

УДК 004.81

# ОНТОЛОГИЧЕСКИЙ ПОДХОД К ОЦЕНКЕ ГРАФОВ ЗНАНИЙ В ДОМЕННОЙ ОБЛАСТИ МАШИНОСТРОИТЕЛЬНЫХ СИСТЕМ ПОЛНОГО ЖИЗНЕННОГО ЦИКЛА

В. В. Гладышев<sup>[0009-0003-8900-3469]</sup>

*Московский физико-технический институт; Центр «Пуск», г. Долгопрудный,  
Россия*

gladyshev.vv@phystech.edu

## ***Аннотация***

Работа посвящена проблеме применения онтологического подхода при построении датасета для оценки и сравнения систем обогащения контекста большой языковой модели с использованием графов знаний в доменной области машиностроительных систем полного жизненного цикла. В доменной области сложно получить необходимое количество текстовых данных с формальной логической структурой для формирования оценочного набора без использования сгенерированных синтетических данных. Для исключения внесения искажений и галлюцинаций при формировании оценочного набора предложено оригинальное решение проблемы дефицита данных за счет извлечения онтологии непосредственно из файлов изделий и сборок, соответствующих стандарту Standard for Exchange of Product model data что потенциально позволяет использовать все данные об изделиях как источник для масштабирования оценочных данных. Целью работы стали создание датасета структурированных текстовых данных в доменной области машиностроительных систем полного жизненного цикла, разработка методики оценки и реализация конвейеров обогащения контекста большой языковой модели с применением и без применения графов знаний для анализа вклада систем с извлечением структуры данных в качество генерируемых ответов. Предложен новый источник оценочных данных, разработана новая методика формирования текстовых оценочных данных с сохранением логической

структуры, реализован конвейер для использования сгенерированных оценочных данных. Получены результаты оценки, подтверждающие положительный вклад систем с извлечением структурированных данных в качество генерируемых ответов в доменной области машиностроительных систем полного жизненного цикла.

**Ключевые слова:** *онтология, датасет, система полного жизненного цикла СПЖЦ/PLM, система автоматизированного проектирования САПР/CAD, большая языковая модель БЯМ/LLM, генерация с обогащением контекста RAG, GraphRAG, Standard for Exchange of Product model data – STEP.*

## **ВВЕДЕНИЕ**

Для построения информационных систем, обеспечивающих использование доменно-ориентированных данных, предназначенных для локального размещения, часто используется подход обогащения контекста (Retrieval-Augmented Generation, RAG) большой языковой модели (БЯМ, Large Language Model, LLM). Стандартный подход обогащения контекста имеет ряд принципиальных ограничений, для преодоления которых применяется генерация графов знаний при помощи БЯМ GraphRAG [1]. Ключевая проблема заключается в том, что основная масса применяемых средств (датасетов, бенчмарков, метрик) оценки качества подхода обогащения контекста БЯМ не предназначена для оценки GraphRAG, так как не учитывает преимущества структурирования данных, а нацелена на оценку качества извлечения отдельных фактов [2]. Особенно остро данная проблема выражена при построении систем в относительно узких доменных областях и, если для медицинского или юридического направлений существует ограниченное количество доменно-ориентированных данных, для домена машиностроительных систем полного жизненного цикла (Product Lifecycle Management – PLM) таких данных найти не удалось. Соответственно, невозможно обосновать (или отвергнуть) необходимость применения более ресурсоемкого GraphRAG-подхода, а также достоверно оценить влияние изменений, вносимых в GraphRAG-систему, на качество ответов [3].

## **Применение GraphRAG-систем**

Традиционный подход RAG решает проблему актуальности знаний БЯМ, извлекая релевантные фрагменты текста (чанки) из внешнего хранилища и передавая их в БЯМ для генерации ответа. Однако этот подход имеет существенный недостаток: он оперирует с изолированными текстовыми фрагментами, не позволяя выполнять обобщение по всему тексту [4]. GraphRAG представляет собой следующий эволюционный шаг, где извлечение происходит из графа знаний [5]. В этом подходе неструктурированный текст сначала преобразуется в структурированное графовое представление: сущности из текста становятся узлами графа, а отношения между ними — ребрами, выполняется многоуровневая кластеризация и обобщение для каждой группы [6]. Это позволяет системе строить ответ на связанных данных, находя цепочки связей между сущностями, которые не упоминаются в одном и том же предложении или документе. Эффективность GraphRAG была продемонстрирована в ряде исследований. В работе [7] сравнивается GraphRAG, построенный на основе экспертно-разработанной онтологической схемы, с базовыми векторными методами RAG. GraphRAG достиг 90%-ной точности в ответах на 20 сложных вопросов в узкой предметной области, в то время как базовый RAG – лишь 60%.

