

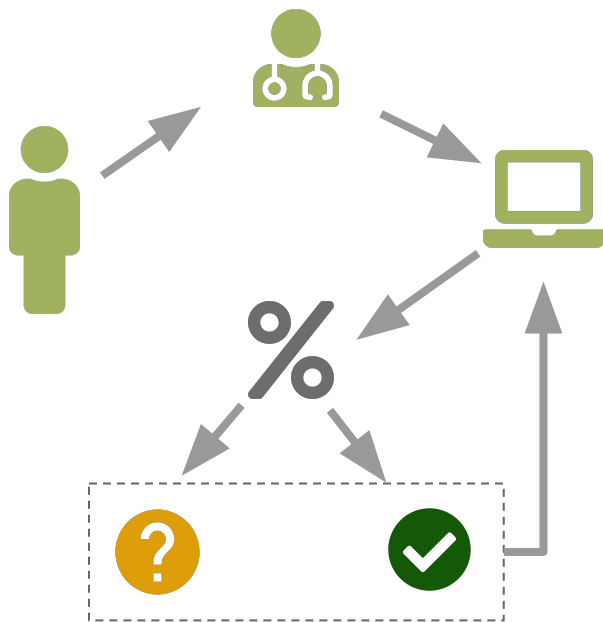
Etapa 4

João Paulo Soubihe RA: 151106
Leandro Ferlin Viana RA: 148729





Problema



A partir de dados coletados em uma consulta clínica, tais como:

- sintomas
- dados geográficos
- comportamentos de risco

inferir informações que possam auxiliar o profissional da saúde no diagnóstico correto de uma doença



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 patient 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 frequency 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, Markate. Automated knowledge acquisition from clinical reports. AMIA Annu Symp Proc. 2008; p. 783-7. PMID: 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:C0392660_shortness of breath
		UMLS:C0012833_dizziness
		UMLS:C0004093_asthenia
		UMLS:C0065636_fat
		UMLS:C0039070_syncope
		UMLS:C0042571_vertigo
		UMLS:C0036992_sweatUMLS:C0700590_sweating increased
		UMLS:C0030252_palpitation
		UMLS:C0027497_nausea
		UMLS:C0002962_angina pectoris
		UMLS:C0439716_pressure chest
UMLS:C0011847_diabetes	1421	UMLS:C0032617_polyuria
		UMLS:C0065602_polydipsia
		UMLS:C0392660_shortness of breath
		UMLS:C0009031_rain chest

2) Diseases Database

[Home](#) | [Index](#) | [Disclaimer](#) | [Contact](#) | [Previous page](#) | [Diseases Database Ver 2.0 : Medical lists and links](#) | [Search](#)

Subject Index:

ABCDEFGHIJKLM
NOPQRSTUVWXYZ

- [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](#)
- [Authentic and funding](#)
- [SNOMED, CTR, JAK, UMLS 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 posted online complete as soon as possible. Please consult your own licensed physician regarding diagnosis and treatment of any medical condition. Please see also our [disclaimer](#).

This site complies with the [Creative Commons](#) for health information [policy](#).
Diseases updated: 2016-10-12.
This [Virus](#) [HTML](#) [1.0](#) revised 2016-10-10 06:04:21. [SASOGL Technology](#)

