



Análise de Relação Doenças-Sintomas

Time Sugoi

Esdras Rodrigues do Carmo

Gabriel Ryo Hioki

RA: 170656

RA: 172434

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by circles of varying sizes, some with concentric rings, and the lines are thin and grey. The diagram is partially cut off by the top and left edges of the slide.

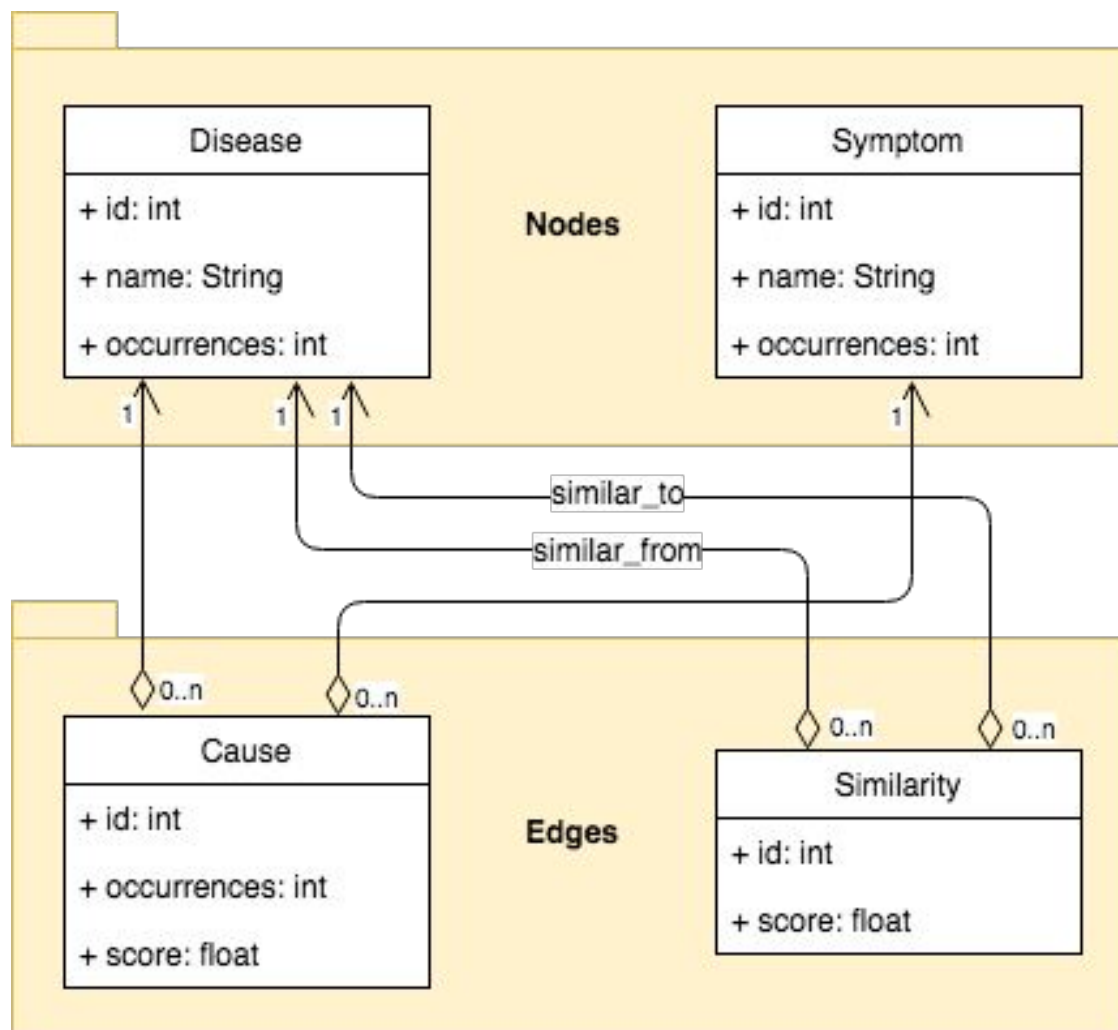
1.

