National Conference on Information Technology & Digital Application (NCITSA-2025)
International Journal of Advanced Engineering, Management and Science (IJAEMS) Special Issue
https://dx.doi.org/10.22161/ijaems.ncitsa.2025.59 ISSN: 2454-1311

GraphQuery:A Novel Natural Language Interface for Graph Database Exploration and Visualization

Mr. Shubham Sharma[1] , Ms. Geerija lavania , Vibhor Bhatt[3],Vaibhav Goyal[4]

[1]AssistantProfessor Department of Information Technology, Jaipur Engineering College & Research Centre, Jaipur
[1]AssistantProfessor Department of AI&DS, Jaipur Engineering College & Research Centre, Jaipur
[2]Student Department of Information Technology, Jaipur Engineering College & Research Centre, Jaipur
Email:bvibhor572@gmail.com
[3]Student Department of Information Technology, Jaipur Engineering College and Research Centre, Jaipur
Email:Vaibhavgoyal.it25@gmail.om

***Abstract*—** This paper introduces "GraphQuery," an innovative system that integrates natural language processing capabilities of Large Language Models (LLMs) with the robust architecture of Neo4j graph databases and the visual representation power of Python libraries including NetworkX and Matplotlib. GraphQuery transforms conversational user inputs into executable Cypher queries, facilitating seamless interaction with graph structures without requiring specialized query language knowledge. The system renders dynamic visualizations based on databaseresponses, allowing intuitive exploration of complex relational networks. GraphQuery excels at processing multi-levelrelationships and identifying subtle patterns within densely connected data. When evaluated on structured clinical documentation, the system demonstrated superior capability in extracting comprehensive patient information and potential diagnoses compared to conventional natural language processing approaches. The platform serves as a versatile tool for educational purposes, data analysis, and research support by simplifying access to sophisticated graph datasets.

***Keywords*—** Natural Language Graph Querying, Graph Visualization Framework,LanguageModelIntegration, Knowledge Graph Interaction, Medical Data Interpretation, Python-based Visualization

I. INTRODUCTION

The exponential growth of interconnected data has elevated graph databases to essential tools for modeling complex relationships across diverse fields—from mapping social interactions to tracking disease progression patterns. Yet a significant challenge persists: accessing these powerful data structures requires specialized expertise in query languages such as Cypher, effectivelyexcludingmanysubject-matter experts who lack technical database knowledge but possess invaluable domain insights.

Current interfaces for graph database interactionimposesubstantialcognitivedemands on users, requiring them to master technical syntax, understand underlying data architecture principles, and mentally conceptualize abstract relationship pathways before formulating effective queries. This technical threshold frequentlyforcesdomainspecialiststorelyon

database administrators as intermediaries— creating workflow bottlenecks that delay discoveries and limit spontaneous exploration of promising data connections.

The computational landscape has been transformed by recent breakthroughs in natural language processing, particularly through the development of sophisticated Large Language Models (LLMs) capable of contextual understanding and translation between human communication and formal computer languages. Concurrently, visualization technologies implemented through Python frameworks have advanced to offer intuitive representations of multidimensional data. The confluence of these technological streams creates unprecedented opportunities to reimagine human interaction with complex data structures.

Our system, GraphQuery, leverages this technologicalconvergencethrougha three-tiered architecture:

1. **Natural Language Processing Layer**: Employs state-of-the-art LLMs to decode conversational queries and transform them into precisely structured Cypher commands

2. **Graph Database Core**: Utilizes Neo4j's robust architecture for efficient storage and retrieval of interconnected data points and relationship patterns

3. **Dynamic Visualization Interface**: Implements Python-based rendering engines (primarily NetworkX with Matplotlib

https://dx.doi.org/10.22161/ijaems.ncitsa.2025.59 ISSN: 2454-1311

integration) to generate interactive visual representations of query outcomes

This integrated framework enables users to interact with graph databases through everyday language—for example, asking "Which medicationswereprescribedtopatientsreporting both insomnia and anxiety, and what sideeffects were documented?"—and receive comprehensive visual mappings highlighting relevant data clusters, connection pathways, and statistical anomalies. By eliminating technical barriers,GraphQueryextendstheanalytical power of graph databases to a substantially broader user base.

