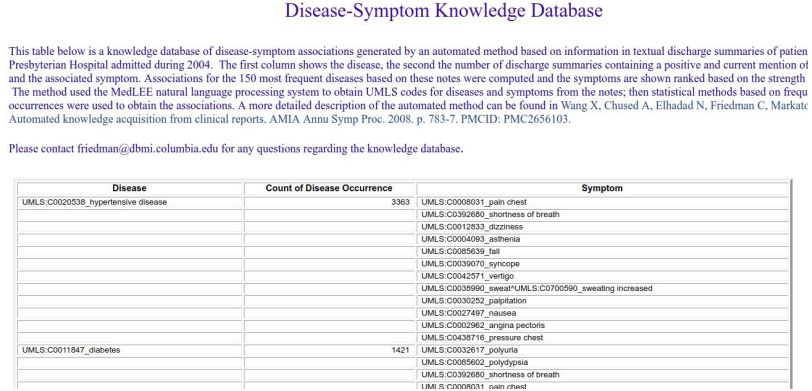
Etapa 4

Banco de dados

1) Disease-Symptom Knowledge Database



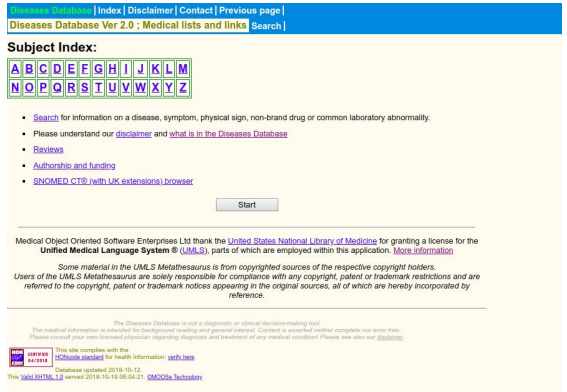
Disease-Symptom Knowledge Database

This table below is a knowledge database of disease-symptom associations generated by an automated method based on information in textual discharge summaries of patien
Presbyterian Hospital admitted during 2004. The first column shows the disease, the second the number of discharge summaries containing a positive and current mention of
and the associated symptom. Associations for the 150 most frequent diseases based on these notes were computed and the symptoms are shown ranked based on the strength
The method used the MedLEE natural language processing system to obtain UMLS codes for diseases and symptoms from the notes; then statistical methods based on frequ
occurrences were used to obtain the associations. A more detailed description of the automated method can be found in Wang X, Chused A, Elhadad N, Friedman C, Markatc
Automated knowledge acquisition from clinical reports. AMIA Annu Symp Proc. 2008. p. 783-7. PMCID: PMC2656103.

Please contact friedman@dbmi.columbia.edu for any questions regarding the knowledge database.

Disease	Count of Disease Occurrence	Symptom
UMLS:C0020538_hypertensive disease	3363	UMLS:C0008031_pain chest
		UMLS:C0392680_shortness of breath
		UMLS:C0012833_dizziness
		UMLS:C0004093_asthenia
		UMLS:C0085639_fall
		UMLS:C0039070_syncope
		UMLS:C0042571_vertigo
		UMLS:C0038990_sweat^UMLS:C0700590_sweating increased
		UMLS:C0030252_palpitation
		UMLS:C0027497_nausea
		UMLS:C0002962_angina pectoris
		UMLS:C0438716_pressure chest
UMLS:C0011847_diabetes	1421	UMLS:C0032617_polyuria
		UMLS:C0085602_polydypsia
		UMLS:C0392680_shortness of breath
		UMLS:C0008031_pain chest

2) Diseases Database



Diseases Database | Index | Disclaimer | Contact | Previous page |

Diseases Database Ver 2.0 ; Medical lists and links Search |

Subject Index:

A B C D E F G H I J K L M

N O P Q R S T U V W X Y Z

- Search for information on a disease, symptom, physical sign, non-brand drug or common laboratory abnormality.
- Please understand our disclaimer and what is in the Diseases Database
- Reviews
- Authorship and funding
- SNOMED CT® (with UK extensions) browser

Start

Medical Object Oriented Software Enterprises Ltd thank the United States National Library of Medicine for granting a license for the **Unified Medical Language System ®** (UMLS), parts of which are employed within this application. More information

Some material in the UMLS Metathesaurus is from copyrighted sources of the respective copyright holders. Users of the UMLS Metathesaurus are solely responsible for compliance with any copyright, patent or trademark restrictions and are referred to the copyright, patent or trademark notices appearing in the original sources, all of which are hereby incorporated by reference.

