УДК 004.85

ИССЛЕДОВАНИЕ ТАКСОНОМИИ С ПОМОЩЬЮ РАССУЖДАЮЩИХ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ И ВЫЗОВА ФУНКЦИЙ

Ф. А. Садковский[1] [0009-0006-4502-9466], **М. М. Тихомиров**[2] [0000-0001-7209-9335], **Н. В. Лукашевич**[3] [0000-0002-1883-4121]

[1–3]*Московский государственный университет имени М. В. Ломоносова, г. Москва, Россия*

[1]feudor987@mail.ru, [2]tikhomirov.mm@gmail.com, [3]louk_nat@mail.ru

Аннотация

Рассмотрена задача пополнения таксономий – иерархических структур для организации понятий. Предложена архитектура на основе подхода ReAct (Reasoning + Acting), позволяющая пополнять таксономию в режиме zero-shot без дообучения больших языковых моделей. Система реализована в двух сценариях: автономная навигация от корневых узлов и верификация гипотез, сгенерированных другими моделями. Эксперименты на материале диахронического датасета RuWordNet показали, что прямое исследование таксономии от корня сталкивается с ограничениями, связанными со сложностью графа (MAP@3 = 24.6%). В то же время использование системы в качестве верификатора позволило улучшить качество предсказаний базовых моделей: прирост MAP@3 составил 9.5 п.п. для FastText и 1.1 п.п. для TaxoYandexGPT-5-Lite. Ключевыми преимуществами подхода являются универсальность, отсутствие необходимости дообучения и интерпретируемость за счет явных цепочек рассуждений.

Ключевые слова*: таксономия, пополнение таксономий, извлечение гиперонимов, LLM, цепочка рассуждений, вызов функций, RuWordNet, переранжирование, компьютерная семантика.*

ВВЕДЕНИЕ

Таксономии – иерархические структуры для организации понятий – требуют постоянного пополнения для поддержки их актуальности, но эта задача является трудоемкой даже для экспертов. В последние годы для ее автоматизации чаще стали применять большие языковые модели (Large Language Model, LLM),

которые показали высокий потенциал в анализе семантических связей [1–7]. Однако большинство существующих подходов требует дообучения моделей, что делает их неуниверсальными, дорогими в адаптации и зачастую неинтерпретируемыми. В то же время современные LLM показывают высокую эффективность в режиме без обучения (zero-shot), особенно при использовании «цепочки рассуждений» (chain-of-thought) в сочетании с вызовом функций.

В настоящей работе мы предлагаем архитектуру для пополнения таксономий, не требующую дообучения. Она основана на подходе ReAct (Reasoning + Acting) [8], в котором LLM после этапа «рассуждения» совершает «действие» – вызов внешнего инструмента для доступа к таксономии. Отсутствие обучения обеспечивает универсальность метода, а использование режима рассуждений повышает интерпретируемость принятого решения. Мы исследуем два сценария применения системы:

1) система-исследователь: автономная навигация «от корня» для оценки возможностей и ограничений метода;

2) система-верификатор: гибридный подход, где система проверяет и уточняет гипотезы, сгенерированные другой моделью.

Сравнение этих дизайнов позволяет оценить как общую эффективность решения, так и фундаментальные ограничения современных LLM при работе со сложными графами без предварительного сужения пространства поиска.

ОБЗОР СУЩЕСТВУЮЩИХ ПОДХОДОВ

Для решения задачи пополнения таксономии широкое применение получили модели-кодировщики. В [1] предложена модель TEMP, использующая дообучение BERT для оценивания пар «определение нового понятия – путь в таксономии». Схожий метод, но для моделей декодировщиков, обсуждается в [2] – это двухэтапный метод обучения, состоящий в сжатии информации о понятии в один токен и обучении векторных представлений слов по их пути в дереве таксономии. В [3] с помощью контрастивного обучения BERT обучены эмбеддинги, сближающие понятия с их позицией ⟨гипероним, гипоним⟩ в графе. По методу TaxoComplete [4] би-кодировщик дообучается на предсказание меры расстояния в графе между двумя узлами, а в методе TacoPrompt [5] на предсказание близости в триплете (запрос, гипероним, гипоним) обучается кросс-кодировщик. В [6]

предложен метод ансамблирования результатов вычисления косинусной близости на эмбеддингах трех LLM. В [7] исследованы возможности дообучения больших языковых моделей для предсказания гиперонима целевого слова, в том числе для решения задачи пополнения таксономии с помощью генерации кандидатов и использования вероятности на выходном линейного слоя.

В работах [9, 10] на материале таксономий лингвистических онтологий WordNet [11] и RuWordNet [12] предложен особый тип датасета – диахронический WordNet, где старая версия ресурса доступна для обучения, а более новая используется для тестирования. Использование датасетов такого типа приближает условия решения задачи пополнения таксономии к более естественным, чем использование датасетов из случайно отобранных узлов. В [13] на материале подобных датасетов наилучшие результаты были достигнуты с помощью объединения мета-эмбеддингов [14] и графовых представлений.

ДАННЫЕ

Набор данных, использованный в работе, состоял из большого корпуса текстов и тестового множества имен существительных, предложенных на соревновании по пополнению таксономий [9] и представляющих дополнение к таксономии в диахроническом датасете на основе RuWordNet [12]. Поскольку тестовые данные содержат одни и те же слова в нескольких значениях, был подготовлен специальный контекстуализированный вариант датасета: из первых 870 тыс. документов корпуса (небольших новостных текстов) были извлечены контексты для всех слов тестовой выборки, встретившихся в них. Далее с помощью модели Qwen3-235B-A22B-Instruct-2507[1] для каждой пары ⟨контекст, слово⟩ было выбрано значение из датасета, наиболее соответствующее контексту. Если ни одно из значений не подходило по контексту, слово исключалось. После завершения автоматической процедуры для обеспечения максимального качества аннотаций в тестовой выборке был проведен ручной этап проверки. В итоговый набор для исследования вошли 1520 текстов и 1360 понятий, где соответствие контекста целевому значению было подтверждено экспертом.

