Question Answering System for predicting a Disease

By: Priyabrata Mallick, Himanshu Punetha, Pranav Nambiar
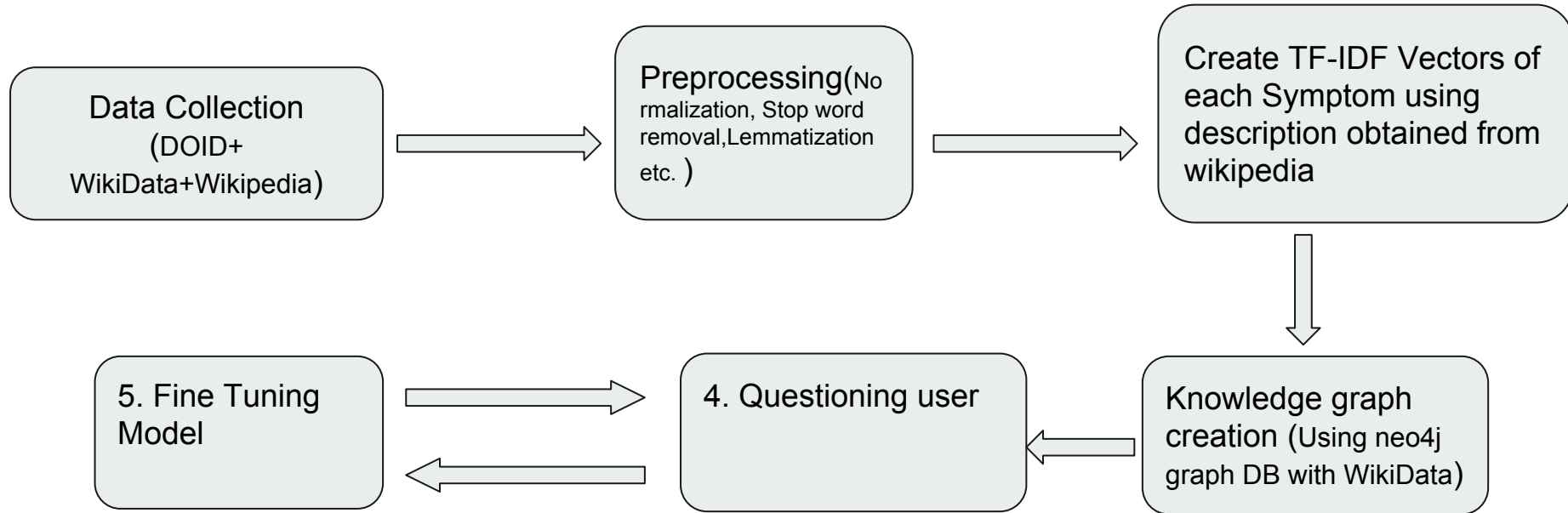
Motivation and Aim

There are various communities for medical question answers in which users asks the questions which are then answered by a Doctor. But for this doctor needs to be present 24x7 to answer the query even to answer some common symptom based disease.
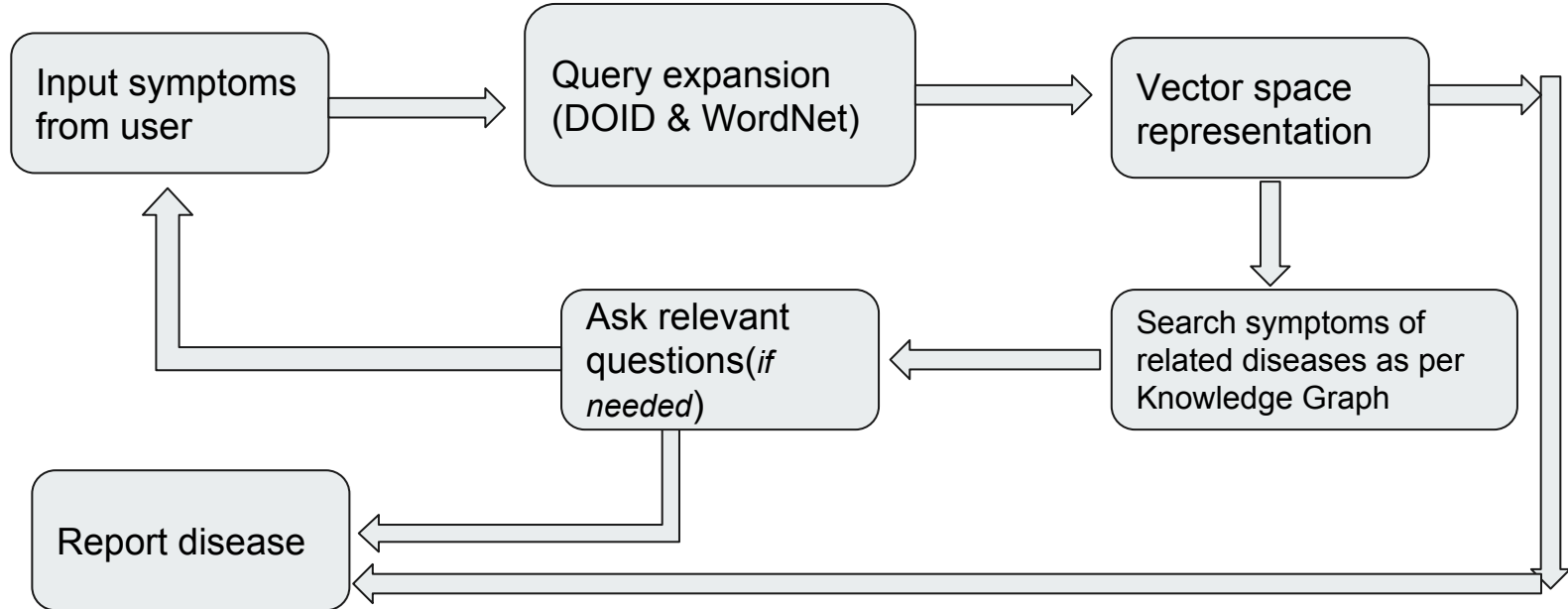
Our aim is to build a system which does following: -