## **Ключевые стандарты данных в системах управления жизненным циклом продукции**

Современные системы управления жизненным циклом продукции (СПЖЦ [8]; Product Lifecycle Management, PLM) в машиностроительной отрасли представляют собой сложные информационные системы, интегрирующие множество модулей: САПР/CAD, CAE, CAM, CAPP, PDM, ERP, MES и др. Данные, накапливаемые в этих системах, имеют специфическую структуру (геометрия деталей, деревья истории построения, составы сборок) и различный формат для каждого модуля. Для решения задачи формирования структурированных оценочных данных для оценки GraphRAG применяют данные в формате ISO 10303 [9] (STandard for Exchange of Product model data, STEP), стандартном для домена машиностроительных PLM-систем. Стандарт ISO 10303 разделяется на

700 базовых стандартов, в его частях 11–18 и 21 описаны язык определения схем данных EXPRESS и STEP-формат. В стандарте также описаны прикладные протоколы (Application Protocols, AP) для представления специфических данных (AP238 для CAM, AP203 и 242 для САПР/CAD, AP 214 для автомобильной промышленности). Национальный институт стандартов и технологий (The National Institute of Standards and Technology, NIST) предоставляет различные инструменты для просмотра и анализа файлов STEP. Схема данных в EXPRESS-формате и соответствующие ей данные конкретной детали или сборки в STEP-формате могут быть преобразованы в виде онтологии в OWL-формате [10]. В NIST разработаны специальные программные средства открытого доступа: OntoSTEP, STEPCode, STP2OWL.

## 1. МЕТОДЫ

### 1.1. Создание датасета для оценки GraphRAG в сравнении с тривиальным RAG

Для формирования оценочного набора корректно структурированных данных нами реализована методика, включающая следующие этапы.

#### Этап 1. Формирование онтологии из инженерных данных STEP

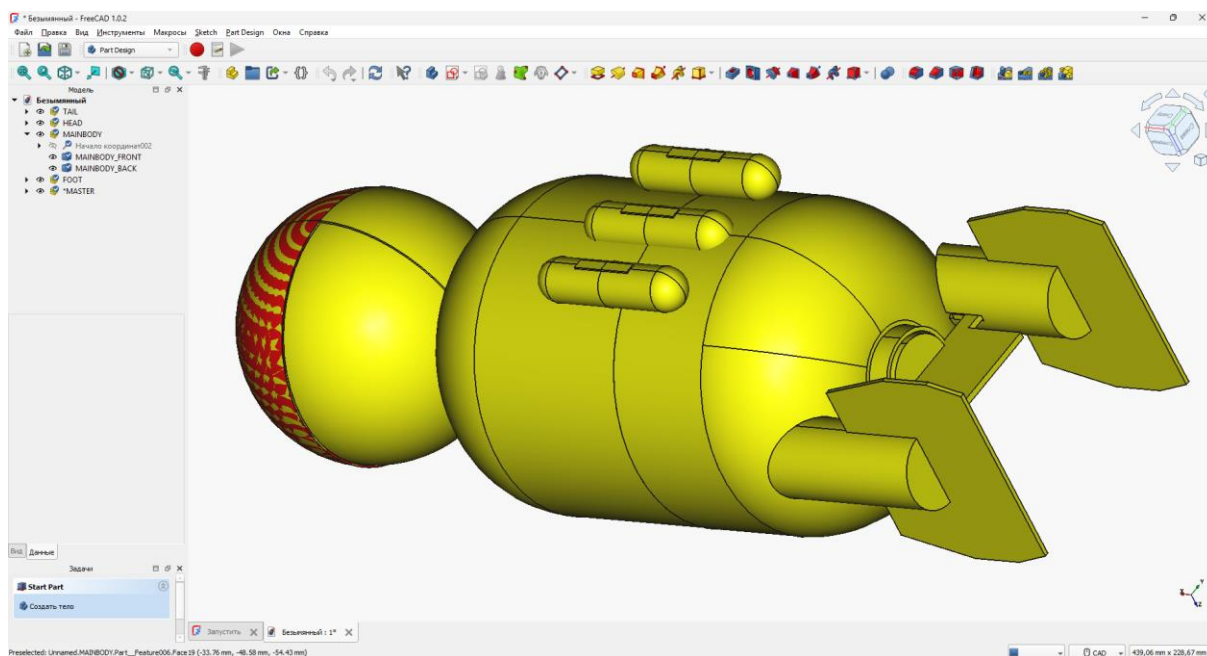


Рис. 1. Сборка s1-c5-214.stp в формате AP214 при просмотре во FreeCAD.

В качестве основы для генерации датасета была использована онтология, автоматически сгенерированная с помощью инструментов Stepcode и STP2OWL. Для извлечения онтологий была использована сборка s1-c5-214 в формате AP214 (STEP-файл с описанием геометрии и сопряжения элементов). Входными данными были файл сборки s1-c5-214.stp в STEP-формате (рис. 1) и логическая схема данных стандарта AP214E3\_2010.exr AP214 в EXPRESS-формате [11]. Преобразование выполнялось при помощи программных средств открытого доступа Stepcode и STP2OWL.

На выходе получены онтологии:

- структуры данных s1-c5-214\_Schema\_DL.owl (Schema – TBox), более 13000 аксиом;
- экземпляра сборки s1-c5-214\_Instances\_DL.owl (Instances – ABox), около 1000 экз. (рис. 2).

