультязычных моделей и моделей, специально адаптированных для русского языка. Эксперименты подтвердили успешную применимость метода к русскоязычным данным и выявили значительное преимущество русскоязычных моделей. После дообучения модель YandexGPT-5-8B-Lite превзошла лучший предыдущий результат, достигнув значения MAP 55.42.

Ключевые слова: таксономия, пополнение таксономий, большие языковые модели, RuWordNet, TaxoLLaMA, RuAdapt.

Введение

Большие языковые модели (LLM) позволили достичь значительного прогресса в решении многих задач обработки естественного языка (NLP) благодаря их способностям к генерации контекстно релевантных терминов и выявлению семантических связей.

¹ Исследование выполнено при финансовой поддержке Междисциплинарной научно-педагогической школы Московского университета (грант № 23-Щ05-11) и государственного задания (регистрационный № 124020100068-4).

LLM показали свою эффективность и в задаче пополнения таксономий, которая ранее решалась с помощью извлечения текстовых шаблонов [Hearst, 1992; Sabirova, Lukanin, 2014], применения векторных моделей или комбинации нескольких методов [Nikishina et al. 2020a,b, 2022]. В недавней работе [Moskvoretskii et al., 2024] авторам удалось достигнуть передовых результатов в задачах извлечения гиперонимов и построения таксономий, а также сопоставимых с передовыми результатов в задаче пополнения таксономий благодаря предложенному методу TaxoLLaMA — обучению модели LLaMA-2-7b на материале пар гипоним–гипероним, извлеченных из тезауруса WordNet [Miller, 1990].

Тем не менее, тестирование подхода происходило почти исключительно на англоязычных данных: другие языки (испанский, итальянский) были представлены лишь в задаче извлечения гиперонимов, но и в этом случае с помощью дообучения удалось превзойти результаты предыдущих подходов.

В то же время, порождающие большие языковые модели и, в частности, подход TaxoLLaMA, ранее не применялись к данным на русском языке. Существующие подходы на основе векторных моделей [Nikishina et al. 2020a,b, 2022] не позволяют обеспечивать качество, достаточное для автоматизации задачи. Это представляется упущением, поскольку имеется ряд современных LLM, которые были специально разработаны для использования на русском языке — и среди них, модели, созданные по методике языковой адаптации RuAdapt [Tikhomirov, Chernyshev, 2023, 2024], эффективность которых в других задачах NLP на русском языке выше, чем у мультязычных аналогов.

В настоящей работе исследуется возможность применения LLM к данным таксономии RuWordNet [Loukachevitch et al., 2016].

- В работе была доказана переносимость метода TaxoLLaMA на русскоязычные данные.
- Сравниваются мультязычные модели и модели, предназначенные для работы с русским языком; вторые показали большую эффективность.
- Дообучение модели YandexGPT-5-Lite-8B-pretrain [Яндекс, 2025] позволило получить передовое качество решения задачи.

1. Обзор литературы

Задача пополнения таксономии изначально формулировалась как задача извлечения гиперонима для нового понятия. Одним из первых подходов к извлечению гиперонимов для слов русского языка был предложен в работе [Sabirova, Lukanin, 2014], в которой текстовые шаблоны из [Hearst, 1992] были переведены на русский язык, дополнены синонимичными выражениями и протестированы на русской версии веб-онтологии DBPedia.

Более поздние работы, посвященные разработке и тестированию моделей на русском материале, связаны с соревнованием RUSSE'2020 [Nikishina et al., 2020a], где был впервые предложен датасет диахронического типа на основе RuWordNet [Loukachevitch et al., 2016]. Лучшие результаты были достигнуты с помощью подхода на основе большого набора признаков из внешних источников (перевод, поиск в интернете, Викисловарь).

В последующих работах исследуется применение разнообразных векторных представлений-источников для предсказания ближайшего гиперонима в RuWordNet [Nikishina et al., 2022]. Самые высокие показатели достигаются использованием мета-эмбедингов, обученных на основе дистрибутивного (контекстного) и графового представления слов.