Proposta Inicial

Planejamento Inicial do Projeto

- Maior eficiência na busca de doenças e sintomas.
 - Identificação de doenças pelos sintomas mais precisa e veloz.
- Análise de redes em grafo de doenças e sintomas.
 - arestas relacionando doenças, sintomas e similaridades entre doenças.
- Criar *clusters* pela similaridade das doenças e assim classificá-las.

Modelo Conceitual



Modelo Lógico

- Primeiro modelo lógico formulado.
- Cada objeto e relação tem *id* como chave primária.

Disease(id, name, occurrences)

Symptom(id, name, occurrences)

Cause(id, disease, symptom, occurrences, score)

- Chave Estrangeira: disease → **Disease**
- Chave Estrangeira: symptom → **Symptom**

Similarity(id, disease_from, disease_to, score)

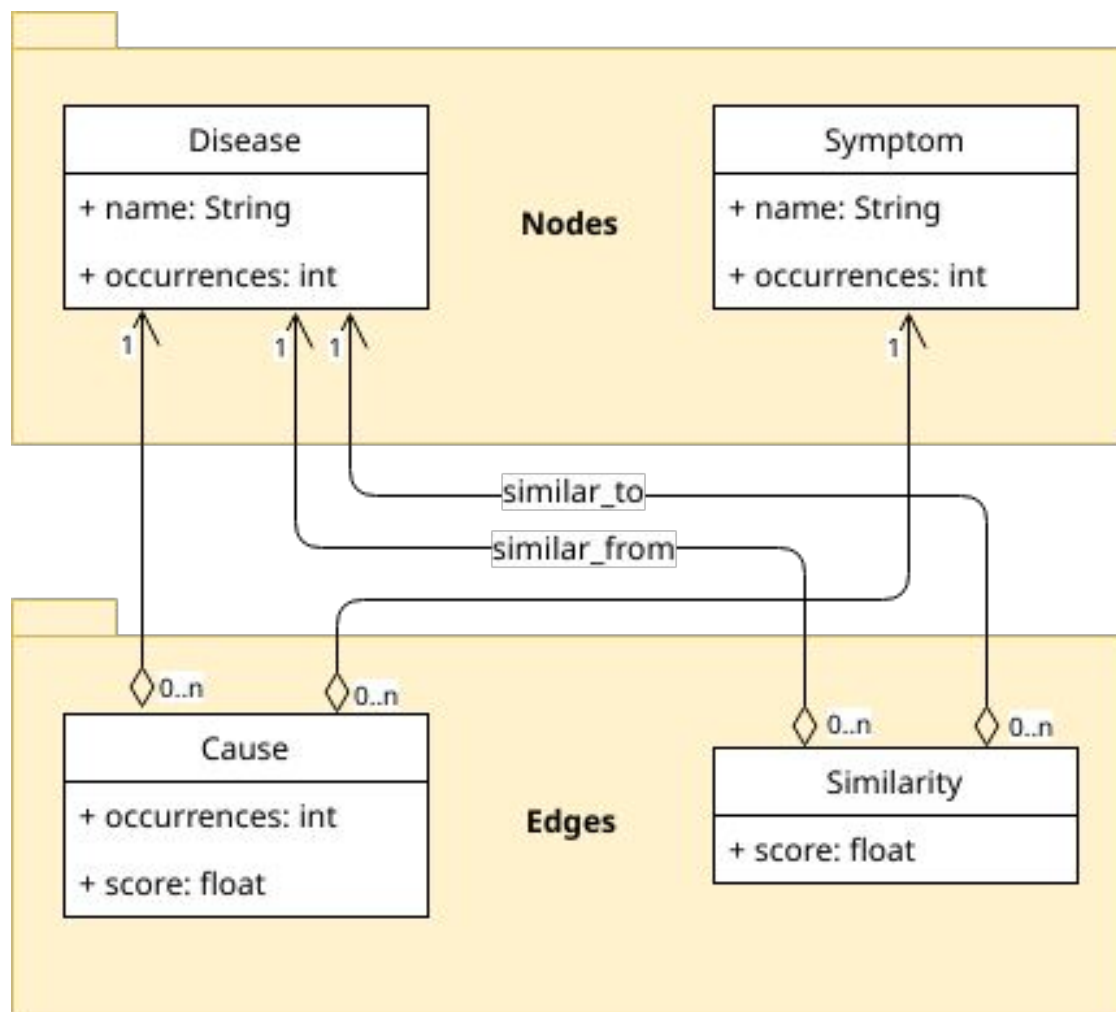
- Chave Estrangeira: disease_from → **Disease**
- Chave Estrangeira: disease_to → **Disease**