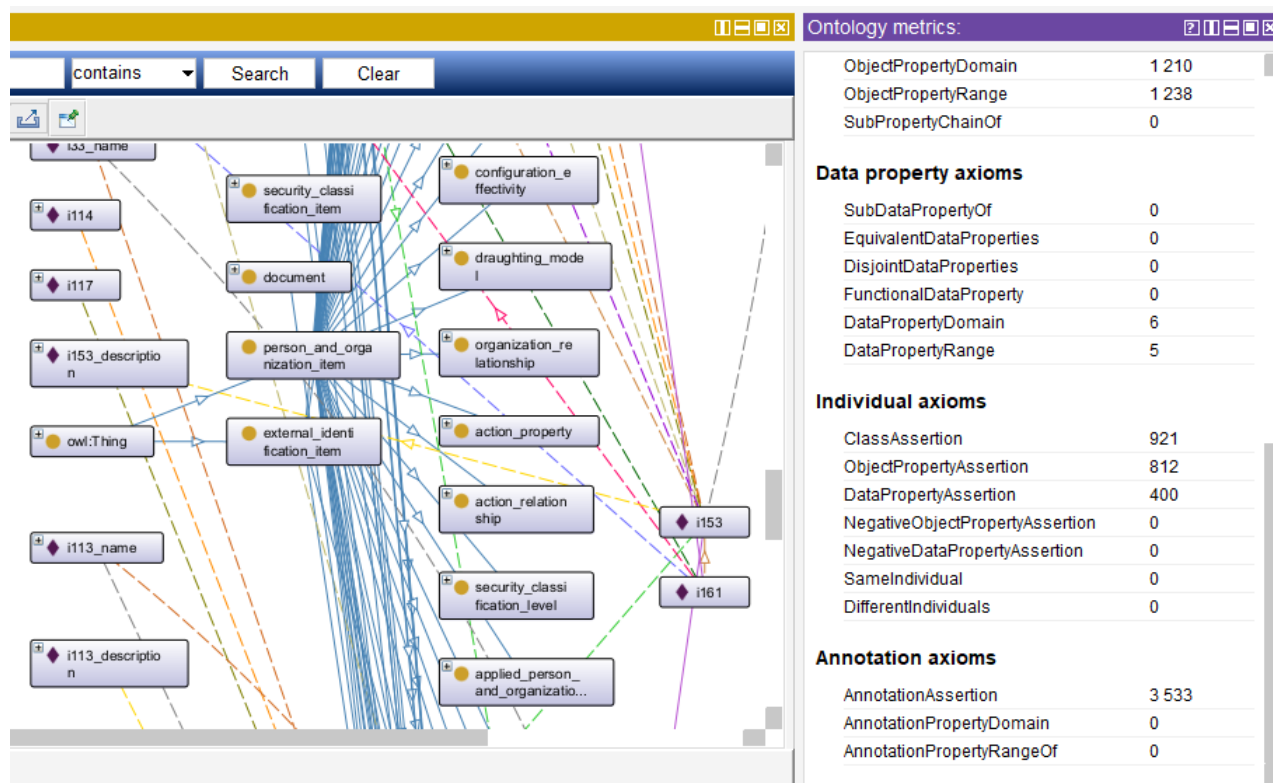


Рис. 2. Связи классов и статистика аксиом экземпляра сборки.

## **Этап 2. Вербализация онтологии в текстовый датасет**

Системы на базе RAG и GraphRAG предназначены для работы с текстом [12]. Поэтому для применения полученной онтологии для построения графа знаний GraphRAG необходимо выполнить преобразование (вербализацию/ verbalization) онтологии в текст. Для сохранения структуры отношений и реализации однозначного преобразования в качестве целевого языка был выбран Attempto Controlled English (ACE) [13], разработанный в Цюрихском университете. Это особое подмножество английского языка, предназначенное для однозначного представления знаний и запросов. Для преобразования использовался OWL Ontology Verbalizer [14]. После преобразования онтологии структуры данных (Schema) получен текстовый файл с аксиомами в ACE-формате. После преобразования онтологии экземпляра сборки (Instances) получен файл с текстами аксиом.

Выбор Attempto Controlled English в качестве целевого языка обусловлен тем, что:

- после преобразования получается интерпретируемый текст – понятный как при чтении человеком, так и корректно интерпретируемый БЯМ [15];
- преобразование сохраняет структурные связи и является формализованным и детерминированным, что исключает вероятностную составляющую при преобразовании с помощью БЯМ;
- преобразование выполняется автоматически и не требует сложной обработки/подготовки;
- преобразование потенциально обратимое.

### **1.2. Возможные ограничения при формировании датасета**

Предлагаемая методика потенциально позволяет произвольно увеличивать размер целевого набора данных при условии разрешения возможного конфликта извлекаемых имен объектов.

Построение датасета на основании онтологий, извлекаемых из схем данных в EXPRESS-формате и STEP-файлов с последующей вербализацией в ACE, имеет одно существенное ограничение в части масштабирования: неполная

поддержка предлагаемого NIST-стандарта на практике. Извлечение ABox выполняется успешно не для всех файлов STEP-формата, что может существенно ограничить базу для формирования целевого набора данных. Кроме того, степень соответствия стандарту в реализациях различных вендоров PLM-систем требует дополнительного исследования.

### **Этап 3. Анализ интерпретируемости ACE-формата и формирование выборки аксиом**

Для анализа интерпретируемости большими языковыми моделями полученного текста в ACE-формате был выполнен ряд экспериментов с увеличением сложности.

#### **1.3. Проверка способности БЯМ выполнять операции с ACE-текстом**

Для проверки способности БЯМ интерпретировать ACE-текст – извлекать аксиомы, логически связанные с некоторым объектом, форматировать ответ в ACE-формате – было подготовлено тестовое задание. Произвольным образом был выбран объект с именем i101 и выполнены поиск и отбор аксиом (в отдельный файл), связанных как с самим объектом, так и с его свойствами. Полученная выборка включает 40 аксиом, что позволяет интерпретировать выводы модели и визуально проанализировать их качество.

