ASSIGNMENT SUBMISSION FORM

Course Name: **Advanced Management Program In Business Analytics**
Assignment Title: **Networks and Graphs**
Submitted by: **Group – 9**

(Student name or group names)

Student Name	PG ID
Vinay Raghava Pothunaku	12310033
Pranjal Kaushik	12310014
Kanishk Gupta	12310008
Sandeep Vudata	12220011

ISB Honour Code

- I will represent myself in a truthful manner.
- I will not fabricate or plagiarise any information with regard to the curriculum.
- I will not seek, receive or obtain an unfair advantage over other students.
- I will not be a party to any violation of the ISB Honour Code.
- I will personally uphold and abide, in theory and practice, the values, purpose and rules of the ISB Honour Code.
- I will report all violations of the ISB Honour Code by members of the ISB community.
- I will respect the rights and property of all in the ISB community.
- I will abide by all the rules and regulations that are prescribed by ISB.

Note: Lack of awareness of the ISB Honour Code is never an excuse for a violation. Please go through the Honour Code in the student handbook, understand it completely. Please also pay attention to the following points:

- Please do not share your assignment with your fellow students under any circumstances if the Honour Code scheme prohibits it. The HCC considers both parties to be guilty of an Honour Code violation in such circumstances.
- If the assignment allows you to refer to external sources, please make sure that you cite all your sources. Any material that is taken verbatim from an external source (website, news article etc.) must be in quotations. A much better practice is to paraphrase the source material (it still must be cited).

Part 1-A:

1. **Context/Background:**
   Healthcare data is often siloed within individual hospitals and clinics, making it challenging for healthcare providers to obtain a comprehensive view of a patient's medical history.
   - **Duplication of Tests:** Without a unified database system, medical tests are often duplicated, wasting patients' time and resources.
   - **Incomplete Medical Histories:** Patients may not always provide a complete and accurate medical history, which can result in wasted time for doctors and potentially incorrect treatments.
   - **Fragmented Data:** The fragmentation of medical data makes it difficult for healthcare providers to access comprehensive patient information, hindering effective treatment.
   - **Research Limitations:** The lack of centralized data restricts researchers from analyzing health trends and patterns across the state, limiting the ability to conduct thorough medical research.
   - **Delayed Detection of Contagious Diseases:** Without a central system, early detection of flues and other contagious diseases is challenging, increasing the risk of outbreaks.

2. **In Scope And Out Scope Use cases:**
   We aim to develop a centralized database system for Telangana, utilizing knowledge graphs to unify the data from all hospitals, clinics, patients, and doctors across the region.
   In-Scope:
   1. **Centralized knowledge model graph creation:**
      - Development of a centralized patient knowledge graph model integrating records from all hospitals and clinics.
      - Inclusion of patient data such as personal information, medical history, diseases, symptoms, medications, doctors, and diet plans.
   2. **Best Doctor Recommendations:**
      - Build a system to recommend the best doctors for consultation based on patient symptoms, medical history, and location.
   3. **Suggesting Pre-Consultation Reports:**
      - Based on the symptoms, the system will suggest necessary medical tests and reports that the patient should obtain before consulting a doctor, saving time for patients and doctors.
   4. **Medical Records Summarization for doctors:**
      - Development of a model to summarize past medical reports, making it easier for doctors to understand a patient's medical history quickly.
   5. **Diet Plan Recommendations:**
      - Recommending diet plans based on the doctor’s prescription, tailored to the patient’s medical needs and conditions.
   6. **Research and Outbreak Detection:**
      - Using the centralized data for research purposes to detect outbreaks of contagious diseases.

Out scope:

1. **Insurance Integration:**
    - The system will not include integration of insurance information or billing processes.
2. **Geographic Limitation:**
    - The system is restricted to handling data only within the Telangana state and does not cover other regions.