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles inside, suggesting a hierarchical or multi-layered structure. The lines are thin and gray, connecting the nodes in a non-linear fashion.

2.

Banco de Dados Relacional

7



Modelo Lógico

- Campo *id* retirado,
 - *name* como chave primária nos nós;
 - Chaves estrangeiras como chave primária nas relações.

Disease(name, occurrences)

Symptom(name, occurrences)

Cause(disease, symptom, occurrences, score)

- Chave Estrangeira: disease → **Disease**
- Chave Estrangeira: symptom → **Symptom**

Similarity(disease from, disease to, score)

- Chave Estrangeira: disease from → **Disease**
- Chave Estrangeira: disease to → **Disease**

Análise dos Dados

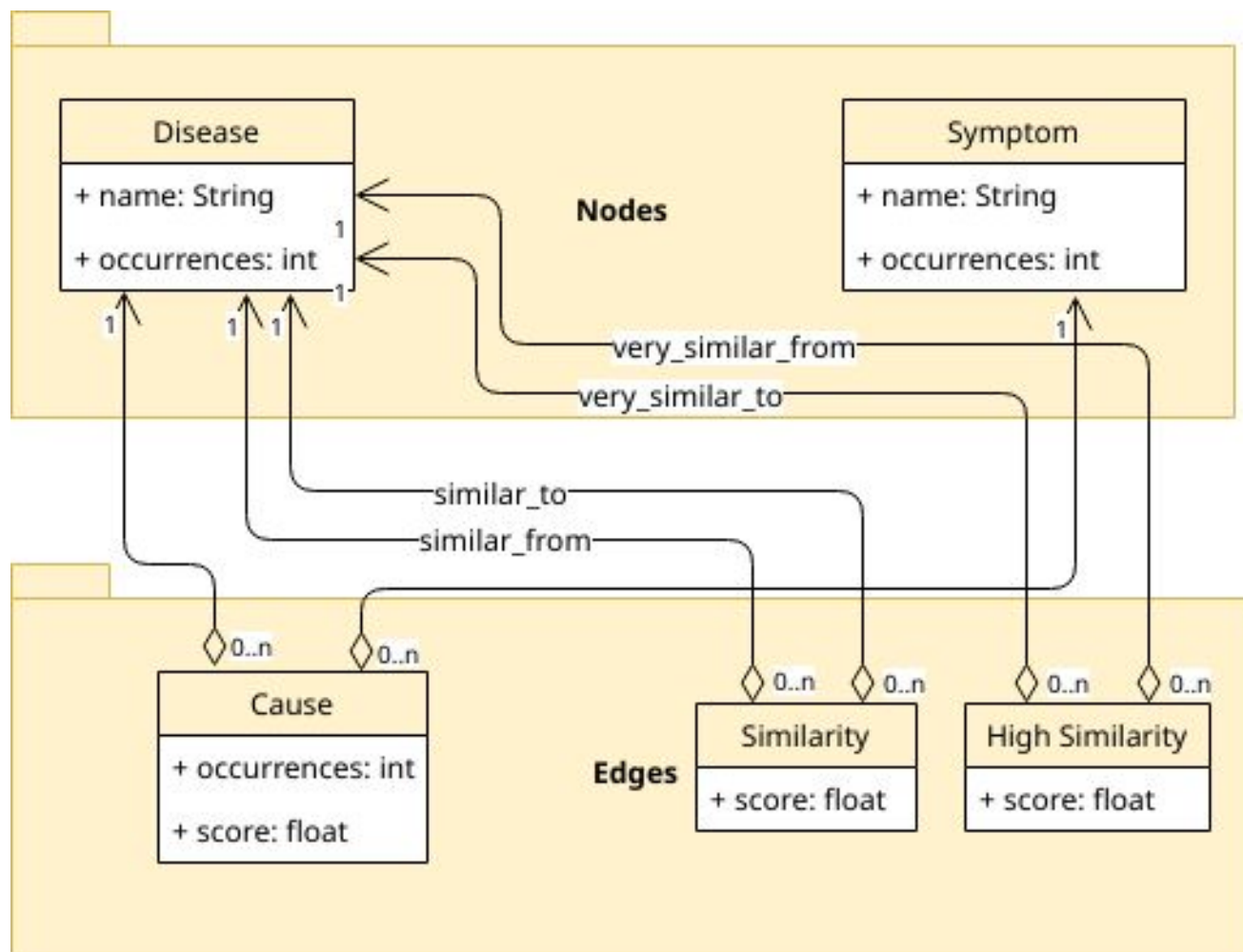
- Retornar o *Score* médio de sintomas de cada doença.
- Sintomas com *Score* acima da média para uma doença específica.
- Dado uma lista de sintomas, retornar as doenças possíveis que contém estes sintomas.
- Retornar doenças similares ($\text{Score} > 0.8$) a uma doença específica.
- Comparação entre sintomas de uma doença e da doença mais similar a esta.