Данная проверка имитирует ситуацию в информационной системе на базе обогащения контекста, когда фазы извлечения и обогащения сформировали качественный (релевантный и полный) контекст.

Проверка выполнялась на трех БЯМ: Anthropic Claude Sonnet 4.5, Qwen 3 Max и OpenAI GPT-5-mini. Первые две модели выбраны, так как используют подход рассуждения и демонстрируют актуальные значения качества в бенчмарках. Модель GPT-5-mini выбрана, поскольку имеет оптимальное соотношение цены и качества и в дальнейшем использовалась для построения графа знаний в библиотеке Microsoft GraphRAG, что потребовало нескольких сотен обращений к модели.

Было выполнено по два запроса к каждой модели. Промпты включали отобранный набор аксиом:

Промпт 1 – команда отобразить набор аксиом, логически связанных с объектом i101: Answer with Attempto Controlled English. Show the axioms logically related to i101;

Промпт 2 – команда сделать логические выводы относительно объекта i101 на основании представленных аксиом: Answer with Attempto Controlled English. Draw logical conclusions based on the axioms regarding i101.

Проведенные эксперименты продемонстрировали способность БЯМ интерпретировать текст в ACE-формате; извлекать контекстно связанные аксиомы; делать отдельные заключения, исходя из аксиом; формировать ответы с использованием аксиом. Соответственно, датасет, формируемый из ACE-аксиом, применим для оценки информационных систем на базе RAG и GraphRAG. Включение аксиом в ответы БЯМ позволяет реализовать систему оценки с опорой на формальные признаки, выраженные в численном виде [16].

#### **1.4. Формирование выборки аксиом для оценки GraphRAG и тривиального RAG**

Специфика реализации фазы извлечения (retrieve) RAG-конвейера заключается в извлечении ограниченного набора документов или чанков для формирования контекста при генерации ответа при помощи БЯМ [17]. В реальных документах информация о каком-либо объекте часто распределена по различным разделам и подразделам. Например, в документации на модуль САПР/CAD описано несколько различных средств построения сечений и также сечения используются как одна из операций при различных построениях [18]. Различные виды сечений и их упоминания распределены по многим разделам документации, что затрудняет формирование контекста при использовании тривиального RAG-подхода, в котором ограничено количество извлекаемых чанков. Соответственно, при построении выборки аксиом выполняется перемешивание для имитации распределения информации по набору документов. Кроме того, было выполнено разделение полученной выборки аксиом на 66 отдельных текстовых документа. Полученный набор файлов с аксиомами использовался для построения библиотекой Microsoft GraphRAG конвейера RAG и соответствующего графа.

Для оценки конвейеров был сформирован набор тестовых вопросов для каждого из 198 объектов ( $i1 - i198$ ). Тексты вопросов одинаковые: необходимо описать связи объекта с другими объектами. Для каждого вопроса при помощи БЯМ был сгенерирован эталонный ответ с использованием гарантированного полного и достоверного набора аксиом (извлечены поиском по полному набору).

## **2. ЭКСПЕРИМЕНТЫ**

### **2.1. Построение конвейера RAG**

Для сравнения с GraphRAG был реализован классический конвейер RAG. В этом конвейере каждый документ разделяется на чанки. Для каждого чанка векторное представление формируется и затем сохраняется в векторную базу данных. При получении запроса пользователя для него также формируется векторное представление и выполняется ранжирование векторов чанков на основании косинусного расстояния. Тексты пяти наиболее близких чанков передаются БЯМ в качестве контекста запроса в фазе генерации.

**Оценка с использованием метрик из библиотеки RAGAS.** Для оценки конвейера RAG брались специальные метрики из библиотеки RAGAS [19]. Метрики для оценки конвейера RAG используют подход LLM-as-a-Judge (LLM-based evaluation) — подход в обработке естественного языка, при котором БЯМ используются как автоматические оценщики текстов или других выходных данных моделей. Такой подход требует большого количества обращений к БЯМ и является ресурсоемким, поэтому был выполнен отбор метрик по критерию ресурсоемкости и длительности вычисления на части оценочного набора вопросов. В табл. 1 приведены результаты оценки времени расчета метрик на фрагменте данных.

Табл. 1. Сравнение длительности вычисления.

Метрика	Среднее время	Общее время	Доля, %
Context Precision	58.06	290.3	30.9
Answer Relevancy	35.40	177.0	18.9
Faithfulness	29.55	147.8	15.8
Context Recall	17.19	86.0	9.2
Context Relevance	16.90	84.5	9.0
Response Groundedness	15.62	78.1	8.3
Answer Accuracy	14.87	74.3	7.9

Были отобраны метрики, оптимальные по соотношению значимости и ресурсоемкости.

Answer Relevancy (релевантность ответа) – насколько сгенерированный ответ соответствует исходному вопросу по смыслу, без оценки фактической точности. Штрафует ответы, неполные или содержащие избыточную информацию. Входные данные: question (вопрос), answer (сгенерированный ответ).