Таксономия RuWordNet, выбранная в качестве пополняемого ресурса, содержит 29296 узлов (синсетов) имен существительных. Максимальная глубина

[1] https://huggingface.co/Qwen/Qwen3-235B-A22B-Instruct-2507

(число топологических поколений) составляет 16, количество корневых узлов – 8.

МЕТОД

Формальные определения

Математически таксономия определяется как ориентированное дерево $\mathcal{T} = (\mathcal{N}, \mathcal{E})$, где каждый узел $n \in \mathcal{N}$ представляет некоторое понятие (класс), а каждое направленное ребро (дуга) $\langle u, v \rangle \in \mathcal{E}$ подразумевает отношение гипонимии $\mathrm{ISA}(v, u)$, которое определяется следующим образом:

$$\mathrm{ISA}(x, y) = \begin{cases} 1, & \text{если } x \text{ – гипоним } y, \\ 0, & \text{иначе.} \end{cases}$$

Однако в практических разработках таксономией чаще называют иерархически организованный набор данных со структурой в виде графа, имеющего менее строгие ограничения, а именно только требование ацикличности. Если в таксономии понятия v и u связаны отношением «являться гипонимом», то понятие v является более конкретным понятием по отношению к понятию u. Отношение ISA транзитивно, поэтому отсутствие циклов должно обеспечивать невозможность вывода из структуры таксономии одновременно пар $\langle u, v \rangle$ и $\langle v, u \rangle$, связанных этим отношением.

Задача пополнения таксономии в лингвистической онтологии формализуется следующим образом. Пусть имеется набор новых слов $W = \{w_1, \dots w_n\}$, этим словам соответствуют набор понятий $\mathcal{C} = \{c_1, \dots c_k\}$ (считаем, что каждому слову w_i может соответствовать набор из по крайней мере одного понятия $[\![w_i]\!] = \{c_{i1}, \dots\}$, где $[\![\dots]\!]$ – функция семантической интерпретации; наборы значений различных слов могут пересекаться) и таксономия $\mathcal{T}_0 = (\mathcal{N}_0, \mathcal{E}_0)$. Пусть в этой таксономии узел $N_i \in \mathcal{N}_0 = \{s_{i1}, \dots\}$ синсетом, содержащим слова s_{im}, объединенные общим значением: $\forall s_{im} \in N_i : \exists c := [\![s_{im}]\!] \cap [\![N_i]\!] = \{c\}$, причем $N_i \in \mathcal{N}_0 : [[\![N_i]\!]] = 1$. Необходимо при данной таксономии $\mathcal{T}_0 = (\mathcal{N}_0, \mathcal{E}_0)$ и наборе понятий $\mathcal{C}$ дополнить существующую таксономию $\mathcal{T}_0$ до более крупной таксономии $\mathcal{T} = (\mathcal{N}_0 \cup \mathcal{N}', \mathcal{E}_0 \cup \mathcal{R})$, где $\mathcal{N}'$ – множество существующих и/или новых синсетов, таких что $\forall w_i \in W : \exists N_i \in \mathcal{N}_0 \cup \mathcal{N}' : [\![w_i]\!] \cap [\![N_i]\!] \neq \emptyset$, а $\mathcal{R}$ – это установленные отношения для каждого синсета из $\mathcal{N}'$:

$$\forall N_i \in \mathcal{N}' : \exists N_k \in \mathcal{N}_0 \cup \mathcal{N}' : \langle N_i, N_k \rangle \in \mathcal{R} \vee \langle N_k, N_i \rangle \in \mathcal{R},$$

причем $\exists N_k \in \mathcal{N}' : N_k \in \mathcal{N}_0 \setminus \mathcal{N}'$, что гарантирует, что добавленные узлы будут достижимы хотя бы из одной вершины $\mathcal{T}_0$. В правильно построенной таксономии при этом гарантируется, что

$$\forall i \neq k : \langle N_k, N_i \rangle \in \mathcal{R} \Leftrightarrow ISA([\![N_i]\!], [\![N_k]\!]).$$

Обычно задачу пополнения таксономии сводят к нахождению оптимальной позиции для всех новых значений c_i некоторого слова либо пары $\langle N_h, N_p \rangle$ такой, что

$$[\![N_h]\!] = h_i, \qquad [\![N_p]\!] = p_i \,\&\, ISA(c_i, p_i) \,\&\, ISA(h_i, c_i),$$

причем h_i может быть нулевым либо в упрощенной постановке только N_p. Пусть существует идеальная таксономия $\mathcal{T}$, определенная выше. Тогда для каждого нового понятия $c \in \mathcal{C}$, $N_i : [\![N_i]\!] = c$ – синсет для этого понятия – требуется найти такой узел $v \in \mathcal{N}_0$, что $\langle v, N_i \rangle \in \mathcal{R} \Leftrightarrow ISA([\![N_i]\!], [\![v]\!])$[2]; добавить в граф ребро $\langle v, N_i \rangle$. В правильно построенной таксономии вместе с этим будет выполняться следующее:

$$\forall i, k, i \neq k : \langle N_k, N_i \rangle \in \mathcal{R} \,\&\, \langle v, N_i \rangle \in \mathcal{R} \rightarrow \langle N_k, v \rangle \in \mathcal{R},$$

т. е. v является самым глубоким подходящим узлом-гиперонимом для N_i.

Архитектура

Для каждого целевого слова, которое нужно включить в таксономию, т.е. найти для него наиболее подходящий синсет-гипероним, из корпуса извлекается пример употребления. Далее модель начинает просматривать информацию об узлах пополняемой таксономии для подбора подходящего синсета.