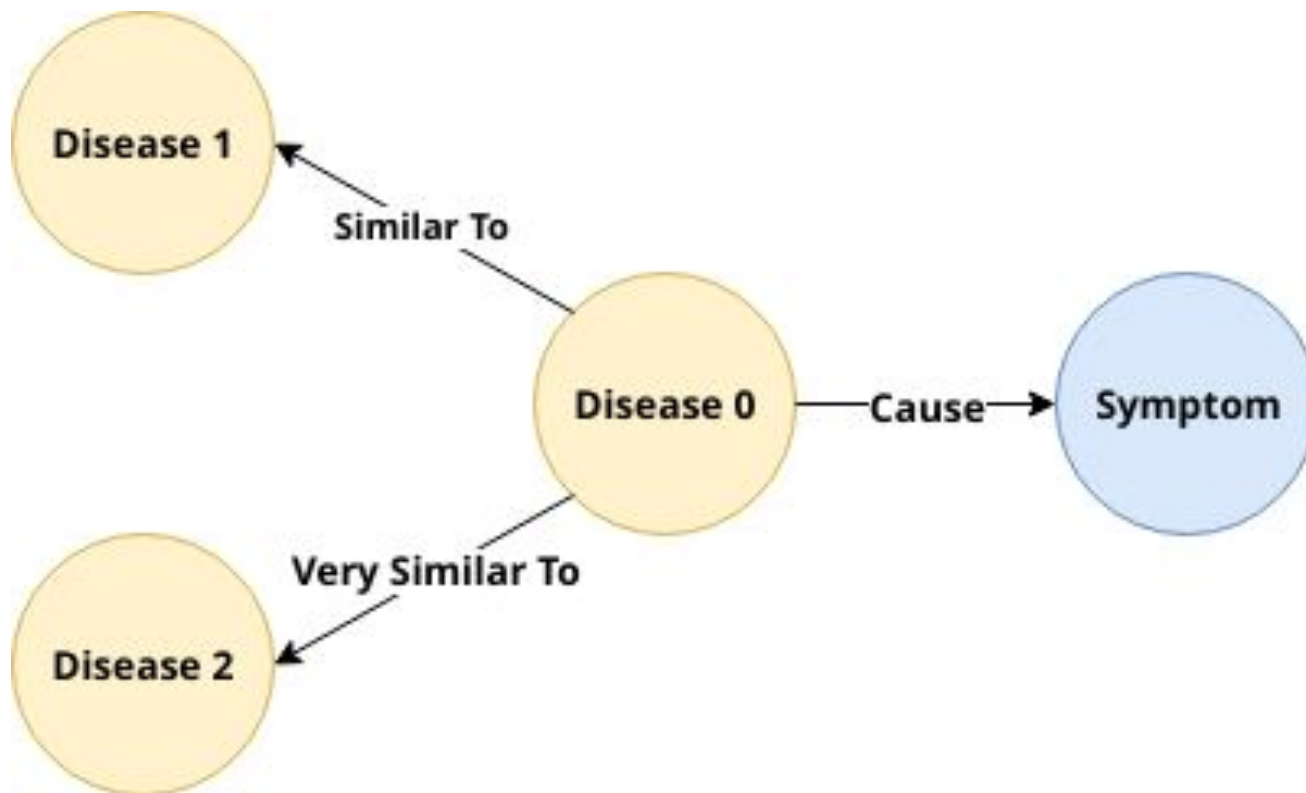
3.