При этом, кроме модели-кодировщика RuBert [Kuratov, Arkhipov 2019], применение больших языковых моделей не было отмечено применительно к русским данным. Это видится серьезным упущением, поскольку в настоящий момент существует как большое количество многоязычных языковых моделей, так и моделей, специально обученных для понимания инструкций на русском языке. Среди них можно отметить модели YandexGPT компании «Яндекс», а также модели, созданные в рамках исследовательского направления по адаптации больших языковых моделей на русский язык RuAdapt [Tikhomirov, Chernyshev, 2023, 2024]. Последний подход предполагает замену токенизатора на униграммный для обеспечения лучшего соответствия между токенами и русскими морфемами [Tikhomirov, Chernyshev, 2023], замену входных представлений и дообучение на коллекции очищенных русскоязычных корпусов DaruLM и датасете инструкций на русском Darumeru [Tikhomirov, Chernyshev, 2024].

2. Данные

Источником данных в настоящей работе служит диахронический датасет на основе RuWordNet, впервые предложенный в [Nikishina et al., 2020a] и позже расширенный в [Nikishina et al., 2020b].

В подобных датасетах в качестве обучающей подвыборки используется более ранняя версия ресурса (RuWordNet 1.0), а в качестве тестовой — слова, добавленные в более позднюю («RuWordNet 2.0–1.0»).

В [Nikishina et al., 2020a] было предложено разделение новых слов RuWordNet на две подвыборки — «public» (валидационная) и «private» (тестовая). Тестовая выборка использовалась для окончательной оценки качества предложенных решений и включает 1 525 синсетов (множеств синонимов); валидационная выборка содержит 763 синсета. В качестве правильных ответов к целевым словам были извлечены не только непосредственные гиперонимы, но и гиперонимы второго порядка.

В RuWordNet узлы графа представляют собой синсеты. Названия синсетов могут быть одним словом, словосочетанием или включать перечисление слов/словосочетаний через запятую, уточняющие слова в скобках. Другая особенность обусловлена тем, что RuWordNet наследует названия синсетов от тезауруса Рутез [Loukachevitch, Dobrov, 2002], семантическая сеть которого не подразделялась на подграфы разных частей речи. Таким образом, в зависимости от семантики слова, названия синсетов для существительных могут быть не только существительными, но и глаголами («буллинг»>«оскорбить») и прилагательными («консонанс»>«благозвучный»).

3. Метод

Оригинальный подход ТахоLLaMA, предложенный в [Moskvoretskii et al., 2024], включает следующие компоненты:

1. подготовка выборки на основе пар гипоним–гипероним из WordNet; тестовые понятия, входящие в датасеты для тестирования, удаляются;
2. дообучение модели LLaMA-2-7b [Touvron et al., 2023] на предсказание списка гиперонимов через запятую;
3. оценка качества на датасетах для тестирования.

В настоящей работе предлагается расширение этого метода на другие данные и модели:

1. В качестве обучающего множества берутся пары гипоним–гипероним из RuWordNet 1.0.
2. Различные 7-8 млрд-ные LLM (перечислены в след. разделе) обучаются предсказывать слова через разделитель «;», поскольку этот знак не встречается в названиях синсетов.
3. Оценка качества на диахроническом датасете «RuWordNet 2.0–1.0» (выборка «nouns-private»).

Перед оцениванием необходимо сопоставить предсказанной последовательности слов названия синсетов. Для этого в работе предложена следующая процедура:

1. В начало списка переносятся последовательности слов, соответствующие существующим в таксономии названиям синсетов.
2. Синсет переводится в векторный вид с помощью L2-нормализации (по [Bollegalo, Bao, 2017]) векторных представлений всех неслужебных слов, входящих в название синсета.
3. Max-Pooling: $k = 15$ ближайших соседей из множества векторных представлений синсетов были найдены для каждого из нерелевантных кандидатов, а затем среди объединенного множества размера $k \times m$ ($m \leq 15$ — число предсказанных кандидатов) были извлечены все названия синсетов, встречающиеся более 1 раза, и упорядочены по

частоте и по позиции первого появления в общем списке $\langle x_{1,1}, \dots x_{1,k}, \dots x_{m,1}, \dots x_{m,k} \rangle$.

Данный метод пост-обработки использует преимущества организации RuWordNet и позволяет достигнуть более высокого итогового качества, чем использование предсказаний в чистом виде. В качестве векторной модели использовалась модель FastText² [Bojanowski et al., 2017].