Context Recall (полнота контекста) – насколько контекст, извлеченный поисковой системой, содержит всю информацию, необходимую для ответа на вопрос. Метрика требует наличия эталонного (ground truth) ответа. Входные данные: question (вопрос), retrieved context (извлеченный контекст), ground truth answer (эталонный ответ).

Context Relevance (релевантность контекста) – измеряет релевантность вопросу выбранного контекста, помогая улучшить выбор контекста для повышения точности ответа. Входные данные: question (вопрос), retrieved context (извлеченный контекст).

Answer Accuracy (точность ответа) – измеряет соответствие между ответом модели и эталонным значением истинности для данного вопроса. Это делается с помощью двух отдельных запросов “LLM-as-a-Judge”, каждый из которых возвращает оценку (0, 2 или 4). Метрика преобразует эти оценки в шкалу [0,1], а затем вычисляет среднее значение двух оценок, полученных от судей. Более

высокие оценки указывают на то, что ответ модели полностью соответствует эталону. Входные данные: question (вопрос), answer (ответ).

Значения метрик конвейера RAG на наборе из 198 вопросов: Answer Relevancy – 0.359, Context Recall – 0.116, Context Relevance – 0.551, Answer Accuracy – 0.367

## 2.2. Построение графа знаний с помощью библиотеки GraphRAG

Для проверки применимости полученных данных для построения графов знаний был построен граф знаний при помощи библиотеки Microsoft GraphRAG [1]: по полному набору аксиом об экземпляре сборки и сокращенному набору контекста объекта i101. Для извлечения графа использовалась локальная модель Google Gemma 3 27B. Полный набор аксиом сборки – s1-c5-214. Граф включает 523 узла – извлеченные сущности и 539 связей (рис. 3).

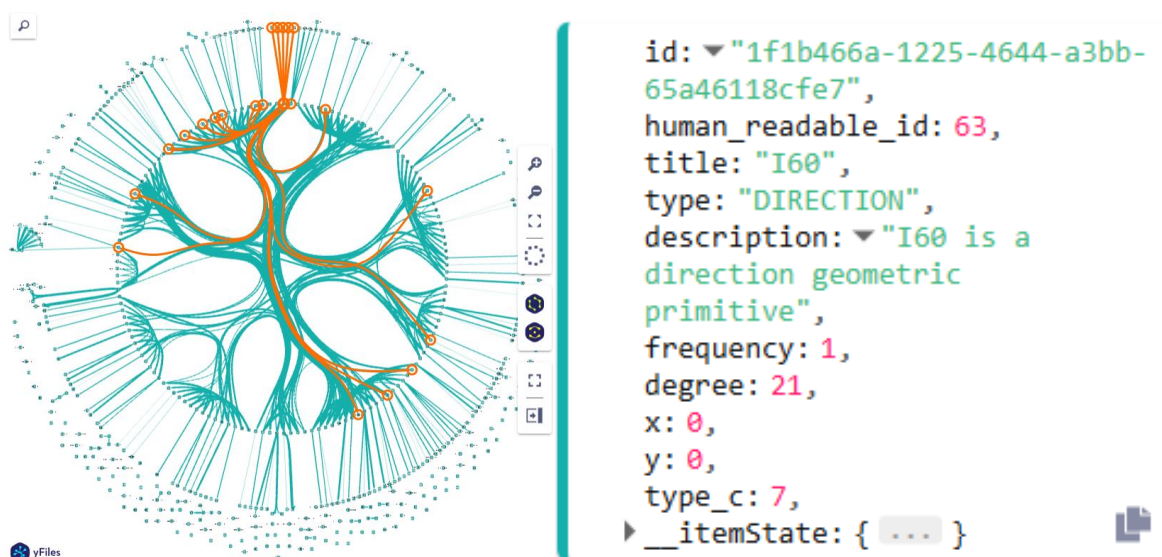


Рис. 3. Граф сборки.

Значения метрик конвейера GraphRAG на наборе из 198 вопросов: Answer Relevancy – 0.483, Context Recall – 0.727, Context Relevance – 0.975, Answer Accuracy – 0.486.

Необходимо сделать уточнение в части алгоритма формирования контекста при расчете метрик Context Recall и Context Relevance. Эти метрики используют в качестве входных данных извлеченные контексты и нацелены на

оценку фазы извлечения конвейера RAG. Однако при формировании ответа GraphRAG фаза извлечения радикально отличается от конвейера RAG и предусматривает формирование выборки релевантных сущностей, связей и обобщений по кластерам [19]. Поэтому для расчета показателей метрик конвейера GraphRAG в контекст были включены значения сущностей и связей, что привело к кратному отличию значений метрики Context Recall и значительному различию в метрике Context Relevance, обусловленному разницей в алгоритме формирования контекста.

### 3. РЕЗУЛЬТАТЫ

Было выполнено сравнение значений метрик конвейеров RAG и GraphRAG на полученном датасете. Проведенное сравнение показателей метрик конвейеров RAG и GraphRAG продемонстрировало положительный вклад извлечения структуры данных в GraphRAG-подходе. Приросты значений метрик составили (табл. 2): Answer Relevancy + 0.124, Context Recall + 0.611, Context Relevance + 0.424, Answer Accuracy + 0.119. Наиболее показательными (используют только вопросы и тексты ответов) являются сравнения метрик Answer Accuracy + 32.3% и Answer Relevancy + 34.7% в пользу GraphRAG.