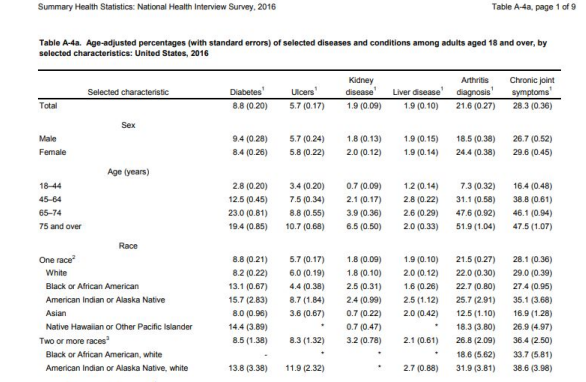
The Diseases Database is not a diagnostic or clinical decision-making tool.
The medical information is intended for background reading and general interest. Content is asserted neither complete nor error free.
Please consult your own licensed physician regarding diagnosis and treatment of any medical condition! Please see also our disclaimer.

This site complies with the HONcode standard for health information: verify here

Database updated 2018-10-12 /

This Valid XHTML 1.0 served 2018-10-19 08:04:21 - GMOOSe Technology

3) National Health Interview Survey

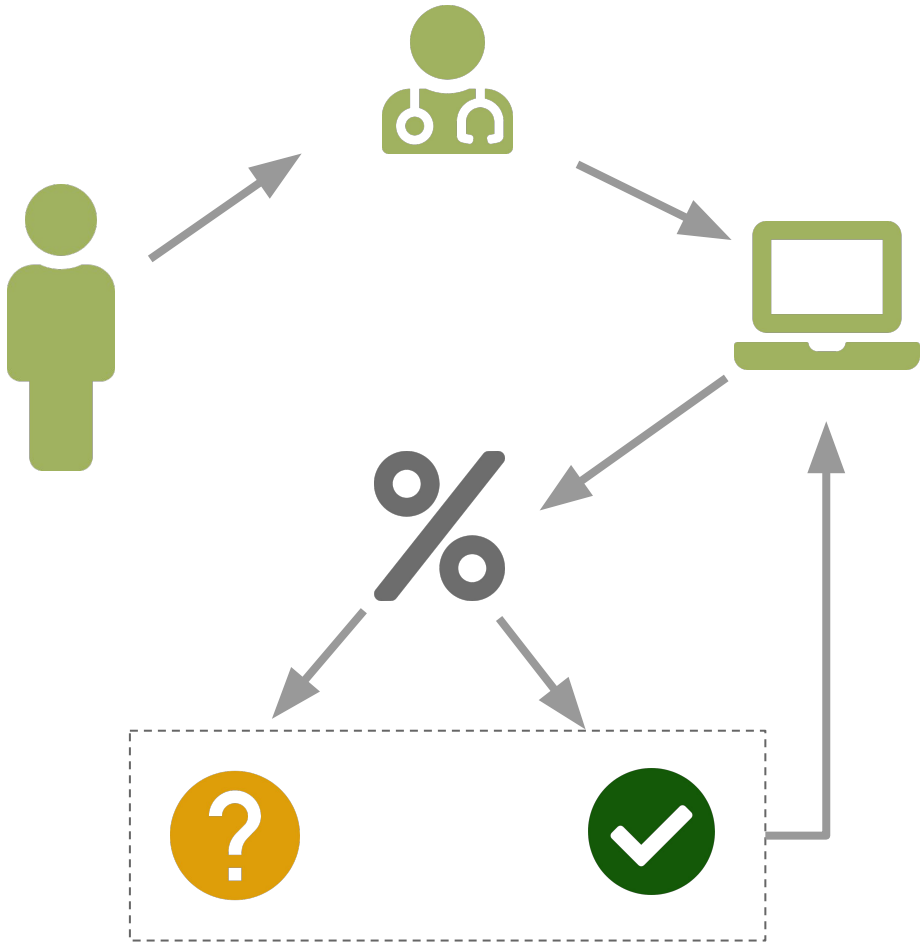


Summary Health Statistics: National Health Interview Survey, 2016 — Table A-4a, page 1 of 9