3) National Health Interview Survey

Summary Health Statistics: National Health Interview Survey, 2016

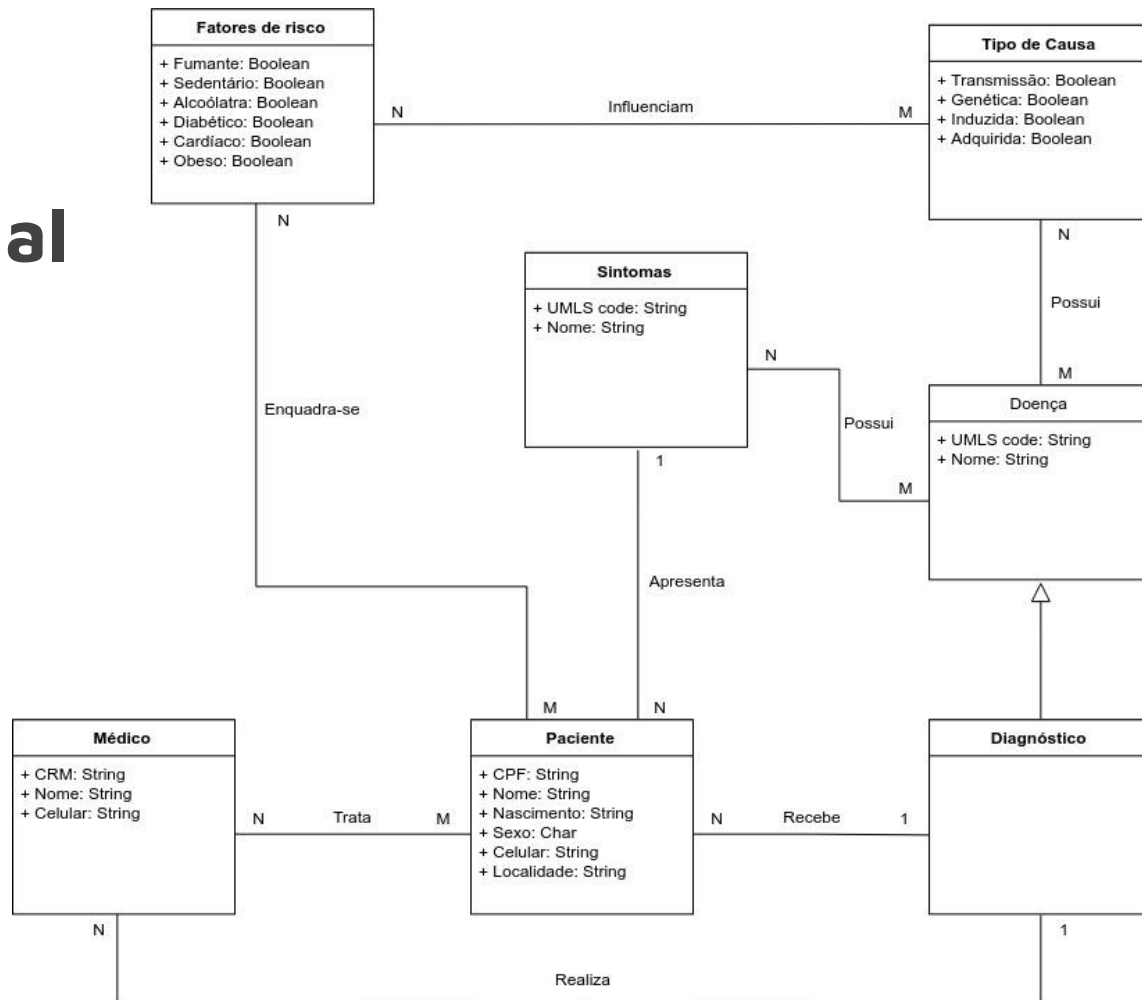
Table A-4a, page 1 of 9

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

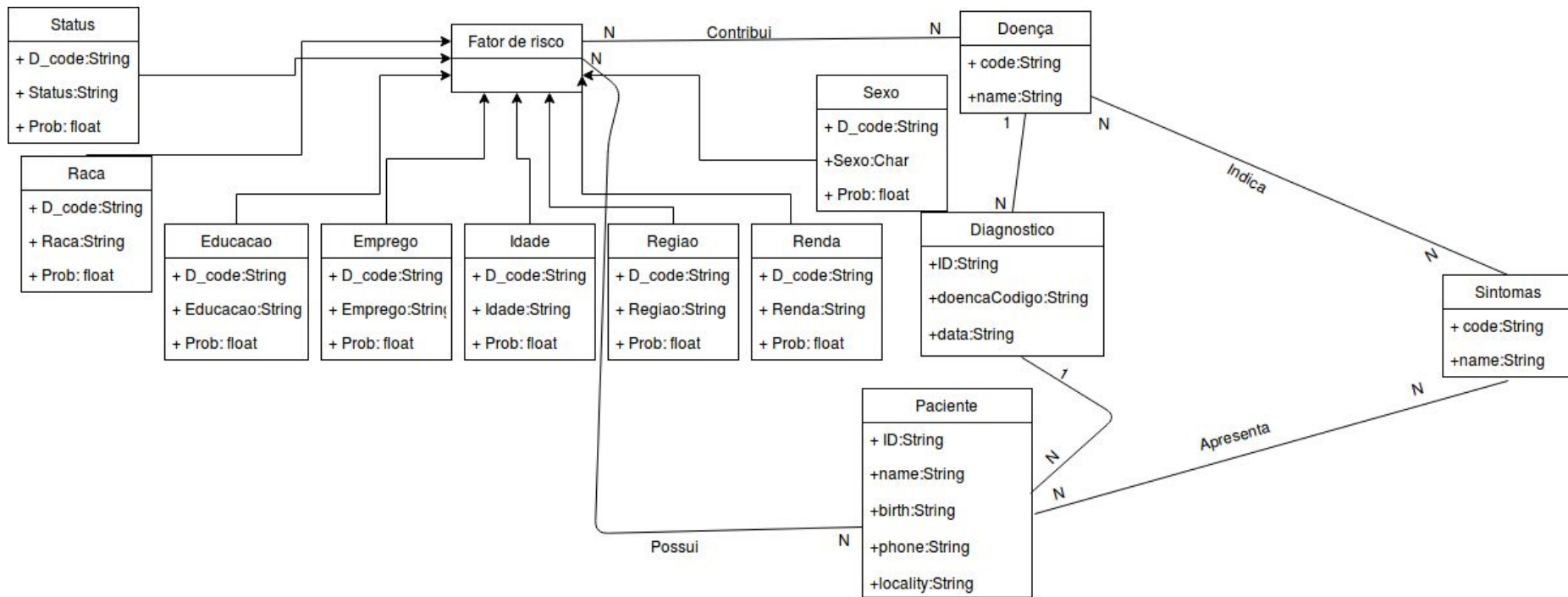
Selected characteristic	Diabetes ¹	Ulcers ¹	Kidney disease ¹	Liver disease ¹	Arthritis diagnosis ¹	Chronic joint symptoms ¹
Total	8.8 (0.26)	5.7 (0.17)	1.9 (0.09)	1.9 (0.10)	21.6 (0.27)	28.3 (0.36)
Sex						
Male	9.4 (0.28)	5.7 (0.24)	1.8 (0.13)	1.9 (0.15)	18.5 (0.38)	26.7 (0.52)
Female	8.4 (0.26)	5.8 (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 (2.53)	8.7 (1.84)	2.4 (0.99)	2.5 (1.12)	25.7 (2.91)	35.1 (3.98)
Asian	8.0 (0.36)	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)	25.9 (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 (3.96)



