

LDBC TEXT2GRAPH Task Force Work Charter

Task Force Name

TEXT2GRAPH Task Force, incorporating existing LEX (Extended GQL Schema) and GQL Implementation Working Groups

Proposed Group Leaders

- Whole project joint leads: Shipeng Qi (Ant Group), Alastair Green (JCC)
- Benchmark (primary) workstream: Songlin Lyu (Ant Group), Oskar van Rest (Oracle)
- Extended GQL Schema (secondary) workstream: Jan Hidders (Birkbeck), Dušan Živković
- GQL Implementation (secondary) workstream: Michael Burbidge, Malcolm Crowe

Task Force Lifetime

Permanent

Charter version date

22 August 2023

Charter version

Draft 0.12 [1.0]

Date charter version agreed by Members Policy Council

[5 September 2025]

Mission

The TEXT2GRAPH Task Force aims to fuse existing graph query language related working groups (GQL Implementation working group, LEX working group), and establish a benchmark for Text to Graph Queries as a representative use case—including Text to GQL, SQL/PGQ, and Cypher, and (when feasible, and as a secondary priority) Text to other possible Graph Query Languages (Gremlin, SPARQL).

To evaluate the capability of Large Language Models (LLMs) in this domain, the task force will build a comprehensive dataset that adapts to industry usage scenarios to make the evaluation representative, reliable and comparable.

The initial goal of this project is a benchmark for the *accuracy* of transformations from text to query languages and the execution of resulting queries. It does not aim to measure and compare the *performance* of the graph data engines used to execute queries.

Motivation

Recent advances in Large Language Models (LLMs) have made industrial-grade Text to Graph Query not only possible but increasingly practical, thanks to their robust natural language understanding and code generation capabilities. As graph interaction evolves, LLM-driven Text to Graph Query capabilities are becoming essential—demonstrated by the adoption of such features in tools like GraphRAG and GraphAgent.

However, the industry currently lacks standardized benchmarks and datasets for evaluating and comparing LLM-based Text to Graph Query solutions. This gap makes it difficult to assess real-world performance and impedes progress in this critical domain.

By establishing the TEXT2GRAPH Benchmark and a comprehensive, industry-relevant dataset, the Task Force seeks to provide a reliable, representative, and comparable foundation for evaluating LLM Text to Graph Query systems. This will serve as a representative use case for existing graph query language-related outcomes, facilitate cooperation between existing working groups, and drive innovation, ensure consistency in measurement, and accelerate the adoption of advanced graph interaction capabilities across the industry.

Scope of Work

The main tasks of the task force are the following:

- Continue and complete the work of the LEX WG, to design an extended schema definition language based on GQL graph types and PG-Schema, which is capable of representing SQL/PGQ tabular graph view schemas. The existing goals of allowing users to easily define content record types using JSON Schema, to easily define subtypes, to produce an Information Schema for graph types, and to access constraints comparable in power to those in SQL and JSON Schema/SHACL Core all remain in place. The LEX language will be used to support the exemplary use case of text-to-graph.
- Continue the ongoing work of the GQL Implementation WG, with an emphasis on tests and tools for the schema aspects of GQL, which are not covered by Neo4j's Cypher TCK work. To maintain the ANTLR grammar, and to extend it to cover LEX and T2G IL requirements.
- Design an Intermediate Language (IL) covering the DDL, DML, DQL, and DCL of GQL/LEX, SQL/PGQ, and Cypher to allow translation between graph query languages. Explore the possibility of making the IR compatible with a wider range of graph query languages (Gremlin, SPARQL, etc.).
- Develop a dataset construction framework: to offer tools and methodologies to help annotators translate, generalize, and generate corpora, and synthesize databases, thereby accelerating dataset construction and improving both quality and quantity.
- Implement the Intermediate Language (IL) as an Intermediate Representation (IR) into the dataset construction framework: to offer graph query languages translation

ability during both the corpora collection stage and the Text to Graph Query translation stage.

- Construct a comprehensive Text to Graph Query dataset: consists of multiple databases with complex business schema and complex graph query statements on them (somewhere between Spider and Spider2.0). Part as a public training set and part as a non-public test set.
- Establish a standardized TEXT2GRAPH benchmark: to develop a standard corpus annotation workflow and define evaluation standards that incorporate different levels of graph query and schema difficulty, as well as multiple evaluation metrics.
- Analyze key choke points in Text to Graph Query: to evaluate the current limitations and strengths of LLM-based Text to Graph Query systems, guide future improvements (also compare with the error analysis in text2sql translation).