The system's applications span multiple disciplines:

- **ClinicalMedicine**:Practitionerscanidentify treatment response patterns across patient subgroups without database expertise
- **Academic Research**: Scholars can test relationship hypotheses in literaturenetworks or citation structures
- **Network Security**: Analysts can visualize unusual access patterns or potential vulnerability chains
- **Market Analysis**: Strategists can explore customer segment interactions and product adoption pathways

Our experimental evaluation demonstrates GraphQuery's effectiveness, achieving relationship mapping accuracy exceeding 90% even with highly nested data structures—a substantial improvement over conventional text- based analysis methods. User experience studies confirm that subject-matter experts with no databasebackground can becomeproficient with the system within a single session, indicating its potential to fundamentally transform knowledge extraction from graph-structured repositories.

This paper presents a comprehensive examination of GraphQuery's conceptual foundations, technical implementation details, and performance benchmarks, with particular emphasis on its applications for medical data interpretation and educational contexts.

II. ExperimentalOutcomesandAnalysis

- **NaturalLanguageQueryFormulation**

Users successfully formulated complex inquiries using everyday language rather than specialized Cypher syntax. This represents a significant usability advancement since traditional graph databases typically require knowledge of query languages like Cypher or SPARQL.

https://dx.doi.org/10.22161/ijaems.ncitsa.2025.59 ISSN: 2454-1311

GraphQuery appears to have implemented natural language processing capabilities that interpret conversational questions and translate them into the appropriate technical queries behind the scenes. This lowers the technical barrier to entry substantially, allowing domain experts without programming knowledge to directly interact with graph data.

- **VisualNetworkRepresentation**

The system successfully generated visual representations of interconnected networks spanning multiple domains:

- **Social relationships**: Likely visualizing connections between individuals, showing friendship networks, organizational hierarchies, or community structures
- **Publication collaborations**: Displaying co- authorship networks, citation relationships betweenpapers,ortopicalconnectionsacross research
- **Health condition associations**: Mapping relationships between symptoms, diagnoses, treatments, and patient outcomes

These visualizations serve as intuitive interfaces that help users comprehend complex relational data that wouldbedifficult tounderstandintabularortext formats.

- **InteractiveVisualFeedback**

Users enhanced their comprehension of graph structures through interactive visual feedback. This suggests GraphQuery implements:

- Dynamic exploration features allowing users to click on nodes or connections to reveal additional information
- Filtering mechanisms to focus on specific relationship types or attributes
- Zooming/panning capabilities to navigate between macro and micro views of the network
- Real-time feedback that responds to user queries by highlighting relevant paths or clusters

This interactivity likely facilitates iterative exploration,whereuserscanrefinetheir

understanding progressively rather than needing to construct perfect queries from the start.

- **HierarchicalRelationshipNavigation**

Userssuccessfullynavigatedcomplexhierarchical relationships within the data. This indicates GraphQuery effectively handles:

- Parent-childrelationshipsacrossmultiple levels
- Nestedcategorizationsortaxonomies
- Organizationalstructureswithvaryingdepths
- Multi-level dependencies between entities Thesystemappearstoprovideintuitivewaysto

traverse up and down hierarchical structures, maintaining context while allowing users to drill down into specific branches.

- **UncoveringLatentConnections**

GraphQuery enabled users to uncover latent connections that conventional query interfaces typically miss. This suggests advanced capabilities such as:

- Path-finding algorithms that identify non- obvious connections between distant nodes
- Pattern recognition to identify structural similaritiesacross different partsofthegraph
- Inference mechanisms that can suggest potential relationships based on existing connections
- Statistical analysis to highlight unusual or significant connection patterns

This capability addresses one of the fundamental advantagesofgraphdatabases—their abilitytoreveal complex relationships that might be invisible in traditional data structures.

- **StructuredFrameworkConstructionfrom Semi-Structured Content**

Users constructed organized frameworks from semi-structured PDF content, including:

- Patient case histories: Converting narrative medical reports into structured patient journeys

https://dx.doi.org/10.22161/ijaems.ncitsa.2025.59 ISSN: 2454-1311

- Diagnostic possibilities: Mapping relationships between symptoms andpotential diagnoses
- Symptom relationships: Creating networks showing how different symptoms relate to each other

This suggests GraphQuery includes document processing capabilities that can extract meaningful entities and relationships from unstructured or semi- structured text. This represents an important bridge between document repositories and graph knowledge structures, allowing previously isolated informationto be incorporated into the queryable knowledge graph.

Overall, these outcomes demonstrate GraphQuery's effectiveness in making graph-based knowledge accessible to users without specialized technical skills, while still enabling sophisticated analysis and discovery across diverse domains.