Modelo Conceitual inicial



Modelo Conceitual - versão 2





SQL



Queries

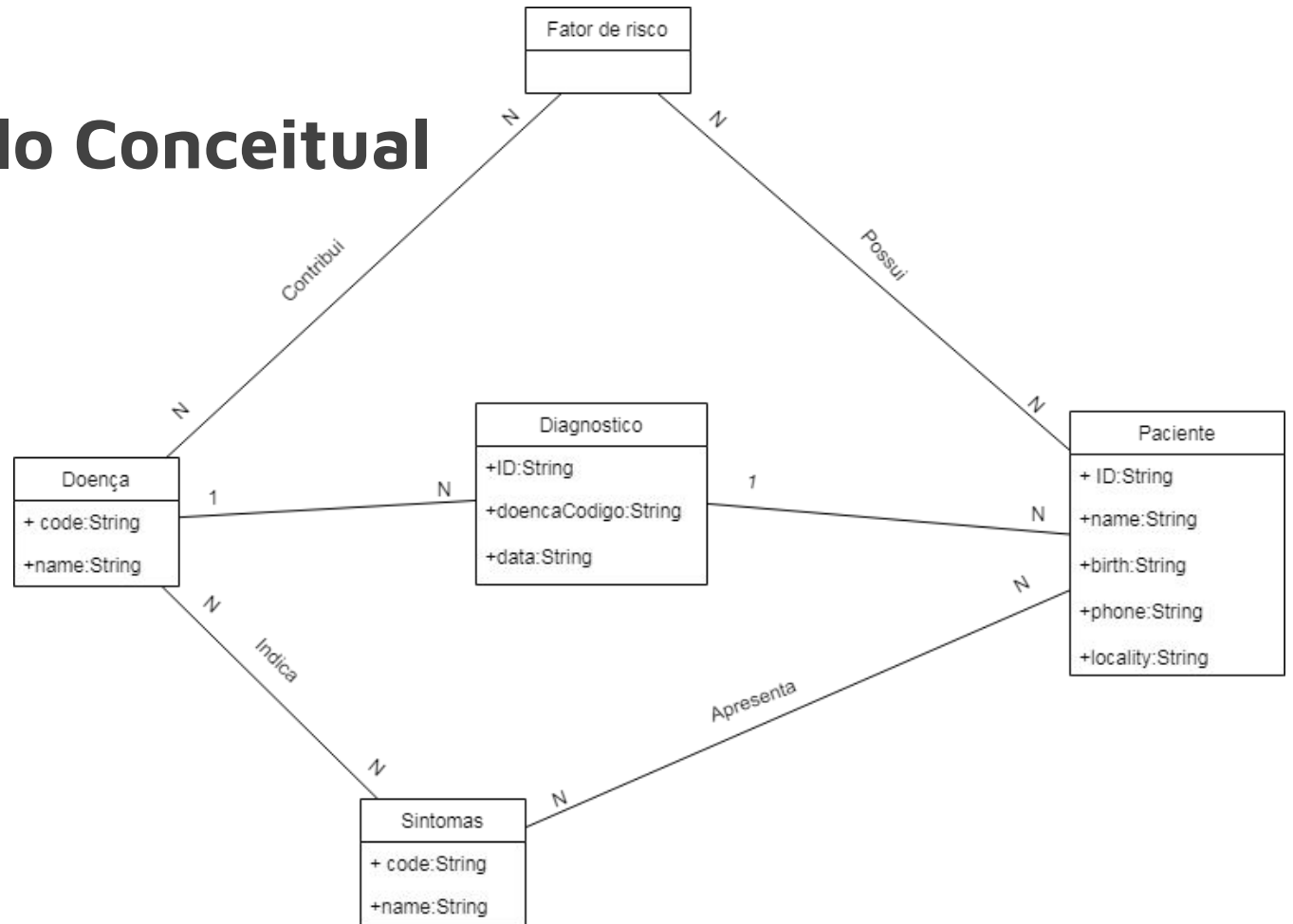
Sintoma mais relatado por pacientes

```
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 Conceitual final





Modelo lógico final

Doenca (codigo, nome)

Sintomas (codigo, nome)

Paciente(ID, nome, nascimento, celular, localidade)

Fator(Codigo, Fator, Tipo)

Diagnostico (ID, doencaCodigo, data)

CHE: doencaCodigo -> Doenca (codigo)

PacienteSintoma(pacienteID, sintomaCodigo)

CHE: IDpaciente -> Paciente (ID)

CHE: CodSintoma -> Sintomas (codigo)

DoencaSintomas (doencaCodigo, sintomaCodigo)

CHE: doencaCodigo -> Doenca (codigo)

CHE: sintomaCodigo -> Sintomas (codigo),

PacienteDiagnostico (pacienteID, diagnosticoID)

CHE: pacienteID -> Paciente (ID)

CHE: diagnosticoID -> Diagnostico(ID)

FatorDoenca(CodFator, Dcode, Prob)

CHE: CodFator -> Fator(Codigo)

CHE: Dcode -> Doenca(codigo)

FatorPaciente(CodFator, IDpaciente)

CHE: CodFator -> Fator(Codigo)

CHE: IDpaciente -> Paciente(ID)