В настоящей работе рассмотрены два варианта процедуры использования системы:

1) исследование: анализ таксономии начинается от корневых узлов, когда модель начинает работу с обзора понятий верхнего уровня;

[2] Из-за сложности задачи различием между случаями $N_i \in \mathcal{N}_0$ и $N_i \notin \mathcal{N}_0$ обычно пренебрегают и на этапе автоматизации сводят их к $N_i \notin \mathcal{N}_0$.

2) верификация: анализ начинается от предзаданных стартовых узлов– заранее предсказанных другой моделью кандидатов – с целью проверить их релевантность и по возможности подобрать более точный ответ.

Архитектура системы состоит из четырех блоков: (1) генератор, (2) критик, (3) обработчик таксономии и (4) переранжировщик. Схема работы системы изображена на рис. 1.

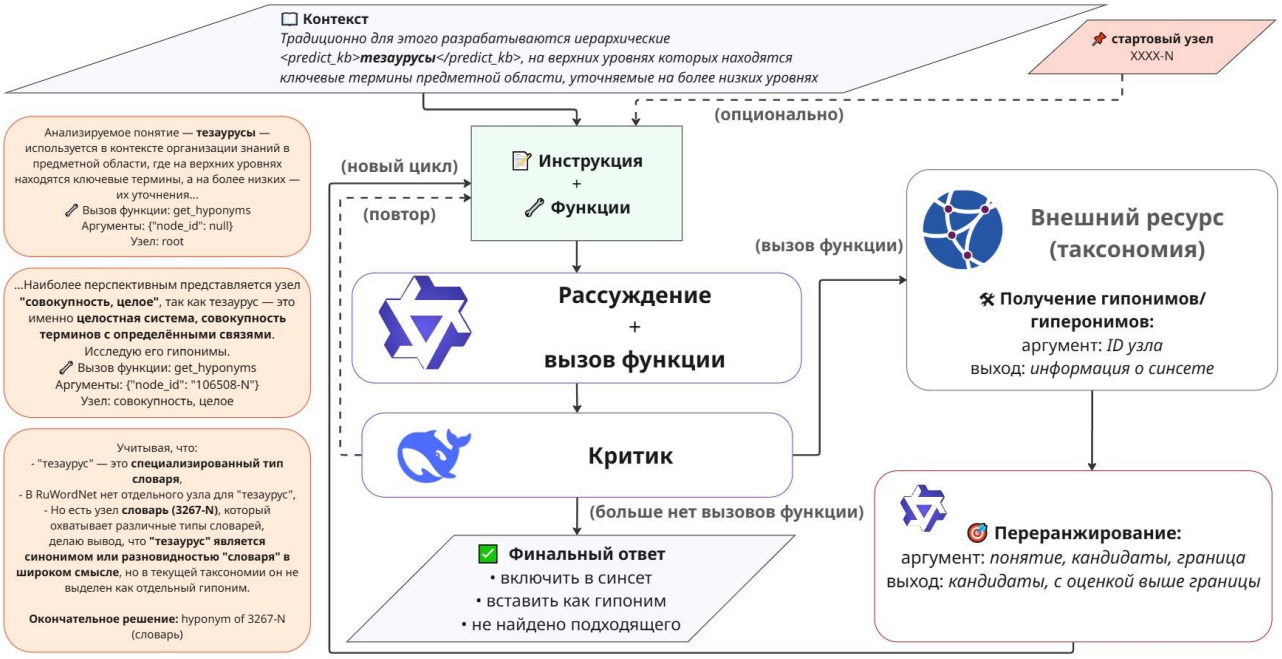

Рис. 1. Схема работы системы, пример запроса и фрагментов рассуждений модели.

Генератор представляет собой главного агента-исполнителя задачи. Он анализирует исходный запрос и стартовый узел при его наличии, порождает рассуждения, формирующие контекст, и определяет следующую функцию для вызова. Генератору доступны одна или две базовые функции:

1) $\text{get\_hyponyms}(n.id) \to (h_1, \dots) \mid n \in \mathcal{N}, h_i \in \mathcal{N}, \langle n, h_i \rangle \in \mathcal{E}$ – извлечение гипонимов из таксономии;

2) $\text{get\_hypernyms}(n.id) \to (p_1, \dots) \mid n \in \mathcal{N}, p_i \in \mathcal{N}, \langle p_i, n \rangle \in \mathcal{E}$ – извлечение гиперонимов из таксономии.

Первая функция доступна при обеих процедурах, вторая – только при движении от стартового узла. При движении от корня вся информация о вышестоящих узлах накапливается в контексте, поэтому при необходимости модели всегда доступна возможность вернуться к анализу верхних узлов.

Информация, которую модель получает о каждом узле, включает:

- *id* – идентификационный номер (ID) узла;
- *name* – название синсета;
- *hyponyms* – список названий и ID синсетов гипонимов;
- *hypernyms* – список названий и ID синсетов гиперонимов;
- *definition* – толкование значения синсета (при наличии).

Вызов функции доступа к таксономии модель может осуществить по ID любого из узлов, представленных в списках *hyponyms/hypernyms.*

Работа генератора происходит в цикле ReAct до тех пор, пока не выполнится одно из двух условий остановки: (1) модель принимает решение о завершении задачи (узел успешно найден) или о невозможности ее выполнить (не удалось найти подходящий узел в ходе исследования) и возвращает ответ без вызова функции; (2) модель достигает лимита итераций или достигается предел максимального количества токенов в контексте – в этих случаях процесс останавливается принудительно и ответом считается последний обработанный узел.

Критик. Задача агента-критика – оценить логику рассуждения генератора и корректность заполнения аргументов функции для вызова. В случае, если критик обнаруживает проблемы в ответе генератора, итерация повторяется заново с учетом замечаний критика. Формат ответа – рассуждение и JSON с полями “status” (“pass”/”fail”) и “comment” (опционально, если необходима обратная связь).