Табл. 2. Сравнение значений метрик конвейеров RAG и GraphRAG.

Метрика	RAG	GraphRAG	Разница	Лучший подход
Answer Accuracy	0.367	0.486	0.119	GraphRAG
Answer Relevancy	0.359	0.483	0.124	GraphRAG
Context Recall	0.116	0.727	0.611	GraphRAG
Context Relevance	0.551	0.975	0.424	GraphRAG

Различия принципов формирования контекста в RAG- и GraphRAG-подходах приводят к объективным различиям в объеме и содержании контекста. В контекст запросов в RAG попадает ограниченное число чанков на основании семантического сходства, что ограничивает полноту извлечения связанных аксиом. В контекст запросов GraphRAG включаются таблицы релевантных сущностей и связей, обобщения, а также чанки документов. Такой подход к наполнению контекста может быть избыточным при формировании

ответов в доменах с ограниченным набором связей (например, в художественной литературе), однако в домене машиностроительной PLM на данных со связной логической структурой обеспечивает прирост качества.

В результате проведенных исследований:

- предложен новый источник оценочных данных в домене машиностроительных систем полного жизненного цикла – файлы изделий и сборок, соответствующих стандарту *STandard for Exchange of Product model data*;
- предложена новая методика формирования текстовых оценочных данных с сохранением логической структуры посредством использования онтологических данных об изделиях в машиностроительных системах полного жизненного цикла и последующей вербализацией с применением однозначного преобразования на управляемый английский – Attempto Controlled English;
- реализован конвейер для использования сгенерированных оценочных данных при сравнении систем обогащения контекста большой языковой модели с применением и без применения графов знаний;
- получены результаты оценки, подтверждающие положительный вклад систем с извлечением структурированных данных в качество генерируемых ответов в доменной области машиностроительных систем полного жизненного цикла.

## **ЗАКЛЮЧЕНИЕ**

Разработана методика генерации синтетического набора данных на основе формальной онтологии, полученной из STEP- файлов – стандарта обмена данными изделий систем автоматизированного проектирования и PLM- систем. Для обеспечения отображения онтологии в текстовую модальность с гарантированным сохранением логической структуры использовано формальное преобразование на язык ACE, что обеспечивает объективную основу для оценки качества извлечения знаний.

Ключевым результатом является формирование методики построения набора данных в доменной области с значительным дефицитом оценочных данных. Такая методика реализует детерминированное формирование набора

данных (свободное от влияния вероятностных факторов при генерации с помощью БЯМ) в автоматическом режиме с возможностью масштабирования.

Практическая значимость работы заключается в реализации программного конвейера для сравнения и оценки извлечения графов знаний с помощью GraphRAG в доменной области машиностроительной PLM. Этот конвейер позволяет получать измеримую оценку для повышения качества обработки структурированной информации, что имеет критическое значение для сохранения логических связей, характерных для данных PLM-домена.

Получены результаты оценки, подтверждающие положительный вклад в качество генерируемых ответов систем, использующих структуру данных (GraphRAG), по сравнению с конвейерами, не использующими извлечение структуры данных (традиционные RAG), в доменной области машиностроительных PLM.

#### **СПИСОК ЛИТЕРАТУРЫ**

1. *Edge D. et al.* From local to global: A graph rag approach to query-focused summarization // arXiv: 2404.16130. <https://doi.org/10.48550/arXiv.2404.16130>
2. *Xiang Z. et al.* When to use Graphs in RAG: A Comprehensive Analysis for Graph Retrieval-Augmented Generation // arXiv: 2506.05690. <https://doi.org/10.48550/arXiv.2506.05690>
3. *Han H. et al.* Rag vs. graphrag: A systematic Evaluation and Key Insights // arXiv: 2502.11371. 2025. <https://doi.org/10.48550/arXiv.2502.11371>
4. *Han H. et al.* Retrieval-Augmented Generation with Graphs (GraphRAG) // arXiv: 2501.00309. 2024. <https://doi.org/10.48550/arXiv.2501.00309>
5. *Gajderowicz B., Bhardwaj A., Fox M.* RAG and Ontologies for Information Retrieval: A Literature Review. 2025. URL: [https://eil.mie.utoronto.ca/wp-content/uploads/2025/09/ONTOLLM\\_2025\\_Aug20v11.pdf](https://eil.mie.utoronto.ca/wp-content/uploads/2025/09/ONTOLLM_2025_Aug20v11.pdf) (дата обращения: 12.02.2026)
6. *Huang Y., Fung T.Y., DeLaurentis D.A.* Addressing Complexity in System of Systems with GraphRAG: An AI-Driven Framework for Dynamic Data Integration // Systems Engineering. 2025. e70012. <https://doi.org/10.1002/sys.70012>

7. *da Cruz T., Tavares B., Belo F.* Ontology Learning and Knowledge Graph Construction: A Comparison of Approaches and Their Impact on RAG Performance // arXiv: 2511.05991. 2025. <https://doi.org/10.48550/arXiv.2511.05991>

8. ГОСТ Р 56862–2016. Национальный стандарт Российской Федерации «Система управления жизненным циклом. Разработка концепции изделия и технологий. Термины и определения».

