

GISphere Knowledge Graph (GISphere-KG) Platform for Geography Programs Recommendation for Future GIS workforce

Zhining Gu, Wenwen Li, Bing Zhou, Yikang Wang, Yanbing Chen, Shan Ye, Kejin Wang, Hongkai Gu, Yuhao Kang

ABSTRACT: The rising global interest in Geographic Information System/Science (GIS) programs has led to an increased demand for higher education in this field. However, students often struggle to find suitable programs and faculty due to the overwhelming options and a lack of personalized guidance. This paper presents GISphere-KG, an AI-powered platform based on the GISphere project. It combines Knowledge Graph (KG) and Large Language Models (LLMs) to aid in the search and recommendation of GIS-related graduate programs. GISphere-KG offers four key features: (1) interactive conversation for applicants to receive responses through a question-answering process in natural language; (2) efficient information retrieval through semantic relationships built in the KG; (3) discovery of professors that match applicants' interests, offering more choices within the same research field; and (4) personalized program recommendations aligned with applicants' academic and career goals. Our platform aims to not only provide a user-friendly tool that assists future students in accomplishing their career aspirations but also enriches the geographical community by attracting more talent and promoting global geography education.

KEYWORDS: *GIS, Education, Large Language Model, Knowledge Graph, Information retrieval, Program recommendation*

Introduction

The increasing importance of Geographic Information System/Science (GIS), cartography, and geospatial analytics has led to an expansion across domains like urban planning, disaster relief, and decision-making. These successful practices evoke rapid flourishing of master's, professional master's, and Ph.D. programs, as well as certifications in this field (Wang et al., 2023; Longley, 2000; Rickles et al, 2017), attracting many students keen to enhance their expertise in this field. However, applicants face challenges due to the overwhelming choices and limited resources, making it difficult to identify programs that align with their specific interests and preferences, such as research compatibility with professors.

To address these challenges, we propose GISphere-KG, an AI-powered platform utilizing the Knowledge Graph (KG) and advanced large language models (LLMs) through LangChain and Neo4j. KG, with its ability to organize diverse data including structured and unstructured data (Li et al., 2022), helps in efficient information retrieval (Li et al., 2023), similar research interest discovery, and answering queries relevant to GIS education. LLMs, capable of semantic recognition for entities and relationships, and intent detection, convert natural language into machine-understandable statements used by graph database (i.e. Neo4j) to support tasks mentioned above. This integration aims to offer a

comprehensive, user-friendly tool that not only provides up-to-date GIS program information worldwide but also uses semantic similarity calculations to match applicants with professors and programs fitting their research interests and preferences. GISphere-KG’s innovative approach has the potential to enhance global GIS education. Furthermore, such useful platform can be extended to support search and recommendation for other kinds of expertise by mining diverse source of data.

Method

This work consists of the following steps, as shown in Figure 1.