Обработчик таксономии. В отличие от двух предыдущих модулей, обработчик таксономии не является агентом, а представляет собой программный модуль для получения информации из графа таксономии. Его задача состоит в проверке корректности вызова функции и их исполнении.

Переранжировщик. Цель агента-переранжировщика состоит в сокращении списка гипонимов/гиперонимов, который может достигать 30–50 элементов для некоторых узлов. Без использования этого модуля происходит быстрое заполнение окна контекста, что не позволяет закончить выполнение задачи. Для оценки кандидатов использовалась 5-балльная шкала, переведенная в интервал [0, 1]. Для фильтрации кандидатов после переранжирования использовался адаптивный порог Th, который становится строже в зависимости от глубины узла в таксономии. Применялась эвристическая формула $Th = \log_2(g(N_i))$, где g –

функция, возвращающая ранг топологического поколения узла N_i. По сути, топологические поколения отражают уровень вложенности узла в иерархии. Формально они представляют собой разбиение множества узлов $\mathcal{N}$ на пересекающиеся подмножества G_k такие, что:

$$G_1 = \{n \in \mathcal{N} \mid \deg^-(n) = 0\};$$

$$\forall r > 1:$$

$$G_r = \left\{ n \in \mathcal{N} \setminus \bigcup_{j=0}^{r-1} G_j \mid \forall \langle v, n \rangle \in \mathcal{E} \Rightarrow v \in \bigcup_{j=0}^{r-1} G_j \ \& \exists \langle v, n \rangle \in \mathcal{E} : v \in G_{r-1} \right\},$$

т. е. в первом поколении находятся все корневые узлы, а в последующих поколениях с рангом r – узлы, все родители которых находятся в предыдущих поколениях, причем существует хотя бы один узел, находящийся в поколении ранга $r-1$.

Соответственно, функция $g: \mathcal{N} \to \mathbb{N}$ определяется следующим образом:

$$g(n) = k : n \in G_k.$$

Таким образом, чем глубже узел в таксономии (т.е. чем больше его ранг $g(n)$), тем выше значение порога Th. Выбор такой оценки обусловлен тем, что на более глубоких уровнях таксономии семантическая гранулярность выше, и требуется более строгий отбор, чтобы из глубоких узлов выбирались только наиболее релевантные. Это способствует тому, что модель быстрее отсеивает нерелевантные пути, затрачивая на анализ меньше времени и токенов.

Усеченный пример работы системы представлен в Приложении.

Эксперименты

Настройка системы включала использование двух LLM: для агентов генератора и переранжировщика использовалась модель Qwen3-235B-A22B-Instruct-2507, а в качестве основы критика – DeepSeek-V3.2-Speciale[3]. Модель другой архитектуры использовалась для того, чтобы уменьшить риск появления однотипных ошибок.

[3] https://huggingface.co/deepseek-ai/DeepSeek-V3.2-Speciale

Для обоих описанных выше вариантов работы системы было проведено по три группы экспериментов, чтобы оценить воспроизводимость. В группе экспериментов с моделью в качестве верификатора гипотез были использованы три наиболее релевантных кандидата, предсказанные двумя базовыми моделями: FastText [10] и YandexGPT-5-Lite[4], обученной по методу TaxoLLaMA [7]. В первой группе экспериментов, помимо независимых попыток, также оценивались два ансамблевых подхода на основе агрегированных предсказаний всех трех попыток, ранжированных по двум принципам: по самом длинному пути (модель просмотрела большее число узлов) и самому глубокому пути (ответы упорядочены по рангу топологического поколения конечного узла).

Для оценивания первой процедуры использованы метрики MAP@1 и MAP@3, для второй – только MAP@3 (по количеству использованных стартовых узлов):

- Mean Average Precision (усредненная средняя точность), MAP по *K*:

$$\mathrm{MAP@}K = \frac{1}{N}\sum_{n=1}^{N} AP@K_n;$$

- Average Precision (средняя точность), AP по *K*:

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

- Precision (точность) по *k*:

$$\mathrm{Precision@}k = \frac{p}{k}.$$

Здесь K – 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.

РЕЗУЛЬТАТЫ

В табл. 1 представлены результаты тестирования системы в режиме исследования таксономии от корневых узлов. Наилучшим оказался метод, заключающийся в объединении трех попыток и упорядочивании итоговых предсказаний

[4] https://huggingface.co/TaxoLLMs/TaxoYandexGPT-5-Lite-8B-pretrain

по длине пути. Это ожидаемый результат, поскольку именно длина пути, или количество итераций, которые система потратила на анализ одной и той же ветки, характеризует степень уверенности в том, что исследование велось в правильном направлении. Объединение попыток с упорядочиванием по глубине привело к несколько более низкому значению метрики. Глубина – связанный с длиной пути параметр, потому что длинные пути обязательно ведут в глубокие узлы. Это объясняет небольшую разницу (всего в 0.6 п.п.) между этими результатами. Но в глубокие узлы также могли прийти и короткие пути, поскольку у одного и того же синсета может быть несколько гиперонимов, в том числе более глубоких, чем те, которые привели модель к рассмотренной ветке, поэтому результат при данном методе ранжирования в среднем оказался хуже.

Табл. 1. Результаты системы в режиме исследования от корня.

Модель и метод	MAP@1	MAP@3
Qwen/Qwen3-235B-A22B-Instruct-2507 (средний)	21.3±1.3	24.1±1.3
Qwen/Qwen3-235B-A22B-Instruct-2507 (сначала длинные)	**21.6**	**24.6**
Qwen/Qwen3-235B-A22B-Instruct-2507 (сначала глубокие)	21.5	24.0

В табл. 2 представлены результаты оценки применения двух процедур в сравнении с базовыми моделями. Эксперимент с навигацией от корня показал относительно низкий результат (MAP 24.1 п.п.). Отсюда следует, что даже при использовании продвинутой техники ReAct современная LLM испытывает сложности при ориентировании в полном пространстве графа таксономии без сильной начальной эвристики. Комбинаторная сложность и большое количество семантически близких, но неверных путей усложняют поиск верного решения.