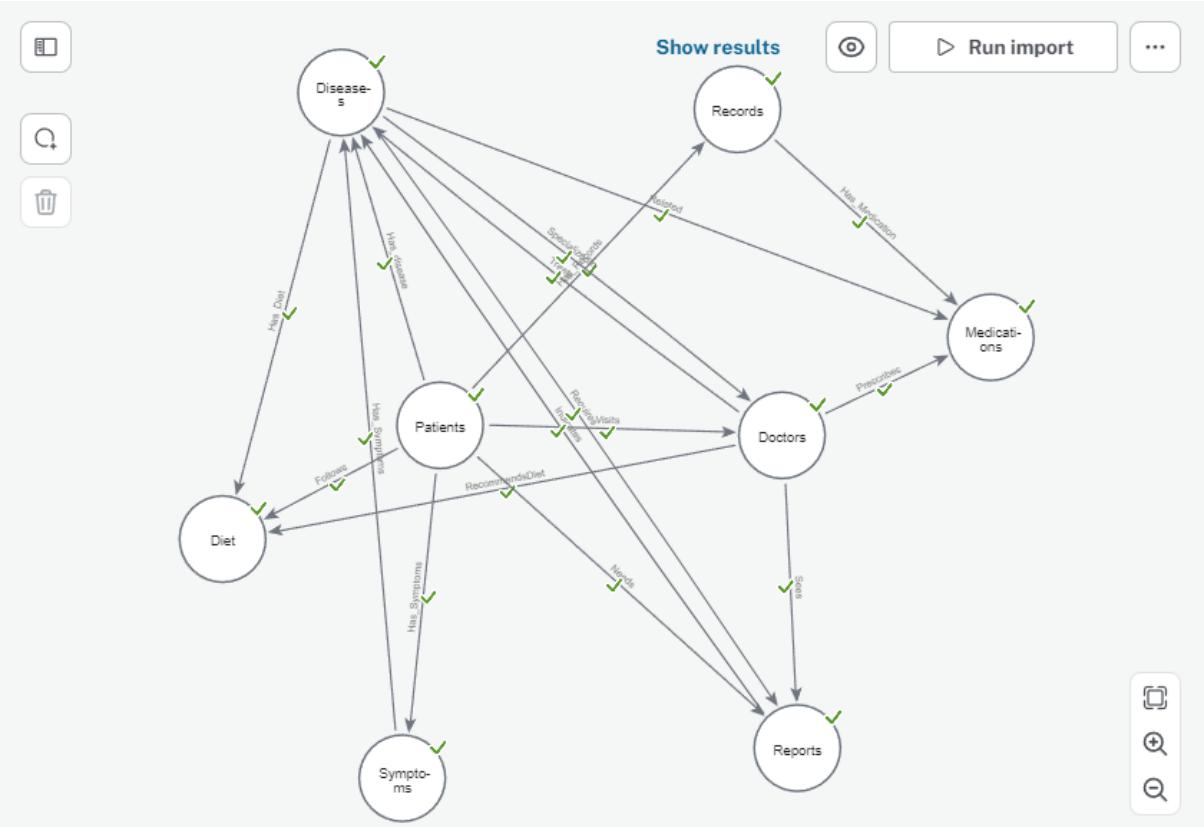
3. **Assumptions**:
    1. **Consent for Participation:**
        - It is assumed that all hospitals, clinics, and patients have consented to participate in the centralized knowledge graph model and share their data for the purposes outlined.
    2. **Research Capabilities:**
        - Researchers will have access to anonymized data to detect outbreaks and analyze health trends, but they will not have access to real-time data feeds.
    3. **Stakeholder Participation:**
        - Healthcare providers, including doctors and hospital administrators, will actively participate in implementing and using the centralized knowledge graph model.
    4. **Data Availability and Quality:**
        - It is assumed that all hospitals and clinics in Telangana will provide accurate and up-to-date patient data for integration into the centralized knowledge graph model.
    5. **Regulatory Compliance:**
        - The project will comply with all relevant healthcare regulations and data protection laws.
    6. **Data Privacy and Security:**
        - Data will not be shared with private pharmaceutical companies, ensuring that patient information remains confidential and is used solely for healthcare improvement and research within the public domain.
4. **Reference Links:**
        - CentralisedHealthData

Part-1B:

Please find the neo4j model in the attachments.

Part 1 – c:

This knowledge graph is designed to facilitate the development of an application that links patient data, diseases, symptoms, records, medications, doctors, and diets. Its primary purpose is to recommend the best doctors for consultations and suggest suitable diet plans for patients based on their medical conditions. Additionally, we have created a new entity for reports to advise patients on the necessary medical tests to obtain before visiting a doctor based on their symptoms and medications.