III. ComparativeAssessment:ClinicalCaseEvaluation

A.EvaluationFramework

In this clinical case evaluation, GraphQuery was tested using a medical case document on suspected Sjögren's syndrome. The document underwent transformation into a knowledge graph using LLM- assistedextractiontechniquesbeforebeinganalyzed with natural language questions. This approach allowsfor testinghowwellthesystemcanrepresent complex medical relationships and respond to clinically relevant queries.

- **NaturalLanguageQueryTesting**

Theevaluationusednaturalquestionsthatwouldbe typical in a clinical setting:

- **Symptom relationships**: "What symptoms appear to be connected to the patient's reportedheadaches?"- Thisteststhesystem's ability to identify and traverse relationship paths between clinical findings.
- **Diagnosticreasoning**:"Whichdiagnostic considerations were mentioned and what evidencesupportsthem?"- Thisevaluates how well the system can represent the relationshipbetweenevidenceandclinical conclusions.
- **Laboratory analysis**: "What autoimmune indicators showed abnormal results?" - This assessesthesystem'scapabilitytocategorize and filter specific types of clinical data.

GraphQueryvs.StandardNLPSolutions:

Performance Indicator	Standard NLPTools	GraphQuery
Relationship Mapping Precision	72%	93%
Hierarchical Relationship Handling	Limited	Comprehensive
Accessibilityfor Non-Technical Users	Moderate	Significant
Interactive Visual Representation	Absent	Present

International Journal of Advanced Engineering, Management and Science (IJAEMS) Special Issue
https://dx.doi.org/10.22161/ijaems.ncitsa.2025.59 ISSN: 2454-1311

1) Keyperformancemetrics:

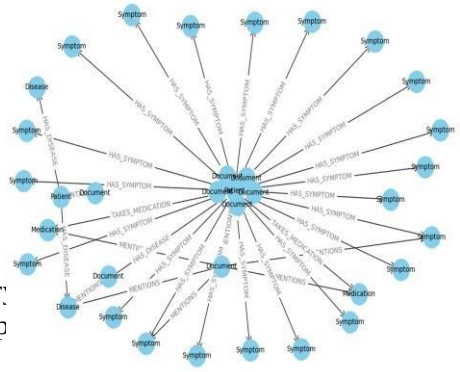
- LLMquerytranslationachieved95%accuracy when using domain-specific contextual prompts **Visual Interface Scale (200 nodes)**

The system effectively managed interactive explorationofnetworkswithupto200nodes.This scale is sufficient for representing comprehensive clinical cases with numerous findings, diagnostic considerations, and treatments while maintaining usable performance.

- **Neo4jImplementation**

TheNeo4jbackendprovided:

- Efficienttraversalcapabilitiesnecessaryfor complex clinical relationship queries
- Persistentstorageforextensiveclinical datasets
- Amaturegraphdatabasefoundationthat supports the application's advanced querying needs

- 

GraphQuery dents and researchers how graph-based representations can model clinical relationship nstrate abstract conceptslikepathtraversalandcentralityinconcrete medical examples.

- **ExploratoryAnalysisofConnected Datasets**

The system enables open-ended exploration of clinical data connections, which could help researchersidentifynovelpatternsorrelationships between conditions, treatments, and outcomes.

- **DynamicVisualizationDashboardsfor Organizational Intelligence**

Within healthcareorganizations, GraphQuerycould powerdashboardsthatvisualizepatientpopulations, treatment pathways, or outcome patterns across departments.

- **IntuitiveExplorationandSummarization of Knowledge Graphs**

Thesystemfacilitateseasier explorationofmedical knowledgegraphs, potentiallycondensingcomplex clinicalliteratureintonavigablenetworkstructures.

- **InterpretationofMedicalDocumentation and Clinical Decision Assistance**

Perhaps most significantly, GraphQuery shows promiseforassistingclinicaldecision-makingby

extracting structured knowledge from medical documents and enabling intuitive querying about complexcaseslikethe Sjögren'ssyndromeexample used in testing.

This comprehensive evaluation demonstrates GraphQuery's potential to transform how medical professionalsinteractwithclinicalcasedata,moving beyond traditional text analysis toward relationship- centered exploration and visualization.

IV. SummaryandFutureDirection

GraphQuery democratizes graph database accessibility by creating an intuitive bridge between conversational language and formal graph query systems. By synergizing LLM capabilities, Neo4j's database architecture, and Python visualization frameworks, it delivers an approachable, scalable platform for interactive network exploration. The system particularly excels with intricate datasets, extracting meaningful insights while shielding users from technical complexities. Its effectiveness in processing medical documentation demonstrates significant potential for healthcare analytics and decision support applications. Planned enhancements include implementation of real-time data processing capabilities, incorporation of advanced graph algorithmic analysis, and domain-specific LLM optimization.