Табл. 2. Результаты моделей, оценены по метрике MAP@3.

Модель и метод	MAP@3
FastText (базовая модель)	38.4
TaxoLLMs/TaxoYandexGPT-5-Lite-8B-pretrain (базовая модель)	54.9
Qwen/Qwen3-235B-A22B-Instruct-2507 (от корня)	24.2
Qwen/Qwen3-235B-A22B-Instruct-2507 (стартовые узлы FastText)	$47.9_{\mathbf{+9.5}}$
Qwen/Qwen3-235B-A22B-Instruct-2507 (стартовые узлы TaxoYandexGPT)	$\mathbf{56.0}_{\mathbf{+1.1}}$

В то же время использование системы для уточнения кандидатов, предложенных базовыми моделями, позволило улучшить результаты этих моделей: на 9.5 п.п. для более слабой базовой модели (FastText) и на 1.1 п.п. для сильной (TaxoYandexGPT-5-Lite). Хотя прирост является умеренным, он важен по двум причинам. Во-первых, он достигнут в zero-shot-режиме, без какого-либо дообучения. Во-вторых, он показывает, что явный, пошаговый процесс рассуждения LLM способен скорректировать и уточнить «интуитивные» предсказания сильной, но непрозрачной нейросетевой модели, что открывает перспективы для создания гибридных, более надежных систем.

Качество компонента переранжирования было отдельно верифицировано на валидационной выборке из [9] (762 понятия): для каждого понятия из выборки был извлечен список узлов, ранг которым приписывался в зависимости от близости к целевому понятию (5 – непосредственные гипронимы, 0 – узлы, связанные только через корень). Качество переанжирования без дообучения оказалось очень высоким: 99.05 по метрике nDCG, коэффициент Кендалла τ = 84.82, коэффициент Спирмена r = 91.04. Высокие показатели свидетельствуют о том, что модель обладает необходимым уровнем понимания семантики для базового различения релевантных и нерелевантных узлов, что является необходимым условием для успешной работы всего пайплайна.

ОБСУЖДЕНИЕ

Предложенный подход показал свою эффективность как метод повышения точности предсказаний других моделей, разработанных для пополнения таксономий, в том числе передовых. Дополнительным преимуществом подхода является его независимость от ресурса и моделей – для применения к новым данным потребуется подготовить только модуль взаимодействия с таксономией и при необходимости настроить гиперпараметры, в первую очередь максимальное число итераций и порог переранжирования. Еще одно существенное достоинство – это возможность отследить цепочки решений, которые привели модель к ответу, что значительно повышает интерпретируемость и позволяет выявлять типовые ошибки.

В частности, анализ рассуждений модели, которые привели к ошибкам в итоговых предсказаниях, показал, что ключевая проблема режима исследования от корня связана с размытостью названий категорий верхнего уровня. Так, в приведенном примере синсет с абстрактной семантикой «физический объект», являющийся гиперонимом 5-го порядка для слова «велосипед» в таксономии RuWordNet, показался генератору более отдаленным по смыслу от целевого слова, чем синсет с названием «изделие»:

> Узел «физический объект» (147133-N) содержит гипоним «предмет, Вещь», но нет более точного узла, связанного с транспортом или изделиями. Этот путь менее информативен, чем «изделие», поскольку не отражает производственную природу велосипеда.
>
> Возвращаюсь к узлу «изделие» (109873-N) – он остается наиболее точным. Хотя у него нет гипонимов, сам он является прямым гиперонимом для «велосипеда»: велосипед – это вид изделия.

Ожидаемый путь *постоянная сущность → физическая сущность → физический объект → предмет, вещь → приспособление, инструмент → техническое устройство* оказался слишком длинным, чтобы модель смогла проанализировать всю ветку и достигнуть наиболее релевантного узла. Из-за этого модель перешла на более короткую ветку, которая производит впечатление более подходящей на меньшей глубине: *постоянная сущность → продукт труда → изделие* (конечный узел).

Заметим, что в настоящей работе впервые на материале RuWordNet задача была поставлена в привязке к корпусному контексту. Успех методов с обучением, взятых в качестве базовых, обусловлен прежде всего тем, что в выборку из онтологии новые слова включены во всех, в том числе самых частотных, значениях, которые передовые модели успешно предсказывают. При этом подобные методы испытывают сложности с предсказанием тех слов, которые не встретились в обучающих данных. Представляется, что предложенный в работе подход способен успешно справляться с такого рода случаям прежде всего за счет умения больших языковых моделей извлекать значение даже незнакомого слова, опираясь на контекст. Однако это предположение требует проверки на дополнительных данных.

Ограничение работы всей системы в целом независимо от режима заключается в высокой длительности выполнения задачи: исследование занимает в среднем по 8 с на итерацию, что существенно дольше, чем при использовании дообученной модели.

Направления будущих исследований будут связаны с решением выявленных проблем, а также с масштабированием подхода на новые данные и ресурсы. В частности, по выходам модели планируется создание датасетов из «успешных» и «неуспешных» размышлений генератора для последующего дообучения более компактной модели, которая может оказаться более эффективной в применении к данным конкретного ресурса. Предполагается также исследовать методы сжатия информации о структуре таксономии, чтобы предотвратить преждевременно отбрасывание релевантных путей.

ЗАКЛЮЧЕНИЕ

Предложена архитектура для пополнения таксономий на основе подхода ReAct, работающая в режиме zero-shot без дообучения. Эксперименты на материале RuWordNet показали, что автономная навигация от корневых узлов сталкивается с ограничениями, связанными с семантической неоднозначностью категорий верхнего уровня и высокой комбинаторной сложностью (наилучший MAP@3 – 24.6%). При этом использование системы в режиме верификатора позволило улучшить предсказания базовых моделей: прирост MAP@3 составил 9.5 п.п. для FastText и 1,1 п.п. для TaxoYandexGPT-5-Lite.