Patients & Doctors Knowledge Graph Network

Entities and Relationships:

In this knowledge graph, we have utilized a total of 8 entities: patient data, diseases, symptoms, records, medications, doctors, diets, and reports. Furthermore, we have developed numerous relationships among these entities to construct a comprehensive and interconnected model.

1. **Patients:** Represents individuals receiving healthcare services. Key attributes include ID, Name, Age, Gender, ContactInfo, and Location.
2. **Diseases:** Represents medical conditions. Key attributes include ID, Name, and Description.
3. **Symptoms:** Represents clinical signs associated with diseases. Key attributes include ID, Name, and Description.
4. **Records:** Represents medical records of patients. Key attributes include ID, Name, Dosage, Frequency, and Notes.
5. **Medications:** Represents pharmaceutical drugs prescribed to patients. Key attributes include ID, Name, Dosage, and Frequency.

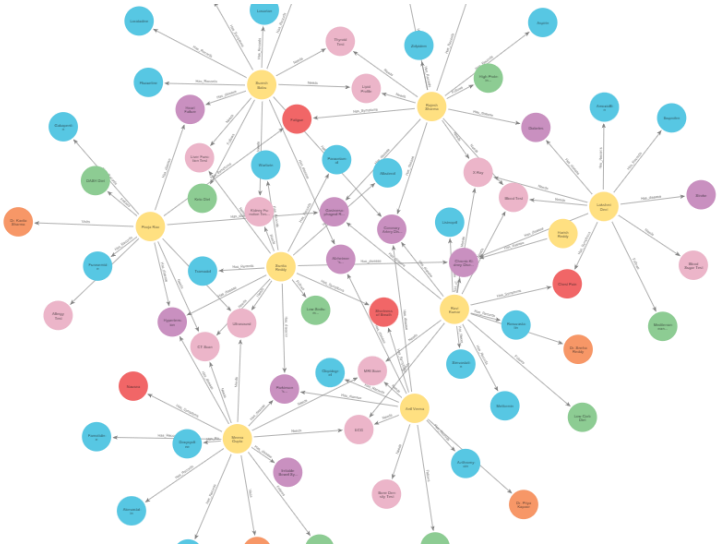
6. **Doctors:** Represents healthcare professionals. Key attributes include ID, Name, Specialty, ContactInfo, Location, and Hospital.
7. **Diets:** Represents dietary plans. Key attributes include ID, Name, and Description.
8. **Reports**: Represents medical reports that indicate various health conditions. Key attributes include ID, Name, and Description.

Relationships:

1. **HAS_DISEASE:** Connects patients to their diagnosed diseases.
2. **HAS_SYMPTOM:** Connects diseases to their associated symptoms.
3. **HAS_RECORD:** Connects patients to their medical records.
4. **INCLUDES_MEDICATION**: Connects records to the medications prescribed.
5. **VISITS:** Connects patients to the doctors they consult.
6. **TREATS:** Connects doctors to the diseases they specialize in treating.
7. **HAS_DIET:** Connects diseases to the recommended dietary plans.
8. **FOLLOWS:** Connect patients to the diets they follow.
9. **SEES:** Connects doctors to the reports they review.
10. **NEEDS_REPORT:** Connects patients to the reports they need based on their symptoms.
11. **REQUIRES:** Connects diseases to the required medications or treatments.
12. **RELATED:** Connects symptoms to related diseases.
13. **SPECIALIZED_IN:** Connects doctors to the diseases they have specialized in.
14. **PRESCRIBES:** Connects doctors to the medications they prescribe.
15. **RECOMMENDS_DIET:** Connects doctors to the diets they recommend to patients.
16. **INDICATES:** Connects reports to the diseases they indicate.

