

# Análise de Relação Doenças-Sintomas

Time Sugoi

Esdras Rodrigues do Carmo - RA: 170656

Gabriel Ryo Hioki - RA: 172434

## Resumo

O problema estudado consiste na eficiência na busca em um conjunto de doenças e sintomas. Com uma busca eficiente, espera-se que a identificação de doenças a partir dos sintomas apresentados em um paciente seja mais precisa e veloz. Será utilizado análise de redes em um grafo de doenças e sintomas, com arestas relacionando doenças, sintomas e similaridades entre doenças. As doenças serão *clusterizadas* de acordo com o *score* de similaridade. Com isso, poderemos classificar as doenças e fazer uma busca mais assertiva no banco de dados. Os sintomas mais comuns serão ordenados utilizando um algoritmo de *PageRank*.

## Requisitos do Modelo Conceitual

O modelo conceitual deve suportar o armazenamento de doenças e sintomas, assim como suas relações. Além disso, deve ter uma boa representação dos relacionamentos existentes, de modo a aumentar a eficiência da análise de dados.

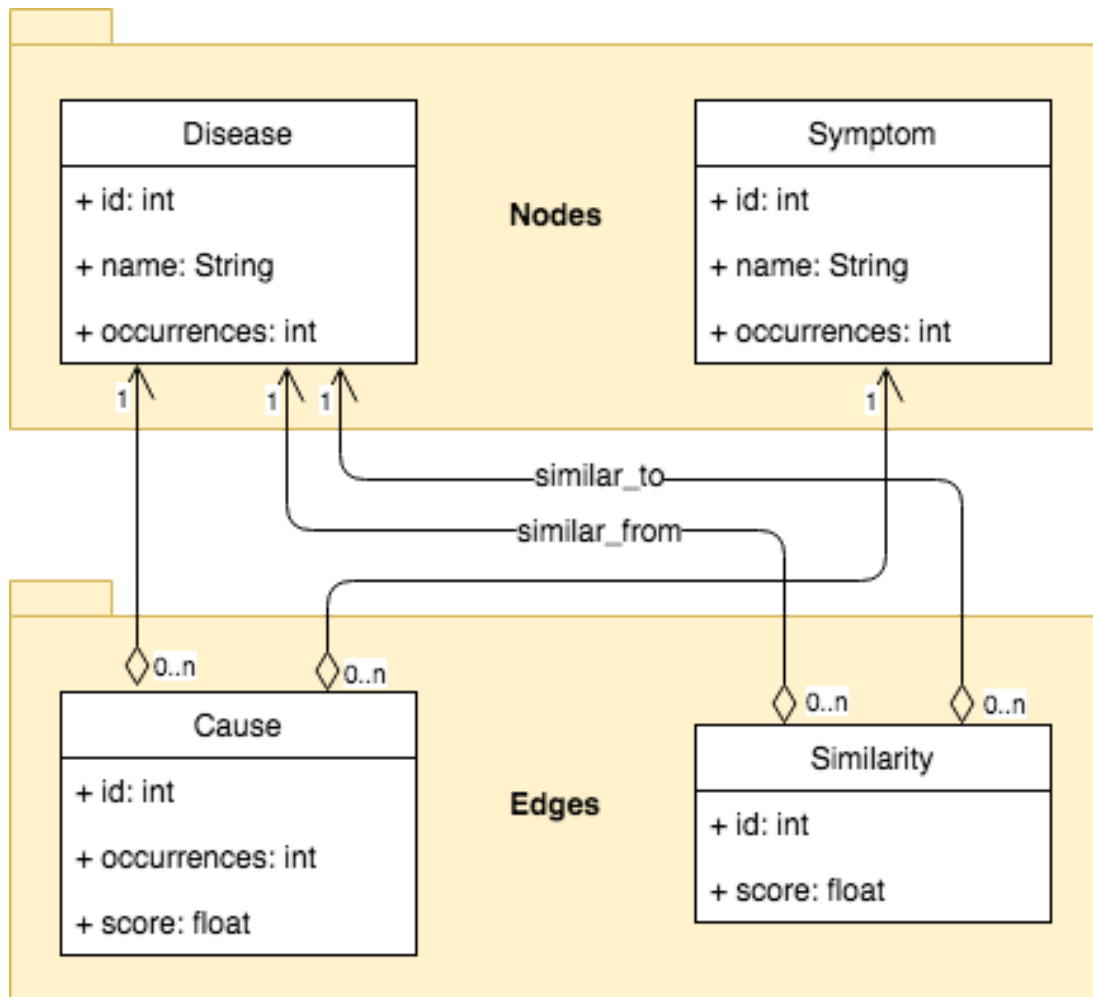
É importante também manter o modelo simples e claro o bastante para que qualquer usuário consiga entendê-lo, mesmo que quando implementado seja utilizado outro modelo lógico mais otimizado, como por exemplo um banco de dados em grafos.