Ключевыми преимуществами подхода являются универсальность (отсутствие необходимости дообучения) и интерпретируемость за счет явных цепочек рассуждений. Основные ограничения связаны с высокой вычислительной стоимостью и чувствительностью к неоднозначности названий узлов. Перспективы дальнейших исследований включают создание датасетов рассуждений для дообучения компактных моделей, разработку методов сжатия структурной информации и масштабирование на новые ресурсы и предметные области.

Благодарности

Исследование выполнено за счет гранта Российского научного фонда № 25-11-00191[5]. Работа выполнена с использованием суперкомпьютера «МГУ-270» МГУ имени М. В. Ломоносова.

СПИСОК ЛИТЕРАТУРЫ

1. *Liu Z. et al.* TEMP: Taxonomy expansion with dynamic margin loss through taxonomy-paths // Proceedings of the 2021 conference on empirical methods in natural language processing. 2021. P. 3854–3863. https://doi.org/10.18653/v1/2021.emnlp-main.313

2. *Xu H. et al.* Compress and mix: Advancing efficient taxonomy completion with large language models // Proceedings of the ACM on Web Conference 2025. 2025. P. 4239–4249. https://doi.org/10.1145/3696410.3714690

3. *Niu Y. et al.* Contrastive representation learning for self-supervised taxonomy completion // IJCAI. 2024. Vol. 8. P. 6442–6450. https://doi.org/10.24963/ijcai.2024/712

4. *Arous I., Dolamic L., Cudré-Mauroux P.* Taxocomplete: Self-supervised taxonomy completion leveraging position-enhanced semantic matching // Proceedings of the ACM Web Conference 2023. 2023. P. 2509–2518. https://doi.org/10.1145/3543507.3583342

5. *Xu H. et al.* Tacoprompt: A collaborative multi-task prompt learning method for self-supervised taxonomy completion //Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. 2023. P. 15804–15817. https://doi.org/10.18653/v1/2023.emnlp-main.979

6. *D'Amico S. et al.* Taxonomy Expansion through Collaborative LLM Mapping //Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing. 2025. P. 1961–1968. https://doi.org/10.1145/3672608.3707906

7. *Moskvoretskii V. et al.* Large Language Models for Creation, Enrichment and Evaluation of Taxonomic Graphs // Semantic Web: Interoperability, Usability, Applicability. 2026. Vol. 17, No. 1. https://doi.org/10.1177/22104968251404186

8. *Yao S. et al.* ReAct: Synergizing Reasoning and Acting in Language Models

[5] https://rscf.ru/project/25-11-00191/

// International Conference on Learning Representations (ICLR). 2023. URL: https://openreview.net/forum?id=WE_vluYUL-X (Accessed: 31.03.2026).

9. *Nikishina I. et al.* Russe'2020: Findings of the first taxonomy enrichment task for the Russian language // Computational Linguistics and Intellectual Technologies. 2020. P. 579–595. URL: https://api.semanticscholar.org/CorpusID:218862855 (Accessed: 31.03.2026).

10. *Nikishina I. et al.* Studying taxonomy enrichment on diachronic WordNet versions // Proceedings of the 28th International Conference on Computational Linguistics. 2020. P. 3095–3106. https://doi.org/10.18653/v1/2020.coling-main.276

11. *Miller G. A. et al.* Introduction to WordNet: An on-line lexical database // International journal of lexicography. 1990. Vol. 3, No. 4. P. 235–244.

12. *Loukachevitch N.V. et al.* Creating Russian wordnet by conversion // Computational Linguistics and Intellectual Technologies. 2016. P. 405–415. URL: https://api.semanticscholar.org/CorpusID:63871134 (Accessed: 31.03.2026).

13. *Nikishina I. et al.* Taxonomy enrichment with text and graph vector representations // Semantic Web. 2022. Vol. 13, No. 3. P. 441–475. https://doi.org/10.1007/978-3-031-16500-9_2

14. *Tikhomirov M., Loukachevitch N.* Meta-embeddings in taxonomy enrichment task // Computational Linguistics and Intellectual Technologies: papers from the Annual conference “Dialogue. 2021. URL: https://dialogue-conf.org/media/5549/tikhomirovmmpluslouka-chevitchnv091.pdf (Accessed: 31.03.2026).

ПРИЛОЖЕНИЕ

Пример текста из тестового корпуса с выделением целевого слова «тезаурус»:

Традиционно для этого разрабатываются иерархические **<target>тезаурусы</ target>**, на верхних уровнях которых находятся ключевые термины предметной области, уточняемые на более низких уровнях.

Размышление модели на шаге 1 (сценарий движения от корня):

Анализируемое понятие – **тезаурусы** – используется в контексте организации знаний в предметной области, где на верхних уровнях находятся ключевые термины, а на более низких – их уточнения. Это указывает на иерархическую структуру, характерную для лингвистических или информационных систем. Ключевой аспект – **иерархический тезаурус** как средство организации терминов. Семантически это – система классификации, основанная на отношениях «род-вид», «целое-часть», «ассоциативные связи» и т.п., применяемая в библиотечном деле, информационном поиске, лингвистике.

В RuWordNet подобные понятия могут находиться в ветке знаний, лингвистики, информационных структур или систем классификации. Начну с корневых узлов, чтобы определить наиболее подходящую ветвь.