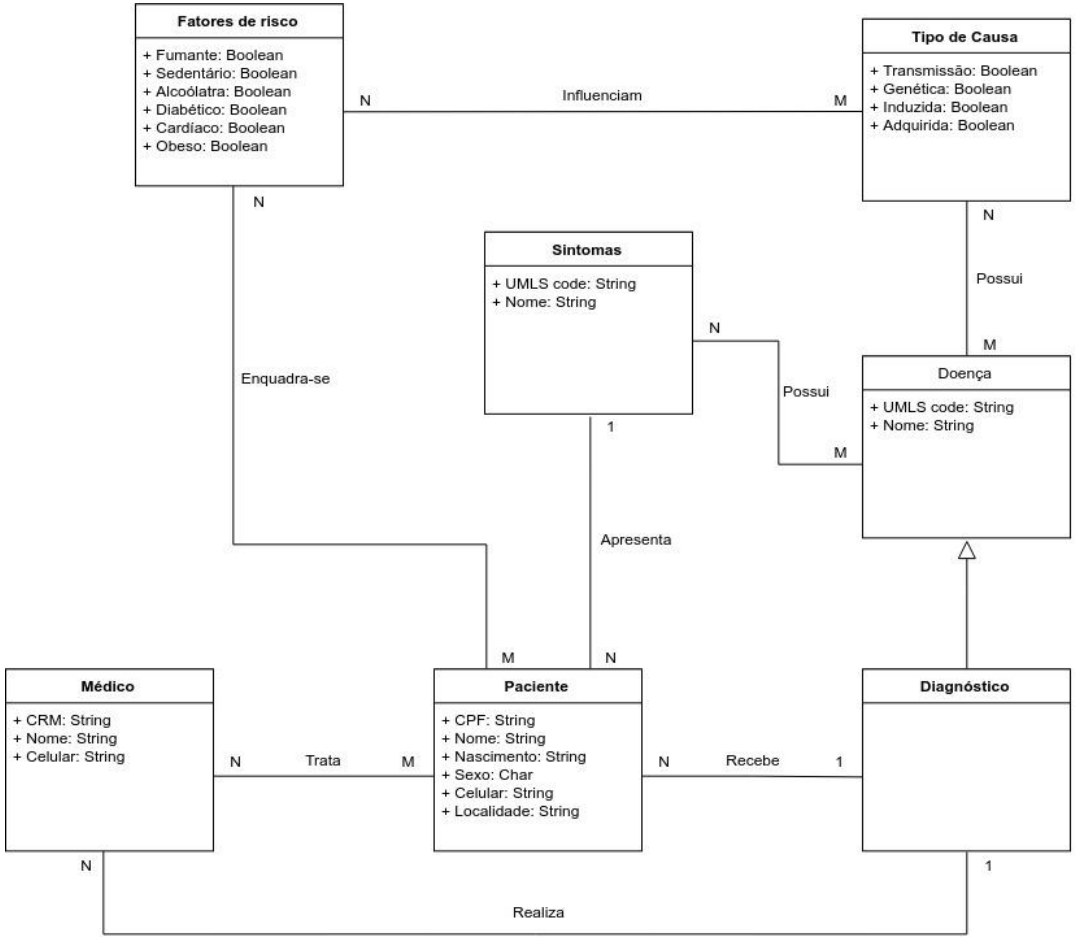
Table A-4a. Age-adjusted percentages (with standard errors) of selected diseases and conditions among adults aged 18 and over, by selected characteristics: United States, 2016