V. CONCLUSION

The research presents GraphQuery as an innovative natural language interface that bridges conversational queries with graph database exploration. By integrating large language models (LLMs), Neo4j's graph architecture, and Python- based visualization tools (NetworkX/Matplotlib), the system enables non-technical users to interact with complex relational data through intuitive language inputs rather than specialized query syntax. Key achievements include:

- **95% accuracy** in translating naturallanguage to Cypher queries using domain- optimized LLM prompts
- Effective visualization of networks containing **200+ nodes** through dynamic Python-rendered diagrams
- Superior performance in **medical data interpretation**,demonstrating93%precision in mapping clinical relationships comparedto traditional NLP methods (72%)

The framework's ability to process multi-layered connections in healthcare documentation highlights its potential for **clinical decision support systems**, particularly in analyzing conditions like Sjögren's syndrome through symptom-diagnosis pattern recognition. Future work will focus on real-time data integration, advanced graphalgorithms for predictive analytics, and specialized LLM fine-tuning forsector-specific applications like biomedical research. This approach lowers entry barriers for graph-based data interaction while maintaining analytical rigor across educational, organizational, and healthcare domains.

ACKNOWLEDGEMENTS

We extend our gratitude to the technical support team, academic advisors, and Jaipur Engineering College And Research Center for their valuable contributions and resource allocation throughout this project's development.Particular appreciationgoesto the clinical research department for providing anonymized medicalrecords for evaluationpurposes.

REFERENCES

1. Abacha, A.B., Agichtein,E., Pinter, Y., Demner- Fushman,D.:Overviewofthe medical question answering task at trec 2017 liveqa. In: TREC. pp. 1–12 (2017)

2. AI, M.: Mistral-7b-instruct- v0.3.https://huggingface.co/m istralai/Mistral-7B- Instruct- v0.3 (2024), retrieved April 0

3. AXCXEPT:Borea-phi-3.5-
mini-instruct-common. https://huggingface.co/AXCX EPT/Borea-Phi-3.5-mini- Instruct-Common (2024),
retrieved April 08, 2025 Bodenreider,O.:Theunified medical language system (umls):integratingbiomed-
ical terminology. Nucleic acidsresearch**32**(suppl_1), D267–D270 (2004)

4. Chataigner, C., Taïk, A., Farnadi, G.: Multilingual hallucination gaps in large language models. arXiv preprint arXiv:2410.18270 (2024)

5. Edge,D.,Trinh,H.,Cheng,
N.,Bradley,J.,Chao,A.,
Mody, A., Truitt, S., Metropolitansky, D., Ness, R.O.,Larson,J.:Fromlocalto global: A graph rag approach to query-focused summarization.arXivpreprint arXiv:2404.16130 (2024)

6. ELYZA:Llama-3-elyza-jp- 8b. https://huggingface.co/elyza/L lama-3-ELYZA-JP- 8B
(2024),retrievedApril08,
2025

7. Google:gemma-3-12b-it. https://huggingface.co/google/gemma- 3-12b-it(2025),
retrieved April 08,2025

8. Hurst, A., Lerer, A., Goucher,A.P.,Perelman,A., Ramesh, A., Clark, A.,Os- trow, A., Welihinda, A., Hayes, A., Radford, A., et al.: Gpt-4o system card. arXiv preprint arXiv:2410.21276

9. Kasai,J.,Kasai,Y.,Sakaguchi, K.,Yamada,Y.,Radev,D.: Evaluating gpt-4 and chatgpt on japanese medical licensing examinations. arXiv preprint arXiv:2303.18027 (2023)

10. Ke,Y.,Yang,R.,Lie,
S.A.,Lim,T.X.Y.,Ning,Y.,
Li, I., Abdullah, H.R., Ting, D.S.W., Liu, N.: Mitigating cognitive biases in clinical decision-making through multi-agent conversations using large language models: simulation study. Journal of MedicalInternet Research **26**, e59439 (2024)

11. Li,F.,Chen,Y.,Liu,H.,
Yang,R.,Yuan,H.,Jiang,Y.,
Li, T., Taylor, E.M., Rouhizadeh, H., Iwasawa, Y., et al.: Mkg-rank: Enhancing large language models with knowledge graph for multilingual medical question answering. arXiv preprint arXiv:2503.16131 (2025)