Вызов функции: get_hyponyms

Аргументы: {"node_id": null}

Узел: root

Размышление модели на шаге 2:

Результат функции:

Найдено гипонимов: 8

1. новшество `132964-N`

Определение: что-либо новое, только что вошедшее в обиход

Слова: новация; новшество

Гипонимы (5): изобретение; нововведение; изыск; техническое новшество; инновация (модернизация)

2. роль, положение, место `106613-N`

Определение: положение кого-либо в обществе, роль, отведенная кому-либо в обществе, в какой-либо деятельности

Слова: место; роль; положение

Гипонимы (26): рабочее место; должность; имущество, собственность; исходный пункт; способ, средство; предмет, объект деятельности; очередь, место в очереди; предвестник; статус; разрушитель *(+16 еще)*

3. вариант, разновидность `106846-N`

Определение: разновидность, видоизменение чего-либо

Слова: вариант; вариация; разновидность

Гипонимы (17): торговая марка; версия; марка, модель, образец; сорт продукции; лекарственная форма; жанр искусства; редакция текста; порода животных; порода деревьев; артикул *(+7 еще)*

4. отношение между сущностями `106646-N`

Слова: отношение между сущность; отношение; связь между сущность; связь

Гипонимы (35): последствие, следствие; ресурс, источник; использовать, употреблять; сходство; ассоциация (связь); обладать свойством; участие, участвовать; краснеть (виднеться); влиять, воздействовать; касаться, дотрагиваться *(+25 еще)*

5. постоянная сущность `153782-N`

Слова: постоянный сущность

Гипонимы (12): средства производства; природа; средства массовой информации; обстоятельства, обстановка, условия; продукт труда; действительность, объективный мир; изображение (результат); запись (то, что записано); наследие прошлого; субъект деятельности *(+2 еще)*

6. совокупность, целое `106508-N`

Определение: совокупность каких-либо объектов, образующих единое целое

Слова: целое; совокупность; состав

Гипонимы (36): партия товаров; комплекс сооружений; тематика; множество, большое количество; набор (совокупность предметов); состав (продукт соединения); объединение (то, что объединено); сочетание, комбинация; сборник (книга); парк технических средств *(+26 еще)*

7. происходящая сущность `153471-N`

Слова: происходящий сущность; длящийся сущность

Гипонимы (25): появиться, возникнуть, начаться; исполниться, осуществиться, претвориться; продолжаться, длиться; повториться, произойти еще раз; совпадение по времени; бушевать (происходить с необычайной

силой); сопровождаться (происходить одновременно); сопровождать (производить одновременно); посчастливиться (удаться); кипеть оживленно *(+15 еще)*

8. составная часть `106509-N`

Определение: одна из составных частей какого-либо целого

Слова: составной часть; часть

Гипонимы (47): комплектующее изделие; полоса на поверхности, в пространстве; излишек, избыток; кусок (отдельная часть); участок, отрезок, кусок (часть расстояния); слой (плоская часть); доля (часть); выступ, выступающая часть; середина, средняя часть; компонент (составная часть) *(+37 еще)*

Анализ

Понятие **«тезаурусы»** относится к структурированным системам организации знаний или терминов. Наиболее перспективные кандидаты среди корневых узлов – **«совокупность, целое»** (поскольку тезаурус – это совокупность терминов) и **«отношение между сущностями»** (так как тезаурусы определяют связи между понятиями). Также возможно, что тезаурус – это **«составная часть»** более крупной системы классификации, но это менее вероятно.

Наиболее перспективным представляется узел **«совокупность, целое»**, так как тезаурус – это именно **целостная система, совокупность терминов с определенными связями**. Исследую его гипонимы.

Вызов функции: get_hyponyms

Аргументы: {"node_id": "106508-N"}

Узел: совокупность, целое

Размышление модели на шаге 5 (шаги 3 и 4 опущены в целях экономии):

Финальный результат:

Гипонимы узла «словарь» – «энциклопедический словарь» и «толковый словарь» – не включают «тезаурус», но это ожидаемо: тезаурус – это отдельный тип словаря, организованный по семантическим связям, а не по алфавиту или толкованию значений.

Учитывая, что:

- «тезаурус» – это **специализированный тип словаря**,
- В RuWordNet нет отдельного узла для «тезаурус»,
- Но есть узел **словарь (3267-N)**, который охватывает различные типы

словарей, делаю вывод, что **«тезаурус» является синонимом или разновидностью «словаря» в широком смысле**, но в текущей таксономии он не выделен как отдельный гипоним.

Альтернативные ветки (например, «система» или «банк данных») менее точны:

- «Система» – слишком широко.
- «База данных» – близко, но тезаурус – не просто база, а семантически структурированная система терминов.

Следовательно, наиболее точное место для «тезаурусы» – как **гипонима «словаря»**.

Окончательная проверка:

1. Понятие понято – да, тезаурус как иерархический словарь терминов.
2. Исследованы релевантные ветки – да, «совокупность», «собрание», «словарь».
3. Узел «словарь» – корректный гипероним – да.
4. Более конкретного узла нет – да, среди гипонимов «словаря» нет «тезауруса», но он подходит как отсутствующий гипоним.
5. Уверен – да.

Ответ: hyponym of 3267-N (словарь)

Релевантные кандидаты в датасете: книга (издание) [3790-N], справочная литература [3619-N], собрание (совокупность собранного) [121692-N], словарь [3267-N]

TAXONOMY EXPLORATION WITH LARGE LANGUAGE MODEL REASONING AND TOOL CALLING

F. A. Sadkovskii[1] [0009-0006-4502-9466], M. M. Tikhomirov[2] [0000-0001-7209-9335],

N. V. Loukachevitch[3] [0000-0002-1883-4121]

[1–3]*Lomonosov Moscow State University, Moscow, Russia*

[1]feudor987@mail.ru, [2]tikhomirov.mm@gmail.com, [3]louk_nat@mail.ru

Abstract

The paper addresses the task of taxonomy expansion – hierarchical structures for organizing concepts. We propose an architecture based on the ReAct (Reasoning + Acting) approach that enables taxonomy expansion in a zero-shot setting without fine-tuning large language models. The system is implemented in two scenarios: autonomous navigation from root nodes and verification of hypotheses generated by other models. Experiments on the diachronic RuWordNet dataset show that direct navigation from the root faces limitations due to graph complexity (MAP@3 = 24.6%). However, using the system as a verifier improves the performance of baseline models: MAP@3 gains of 9.5 pp for FastText and 1.1 pp for TaxoYandexGPT-5-Lite. The key advantages of the approach are its universality, the absence of fine-tuning requirements, and interpretability through explicit reasoning chains.