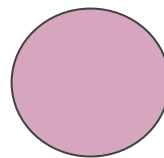
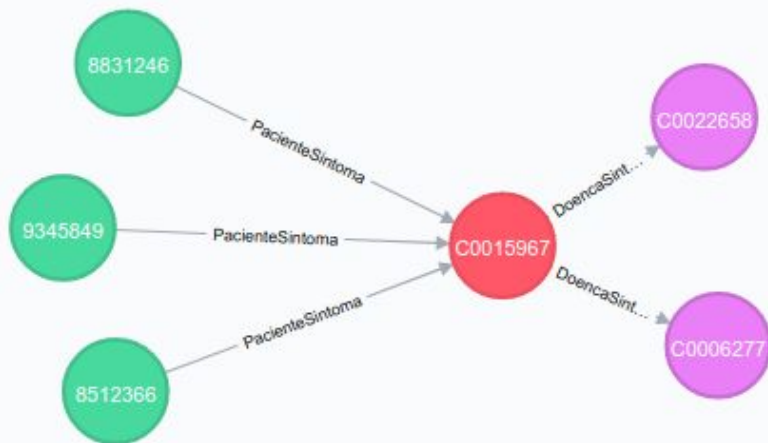


Grafos - Cypher

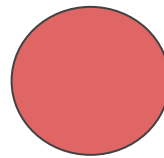


Queries

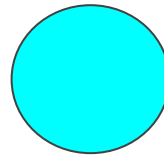
```
// retorna o caminho para um paciente que apresenta como sintoma 'febre'  
MATCH p = (l:Paciente)-[:PacienteSintoma]->(m:Sintomas {nome: 'fever'})-[t:DoencaSintoma]->(d:Doenca)  
RETURN p
```



Doença



Sintomas



Paciente



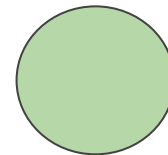
Queries

Dados certos sintomas, quais doenças são possíveis? (Dado nosso banco de dados)

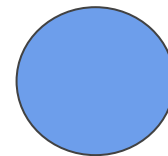
```
1 MATCH p2 = (m:Sintomas)-[t:DoencaSintoma]->(d:Doenca)
2 WHERE m.nome = 'vertigo' OR m.nome = 'pain chest' OR m.nome = 'asthenia' OR m.nome = 'nausea'
3 RETURN p2
```



Análise



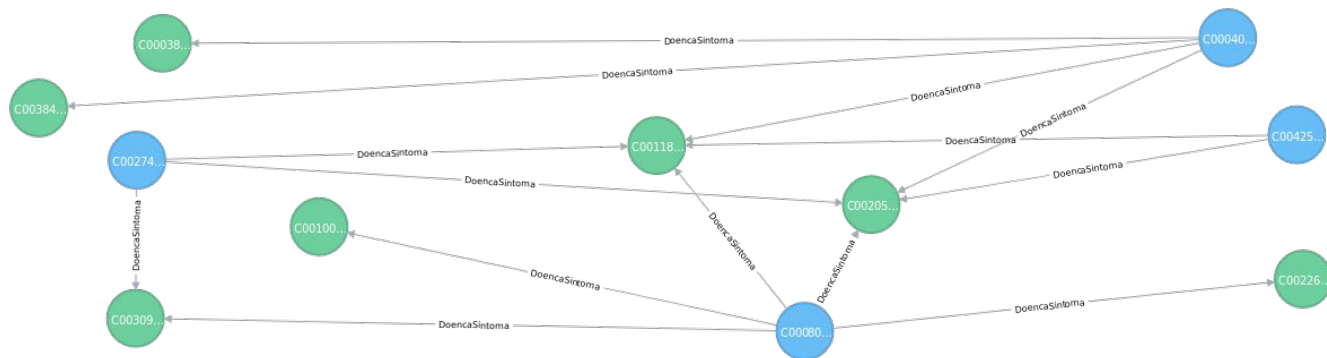
Doença



Sintomas

Análise gráfica das doenças influenciadas pelos determinados

sintomas





Queries

Dado os sintomas do paciente, qual a probabilidade dele ter determinada doença?