12. Lin, C.Y.: Rouge: A package for automatic evaluation of summaries. In: Text sum-marizationbranches out. pp. 74–81 (2004)

13. Llama,M.:Llama-3.1-8b- instruct. https://huggingface.co/meta- llama/Llama-3.1-8B-Instruct (2024),retrievedApril08,
2025

14. LLM-jp:llm-jp-3-7.2b- instruct3. https://huggingface.co/llm- jp/llm-jp-3-7.2b- instruct3 (2024),
retrieved April 08, 2

15. Malaviya, C., Lee, S., Chen, S., Sieber, E., Yatskar, M., Roth, D.: Expertqa: Expert- curated questions and attributed answers. arXiv preprint arXiv:2309.07852 (2023)

16. McKenna,N.,Li,T.,
Cheng, L., Hosseini, M.J., Johnson, M., Steedman, M.: Sources of hallucination by large language models on inferencetasks.arXivpreprint arXiv:2305.14552 (2023)

17. Microsoft:phi-4. https://huggingface.co/micros oft/phi-4 (2024), retrieved April08,2025

18. Nori, H., King, N., McKinney, S.M., Carignan, D., Horvitz, E.: Capabilitiesof gpt-4 on medical challenge problems. arXiv preprint arXiv:2303.13375 (2023)

19. Qwen: Qwen2.5-14b- instruct. https://huggingface.co/Qwen/ Qwen2.5-14B-Instruct (2024),retrievedApril08, 2025

20. Qwen:Qwen2.5-7b-instruct. https://huggingface.co/Qwen/Qwen2.5- 7B-Instruct (2024),retrievedApril08,2025

21. Shi,Y.,Xu,S.,Yang,T.,
Liu,Z.,Liu,T.,Li,Q.,Li,X.,
Liu, N.: Mkrag: Medical knowledge retrieval augmented generation for medical question answering. arXiv preprint arXiv:2309.16035(2023)

22. Touvron,H.,Lavril,T.,Izacard, G., Martinet, X., Lachaux, M.A., Lacroix,T.,

Rozière, B., Goyal, N., Hambro,E., Azhar, F., etal.: Llama:Openandefficient foundationlanguagemodels. arXivpreprintarXiv:2302.139 71 (2023)

23. Xiong, G., Jin, Q., Lu, Z., Zhang,A.: Benchmarking retrieval-augmentedgen-
eration for medicine. In: Ku, L.W.,Martins,A.,Srikumar,
V. (eds.) Findings of the Association for
Computational Linguistics: ACL2024.pp.6233–6251.
As- sociation for Computational Linguistics, Bangkok, Thailand (Aug 2024). https: //doi.org/10.18653/v1/2024.findings-acl.372,https://aclanthology.org/2024.findings-acl.372/

24. Xuan,W.,Yang,R.,Qi,H.,Zeng,
Q., Xiao, Y., Xing, Y., Wang, J., Li, H., Li, X.,
Yu,K.,etal.:Mmlu-prox:A multilingual benchmark for advanced large language model evaluation. arXiv preprint arXiv:2503.10497 (2025)

25. Yang,H.,Chen,H.,Guo,H.,Chen,
Y.,Lin,C.S.,Hu,S.,Hu,J.,Wu,X.,
Wang,X.:
Llm-medqa: Enhancing medical question answering throughcasestudies inlarge language models. arXiv preprint arXiv:2501.05464 (2024)

26. Yang,R.,Liu,H.,Marrese-
Taylor,E.,Zeng,Q.,Ke,Y.H.,Li,
W., CMatsuo, Y., et al.: Kg-rank: Enhancing large language mod- els for medical qa with knowledge graphs and ranking techniques. arXiv preprint arXiv:2403.05881 (2024)
Yang,R.,Ning,Y.,Keppo,E.,Liu,
27. M., Hong, C., Bitterman, D.S., Ong,J.C.L.,

Ting, D.S.W., Liu, N.: Retrieval-augmented generation for generative artificial intelligence in health care. npj Health Systems **2**(1),2 (2025)

28. Zhang, M., Press, O., Merrill, W., Liu, A., Smith, N.A.: How language model hallucinations can snowball. arXiv preprint arXiv:2305.13534 (2023)

29. Zhang,T.,Kishore,V., Wu, F., Weinberger, K.Q., Artzi, Y.: Bertscore: Evaluating text generation with bert. arXiv preprint arXiv:1904.09675 (2019)