Selected characteristic	Diabetes[1]	Ulcers[1]	Kidney disease[1]	Liver disease[1]	Arthritis diagnosis[1]	Chronic joint symptoms[1]
Total	8.8 (0.20)	5.7 (0.17)	1.9 (0.09)	1.9 (0.10)	21.6 (0.27)	28.3 (0.36)
Sex						
Male	9.4 (0.28)	5.7 (0.24)	1.8 (0.13)	1.9 (0.15)	18.5 (0.38)	26.7 (0.52)
Female	8.4 (0.26)	5.8 (0.22)	2.0 (0.12)	1.9 (0.14)	24.4 (0.38)	29.6 (0.45)
Age (years)						
18–44	2.8 (0.20)	3.4 (0.20)	0.7 (0.09)	1.2 (0.14)	7.3 (0.32)	16.4 (0.48)
45–64	12.5 (0.45)	7.5 (0.34)	2.1 (0.17)	2.8 (0.22)	31.1 (0.58)	38.8 (0.61)
65–74	23.0 (0.81)	8.8 (0.55)	3.9 (0.36)	2.6 (0.29)	47.6 (0.92)	46.1 (0.94)
75 and over	19.4 (0.85)	10.7 (0.68)	6.5 (0.50)	2.0 (0.33)	51.9 (1.04)	47.5 (1.07)
Race						
One race[2]	8.8 (0.21)	5.7 (0.17)	1.8 (0.09)	1.9 (0.10)	21.5 (0.27)	28.1 (0.36)
White	8.2 (0.22)	6.0 (0.19)	1.8 (0.10)	2.0 (0.12)	22.0 (0.30)	29.0 (0.39)
Black or African American	13.1 (0.67)	4.4 (0.38)	2.5 (0.31)	1.6 (0.26)	22.7 (0.80)	27.4 (0.95)
American Indian or Alaska Native	15.7 (2.83)	8.7 (1.84)	2.4 (0.99)	2.5 (1.12)	25.7 (2.91)	35.1 (3.68)
Asian	8.0 (0.96)	3.6 (0.67)	0.7 (0.22)	2.0 (0.42)	12.5 (1.10)	16.9 (1.28)
Native Hawaiian or Other Pacific Islander	14.4 (3.89)	*	0.7 (0.47)	*	18.3 (3.80)	26.9 (4.97)
Two or more races[3]	8.5 (1.38)	8.3 (1.32)	3.2 (0.78)	2.1 (0.61)	26.8 (2.09)	36.4 (2.50)
Black or African American, white	-	*	*	*	18.6 (5.62)	33.7 (5.81)
American Indian or Alaska Native, white	13.8 (3.38)	11.9 (2.32)	*	2.7 (0.88)	31.9 (3.81)	38.6 (3.98)