Keywords*: taxonomy, taxonomy completion, hypernym extraction, LLM, chain-of-thought, function calling, RuWordNet, reranking, computational semantics.*

REFERENCES

1. *Liu Z. et al.* TEMP: Taxonomy expansion with dynamic margin loss through taxonomy-paths // Proceedings of the 2021 conference on empirical methods in natural language processing. 2021. P. 3854–3863. https://doi.org/10.18653/v1/2021.emnlp-main.313

2. *Xu H. et al.* Compress and mix: Advancing efficient taxonomy completion with large language models // Proceedings of the ACM on Web Conference 2025. 2025. P. 4239–4249. https://doi.org/10.1145/3696410.3714690

3. *Niu Y. et al.* Contrastive representation learning for self-supervised taxonomy completion // IJCAI. 2024. Vol. 8. P. 6442–6450.

https://doi.org/10.24963/ijcai.2024/712

4. *Arous I., Dolamic L., Cudré-Mauroux P.* Taxocomplete: Self-supervised taxonomy completion leveraging position-enhanced semantic matching //Proceedings of the ACM Web Conference 2023. 2023. P. 2509–2518. https://doi.org/10.1145/3543507.3583342

5. *Xu H. et al.* Tacoprompt: A collaborative multi-task prompt learning method for self-supervised taxonomy completion //Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. 2023. P. 15804–15817. https://doi.org/10.18653/v1/2023.emnlp-main.979

6. *D'Amico S. et al.* Taxonomy Expansion through Collaborative LLM Mapping //Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing. 2025. P. 1961–1968. https://doi.org/10.1145/3672608.3707906

7. *Moskvoretskii V. et al.* Large Language Models for Creation, Enrichment and Evaluation of Taxonomic Graphs // Semantic Web: Interoperability, Usability, Applicability. 2026. Vol. 17, No. 1. https://doi.org/10.1177/22104968251404186

8. *Yao S. et al.* ReAct: Synergizing Reasoning and Acting in Language Models // International Conference on Learning Representations (ICLR). 2023. URL: https://openreview.net/forum?id=WE_vluYUL-X (Accessed: 31.03.2026).

9. *Nikishina I. et al.* Russe'2020: Findings of the first taxonomy enrichment task for the Russian language // Computational Linguistics and Intellectual Technologies. 2020. P. 579–595. URL: https://api.semanticscholar.org/CorpusID:218862855 (Accessed: 31.03.2026).

10. *Nikishina I. et al.* Studying taxonomy enrichment on diachronic WordNet versions // Proceedings of the 28th International Conference on Computational Linguistics. 2020. P. 3095–3106. https://doi.org/10.18653/v1/2020.coling-main.276

11. *Miller G.A. et al.* Introduction to WordNet: An on-line lexical database // International journal of lexicography. 1990. Vol. 3, No. 4. P. 235–244.

12. *Loukachevitch N.V. et al.* Creating Russian wordnet by conversion // Computational Linguistics and Intellectual Technologies. 2016. P. 405–415. URL: https://api.semanticscholar.org/CorpusID:63871134 (Accessed: 31.03.2026).

13. *Nikishina I. et al.* Taxonomy enrichment with text and graph vector representations // Semantic Web. 2022. Vol. 13, No. 3. P. 441–475. https://doi.org/10.1007/978-3-031-16500-9_2

14. *Tikhomirov M., Loukachevitch N.* Meta-embeddings in taxonomy enrichment task // Computational Linguistics and Intellectual Technologies: papers from the Annual conference “Dialogue. 2021. URL: https://dialogue-conf.org/media/5549/tikhomirovmmplusloukachevitchnv091.pdf (Accessed: 31.03.2026).

СВЕДЕНИЯ ОБ АВТОРАХ

САДКОВСКИЙ Федор Алексеевич. — аспирант кафедры теоретической и прикладной лингвистики филологического факультета МГУ им. М. В. Ломоносова, г. Москва, Россия.

Fedor A. SADKOVSKII is a PhD student of the Department of Theoretical and Applied Linguistics, Faculty of Philology, Lomonosov Moscow State University, Moscow, Russia.

email: feudor987@mail.ru

ORCID: 0009-0006-4502-9466

ТИХОМИРОВ Михаил Михайлович — кандидат физ. мат. наук, старший научный сотрудник НИВЦ МГУ, руководитель проекта Ruadapt по адаптации больших языковых моделей на русский язык.

Mikhail M. TIKHOMIROV – PhD, senior researcher at the Lomonosov Moscow State University Research Computing Center and head of the Ruadapt project on adapting large language models to Russian.

email: tikhomirov.mm@gmail.com

ORCID: 0000-0001-7209-9335

ЛУКАШЕВИЧ Наталья Валентиновна — доктор технических наук, ведущий научный сотрудник НИВЦ МГУ имени М. В. Ломоносова, заведующая кафедрой факультета ВМК, профессор ВМК, филологического факультета МГУ. Область научных интересов: представление знаний, автоматическая обработка текстов, большие языковые модели

Natalia V. LOUKACHEVITCH is a Doctor of Technical Sciences, leading researcher at the Research Computing Center of Lomonosovv Moscow State University, head of Department at the Faculty of Computational Mathematics and Cybernetics, professor at the Faculty of Computational Mathematics and Cybernetics, Faculty of Philology at Moscow State University. Research interests: knowledge representation, automated text processing, large language models.

email: louk_nat@mail.ru

ORCID: 0000-0002-1883-4121

Материал поступил в редакцию 6 апреля 2026 года