Banco de Dados em Grafos

Modelo Conceitual em UML



Modelo Conceitual em Grafo



Modelo Lógico

- Banco de Grafos, ferramenta Neo4j para a análise.
- Criada nova relação *High Similarity*.

Disease(name, occurrences)

Symptom(name, occurrences)

Cause(disease, symptom, occurrences, score)

- Chave Estrangeira: disease → **Disease**
- Chave Estrangeira: symptom → **Symptom**

Similarity(disease from, disease to, score)

- Chave Estrangeira: disease from → **Disease**
- Chave Estrangeira: disease to → **Disease**

High Similarity(very similar from, very similar to, score)

- Chave Estrangeira: very similar from → **Disease**
- Chave Estrangeira: very similar to → **Disease**

Análise dos Dados

- Doenças que possuem o sintoma “Amnesia” e suas relações.
 - Doenças com score > 200 da relação *cause* para o sintoma.
- Similaridade de doenças com “Diabetes Tipo 1” em dois níveis de hierarquia.
 - Aresta “VERY_SIMILAR_TO” para score de similaridade ≥ 0.8
- Listar todas as doenças da componente conexa da “Diabetes Tipo 1”

Análise dos Dados

- Encontrar *clusters* de doenças altamente similares
→ Foi utilizado o algoritmo “Label Propagation”.
- Algoritmo “PageRank” para identificar as doenças com mais conexões entre similares.