URL: <https://protect.gost.ru/document.aspx?control=7&id=202875> (дата обращения: 12.02.2026)

9. ГОСТ Р ИСО 10303-1 – 2022. Национальный стандарт Российской Федерации «Системы автоматизации производства и их интеграция. Представление данных об изделии и обмен этими данными. Часть 1. Общие представления и основополагающие принципы (ISO 10303-1:2021, IDT)»

URL: <https://meganorm.ru/Data/792/79232.pdf> (дата обращения: 12.02.2026)

10. *Kwon S., Monnier L.V., Barbau R., Bernstein W.Z.* A New Implementation of OntoSTEP: Flexible Generation of Ontology and Knowledge Graphs of EXPRESS-Driven Data // ASME Journal of Computing and Information Science in Engineering. 2022. Vol. 22 (2). 024502. <https://doi.org/10.1115/1.4053079>

11. Stepcode.

URL: <https://github.com/stepcode/stepcode/tree/develop/data/ap214e3> (дата обращения: 12.02.2026)

12. *Zhang Q. et al.* A Survey of Graph Retrieval-Augmented Generation for Customized Large Language Models // arXiv: 2501.13958. 2025.

<https://doi.org/10.48550/arXiv.2501.13958>

13. *Fuchs N.E., Kaljurand K., Kuhn T.* Attempto controlled english for knowledge representation // Reasoning Web: 4th International Summer School 2008, Venice, Italy, September 7–11, 2008, Tutorial Lectures. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. P. 104–124. <https://doi.org/10.1007/978-3-540-85658-0>

14. *Zaitoun A., Sagi T., Peleg M.* Generating ontology-learning training-data through verbalization // Proc. of the AAI Symposium Series. 2024. Vol. 4. No. 1. P. 233–241. <https://doi.org/10.1609/aaais.v4i1.31797>

15. *Fuchs N.E. et al.* Attempto controlled english: A knowledge representation language readable by humans and machines // Reasoning Web: First International Summer School 2005. Springer Berlin Heidelberg, 2005. P. 213–250.  
[https://doi.org/10.1007/11526988\\_6](https://doi.org/10.1007/11526988_6)
16. *Dong S. et al.* Knowledge-Graph Based RAG System Evaluation Framework // arXiv: 2510.02549. 2025. <https://doi.org/10.48550/arXiv.2510.02549>
17. *Chen E. et al.* Comparing RAG and GraphRAG for Page-Level Retrieval Question Answering on Math Textbook // arXiv: 2509.16780. 2025.  
<https://doi.org/10.48550/arXiv.2501.13958>
18. *Xiao Y. et al.* GraphRAG-Bench: Challenging Domain-Specific Reasoning for Evaluating Graph Retrieval-Augmented Generation // arXiv: 2506.02404. 2025.  
<https://doi.org/10.48550/arXiv.2506.02404>
19. *Es S. et al.* Ragas: Automated evaluation of retrieval augmented generation // Proc. of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations. 2024. P. 150–158.  
<https://doi.org/10.48550/arXiv.2309.15217>

## **ONTOLOGICAL APPROACH TO KNOWLEDGE GRAPH ASSESSMENT IN THE DOMAIN OF MECHANICAL PRODUCT LIFECYCLE MANAGEMENT SYSTEMS**

**V. V. Gladyshev**<sup>[0009-0003-8900-3469]</sup>

*Moscow Institute of Physics and Technology, The Pusk Center, Dolgoprudny, Russia*

gladyshev.vv@phystech.edu

### **Abstract**

This paper examines the application of an ontological approach to constructing a dataset for evaluating and comparing context enrichment systems for large language models using knowledge graphs in the domain of mechanical product lifecycle management systems. In this domain, obtaining the required amount of textual data with a formal logical structure to form an evaluation set without using generated synthetic data is challenging. To avoid introducing distortions and hallucinations when generating evaluation data, a novel solution to the data deficiency is proposed. This solution involves extracting ontology directly from product and assembly files compliant with the STandard for Exchange of Product Model Data. This potentially enables the use of all product data as a source for scaling evaluation data. The goal of this paper is to create a dataset of structured textual data in the domain of mechanical product lifecycle management systems, develop an evaluation methodology, and implement context enrichment pipelines for large language models with and without knowledge graphs to analyze the contribution of data-structure-extracting systems to the quality of generated responses. In this paper: a new source of evaluation data is proposed, a new methodology for generating text evaluation data while preserving the logical structure is developed, a pipeline for using the generated evaluation data is implemented, and evaluation results are obtained that confirm the positive contribution of systems with the extraction of structured data to the quality of generated responses in the domain of mechanical product lifecycle management systems.

**Keywords:** *ontologies, dataset, product lifecycle management – PLM, computer-aided design – CAD, large language model – LLM, retrieval-augmented generation – RAG, GraphRAG, STandard for Exchange of Product model data – STEP.*