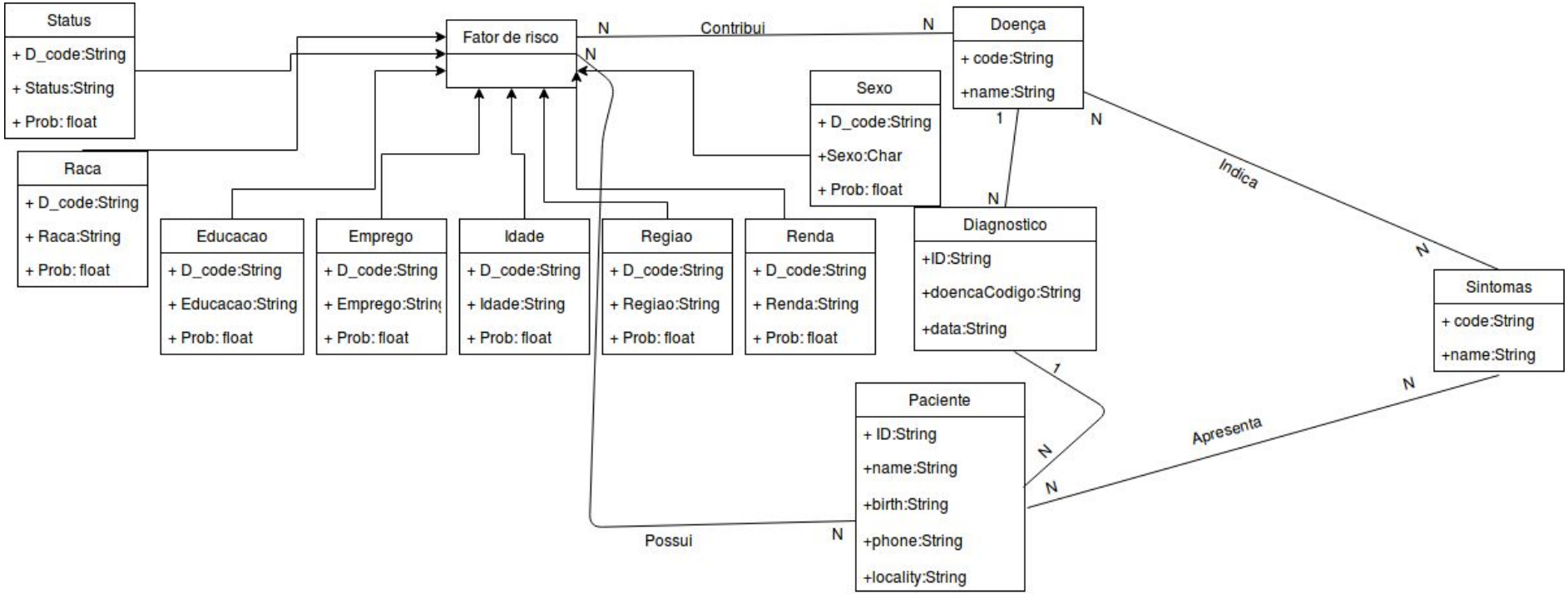
Requisitos



Modelo inicial



Modelo etapa 2



Queries

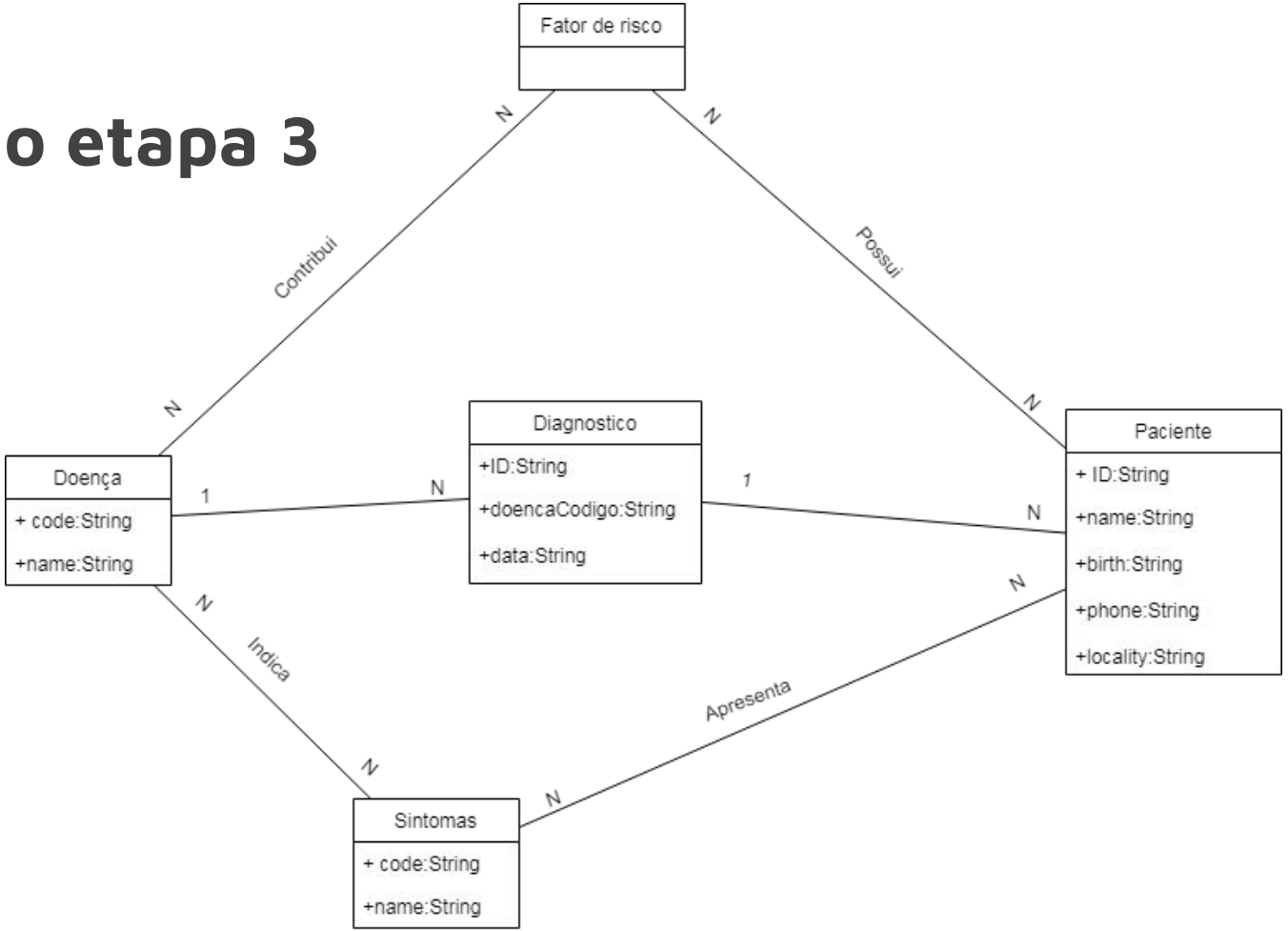
Sintoma mais relatado por pacientes

In [ ]:
```
SELECT nome
FROM (SELECT S.nome, COUNT(*) contagem
        FROM Sintomas S, DoencaSintomas DS, Doenca D
        WHERE DS.sintomaCodigo = S.codigo AND DS.doencaCodigo = D.codigo GROUP BY S.nome)
WHERE contagem = (SELECT MAX(contagem) FROM (SELECT S.nome, COUNT(*) contagem