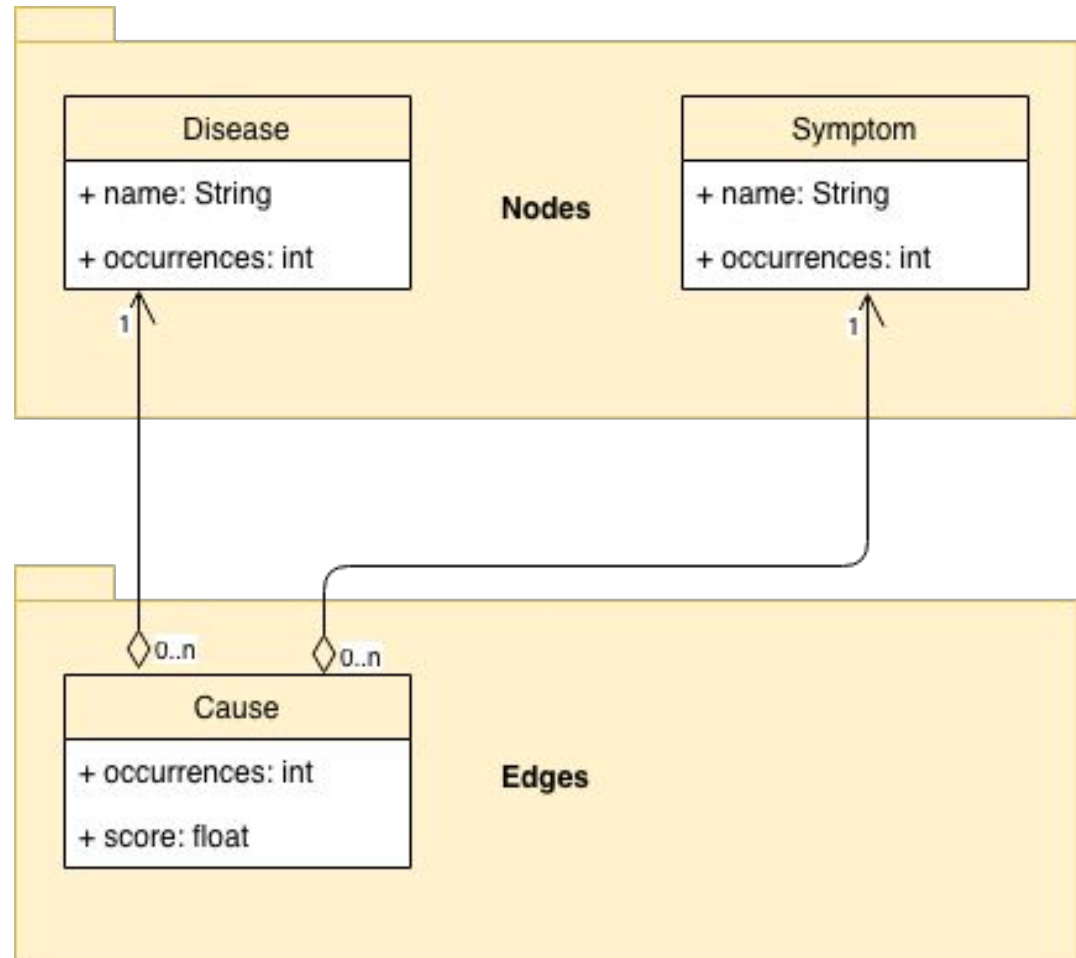


4.