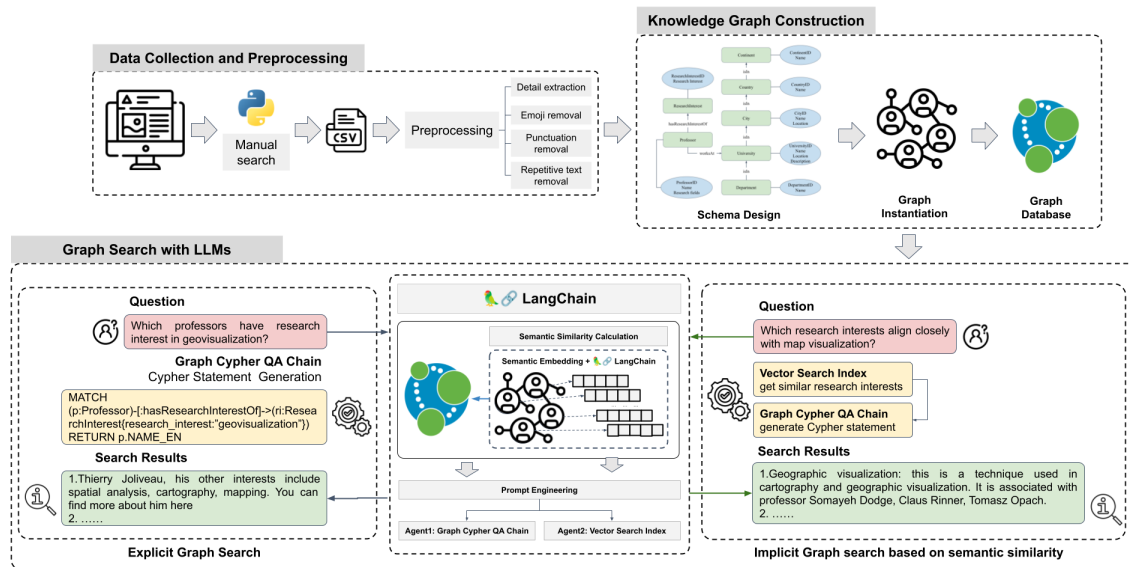


Figure 1: workflow of GISphere-KG

Data collection and preprocessing

The first step is data collection and preprocessing. From the year 2019 to 2023, over 200 volunteers have contributed to the GISphere project, collecting information from over 600 GIS-related programs worldwide from 97 countries or regions as of December 2023. Such an extensive dataset covers faculty information from about 2000 professors in over 400 academic institutions. To obtain this rich dataset, volunteers utilized web crawling techniques and manual searches. The raw datasets were preprocessed and cleaned by extracting specific details and removing emojis, punctuation, and repetitive texts. An annual validation and revision process is in place to ensure the data’s accuracy and relevance.

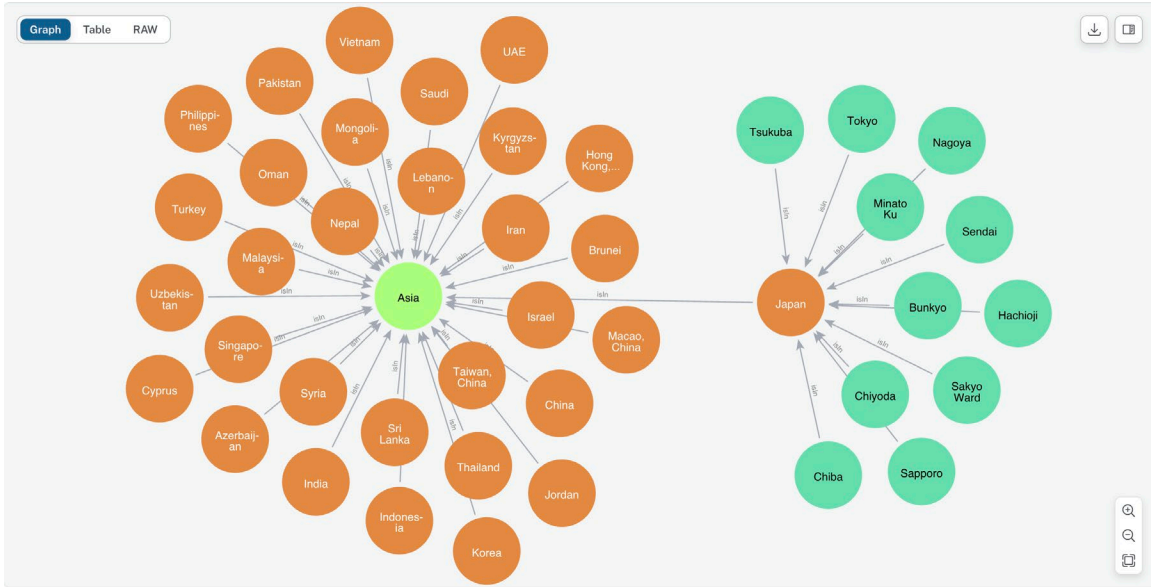
Knowledge graph construction

The second step is KG construction. We designed an ontology for GISphere-KG, defining the semantic structure among concepts, properties, and their interrelationships. GISphere-KG ontology contains seven classes (*Continent*, *Country*, *City*, *University*, *Department*, *Professor*, and *Research Interest*). *Continent*, *Country*, and *City* indicate the administrative

regions that each program is from, containing name attribute. *University* not only has its name property but also contains the textual “description” of the university, “latitude”, and “longitude”. *Department* has the property of the department name related to the GIS program. For *Professor* class, it contains name attribute and expertise scopes (i.e. Physical Geography, Human Geography, Urban Planning, GIS). *Research Interest* is the most important entity in our knowledge graph. It defines different research interests, which can be further leveraged to analyze similar expertise among professors and provide the most suitable program recommendation for applicants.