4. Эксперименты

4.1. Метрики

В задаче пополнения таксономии используются две основные метрики: Mean Average Precision, MAP (4.1) и Mean Reciprocal Rank, MRR (4.2):

$$MAP@K = \frac{1}{N} \sum_{n=1}^N AP@K_n, \quad (4.1)$$

$$AP@K = \frac{1}{R} \sum_{k=1}^K (Precision@k \cdot I[y_k = 1]),$$

$$Precision@k = \frac{p}{k},$$

$$MRR@K = \frac{1}{N} \sum_n \frac{1}{rank_i} \quad (4.2)$$

Обозначения: $K = 15$ — максимальное число предсказаний; N — размер выборки; R — количество правильных ответов (гиперонимов) для целевого слова; k — индекс элемента в списке предсказаний; I — индикаторная функция; y_k — класс (верный ответ/неверный ответ) на позиции k , p — количество релевантных элементов среди k первых предсказанных; $rank_i$ — позиция первого релевантного элемента.

Используемая в ряде работ (в частности, [Moskvoretskii et al., 2024]) модификация Scaled MRR не используется, поскольку она предназначена для выборок, включающих только непосредственные гиперонимы.

4.2. Базовые методы

В качестве базовых методов используются базовый метод на основе векторной модели FastText (*fasttext*) и лучший метод на основе мета-эмбеддингов (*AAEME (words + TADW)*), усредненная сумма графовых + контекстных представлений) из [Nikishina et al., 2022].

² Предобученные векторы были взяты с официального сайта <https://dl.fbaipublicfiles.com/fasttext/vectors-crawl/cc.ru.300.bin.gz>

4.3. Модели

Для экспериментов были отобраны большие языковые модели типа декодирующей с числом параметров, приблизительно соответствующим числу параметров исходной модели LLaMA-2 — 7–8 млрд. Выбирались модели в открытом доступе, не проходившие инструктивное дообучение:

- LLaMA-2-7B-hf³;
- LLaMA-3.1-8B⁴;
- Mistral-7B-v0.1 [Jiang et al., 2023];
- Mistral-7B-v0.3⁵;
- Qwen2.5-7B [Yang et al., 2024];
- Qwen3-8B-Base⁶;
- Gemma-7b [Team Google, 2024].

В список также была включена оригинальная модель TaxoLLaMA: результаты ее дообучения позволяют судить о том, насколько модель способна переносить знания о таксономических отношениях с одного языка на другой.

Другая группа моделей включала базовые модели, специально предназначенные для работы с русским языком и примерно с тем же количеством параметров:

- YandexGPT-5-Lite-8B-pretrain [Яндекс, 2025]
- Модели из серии RuAdapt [Tikhomirov, Chernyshev, 2023, 2024]:
 - RuAdaptQwen2.5-7B-Lite-Beta;
 - RuAdaptLLaMA-2-7b;
 - RuAdaptMistral-v0.1;
 - RuAdaptLLaMA-3.1-8B.

4.3. Результаты

Результаты моделей⁷ приводятся в таблице 1⁸. Среди мультиязычных моделей, наилучшие результаты продемонстрировали модели семейства Qwen, в особенности Qwen3-8B-Base. Тем не менее, по метрике MAP даже лучшие модели не смогли превзойти базовые методы.

Модели для русского языка в среднем показали результаты выше: 45.02 против 38.12 по MRR и 32.8 против 23.21 по MAP. Все модели, прошедшие адаптацию на русский язык, показали более высокое качество, чем их оригинальные неадаптированные версии, см. таблицу 2.

³ <https://huggingface.co/meta-llama/Llama-2-7b-hf>

⁴ <https://huggingface.co/meta-llama/Llama-3.1-8B>

⁵ <https://huggingface.co/mistralai/Mistral-7B-v0.3>

⁶ <https://huggingface.co/Qwen/Qwen3-8B-Base>; <https://qwenlm.github.io/blog/qwen3/>

⁷ Веса дообученных моделей доступны на сайте: <https://huggingface.co/TaxoLLMs>

⁸ Здесь и далее значения метрик умножены на 100

Модель на основе Qwen2.5-7B оказалась лидером как по абсолютному результату, так и по приросту метрик, который дает методика адаптации. Следующей моделью по значению метрик оказалась модель на основе LLaMA-3.1-8B, а следующей моделью по приросту качества — модель на основе Mistral-7B-v0.1.

