**CYPHER CODE**

Creation of nodes and relationships

**Creation of constraints:**

CREATE CONSTRAINT diseaseIdConstraint FOR (d:Disease) REQUIRE d.id IS UNIQUE

Added 1 constraint, completed after 267 ms.

CREATE CONSTRAINT symptomIdConstraint FOR (s:Symptom) REQUIRE s.id IS UNIQUE

Added 1 constraint, completed after 91 ms.

Index for both disease and symptom node kind is directly created by the creation of constraints.

**Creation of nodes of symptoms and diseases:**

Python Code for generating desired cypher code is uploaded on GitHub.

Added 894 labels, created 894 nodes, set 3739 properties, completed after 4815 ms.

Added 3316 labels, created 3316 nodes, set 16580 properties, completed after 41574 ms.

Added 4705 labels, created 4705 nodes, set 23525 properties, completed after 83401 ms.

Added 3529 labels, created 3529 nodes, set 14116 properties, completed after 31235 ms

In total, 11550 diseases are added.

**Creation of relationships between diseases:**

//To delete any pre-existing relationships

MATCH ()-[r]->() DELETE r

:auto LOAD CSV WITH HEADERS FROM 'file:////right\_disease\_relationships.csv' AS csvLine

CALL {

WITH csvLine

MATCH (parent:Disease {id:csvLine.parent\_id}), (child:Disease {id: csvLine.child\_id})

CREATE (child)-[:Parent\_Disease]->(parent)

} IN TRANSACTIONS OF 2 ROWS

Created 16084 relationships, completed after 14818 ms.

**Creation of relationships between symptoms:**

:auto LOAD CSV WITH HEADERS FROM 'file:////symptom\_relationships.csv' AS csvLine

CALL {

WITH csvLine

MATCH (parent:Symptom {id:csvLine.parent\_id}), (child:Symptom {id: csvLine.child\_id})

CREATE (child)-[:Parent\_Symptom]->(parent)

} IN TRANSACTIONS OF 2 ROWS

Created 895 relationships, completed after 4305 ms.

**Creation of relationships between symptoms and diseases:**

:auto LOAD CSV WITH HEADERS FROM 'file:////final\_dis\_symp\_relations.csv' AS csvLine

CALL {

WITH csvLine

MATCH (corres\_disease:Disease {id:csvLine.Disease\_id}), (corres\_symptom:Symptom {id: csvLine.Symptom\_id})

CREATE (corres\_disease)-[:corresponding\_Symptom]->(corres\_symptom)

} IN TRANSACTIONS OF 2 ROWS

Created 27793 relationships, completed after 24885 ms.

Validation of the graph

**Single Symptom Validation:**

Sample Code provided for finding associated diseases given a symptom called "Low Back Pain".

1. **For fetching directly related diseases:**
2. The diseases which are directly related to the given symptom are retrieved.

MATCH (s:Symptom)

WHERE toLower(s.name) CONTAINS toLower("Low Back Pain")

OR any(synonym IN s.synonyms WHERE toLower(synonym) CONTAINS toLower("Low Back Pain"))

OPTIONAL MATCH (s)-[r]-(related:Disease)

RETURN s, r, related

1. To check if the most commonly diagnosed diseases for the given symptom are retrieved by the above code.

First, check if each disease in the given disease list is a disease node.

WITH ["Lumbosacral radiculopathy", "Inflammatory low back pain", "Degenerative Disc Disease", "Spinal Stenosis", "Spondylolisthesis"] AS checkList

// Match any nodes that are either Disease or Symptom and check if their name or synonyms match the list

MATCH (n)

WHERE (n:Disease OR n:Symptom) AND

(toLower(n.name) IN [name IN checkList | toLower(name)]

OR any(synonym IN n.synonyms WHERE toLower(synonym) IN [name IN checkList | toLower(name)]))

RETURN n

To check if the most commonly diagnosed diseases for the given symptom are retrieved.