        FROM Sintomas S, DoencaSintomas DS, Doenca D
        WHERE DS.sintomaCodigo = S.codigo AND DS.doencaCodigo = D.codigo GROUP BY S.nome))
```

Nome dos constatados com a doença X (no caso abaixo: diabetes)

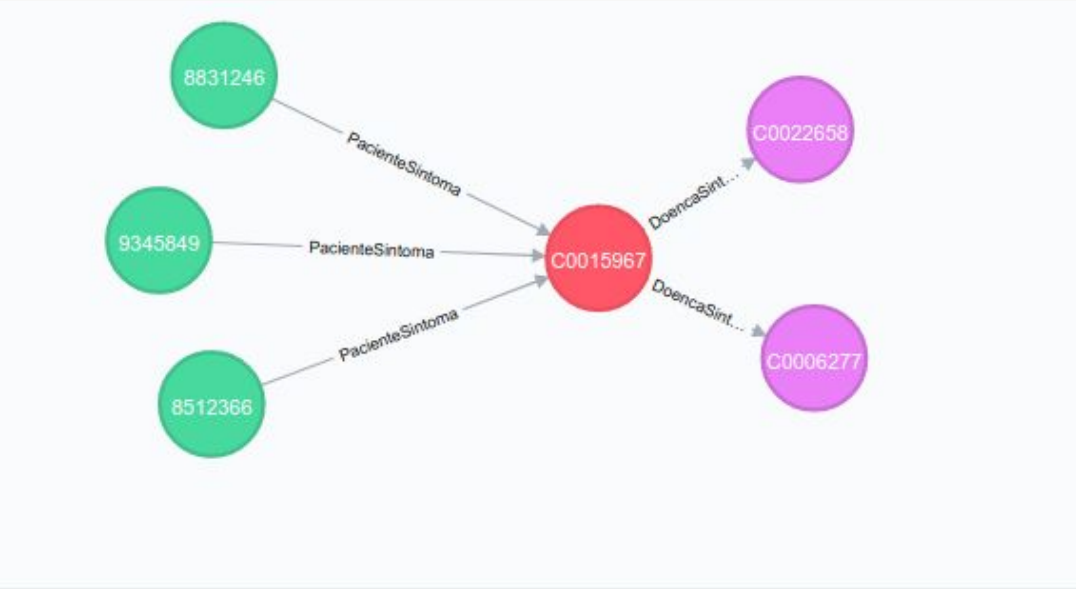
In [ ]:
```
SELECT P.nome
FROM PacienteDiagnostico PD, Diagnostico D, Doenca Do, Paciente P
WHERE PD.pacienteID = P.ID AND PD.diagnosticoID = D.ID AND D.doencaCodigo = Do.codigo AND Do.nome = 'diabetes'
```