```
1 MATCH p = (l:Paciente {ID: 'patV89MMK'})-[:PacienteSintoma]->(m:Sintomas)-[t:DoencaSintoma]->(d:Doenca{nome: 'hypertensive disease'})
2 WITH COUNT(1) AS prob
3 MATCH q = (l:Paciente {ID: 'patV89MMK'})-[:PacienteSintoma]->(m:Sintomas)-[t:DoencaSintoma]->(d:Doenca) WITH COUNT(1)/prob AS tot
4 RETURN toFloat(1)/tot
```

toFloat(1)/tot

0.3333333333333333



Análise

Levando em conta somente os sintomas apresentados pelos pacientes

Idpaciente/Doenca	hypertensive disease	diabetes	myocardial infarction	coronary arteriosclerosis
patV89MMK	33,33%	33,33%	0,00%	7,14%
patKMPV11	20,00%	20,00%	0,00%	20,00%
pat1GY4KP	20,00%	20,00%	0,00%	20,00%
patO5BV5Q	25,00%	16,66%	0,00%	33,33%



Queries

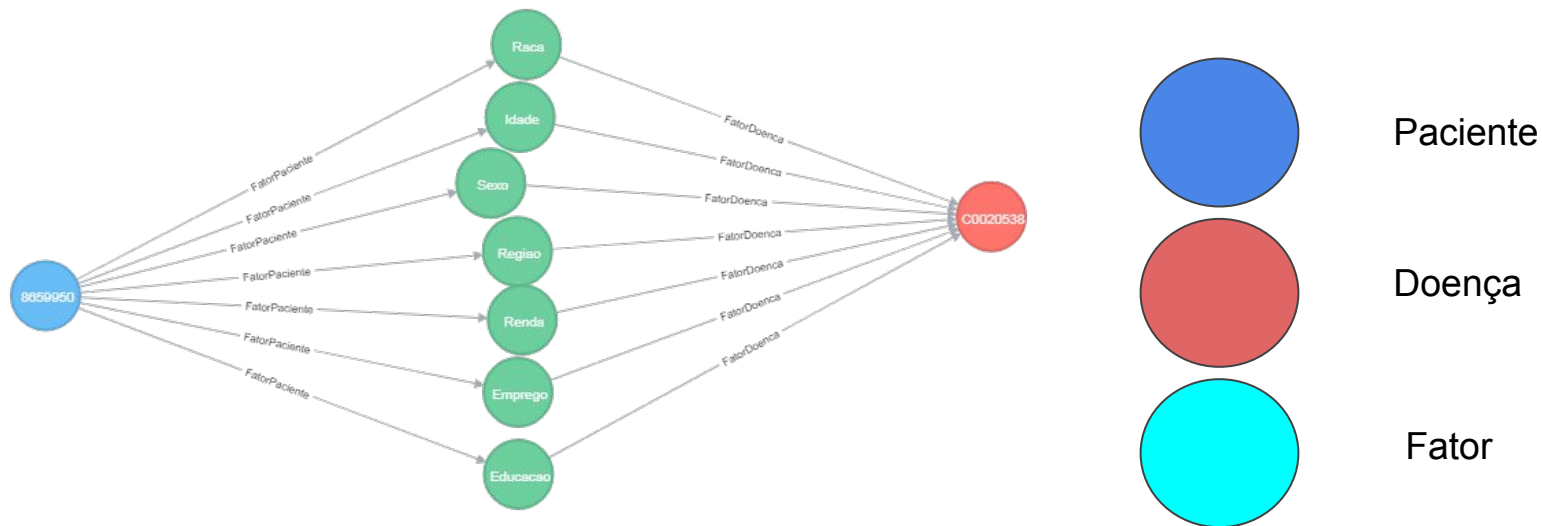
Quais fatores influenciam a doença?(Conforme dados fornecidos pelo paciente)