We also have defined four types of relationships among these seven classes: *isIn*, *worksAt*, *hasResearchInterestOf*, and *isSimilarTo*. *isIn* connects administrative regions or places, such as using $\langle \textit{Country}, \textit{isIn}, \textit{Continent} \rangle$ to denote that a *Country* is part of a *Continent*. The “*worksAt*” relationship specifies the *University* where a *Professor* is affiliated, represented by $\langle \textit{Professor}, \textit{worksAt}, \textit{University} \rangle$. The “*hasResearchInterestOf*” relationship indicates the research interests of a *Professor*, as in $\langle \textit{Professor}, \textit{hasResearchInterestOf}, \textit{Research Interest} \rangle$. “*isSimilarTo*” relationship is built to measure research interest similarity between each pair of *Professors*. Specifically, the relationship between *Professor* and *Research Interest* is one-to-many, meaning that a *Professor* can have multiple research interests.

Therefore, we compiled the dataset into the knowledge graph stored in Neo4j graph database according to the designed ontology. Figure 2 displays examples of the constructed knowledge graph in Neo4j built-in visualization platform. Circles with different colors represent different entities, such as *City*, *Country*, *Continent*, *Department*, *People*, *Research Interest*, and *University*. Arrows pointing from one entity to another one show the semantic relationships between a pair of entities, including *hasResearchInterestOf*, *isIn*, *isSimilarTo*, and *WorksAt*. Figure 2(a) shows a subgraph that contains $\langle \textit{City}, \textit{isIn}, \textit{Country} \rangle$ and $\langle \textit{Country}, \textit{isIn}, \textit{Continent} \rangle$, in which orange circles represent *City* entities, light green circles represent *Country*, and light-yellow circles are *Continent* node. In this figure, different types of nodes are connected by an edge of “*isIn*”. Figure 2(b) shows a few examples of $\langle \textit{Professor}, \textit{hasResearchInterestOf}, \textit{Research Interest} \rangle$ and $\langle \textit{Professor}, \textit{worksAt}, \textit{University} \rangle$. Specifically, the *Professor* node is in blue, the *Research Interest* nodes are in orange, and the node with pink color is the *University* node.



(a)



(b)

Figure 2: GISphere-KG on Neo4j built-in visualization function. (a) shows the subgraph for $\langle City, isIn, Country \rangle$ and $\langle Country, isIn, Continent \rangle$; (b) shows the subgraph for $\langle Professor, hasResearchInterestOf, Research Interest \rangle$ and $\langle Professor, worksAt, University \rangle$.

Integration of LLMs with KG

The third step is the integration of LLMs and KG through the LangChain framework to support advanced question-answering-based functionalities. LLMs take questions as inputs, detect the question intention, and convert natural language question into machine-understandable format, namely Cypher statement for Neo4j. It supports two types of inquires satisfying applicants' diverse needs through question-answering process, including explicit graph search and implicit graph search. In particular, the explicit graph

search retrieves information existed in the graph database with explicit entities and relations extracted from questions. It also supports the exploration of research similarity among professors. In contrast, implicit graph search can return inferred entities and relations by using semantic similarity measurement of research interests, supporting program recommendation. The following subsections will introduce the conversion from natural language to Cypher statement in subsection (1), research similarity measurement among professors in subsection (2), and two types of inquiries in subsection (3) and (4).

(1) Query with natural language: GISphere-KG provides a natural language interface which receives questions in the natural language format. Since Neo4j database uses Cypher statements to retrieve information from graphs by following syntax standards, LLMs are leveraged to convert natural language question into Cypher statement, which is further leveraged to retrieve information from graph database. Traditionally, a few Cypher statements are predefined in the platform. However, predefined statements are not sufficient and flexible enough to handle various querying demands. Especially, queries with more complicated conditions might not even be considered and included in the platform. Moreover, it is more convenient to use natural language for questioning and querying without any prior knowledge required. Instead of setting up several Cypher statements that cannot cover all possible inquiries, the incorporation of the conversion to Cypher statement provides a flexible solution for both users and the following querying process. Users can more conveniently provide natural language questions with less trouble. Additionally, Cypher statements can be generated on-spot according to different questions.