Modelo etapa 3



Queries

```
// retorna o caminho para um paciente que apresenta como sintoma 'febre'
MATCH p = (l:Paciente)-[:PacienteSintoma]->(m:Sintomas {nome: 'fever'})-[t:DoencaSintoma]->(d:Doenca)
RETURN p
```



Queries

```
// Dado um sintoma X, retorna a probabilidade de um paciente ter a doença Y
MATCH p = (l:Paciente)-[:PacienteSintoma]->(m:Sintomas {nome: 'pain chest'})-[t:DoencaSintoma]->
(d:Doenca{nome: 'hypertensive disease'})
WITH COUNT(l) AS prob
MATCH q = (l:Paciente)-[:PacienteSintoma]->(m:Sintomas {nome: 'pain chest'})-[t:DoencaSintoma]->(d:Doenca)
RETURN toFloat(prob)/COUNT(l)
```

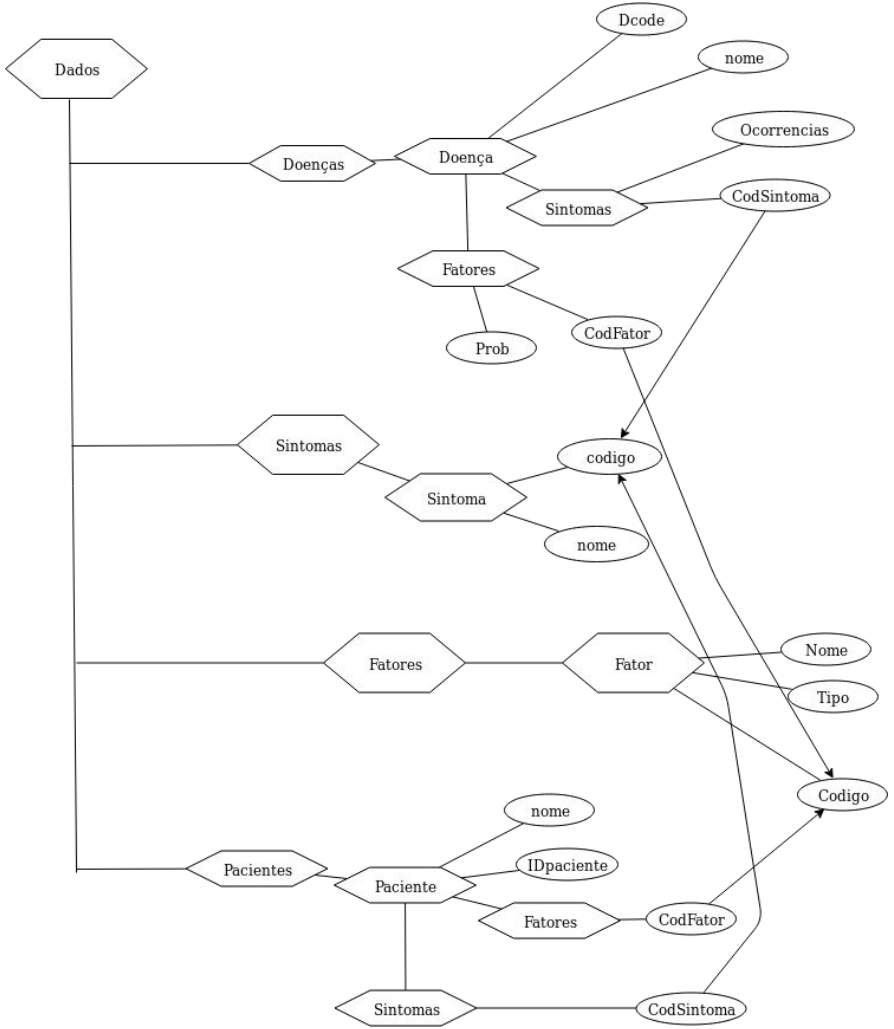
toFloat(prob)/COUNT(l)
0.2

Queries

```
// Dado um Paciente X, retorna a probabilidade do paciente ter a doença Y levando em conta os
// seus fatores de risco e os seus sintomas
MATCH p = (l:Paciente {ID: 'patKMPV11'})-[:PacienteSintoma]->(m:Sintomas)-[t:DoencaSintoma]->
(d:Doenca{nome: 'hypertensive disease'})
WITH COUNT(l) AS prob
MATCH q = (l:Paciente {ID: 'patKMPV11'})-[:PacienteSintoma]->(m:Sintomas)-[t:DoencaSintoma]->
(d:Doenca) WITH COUNT(l)/prob AS tot
MATCH o = (n:Paciente{ID: 'patKMPV11'})-[FatorPaciente]->(Fator)-[t:FatorDoenca]->
(d:Doenca{nome: 'hypertensive disease'})
RETURN AVG(toFloat(replace(t.Probabilidade, ",", ".")))*toFloat(1)/tot
```

AVG(toFloat(replace(t.Probabilidade, ",", ".")))*toFloat(1)/tot
5.4175

Modelo etapa 4



Queries

```
# Get definitions for Neoplasms

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX meshv: <http://id.nlm.nih.gov/mesh/vocab#>
PREFIX mesh: <http://id.nlm.nih.gov/mesh/>
PREFIX mesh2015: <http://id.nlm.nih.gov/mesh/2015/>
PREFIX mesh2016: <http://id.nlm.nih.gov/mesh/2016/>
PREFIX mesh2017: <http://id.nlm.nih.gov/mesh/2017/>

SELECT *
FROM <http://id.nlm.nih.gov/mesh>
WHERE {
  mesh:D009369 rdfs:label ?DescriptorLabel .
  mesh:D009369 meshv:preferredConcept ?concept .