```
1 MATCH p = (n:Paciente{ID: 'pat0JATVU'})-[r:FatorPaciente]->(m:Fator)-[t:FatorDoenca]->(d:Doenca{nome: 'hypertensive disease'})
2 RETURN p
```




Análise

Análise gráfica dos fatores de risco fornecidos pelo paciente que influenciam determinada doença





Queries

Finalmente, qual a chance do paciente ter contraído uma determinada doença?

```
1 MATCH p = (l:Paciente {ID: 'patV89MMK'})-[:PacienteSintoma]->(m:Sintomas)-[t:DoencaSintoma]->(d:Doenca{nome: 'coronary arteriosclerosis'})
2 WITH COUNT(1) AS prob
3 MATCH q = (l:Paciente {ID: 'patV89MMK'})-[:PacienteSintoma]->(m:Sintomas)-[t:DoencaSintoma]->(d:Doenca) WITH COUNT(1)/prob AS tot
4 MATCH o = (n:Paciente{ID: 'patV89MMK'})-[FatorPaciente]->(Fator)-[t:FatorDoenca]->(d:Doenca{nome: 'coronary arteriosclerosis'})
5 RETURN (AVG(toFloat(replace(t.Probabilidade, ",", ".")))/100)*toFloat(1)/tot
```

(AVG(toFloat(replace(t.Probabilidade, ",", ".")))/100)*toFloat(1)/tot

0.004892857142857142



Análise

Levando em conta a média das influências dos fatores de risco e os sintomas do paciente

Idpaciente/Doença	hypertensive disease	diabetes	myocardial infarction	coronary arteriosclerosis
patV89MMK	8,45%	3,45%	0,00%	0,50%
patKMPV11	5,40%	1,90%	0,00%	1,20%
pat1GY4KP	4,80%	1,50%	0,00%	0,85%
patO5BV5Q	5,75%	1,40%	0,00%	1,55%



MeSH e SPARQL



Queries

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

```
WHERE {
  mesh:D009369 meshv:treeNumber ?treeNum .
  ?childTreeNum meshv:parentTreeNumber+ ?treeNum .
  ?descriptor meshv:treeNumber ?childTreeNum .
  ?descriptor rdfs:label ?label .
}
```

```
ORDER BY ?label
```

```
# Descendants of Heart Diseases (D006331)
```

descriptor	label
mesh:D000182	ACTH Syndrome, Ectopic
mesh:D049913	ACTH-Secreting Pituitary Adenoma
mesh:D000008	Abdominal Neoplasms
mesh:D058739	Aberrant Crypt Foci
mesh:D049309	Acanthoma
mesh:D018250	Acrospiroma
mesh:D050398	Adamantinoma
mesh:D000230	Adenocarcinoma



Queries

```
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 ?d ?dName ?c ?cName
FROM <http://id.nlm.nih.gov/mesh>