(2) research similarity measure: When searching for a specific professor with certain research interests, applicants may want to discover other professors who have similar research interests in the research field. We have utilized LLMs to encode research interests into vectors as additional attributes, where the semantic similarity calculation directly uses the existing embeddings (Wang et al., 2018; Mai et al., 2018; Zhang et al., 2016; Sun et al., 2018) without needing encoding every time there is a request for similarity calculation, which further speeds up the efficiency. For semantic similarity calculation, the input question is embedded into a vector. It is further used to calculate semantic similarity between the input embedding and each existing vector. As a result, the system returns the top K results according to different information requests. We used “*isSimilarTo*” relationship to indicate research similarity between each pair of professors to discover professors who share the most similar research interests.

(3) Explicit graph search: GISphere-KG supports explicitly retrieving information in knowledge graph. Natural language question is converted into Cypher statement which further is fed into graph database (Neo4j) to search for knowledge from the database. Searched information can be pre-defined entities and their attributes, and relationships between entities. It also supports to search for professors sharing the similar research with the one the user provides by using research similarity measure. Below is an example of query for graph knowledge retrieval:

Question 1: Which professors have research interests in geovisualization?

(4) Implicit graph search for research-interest-based recommendation: Except for explicit graph search, GISphere-KG also supports research-interest-based recommendation and relevant graph retrieval which takes semantic similarity into account. Different from the semantic similarity to calculate professors' research similarity in previous subsection, this semantic similarity refers to each pair of research interests. It retrieves information that can be inferred by using similarity calculation for each pair of research interest embeddings. As a result, it recommends programs which are relevant to applicants' interested research interests. Below is an example:

Question 2: Which research interests align closely with map visualization?

Results

Figure 3 shows the user interface of GISphere-KG. It demonstrates the retrieval of existing information from the graph. By typing the question "Which professors have research interests in geovisualization?" the system first summarizes the question, then searches the keyword in our database, and finally lists out ten different professors in the answer. Moreover, with the support of Cypher generation prompt, it also briefly introduces each professor's other research interests and provides external links to their personal/institutional websites, as shown in Figure 3(a).

Figure 3(b) displays the search results of finding similar research interests based on applicants' input. Since the same research interest can be expressed in different ways using different words, the research interest passed by applicants to the system may have other similar research interests existing in the graph database. As a result, applicants can find out more research interests that they are interested in. In Figure 3(b), the platform interprets the question "What research interests align closely with map visualization?". First, it identifies the specific research interests, "map visualization". Then, it translates into Cypher statement under the guidance of designed prompt template and retrieves similar research directions from the graph database. Specifically, it embeds the question into a vector and fetches embeddings from the graph database to calculate the semantic similarity between the input and each existing embedding from the graph. Then the system can find the most similar research interests in the KG. After that, the system wraps up results and shows answers in a human-understandable way, providing applicants with the research interests similar to map visualization, including geographic visualization, geographic

visualization/visual analytics, cartographic visualization, and cartography and geovisualization.



which professors have research interest in geovisualization?



The people who have a research interest in geovisualization are:

1. Thierry Joliveau, his other interests include spatial analysis, cartography, mapping. You can find more about him [here](#).
2. Christian Kaiser, he is also interested in Spatial Analysis, Spatial databases, Socio-economic simulation, and modeling. More about him can be found [here](#).
3. Ron van Lammeren, his additional interests are GIS Education, Geo-information in landscape. For more information, visit his profile [here](#).
4. Olivier Klein, he also researches Daily Mobility, Urban Geography, Spatial Analysis, GIS. His profile can be found [here](#).
5. Phil James, his research interests also include smart cities, urban data analytics, geospatial database systems. More about him can be found [here](#).
6. Jon Mills, he is also interested in 3D Modelling, LiDAR, geospatial analytics. You can learn more about him [here](#).
7. Mairead de Roiste, she researches GIS incorporating usability, public participation GIS, GIS pedagogy, and capability building. More about her is available [here](#).
8. Antoni Moore, his other interests are GIS, spatial analysis, geo-database, spatial modeling. You can find more about him [here](#).
9. André Skupin, he is also interested in geocomputation, spatial analysis, applied GIS, data science. His profile can be found [here](#).
10. Xiao Huang, he also researches GeoAI, deep learning, natural hazards, human-environment

What is up?