  ?concept meshv:scopeNote ?ScopeNote
}
```

Queries

```
# Descendants of Heart Diseases (D006331)

PREFIX mesh: <http://id.nlm.nih.gov/mesh/>
PREFIX mesh2015: <http://id.nlm.nih.gov/mesh/2015/>
PREFIX mesh2016: <http://id.nlm.nih.gov/mesh/2016/>
PREFIX mesh2017: <http://id.nlm.nih.gov/mesh/2017/>

 SELECT DISTINCT ?descriptor ?label
 FROM <http://id.nlm.nih.gov/mesh>

 WHERE {
   mesh:D006331 meshv:treeNumber ?treeNum .
   ?childTreeNum meshv:parentTreeNumber+ ?treeNum .
   ?descriptor meshv:treeNumber ?childTreeNum .
   ?descriptor rdfs:label ?label .
 }

 ORDER BY ?label
```

Queries

```
-------------------------------------------------------------------------------
Sintomas em comum entre duas doenças específicas

let $documento := doc('mydoc.xml')

for $i in ($documento//Doenca[nome = "diabetes"]/Sintomas)
for $j in ($documento//Doenca[nome = "hypertensive disease"]/Sintomas)
where $i/CodSintoma = $j/CodSintoma
for $k in ($documento//Sintoma)
where $k/codigo = $j/CodSintoma
return data($k/nome)


-------------------------------------------------------------------------------
Classicacao dos sintomas mais frequentes (com maior numero de Ocorrencias)
segundo nosso banco de dados

let $documento := doc('mydoc.xml')

return
<classificacao>
{
   for $i in ($documento//Doenca/Sintomas)
   where $i[Ocorrencias > 2000] order by $i/@Ocorrencias
   return <frequente>{data($i/CodSintoma)}</frequente>
}
</classificacao>
```

Queries

```
-------------------------------------------------------------------------------
Avalia qual doenca tem mais propensao a ocorrer em mulheres ou homens

let $documento := doc('mydoc.xml')

for $i in ($documento//Doenca)
return if ($i/Fatores[CodFator="24"]/Prob < $i/Fatores[CodFator="25"]/Prob)
then <Masculino>{distinct-values($i/nome)}</Masculino>
else <Feminino>{distinct-values($i/nome)}</Feminino>
```