Среди всех моделей самого высокого результата по метрикам MRR и MAP достигла модель YandexGPT-5-Lite-8B-pretrain. Значение метрики MAP 55.42 превосходит значение 47.4 лучшего подхода из [Nikishina et al., 2022b]. Вторую позицию занимает модель RuAdapt-Qwen2.5-7B с результатом 41.69. Этот результат не смог превзойти лучший предыдущий результат, однако оказался выше, чем у базового решения (41.4).

Самого низкого качества достигла дообученная на русских данных модель TaxoLLaMA. Значения метрик оказались даже ниже, чем у исходной для нее LLaMA-2-7b-hf, что говорит о неспособности модели эффективно переносить таксономические знания из WordNet на данные RuWordNet.

Табл. 1. Сравнение качества моделей после дообучения на «RuWordNet 1.0 – 2.0». Жирный шрифт — лучший результат, подчеркивание — второй после лучшего.

модели	MRR	MAP
LLaMA-2-7b-hf	25.01	17.35
LLaMA-3.1-8B	35.11	25.35
Mistral-7B-v0.1	32.94	22.32
Mistral-7B-v0.3	32.73	21.62
Qwen2.5-7B	38.2	29.43
Qwen3-8B-Base	40.06	30.06
Gemma-7b	30.79	23.44
TaxoLLaMA	24.10	16.09
YandexGPT-5-Lite-8B-pretrain	61.55	55.42
RuadaptQwen2.5-7B-Lite-Beta	<u>49.47</u>	41.69
RuAdaptLLaMA-2-7b	30.82	25.11
RuAdaptLLaMA-3.1-8B	42.32	34.79
RuAdaptMistral-v0.1	40.95	33.57
<i>fasttext</i>	—	41.40
<i>AAEME triplet loss (words)</i>	—	<u>47.40</u>

Табл. 2. Сравнение результатов моделей с адаптацией на русский язык по методике RuAdapt [Tikhomirov, Chernyshev, 2023, 2024] и без.

модель	оригинал		RuAdapt	
	MRR	MAP	MRR	MAP
LLaMA-2-7b-hf	25.01	17.35	30.82 ^{+5.81}	25.11 ^{+7.76}
LLaMA-3.1-8B	35.11	25.35	<u>42.32</u> ^{+7.21}	<u>34.79</u> ^{+9.44}
Mistral-7B-v0.1	32.94	22.32	40.95 ^{+8.01}	33.57 ^{+11.25}
Qwen2.5-7B	38.2	29.43	49.47 ^{+11.27}	41.69 ^{+12.26}

5. Анализ ошибок

В тестировании на русском языке можно выделить 7 основных типов ошибок. Эти типы перечислены ниже и проиллюстрированы примерами.

1. Неверно определена предметная область. Слово «авуары» (первое значение по [BTC, 2003] — «средства, за счёт которых производятся платежи и погашаются обязательства») относится к сфере финансов, однако модель Qwen3-8B-Base предсказывает гиперонимы из области этнологии: «этническая общность; этнические народы; африканцы (население); население государства; нация» при правильных «денежная единица; свободно конвертируемая валюта; банковский вклад».

2. Другие семантические отношения. Примером может служить ответ модели RuAdaptQwen2.5-7B для целевого слова «пивная»: модель через корень связала это слово с алкогольным напитком, но предсказала кандидаты, подходящие для других однокоренных слов — «пивоварни» и «пива»: «предприятие пищевой промышленности; алкогольный напиток; спиртосодержащая продукция».

3. Слишком абстрактные кандидаты. К таким примерами относится, в частности, предсказание YandexGPT-5-Lite для слова «перевязь» — «изделие легкой промышленности; изделие; полоса (кусок); кусок (отдельная часть); повязка, повязка на теле;...» при верном ответе «медицинская повязка; медицинская продукция».