WITH ["Lumbosacral radiculopathy", "Inflammatory low back pain", "Degenerative Disc Disease", "Spinal Stenosis", "Spondylolisthesis"] AS diseaseList

MATCH (s:Symptom)

WHERE toLower(s.name) CONTAINS toLower("Low Back Pain")

OR any(synonym IN s.synonyms WHERE toLower(synonym) CONTAINS toLower("Low Back Pain"))

OPTIONAL MATCH (s)-[r]-(related:Disease)

WHERE toLower(related.name) IN [term IN diseaseList | toLower(term)]

OR any(synonym IN related.synonyms WHERE toLower(synonym) IN [term IN diseaseList | toLower(term)])

RETURN s, r, related

1. **For fetching directly related diseases and directly related symptoms:**
2. The diseases and symptoms which are directly related to the given symptom are retrieved.

MATCH (s:Symptom)

WHERE toLower(s.name) CONTAINS toLower("Low Back Pain")

OR any(synonym IN s.synonyms WHERE toLower(synonym) CONTAINS toLower("Low Back Pain"))

OPTIONAL MATCH (s)-[r]-(related)

RETURN s, r, related

1. To check if the most commonly diagnosed diseases for the given symptom are retrieved by the above code.

// List of diseases to match against

WITH ["Lumbosacral radiculopathy", "Inflammatory low back pain", "Degenerative Disc Disease", "Spinal Stenosis", "Spondylolisthesis"] AS diseaseList

MATCH (s:Symptom)

WHERE toLower(s.name) CONTAINS toLower("Low Back Pain")

OR any(synonym IN s.synonyms WHERE toLower(synonym) CONTAINS toLower("Low Back Pain"))

OPTIONAL MATCH (s)-[r]-(related)

// Check if the related node is a Disease and matches the given list

WITH s, r, related, diseaseList

WHERE related:Disease AND (

toLower(related.name) IN [term IN diseaseList | toLower(term)] OR

any(synonym IN related.synonyms WHERE toLower(synonym) IN [term IN diseaseList | toLower(term)])

)

// Return only the diseases from the list that matched

RETURN s, r, related

1. **For fetching directly related diseases, directly related symptoms, and the sub-diseases to the aforementioned diseases:**
2. The diseases and symptoms which are directly related to the given symptom are retrieved.

The diseases which are sub-diseases or child diseases to the diseases which were directly related to the given symptom are also retrieved.

MATCH (s:Symptom)

WHERE toLower(s.name) CONTAINS toLower("Low Back Pain")

OR any(synonym IN s.synonyms WHERE toLower(synonym) CONTAINS toLower("Low Back Pain"))

OPTIONAL MATCH (s)-[r]-(d:Disease)

OPTIONAL MATCH (s)-[r1]-(relatedSymptom:Symptom)

OPTIONAL MATCH (d)<-[r2]-(relatedDisease1:Disease)

RETURN s, r, d, r1, relatedSymptom, r2, relatedDisease1

1. To check if the most commonly diagnosed diseases for the given symptom are retrieved by the above code.

MATCH (s:Symptom)

WHERE toLower(s.name) CONTAINS toLower("Low Back Pain")

OR any(synonym IN s.synonyms WHERE toLower(synonym) CONTAINS toLower("Low Back Pain"))

OPTIONAL MATCH (s)-[r]-(d:Disease)

OPTIONAL MATCH (s)-[r1]-(relatedSymptom:Symptom)

OPTIONAL MATCH (d)<-[r2]-(relatedDisease1:Disease)

WITH s, r, d, r1, relatedSymptom, r2, relatedDisease1,

[disease IN [d.name, d.synonyms] WHERE toLower(disease) IN [toLower("Lumbosacral radiculopathy"), toLower("Inflammatory low back pain"), toLower("Degenerative Disc Disease"), toLower("Spinal Stenosis"), toLower("Spondylolisthesis")] | disease] AS matchedDiseases

RETURN DISTINCT d

1. **For fetching directly related diseases, directly related symptoms, and diseases related to all the symptoms and sub-diseases to all the diseases:**
2. The diseases and symptoms which are directly related to the given symptom are retrieved.

The diseases related to these symptoms (the symptoms which are related to the given symptom) are also retrieved.

The diseases which are sub-diseases or child diseases to all these diseases (to the

diseases which were directly related to the given symptom and directly related to other

related symptoms) are also retrieved.