WHERE {
  ?d a meshv:Descriptor .
  ?d meshv:active 1 .
  ?d meshv:concept ?c .
  ?d rdfs:label ?dName .
  ?c rdfs:label ?cName
  FILTER(REGEX(?dName,"Dermatoses","i") || REGEX(?cName,"Dermatoses","i"))
}
ORDER BY ?d
```

Find a set of descriptors and terms that contain "Dermatoses."

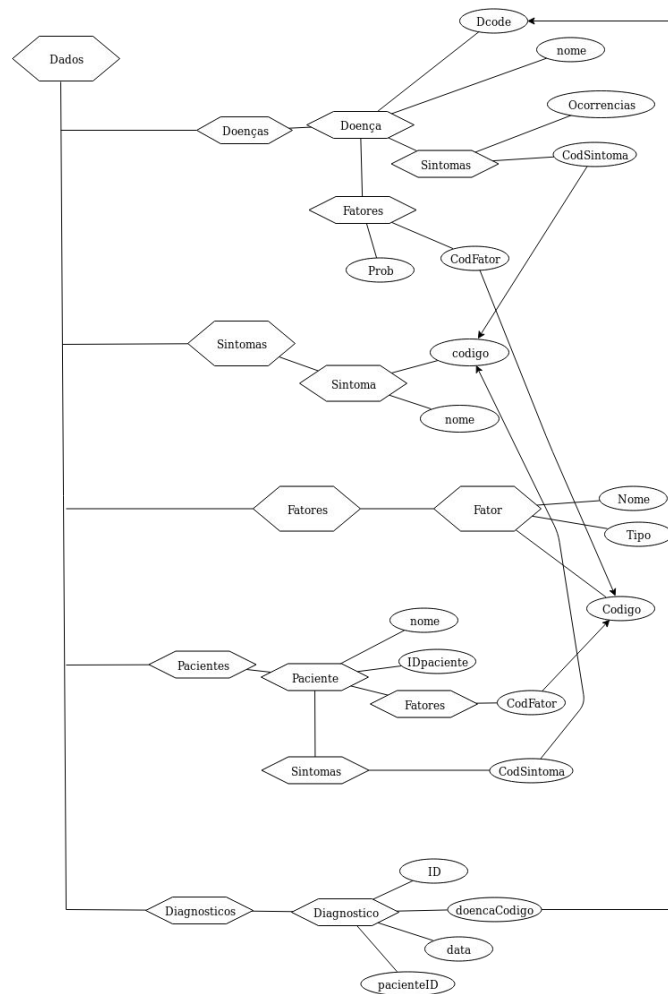
d	dName
mesh:D005148	Facial Dermatoses
mesh:D005148	Facial Dermatoses
mesh:D005148	Facial Dermatoses
mesh:D005533	Foot Dermatoses
mesh:D006229	Hand Dermatoses
mesh:D007868	Leg Dermatoses
mesh:D012536	Scalp Dermatoses



XML



Modelo etapa 4





Foram analisadas 9 doenças a partir de nosso banco de dados:

Hipertensão

Úlcera péptica

Diabetes

Doença renal

Arteriosclerose coronariana

Enfisema pulmonar

acidente cerebrovascular

Artrite

Bronquite



Queries

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

```
for $i in ($documento//Doenca/Sintomas)
let $med := sum($documento//Doenca/Sintomas[CodSintoma = $i/CodSintoma]/Ocorrencias)
let $ord := ($documento//Doenca/Sintomas[CodSintoma = $i/CodSintoma])
order by $i/CodSintoma
return data($ord/CodSintoma)
```

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

```
for $i in ($documento//Sintoma)
let $ord := sum($documento//Doenca/Sintomas[CodSintoma = $i/codigo]/Ocorrencias)
let $gr := $documento//Sintomas/codigo
group by $gr
order by $gr
return ($ord)
```

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

```
for $i in ($documento//Doenca/Sintomas)
let $med := sum($documento//Doenca/Sintomas[CodSintoma = $i/CodSintoma]/Ocorrencias)
let $ord := $documento//Doenca/Sintomas/Ocorrencias
let $gr := $documento//Sintomas/codigo
group by $gr
order by $gr
for $k in ($documento//Sintoma)
where $k/codigo = $i/CodSintoma
return data($k/nome)
```



Análise



Sintoma menos seletivo



Sintoma mais seletivo

Nome sintoma	Codigo Sintoma	Nº Ocorrencias
pain chest	C0008031	6307
asthenia	C0004093	5848
vertigo	C0042571	3363
nausea	C0027497	4927
sweat	C0038990	6068
shortness of breath	C0392680	6515
fever	C0015967	268
hypokinesia	C0086439	1284
orthopnea	C0085619	2705
pressure chest	C0438716	4647
hemodynamically stable	C0578150	179
palpitation	C0030252	3363
nervousness	C0027769	143
decreased body weight	C0043096	172
dizziness	C0012833	3363
distended abdomen	C0000731	179
rale	C0034642	1501
fall	C0085639	3363
pain abdominal	C0000737	143
dyspnea	C0013404	252
cough	C0010200	172
productive cough	C0239134	172



Queries

Sintomas em comum entre duas doenças

```
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)
```



Análise

Sintomas em comum entre diabetes e hipertensão:

Dor no peito

Náusea

Suor

Falta de ar

Astenia



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>
```



Resultado

Homens:

bronchitis

ulcer peptic

kidney disease

arthritis

Mulheres:

hypertensive disease

diabetes

coronary arteriosclerosis

accident cerebrovascular

emphysema pulmonary



Queries

Fatores que mais influenciam em determinada Doença (levando em conta a media de probabilidades)



Análise