- Takes unstructured symptom as input Uses some existing Medical ontology (example - UMLS, Human Disease Ontology) to find out synonymous medical terms for different symptoms
- Takes unstructured query from user and asks questions to identify his symptoms in order to narrow down possible diseases
- Ranks the diseases based on the relevance based on symptoms identified and returns result.

Workflow:

Data Collection (DOID+ WikiData+Wikipedia) → Preprocessing(Normalization, Stop word removal,Lemmatization etc.) → Create TF-IDF Vectors of each Symptom using description obtained from wikipedia → Knowledge graph creation (Using neo4j graph DB with WikiData) → 4. Questioning user → 5. Fine Tuning Model

Querying Process

Model Architecture

Input query: An unstructured user query with some relevant information is taken as input from the user.

Query expansion: Unstructured query is expanded to find the most appropriate match in the knowledgebase.

- DOID: Disease Ontology is used to enrich the knowledge graph.
- WordNet: Used to get more semantic similarity between user query and existing knowledgebase

Vector space representation:

- TFIDF
- Word2Vec

Model Architecture

Search symptoms in KG:

- TF IDF Matrix for unstructured description of each symptom
- TF IDF vector representation for unstructured query
- Cosine similarity between them
- Find diseases related to input symptoms and retrieve other Symptoms of those diseases from Knowledge graph (Using **CYPHER** Queries)
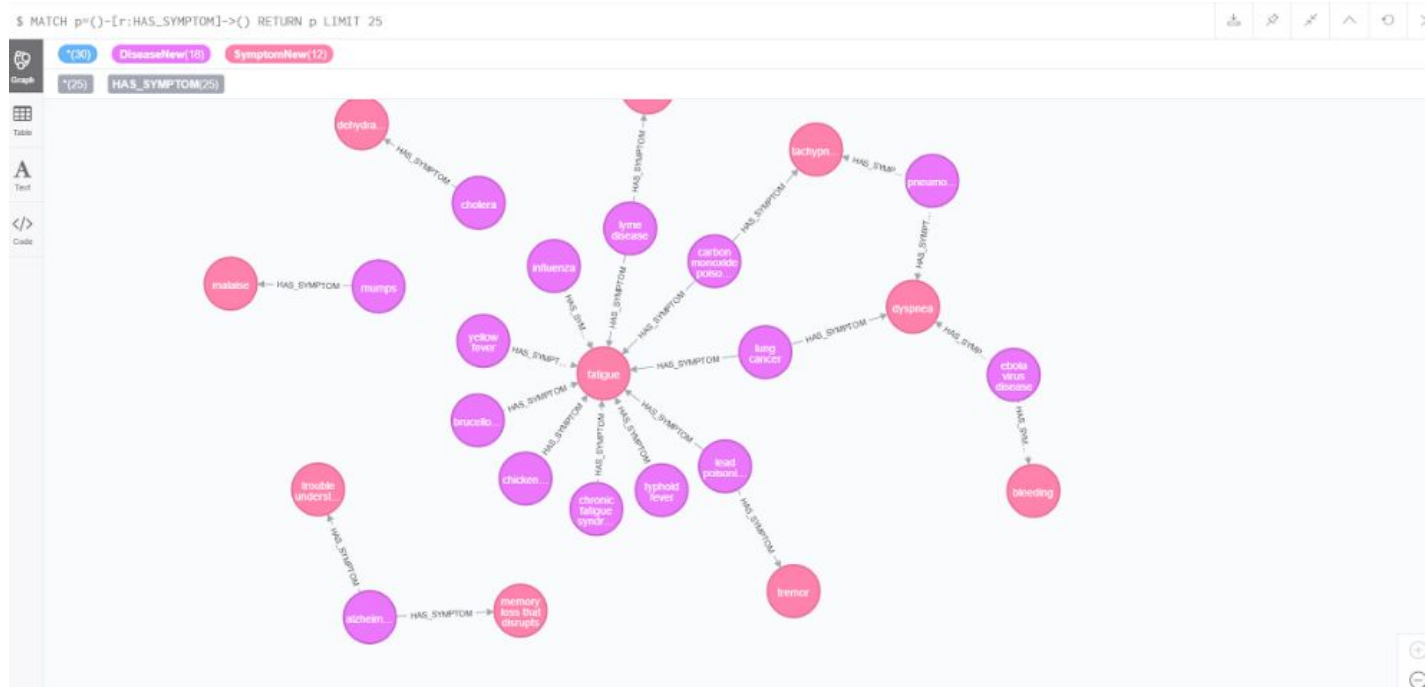
Ask relevant questions to user:

- Question is framed by identifying the symptoms of the nearest disease to our symptoms.

Report disease:

- Report the disease for which max number of symptoms matches.

Knowledge Graph

Evaluation

- **Manual Identification of symptom specific questions from WebMD**
- **Selected questions are provided as input to the model**
- **Results of the model compared with the WebMD's answer manually**
- **We observed an accuracy of (50-60)% of the cases**

Sample result:


```
C:\Users\hpunetha\PycharmProjects\Assignment4\Anaconda\Scripts\python.exe C:/Users/hpunetha/PycharmProjects/ProjectIR/queryExpansion.py
Please enter your age group:(1:less than 10, 2: 10-50 , 3: 50 & above): 2
Please enter your gender(m/f): m
Please enter the list of symptoms you are facing: Heart palpitations, shortness of breath, weakness
Q217111 bradycardia
Q189331 Heart arrhythmia
Q5445 anemia
['Q217111', 'Q189331']
'Q217111'
(typhoid fever)-[ Has_Symptom ]->(bradycardia)
(yellow fever)-[ Has_Symptom ]->(bradycardia)
==============================================================
(typhoid fever)-[ Has_Symptom ]->(Q29644032, continuous fever),  (a type of fever)
(typhoid fever)-[ Has_Symptom ]->(Q7127114, paleness),  ()
(typhoid fever)-[ Has_Symptom ]->(Q18554982, intestinal perforation),  (human disease)
(typhoid fever)-[ Has_Symptom ]->(Q1495657, gastrointestinal bleeding),  (bleeding from any part of the gastrointestinal tract)
(typhoid fever)-[ Has_Symptom ]->(Q653197, rash),  ()
(typhoid fever)-[ Has_Symptom ]->(Q9690, fatigue),  (range of afflictions, usually associated with physical and/or mental weakness)
(typhoid fever)-[ Has_Symptom ]->(Q178436, constipation),  (bowel dysfunction that is characterized by infrequent or difficult evacuation of feces.)
(typhoid fever)-[ Has_Symptom ]->(Q86, headache),  (pain in the head or neck)
(yellow fever)-[ Has_Symptom ]->(Q192120, cyanosis),  (medical diagnosis)
(yellow fever)-[ Has_Symptom ]->(Q38933, fever),  (common medical sign characterized by elevated body temperature)
(yellow fever)-[ Has_Symptom ]->(Q127076, vomiting),  (involuntary, forceful expulsion of stomach contents through the mouth and sometimes the nose)
(yellow fever)-[ Has_Symptom ]->(Q9690, fatigue),  (range of afflictions, usually associated with physical and/or mental weakness)
(yellow fever)-[ Has_Symptom ]->(Q86, headache),  (pain in the head or neck)
'Q189331'
(lyme disease)-[ Has_Symptom ]->(heart arrhythmia)
==============================================================
(lyme disease)-[ Has_Symptom ]->(Q213433, vertigo),  (type of dizziness)
(lyme disease)-[ Has_Symptom ]->(Q1397878, fasciculation),  ()
```