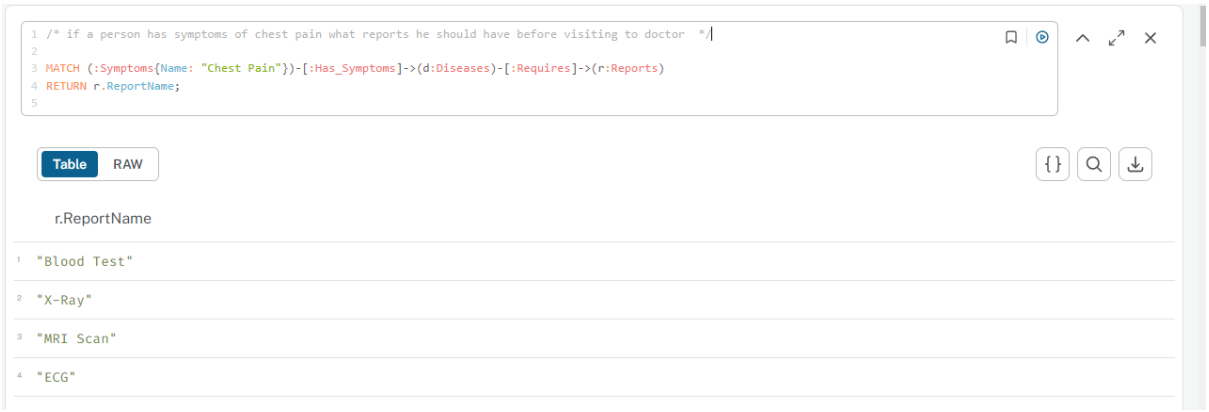
Utilizing the above entities and relationships, we have developed a system that recommends doctors and medications to patients based on their symptoms. Through querying, the system can also summarize a patient's medical history for doctors. Furthermore, we can apply an LLM RAG-based approach to generate comprehensive summaries of patient data for a doctor.

Here is the resultant visualization of the knowledge graph.

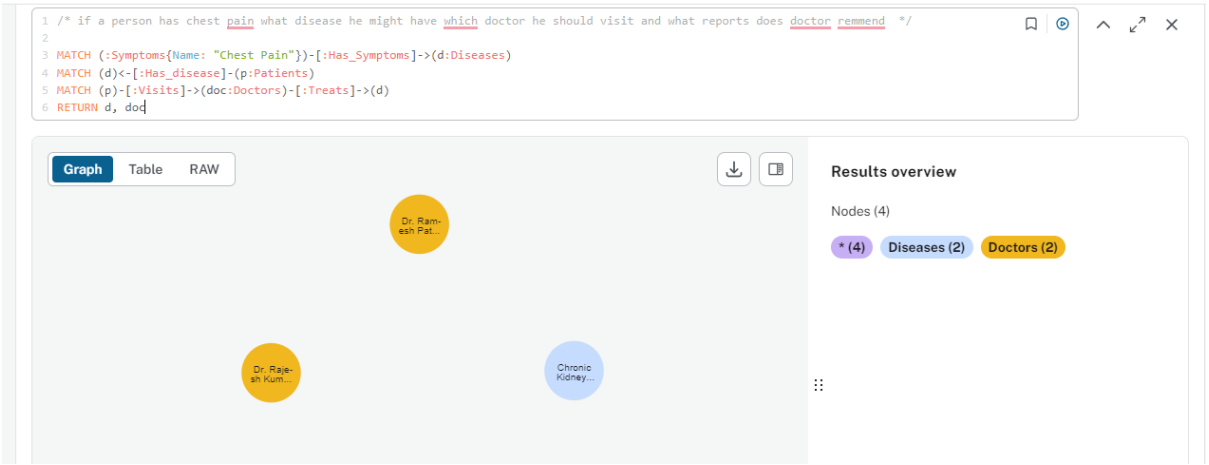


Here are some of the query results.

Query – 1:



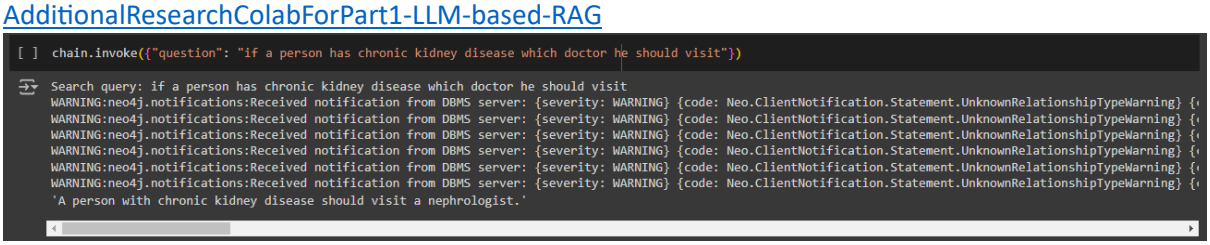
Query-2:



The screenshots above demonstrate that our knowledge graph has successfully generated recommendations for the appropriate doctors to consult and the necessary reports to obtain based on the presented symptoms.