(a) Explicit graph search



Which research interests align closely with map visualization?



The research interests that align closely with map visualization are:

1. Geographic visualization: This is a technique used in cartography and geographic visualization. It is associated with professors Somayeh Dodge, Claus Rinner, Tomasz Opach.
2. Geographic visualization / visual analytics: This is also a form of map visualization, associated with professor Jukka Krisp.
3. Cartographic visualization: This involves the use of cartography to create visual representations of geographic data. It is associated with professor PUN, Cheng Shuk-ching, Lilian.
4. Cartography and geo-visualization: This is a combination of cartography and geographic visualization. It is associated with professor Martijn Meijers.

(b) Implicit graph search

Figure 3: Query results visualization. (a) shows explicit graph search. (b) gives one query example to show implicit graph search

References

- Wang, Y., Kang, Y., Liu, H., Hou, C., Zhou, B., Ye, S., ... & Gao, S. (2023). Choosing GIS graduate programs from afar: Chinese students' perspectives. *Transactions in GIS*, 27(2), 450-475.
- Longley, P. A. (2000). The academic success of GIS in geography: Problems and prospects. *Journal of Geographical Systems*, 2(1), 37-42. <https://doi.org/10.1007/s101090050027>
- Rickles, P., Ellul, C., & Haklay, M. (2017). A suggested framework and guidelines for learning GIS in interdisciplinary research. *Geo: Geography and Environment*, 4(2), e00046. <https://doi.org/10.1002/geo2.46>
- Li, W., Wang, S., Wu, S., Gu, Z., & Tian, Y. (2022). Performance benchmark on semantic web repositories for spatially explicit knowledge graph applications. *Computers, environment and urban systems*, 98, 101884.
- Li, W., Wang, S., Chen, X., Tian, Y., Gu, Z., Lopez-Carr, A., ... & Zhu, R. (2023). Geographvis: a knowledge graph and geovisualization empowered

- cyberinfrastructure to support disaster response and humanitarian aid. *ISPRS International Journal of Geo-Information*, 12(3), 112.
- Wang, Meng, Ruijie Wang, Jun Liu, Yihe Chen, Lei Zhang, and Guilin Qi. Towards empty answers in SPARQL: approximating querying with RDF embedding. *In The Semantic Web–ISWC 2018: 17th International Semantic Web Conference*, Monterey, CA, USA, October 8–12, 2018, Proceedings, Part I 17, pp. 513-529. Springer International Publishing, 2018.
- Mai, Gengchen, Krzysztof Janowicz, and Bo Yan. Combining Text Embedding and Knowledge Graph Embedding Techniques for Academic Search Engines. *In Semdeep/NLIWoD@ ISWC*, pp. 77-88. 2018.
- Zhang, Fuzheng, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. Collaborative knowledge base embedding for recommender systems. *In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 353-362. 2016.
- Sun, Haitian, Bhuwan Dhingra, Manzil Zaheer, Kathryn Mazaitis, Ruslan Salakhutdinov, and William Cohen. "Open Domain Question Answering Using Early Fusion of Knowledge Bases and Text." *In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 4231-4242. 2018.
- Zhining Gu**, PhD student, School of Geographical Sciences & Urban Planning, Arizona State University, Tempe, AZ, USA
- Wenwen Li**, Professor, School of Geographical Sciences & Urban Planning, Arizona State University, Tempe, AZ, USA
- Bing Zhou**, PhD student, Department of Geography, Texas A&M University, College Station, TX 77843
- Yikang Wang**, PhD student, Centre of Advanced Spatial Analysis, University College London, London, UK
- Yanbing Chen**, PhD student, Department of Geography, University of Wisconsin – Madison, Madison, WI, USA
- Shan Ye**, Assistant Professor, School of Information Engineering, China University of Geosciences, Beijing, China
- Kejin Wang**, PhD student, Department of Environmental Sciences, Louisiana State University, Baton Rouge, LA, USA
- Hongkai Gu**, PhD student, Department of Geography, Ghent University, Ghent, Belgium
- Yuhao Kang**, Assistant Professor, GISense Lab, Department of Geography, University of South Carolina, Callcott, Columbia, SC, USA