4. Слишком конкретные кандидаты. Такие случаи связаны с неполнотой ресурса, а не с тем, как модель выявляет семантические связи. Примером может служить предсказание Qwen3-8B-Base для слова «бойскаут» — «член организации; участник; человек по роли; учащийся; подросток;...» при правильном ответе «мальчик; мужчина, человек мужского пола; скаут; несовершеннолетние дети». Определение этого слова по словарю [Крысин, 2018] — «мальчик или подросток — член скаутской (см. скаут) организации», из чего можно сделать вывод о релевантности кандидата на первой позиции.

5. Ошибки, связанные с многозначностью. Предсказание гиперонимов для слов, обозначающих несколько понятий одновременно, вне контекста

использования также составляет проблему. В частности, в тестовую выборку входит слово «лигатура», имеющее следующие гиперонимы: (1) «хирургическое оборудование», «нить (предмет)», (2) «металл», «добавление (то, что добавлено)» и (3) «графический знак», «сочетание, комбинация». Моделям YandexGPT-5-8B-Lite и Qwen3-8B-Base удалось предсказать только последнее значение: модель серии YandexGPT смогла предсказать непосредственный гипероним «графический знак», а также «знак, обозначение», а модели серии Qwen3 — гипероним только второго порядка («изображение (результат)»).

6. Интерференция другого языка в русский текст. До обработки вывода для извлечения синсетов, среди кандидатов могут попадаться как частичные, так и полные замены слов: высший «законодательный орган» (LLaMA-3.1-8B); «органическое glass», «geometric solid», «combinatorial analysis», «травянистое植物» (Qwen2.5-7B).

7. Неверно предсказана часть речи в названии синсета. Например, модель RuAdapt-Qwen2.5-7B для слова «осиплость» предсказала гиперонимы «измениться, изменение; изменить, сделать иным;...», вероятно, связав их с событийной семантикой корня «осип-» — ‘стать хриплым’. При этом в RuWordNet для данного слова требовалось предсказать признак: «хриплый; глухой, глухо звучащий».

6. Обсуждение результатов

Предложенные методы использования мультязычных LLM с числом параметров 7–8 млрд оказались недостаточно эффективными в применении к таксономии на русском языке. В то же время, использование специализированных русскоязычных моделей продемонстрировало свой потенциал. В частности, максимальное качество, значительно превосходящее предыдущие векторные подходы, удалось получить благодаря дообучению модели YandexGPT-5-Lite-8B-pretrain.

Достоверно неизвестно, насколько этот результат является «чистым» в смысле проникновения данных RuWordNet в обучающую выборку этой модели. Однако можно привести ряд доводов против такого предположения.

- Разница в качестве между предсказанием до сопоставления с синсетами с помощью модели FastText и после (около 3 по MAP) не отличается от разницы, наблюдаемой у моделей RuAdapt, про которые достоверно известно, что RuWordNet не был включен в обучающие данные. Если бы модель специально дообучалась на RuWordNet, можно было бы ожидать меньшего выигрыша от использования нерелевантных (по форме) предсказаний.
- В официальной статье [Яндекс, 2025] нет сведений о том, что RuWordNet как-либо использовался для обучения или тестирования.

- В той же статье показано, что на основных датасетах для сравнения качества моделей YandexGPT-5-Lite достигает паритета с аналогами или превосходит их.

Таким образом, естественно ожидать, что дообучение данной модели в целом эффективно для широкого спектра задач обработки русского языка.

Лучший результат среди протестированных моделей серии RuAdapt достигнут моделью на основе Qwen2.5-7B, которая превзошла базовый метод *fasttext* по метрике MAP. Из неадаптированных моделей самого высокого качества достигла модель Qwen3-8B-Base, для которой адаптированная на русский язык версия на момент написания работы еще не была выпущена. Из адаптированных моделей ей удалось обойти только RuAdapt-LLaMA-2-7b, однако исходя из наблюдаемой тенденции можно ожидать, что качество адаптированной модели может быть сопоставимо с результатами модели от компании Yandex.

Что касается ошибок моделей, то более слабым моделям свойственно делать серьезные семантические ошибки (1–2 типы), а более сильным моделям — менее серьезные (3–4 типы), которые не являются ошибками в идеологическом смысле. Это указывает на то, что применение LLM к данной задаче весьма перспективно, хотя нуждается в дальнейшем совершенствовании методологии.