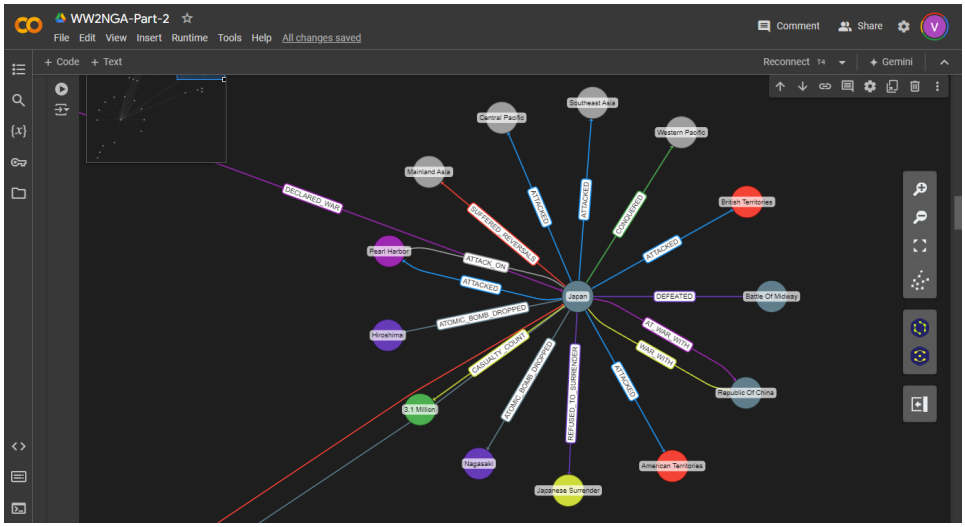
Additional:

Furthermore, we have enhanced the system by implementing an LLM-based RAG approach. The Colab notebook can be accessed here:
[AdditionalResearchColabForPart1-LLM-based-RAG]

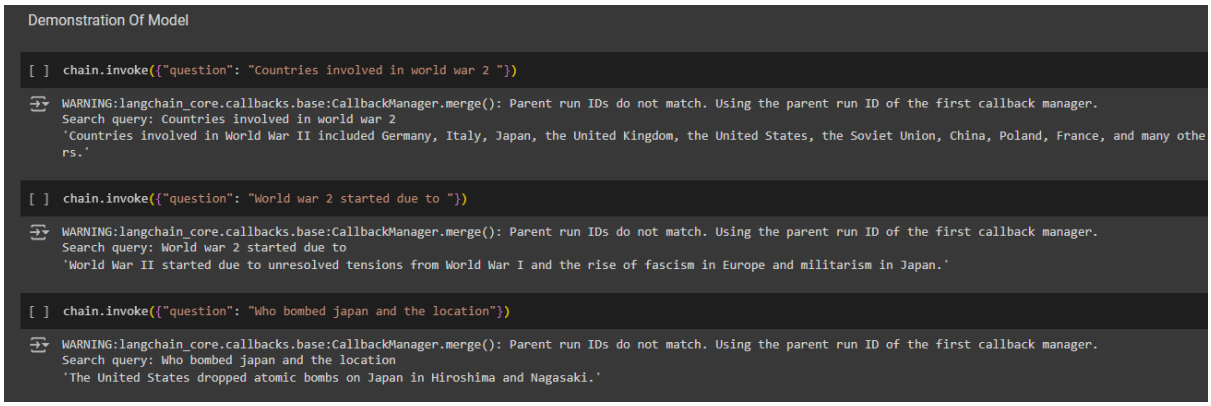


Part – 2:

We selected World War II as the focus for our Knowledge Graph-enhanced RAG implementation using Wikipedia data. We employed the OpenAI API as the LLM and used Neo4J to construct the knowledge graph. We built a corresponding knowledge graph after extracting data from the Wikipedia page. We then applied an LLM-based RAG to generate answers to relevant queries.



Here are few some of the results of the queries:



The Python notebook is attached in the attachments.