Banco de dados em RDF e XML

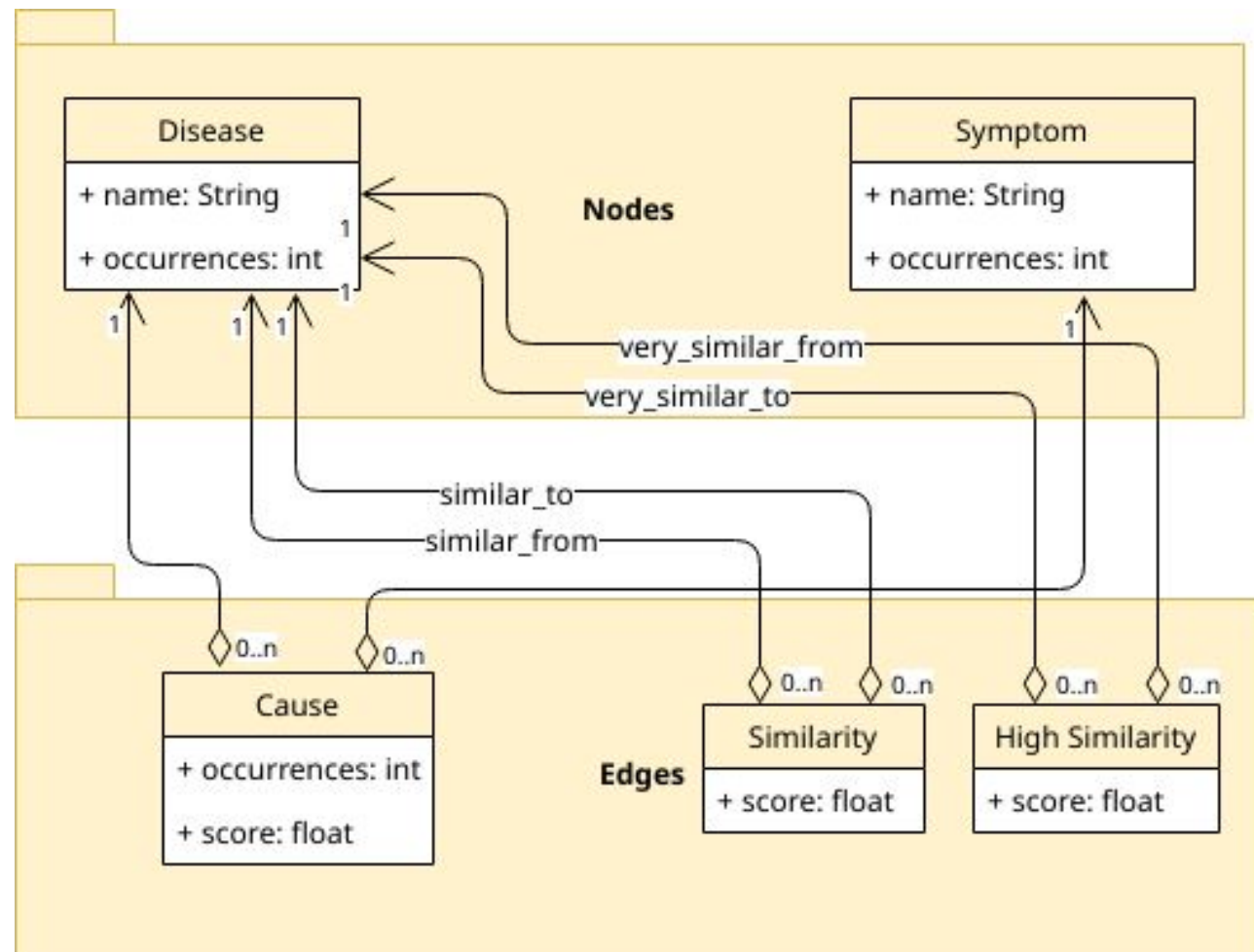
Modelo Conceitual

Modelo Conceitual em UML para XML



Modelo Conceitual

Modelo Conceitual em UML para RDF



Modelo Lógico

- Modelo XML, ferramenta SPARQL para a análise.
- Relação Similarity retirada.
 - Zorba não suporta a quantidade de dados da fonte.

Disease(name, occurrences)

Symptom(name, occurrences)

Cause(disease, symptom, occurrences, score)

- Chave Estrangeira: disease → **Disease**
- Chave Estrangeira: symptom → **Symptom**

Modelo Lógico

- Modelo RDF, ferramenta XQuery para a análise.
- Mesma modelagem do bando de grafos (Neo4j).
 - RDF é uma canonização de um grafo dirigido.

Disease(name, occurrences)

Symptom(name, occurrences)

Cause(disease, symptom, occurrences, score)

- Chave Estrangeira: disease → **Disease**
- Chave Estrangeira: symptom → **Symptom**

Similarity(disease from, disease to, score)

- Chave Estrangeira: disease from → **Disease**
- Chave Estrangeira: disease to → **Disease**

High Similarity(very similar from, very similar to, score)

- Chave Estrangeira: very similar from → **Disease**
- Chave Estrangeira: very similar to → **Disease**

Análise dos Dados

Análise XQuery

- Doenças rotuladas como “Diabetes”.
- Doenças e os sintomas acima da média de score.
- Doenças que possuem o sintoma “Obesity”.
- Retornar a média de *occurrences* das doenças.
- Quantidade de doenças que estão acima da média de *occurrences*.

Análise dos Dados

Análise SPARQL

- Retornar todas as doenças do sistema digestivo.
- Doenças altamente similares a “Hepatitis B” em todos os níveis.
- Doenças do sistema digestivo que são infecções bacterianas.
- Relações entre sintomas de infecções bacterianas do sistema digestivo
- Doenças do sistema digestivo que são intestinais



Obrigado!