Initial Members

Num	Company/University	Representatives
1	Ant Group (TU Graph)	
2	Oracle	
3	Microsoft	
4	Google	
5	Neo4j	
6	Birkbeck	
7		
8	Individuals	

Intended output

The primary output of the task force is the TEXT2GRAPH benchmark specification [TEXT2GRAPH specification].

For generating the dataset and performing audits, we maintain the following software components and the dataset:

Graph-IL

Graph-IL is a new IL which will be based on the common core of GQL and PGQ called GPM, and which will deviate from GQL/PGQ only when explicitly agreed by the WG.

The feature set of Graph-IL will therefore include GQL insert/update/delete (which has no equivalent in PGQ), and GQL CREATE GRAPH TYPE alongside SQL/PGQ CREATE PROPERTY GRAPH, for graph schema definitions.

As the WG subsumes LEX, and there is no intention to change LEX's charter goals, it is taken as read that the LEX sub-stream will extend GQL graph types, so that LEX (LDBC Extended GQL) schemas will broadly match the functionality of GQL graph types plus SQL schema constraints and SHACL Core Constraints. This extension includes as an early goal that GQL graph types are capable of fully representing the schema information content of a PGQ tabular property graph.

The baseline (starting point) for property graph schema in the main T2G project is the graph schema functionality of SQL/PGQ tabular property graphs, which are, at a high level of abstraction, equivalent to GQL graph types extended to include keys. The issue of JSON data, which is well-handled by e.g. Oracle's product set, (extending SQL's JSON support to incorporate JSON Schema) is a second area where extension to the existing GQL and PGQ standards may well be immediately relevant for the T2G benchmark. The extent to which LEX schema extensions which exceed current GQL and SQL/PGQ are directly used by the T2G project is an example of where explicit WG decisions will need to be made.

Dataset Construction Framework

A dataset construction framework as a Python library that offers tools for corpora translation, generalization, and generation, and database synthesis. Accelerate the construction progress and improve the quantity and quality of the dataset.

The framework will implement Graph-IL as an Intermediate Representation (IR) for the parsing and translation of different graph query languages.

Benchmark Evaluation Framework

An evaluation framework to evaluate LLM's Text2Graph ability from multiple metrics.

Text2Graph Dataset

A Text2Graph dataset consists of multiple databases with complex business schema and complex query statements in different graph query languages on them (somewhere between Spider and Spider2.0). Part as a public training set and part as a non-public test set.

Other intended output/work product

- Presentations explaining the benchmark and the dataset given at industry meetups, academic workshops, LDBC TUC meetings, etc.
- Auditing guidelines
- Auditor exam sheets

Intended timescales

- Q4 2025:
 - § IR covering GQL, SQL/PGQ, and Cypher
 - § Dataset construction framework integrating GQL, SQL/PGQ, and Cypher parser
 - § Benchmark evaluation framework
- Q1 2026:
 - § LEX schema language draft
 - § Dataset construction
 - § Test results and analysis on baseline models and optimized models
 - § Graph query and schema hardness criteria design and discussion
 - § Draft of the TEXT2GRAPH benchmark v0.1.0 specification
- Q2 2026:
 - § IR compatibility exploration on Gremlin and SPARQL
 - § Dataset construction framework integrating Gremlin and SAPRQL parser
 - § Final draft of the TEXT2GRAPH benchmark v0.1.0 specification

Related Task Forces or Working Groups

The LDBC GQL Implementation Working Group works on creating tooling and documentation to assist in the implementation and adoption of GQL, while the LDBC Extended GQL Schema (LEX) Working Group works on proposing a concrete extended GQL schema language design.

These two Working Groups are now incorporated into this Task Force, as subordinate work streams.

The working tools (rolling agenda, document registry) and standards for registered, numbered working documents developed for LEX will now be extended for use by this project.

Documents relating to the whole project and to the primary Benchmark workstream will be numbered with prefix T2G, for the LEX workstream we will continue to use LEX, and we will use IMP for the Implementation workstream. Each prefix has an independent number fountain.

References to relevant documents, standards, etc

- [LDBC Extended GQL Schema \(LEX\) Work Charter](#)
- [LDBC GQL Implementation Work Charter](#)