## Fonte de Dados

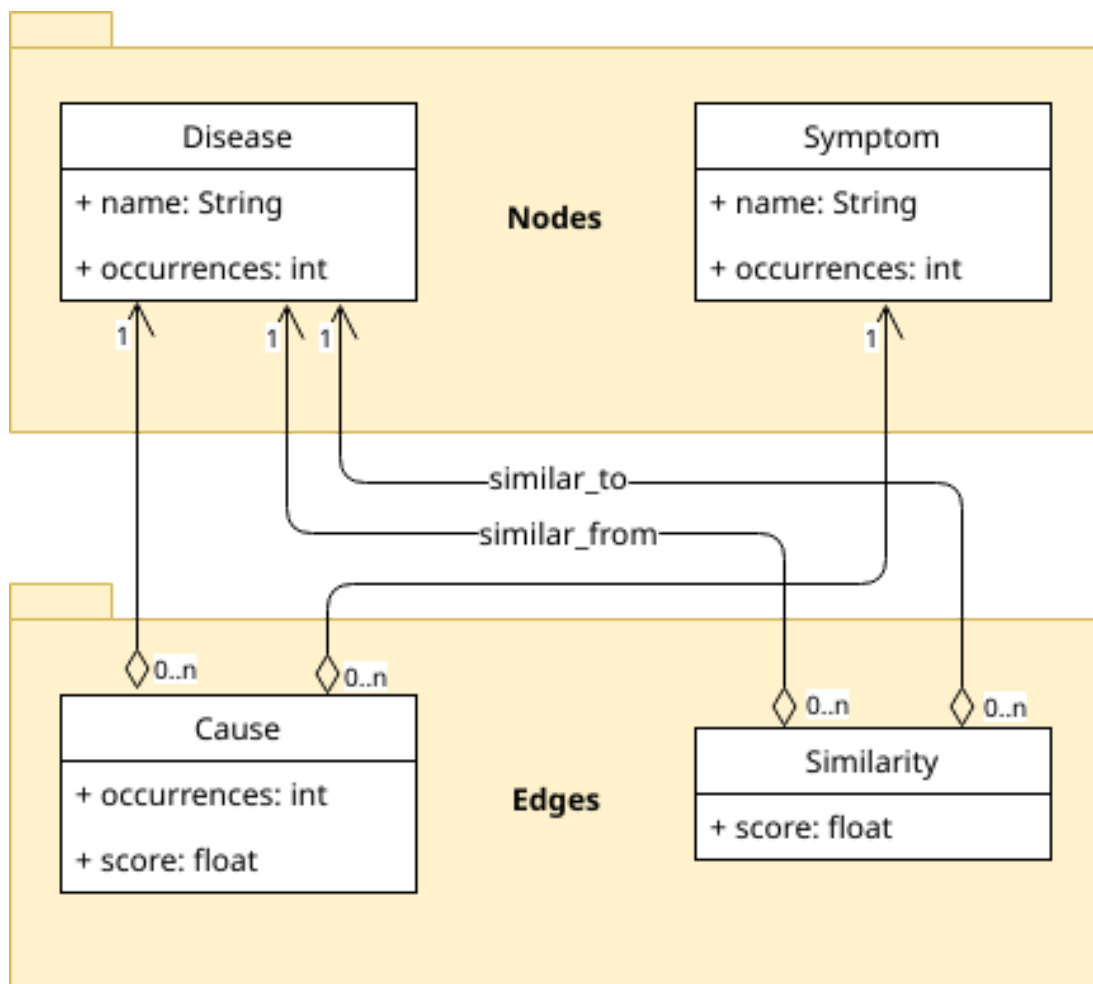
O conjunto de dados é formado por 4 arquivos, sendo cada um deles uma tabela: doenças, sintomas, relacionamento entre doenças e sintomas, relacionamento de similaridade entre doenças.

Zhou, XueZhong et al. (2014). Human symptoms-disease network.  
<https://www.nature.com/articles/ncomms5212#s1>

## Modelagem Conceitual



**Figura 1:** Modelo conceitual inicial



**Figura 2:** Modelo conceitual após alterações

A mudança realizada entre os modelos consiste na otimização do atributo *name* dos objetos, que passou a ser a chave primária, removendo o atributo *id*. O mesmo foi feito nas relações, mas ao invés de colocar um atributo como chave primária, recebem chaves estrangeiras dos objetos.

## Modelagem Lógica

Como na modelagem conceitual o relacionamento N-M entre as tabelas *Disease* e *Symptom* foram explicitados como uma nova tabela *Cause*, foi possível mapear diretamente o modelo conceitual para o lógico, apenas adicionando chaves estrangeiras nos relacionamentos.

- **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**