// Match symptom first

MATCH (s:Symptom)

WHERE toLower(s.name) CONTAINS toLower("Low Back Pain")

OR any(synonym IN s.synonyms WHERE toLower(synonym) CONTAINS toLower("Low Back Pain"))

// Match directly related disease nodes

OPTIONAL MATCH (s)-[r]-(relatedDisease:Disease)

// Match symptoms related to the initial symptom

OPTIONAL MATCH (s)-[r1]-(relatedSymptom:Symptom)

// Match diseases related to these related symptoms

OPTIONAL MATCH (relatedSymptom)-[r2]-(relatedDisease1:Disease)

// Match diseases related to the directly related diseases

OPTIONAL MATCH (relatedDisease)<-[r3:Parent\_Disease]-(relatedDisease2:Disease)

// Match diseases related to the diseases from related symptoms

OPTIONAL MATCH (relatedDisease1)<-[r4:Parent\_Disease]-(relatedDisease3:Disease)

// Return all matched nodes and relationships

RETURN s, r, relatedDisease, r1, relatedSymptom, r2, relatedDisease1, r3, relatedDisease2, r4, relatedDisease3

1. To check if the most commonly diagnosed diseases for the given symptom are retrieved by the above code.

WITH ["Lumbosacral radiculopathy", "Inflammatory low back pain", "Degenerative Disc Disease", "Spinal Stenosis", "Spondylolisthesis"] AS diseaseList

// Match the symptom first

MATCH (s:Symptom)

WHERE toLower(s.name) CONTAINS toLower("Low Back Pain")

OR any(synonym IN s.synonyms WHERE toLower(synonym) CONTAINS toLower("Low Back Pain"))

// Match directly related disease nodes

OPTIONAL MATCH (s)-[r]-(relatedDisease:Disease)

// Match symptoms related to the initial symptom

OPTIONAL MATCH (s)-[r1]-(relatedSymptom:Symptom)

// Match diseases related to these related symptoms

OPTIONAL MATCH (relatedSymptom)-[r2]-(relatedDisease1:Disease)

// Match diseases related to the directly related diseases

OPTIONAL MATCH (relatedDisease)<-[r3:Parent\_Disease]-(relatedDisease2:Disease)

// Match diseases related to the diseases from related symptoms

OPTIONAL MATCH (relatedDisease1)<-[r4:Parent\_Disease]-(relatedDisease3:Disease)

// Collect all related diseases

WITH s, r, relatedDisease, r1, relatedSymptom, r2, relatedDisease1, r3, relatedDisease2, r4, relatedDisease3, diseaseList

// Filter to return only the diseases or their synonyms that match the disease list

WITH s, r, r1, r2, r3, r4, relatedSymptom,

CASE

WHEN toLower(relatedDisease.name) IN [diseaseList IN diseaseList | toLower(diseaseList)] THEN relatedDisease

WHEN any(synonym IN relatedDisease.synonyms WHERE toLower(synonym) IN [diseaseList IN diseaseList | toLower(diseaseList)]) THEN relatedDisease

END AS matchedRelatedDisease,

CASE

WHEN toLower(relatedDisease1.name) IN [diseaseList IN diseaseList | toLower(diseaseList)] THEN relatedDisease1

WHEN any(synonym IN relatedDisease1.synonyms WHERE toLower(synonym) IN [diseaseList IN diseaseList | toLower(diseaseList)]) THEN relatedDisease1

END AS matchedRelatedDisease1,

CASE

WHEN toLower(relatedDisease2.name) IN [diseaseList IN diseaseList | toLower(diseaseList)] THEN relatedDisease2

WHEN any(synonym IN relatedDisease2.synonyms WHERE toLower(synonym) IN [diseaseList IN diseaseList | toLower(diseaseList)]) THEN relatedDisease2

END AS matchedRelatedDisease2,

CASE

WHEN toLower(relatedDisease3.name) IN [diseaseList IN diseaseList | toLower(diseaseList)] THEN relatedDisease3

WHEN any(synonym IN relatedDisease3.synonyms WHERE toLower(synonym) IN [diseaseList IN diseaseList | toLower(diseaseList)]) THEN relatedDisease3

END AS matchedRelatedDisease3

// Return only the matched diseases

RETURN s, r, r1, r2, r3, r4, relatedSymptom,

matchedRelatedDisease AS relatedDisease,

matchedRelatedDisease1 AS relatedDisease1,

matchedRelatedDisease2 AS relatedDisease2,

matchedRelatedDisease3 AS relatedDisease3