## REFERENCES

1. *Edge D. et al.* From local to global: A graph rag approach to query-focused summarization // arXiv: 2404.16130. <https://doi.org/10.48550/arXiv.2404.16130>
2. *Xiang Z. et al.* When to use Graphs in RAG: A Comprehensive Analysis for Graph Retrieval-Augmented Generation // arXiv: 2506.05690. <https://doi.org/10.48550/arXiv.2506.05690>
3. *Han H. et al.* Rag vs. graphrag: A systematic Evaluation and Key Insights // arXiv: 2502.11371. 2025. <https://doi.org/10.48550/arXiv.2502.11371>
4. *Han H. et al.* Retrieval-Augmented Generation with Graphs (GraphRAG) // arXiv: 2501.00309. 2024. <https://doi.org/10.48550/arXiv.2501.00309>
5. *Gajderowicz B., Bhardwaj A., Fox M.* RAG and Ontologies for Information Retrieval: A Literature Review. 2025. URL: [https://eil.mie.utoronto.ca/wp-content/uploads/2025/09/ONTOLLM\\_2025\\_Aug20v11.pdf](https://eil.mie.utoronto.ca/wp-content/uploads/2025/09/ONTOLLM_2025_Aug20v11.pdf) (last access: 12.02.2026)
6. *Huang Y., Fung T.Y., DeLaurentis D.A.* Addressing Complexity in System of Systems With GraphRAG: An AI-Driven Framework for Dynamic Data Integration // Systems Engineering. 2025. e70012. <https://doi.org/10.1002/sys.70012>
7. *da Cruz T., Tavares B., Belo F.* Ontology Learning and Knowledge Graph Construction: A Comparison of Approaches and Their Impact on RAG Performance // arXiv: 2511.05991. 2025. <https://doi.org/10.48550/arXiv.2511.05991>
8. GOST R 56862-2016. Nacional'nyj standart Rossijskoj Federacii “Sistema upravlenija ziznennym ciklom. Razrabotka koncepcii izdelija i tehnologij. Terminy i opredelenija”. URL: <https://protect.gost.ru/document.aspx?control=7&id=202875> (last access: 12.02.2026)
9. GOST R ISO 10303-1-2022. Nacional'nyj standart Rossijskoj Federacii “Sistemy avtomatizacii proizvodstva i ih integracija. Predstavlenie dannyh ob izdelii i obmen etimi dannymi. Cast' 1. Obsie predstavlenija i osnovopolagausie principy (ISO 10303-1:2021, IDT)” URL: <https://meganorm.ru/Data/792/79232.pdf> (last access: 12.02.2026)

10. *Kwon S., Monnier L.V., Barbau R., Bernstein W.Z.* A New Implementation of OntoSTEP: Flexible Generation of Ontology and Knowledge Graphs of EXPRESS-Driven Data // ASME Journal of Computing and Information Science in Engineering. 2022. Vol. 22 (2). 024502. <https://doi.org/10.1115/1.4053079>
  11. Stepcode. URL: <https://github.com/stepcode/stepcode/tree/develop/data/ap214e3> (last access: 12.02.2026)
  12. *Zhang Q. et al.* A Survey of Graph Retrieval-Augmented Generation for Customized Large Language Models // arXiv: 2501.13958. 2025. <https://doi.org/10.48550/arXiv.2501.13958>
  13. *Fuchs N.E., Kaljurand K., Kuhn T.* Attempto controlled english for knowledge representation // Reasoning Web: 4th International Summer School 2008, Venice, Italy, September 7–11, 2008, Tutorial Lectures. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. P. 104–124. <https://doi.org/10.1007/978-3-540-85658-0>
  14. *Zaitoun A., Sagi T., Peleg M.* Generating ontology-learning training-data through verbalization // Proc. of the AAAI Symposium Series. 2024. Vol. 4. No.1. P. 233–241. <https://doi.org/10.1609/aaais.v4i1.31797>
  15. *Fuchs N.E. et al.* Attempto controlled english: A knowledge representation language readable by humans and machines // Reasoning Web: First International Summer School 2005. Springer Berlin Heidelberg, 2005. P. 213–250. [https://doi.org/10.1007/11526988\\_6](https://doi.org/10.1007/11526988_6)
  16. *Dong S. et al.* Knowledge-Graph Based RAG System Evaluation Framework // arXiv: 2510.02549. 2025. <https://doi.org/10.48550/arXiv.2510.02549>
  17. *Chen E. et al.* Comparing RAG and GraphRAG for Page-Level Retrieval Question Answering on Math Textbook // arXiv: 2509.16780. 2025. <https://doi.org/10.48550/arXiv.2501.13958>
  18. *Xiao Y. et al.* GraphRAG-Bench: Challenging Domain-Specific Reasoning for Evaluating Graph Retrieval-Augmented Generation // arXiv: 2506.02404. 2025. <https://doi.org/10.48550/arXiv.2506.02404>
  19. *Es S. et al.* Ragas: Automated evaluation of retrieval augmented generation // Proc. of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations. 2024. P. 150–158. <https://doi.org/10.48550/arXiv.2309.15217>
-

## СВЕДЕНИЯ ОБ АВТОРЕ



**ГЛАДЫШЕВ Виталий Владимирович** – магистрант МФТИ, Центр «Пуск». Область научных интересов: Средства обработки естественного языка и автоматизации структурирования данных, онтологии, применение больших языковых моделей и технологии их адаптации, RAG, GraphRAG, агентные системы.

**Vitaly Vladimirovich GLADYSHEV** – master's student at MIPT, the Pusk Center. Research interests: Natural language processing and automation of data structuring, ontologies, application of large language models and technologies for their adaptation, RAG, GraphRAG, agent systems.

email: [gladyshev.vv@phystech.edu](mailto:gladyshev.vv@phystech.edu)

ORCID: 0009-0003-8900-3469

*Материал поступил в редакцию 15 апреля 2026 года*