Можно предложить следующие поправки к методологии, которые будут реализованы в дальнейших исследованиях:

1. использовать в инструкции контекст с новым словом; многозначным словам должны соответствовать разные контексты;

2. использовать модель не как генератор, а как переранжировщик (англ. *Reranker*): согласно ряду исследований, эта роль — более сильная стороны генеративных больших языковых моделей, чем извлечение закономерностей из примеров (*In-context learning*) [Ma et al., 2023];

3. подходы 1 и 2 можно совмещать, представляя задачу пополнения таксономии как задачу генерации, дополненной поиском (англ. *Retrieval Augmented Generation*, RAG): LLM получает на вход источники информации (список потенциальных гиперонимов, контексты из корпуса, определения из внешних ресурсов), принимает решения о релевантности и выдает итоговый упорядоченный список.

Заключение

В статье была впервые рассмотрена возможность применения порождающих больших языковых моделей к решению задачи пополнения таксономии на русском языке. На диахроническом датасете RuWordNet были обучены и протестированы как применявшиеся в предыдущих частях модели, так и специально адаптированные под русский язык версии некоторых из них, а также модель YandexGPT-5-8B-Lite, также

предназначенная для использования преимущественно на русском. В работе использована методология дообучения, разработанная на основе передового подхода TaхоLLaMA с учетом специфики данных.

В результате двум дообученным моделям — YandexGPT-5-8B-Lite и RuAdaptQwen2.5-7B-Lite — удалось превзойти показатели предыдущего базового метода, а модели YandexGPT-5-8B-Lite, достигшей результата 55.42 по метрике MAP, также превзойти и результат наилучшего предыдущего метода с результатом 47.4. Также на материале настоящей работы удалось дополнительно продемонстрировать преимущество методики RuAdapt для русского языка, поскольку все модели, созданные по этой методике, после дообучения на датасете RuWordNet продемонстрировали лучшее качество, чем их оригинальные версии, после аналогичного дообучения.

Список литературы

- [БТС, 2003] Большой толковый словарь / Гл. ред. С. А. Кузнецов. – СПб.: Норинт, 2004. – 1534 с.
- [Крысин, 2018] Крысин Л.П. Современный словарь иностранных слов. – М.: АСТ, 2018. – 416 с.
- [Яндекс, 2025] Встречаем YandexGPT 5 — в Алисе, облаке и opencore / Блог компании Яндекс [Электронный ресурс] // Хабр. 2025. URL: <https://habr.com/ru/companies/yandex/articles/885218s> (дата обращения 15.06.2025).
- [Bollegala, Bao, 2017] Bollegala D., Bao C. Learning word meta-embeddings by autoencoding. In: Proceedings of the 27th international conference on computational linguistics, 2018, pp. 1650-1661.
- [Bojanowski et al., 2017] Bojanowski, P., Grave, E., Joulin, A., Mikolov, T. Enriching word vectors with subword information. Transactions of the association for computational linguistics. 2017. Vol. 5. pp. 135-146. doi: 10.1162/tacl_a_00051
- [Hearst, 1992] Hearst M. A. Automatic acquisition of hyponyms from large text corpora In: Proc. COLING 1992 volume 2: The 14th international conference on computational linguistics, Nantes, France, August 1992, pp. 539–545.
- [Jiang et al., 2023] Jiang A. Q. et al. Mistral 7b. Computer research repository. 2023. doi: 10.48550/arXiv.2310.06825
- [Kuratov, Arkhipov, 2019] Kuratov, Y., Arkhipov, M. Adaptation of deep bidirectional multilingual transformers for Russian language. Computer research repository. doi: 10.48550/arXiv.1905.0721
- [Loukachevitch et al., 2016] Loukachevitch N. V., Gerasimova, A. A., Dobrov, B. V., Lashevich, G., & Ivanov, V. V. Creating Russian wordnet by conversion. In: Computational Linguistics and Intellectual Technologies, Moscow, Russia, June, 2016, pp. 405-415.
- [Ma et al., 2023] Ma, Y., Cao, Y., Hong, Y., & Sun, A. Large language model is not a good few-shot information extractor, but a good reranker for hard samples!. Arxiv. 2023. doi: 10.48550/arXiv.2303.08559

