

Option 1: Data Analysis – Project

This is an exploratory project focusing on practical data analysis in the context of modern tools and technologies such as SQL, social networks analytics, knowledge graphs, Neo4j, LLMs, and vector databases. You are encouraged to select an interesting problem or application domain that leverages these technologies. Your project will consist of the following three phases:

Data Collection and Preparation: Identify and obtain datasets relevant to your selected domain. This could involve finding datasets that fit in a relational database, social network datasets, or working with graph databases like Neo4j.

Data Analysis and Visualization: Process and visualize the data to uncover meaningful insights. This could include running SQL queries, visualizing social network connections, or querying knowledge graphs to uncover relationships.

Insight Discovery: Pose research questions and answer them using your analysis.

At least one phase should be non-trivial. For example, preprocessing large-scale graph datasets or working with APIs to extract complex social network data could be challenging. Similarly, using or customizing advanced algorithms for large-scale social or knowledge graph analysis or vector database indexing in retrieval augmented generation (RAG) using vector databases may also pose significant challenges.

You should submit a report describing your work. The report should be around 14 pages in LNCS format or 6 pages in IEEE conference format. Google overleaf lncs or ieee for latex templates.

Option 2: Research Paper – Project

This is an exploratory research project aimed at studying, implementing, and experimenting with modern data analysis techniques using technologies like SQL, social networks, knowledge graphs, Neo4j, LLMs, or vector databases.

Your project can involve the following phases:

Literature Review and Problem Exploration: Study **key recent** (last 2-3 years) research papers or articles related to your chosen domain. Select one paper for further analysis. Explain the problem the paper uses, the datasets it analyses, and the key algorithms or tools involved.

Implementation and Development: Implement the main techniques discussed in your chosen paper. This could include writing code for analysis, implementing knowledge graph queries in Cypher, or training vector embeddings for semantic search tasks.

Experimentation and Analysis: Experiment with the implemented techniques on real-world datasets. Analyze their performance, compare them with alternative methods, and highlight interesting findings.

At least one phase should be non-trivial. For example, a paper on advanced knowledge graph algorithms might require significant effort to implement, or experimenting with large-scale vector databases could reveal important performance considerations.

You should submit a report describing your work. The report should be around 14 pages in LNCS format or 6 pages in IEEE conference format. Google overleaf lncs or ieee for latex templates.