- [**Miller et al., 1990**] Miller G. A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K. J. Introduction to WordNet: An on-line lexical database. *International journal of lexicography*. 1990. Vol. 3(4). pp. 235–244. doi: 10.1093/ijl/3.4.235
- [**Moskvoretskii et al., 2024**] Moskvoretskii, V., Neminova, E., Lobanova, A., Panchenko, A., Nikishina, I. Large Language Models for Creation, Enrichment and Evaluation of Taxonomic Graphs. *Semantic Web Journal* (forthcoming). 2024.
- [**Nikisina et al., 2020a**] Nikishina I., Logacheva V., Panchenko A., Loukachevitch N. RUSSE'2020: Findings of the First Taxonomy Enrichment Task for the Russian language. In: *International Conference on Computational linguistics and intellectual technologies Dialog-2020*, Moscow, Russia, June, 2020, pp. 579–595.
- [**Nikisina et al., 2020b**] Nikishina, I., Panchenko, A., Logacheva, V., & Loukachevitch, N. Studying taxonomy enrichment on diachronic wordnet versions. In: *Proc.of the 28th International Conference on Computational Linguistics*, Barcelona, Spain, December, 2020, pp. 3095–3106.
- [**Nikisina et al., 2022**] Nikishina, I., Tikhomirov, M., Logacheva, V., Nazarov, Y., Panchenko, A., Loukachevitch, N. Taxonomy enrichment with text and graph vector representations. *Semantic Web*. 2022. 13(3). pp. 441–475. doi: 10.3233/SW-212955
- [**Sabirova, Lukanin, 2024**] Sabirova K., Lukanin A. Automatic Extraction of Hypernyms and Hyponyms from Russian Texts. In: *AIST (supplement)*. Yekaterinburg, Russia, 2014, pp. 35–40.
- [**Team Google, 2024**] Team Google. Gemma: Open models based on gemini research and technology. *Arxiv*. 2024. doi: 10.48550/arXiv.2403.08295
- [**Tikhomirov, Chernyshev, 2023**] Tikhomirov M., Chernyshev D. Impact of tokenization on LLaMa Russian adaptation. In: *Ivannikov Ispras Open Conference (ISPRAS)*, Moscow, Russia, December, 2023, pp. 163–168.
- [**Tikhomirov, Chernyshev, 2024**] Tikhomirov M. M., Chernyshev D. I. Improving Large Language Model Russian adaptation with preliminary vocabulary optimization. *Lobachevskii Journal of Mathematics*. 2024. Vol 45(7). pp. 3211–3219. doi: 10.1134/S1995080224604120
- [**Touvron et al., 2023**] Touvron H. et al. Llama: Open and efficient foundation language models. *Arxiv*. 2023. doi: 10.48550/arXiv.2302.13971
- [**Yang et al., 2023**] Yang A. et al. Qwen2. 5 technical report. *Arxiv*. 2024. doi: 10.48550/arXiv.2412.15115

Automatic taxonomy enrichment in Russian language with large language models

F.A. Sadkovsky (*sadkovsky@iling-ran.ru*)^{A,B}

N.V. Loukachevitch (*louk_nat@mail.ru*)^A

I.Yu. Grishin (*grishin@sev.msu.ru*)^A

^A Lomonosov Moscow State University, Moscow

^B Institute of Linguistics RAS, Moscow

This paper presents the first investigation into the potential of large language models for the task of automatic taxonomy enrichment using Russian language material. The authors have adapted the TaxoLLaMA methodology, which had previously demonstrated high efficacy for the English language, by utilizing data from the Russian-language thesaurus, RuWordNet. As part of this research, a comparative performance analysis was conducted between multilingual models and models specifically adapted for the Russian language. The experiments confirmed the successful applicability of the method to Russian-language data and revealed a significant advantage of the Russian-specific models. After fine-tuning, the YandexGPT-5-8B-Lite model surpassed the previous best result, achieving a Mean Average Precision (MAP) score of 55.42.

Keywords: taxonomy, taxonomy enrichment, large language models, TaxoLLaMA, RuWordNet, RuAdapt.