

MC536

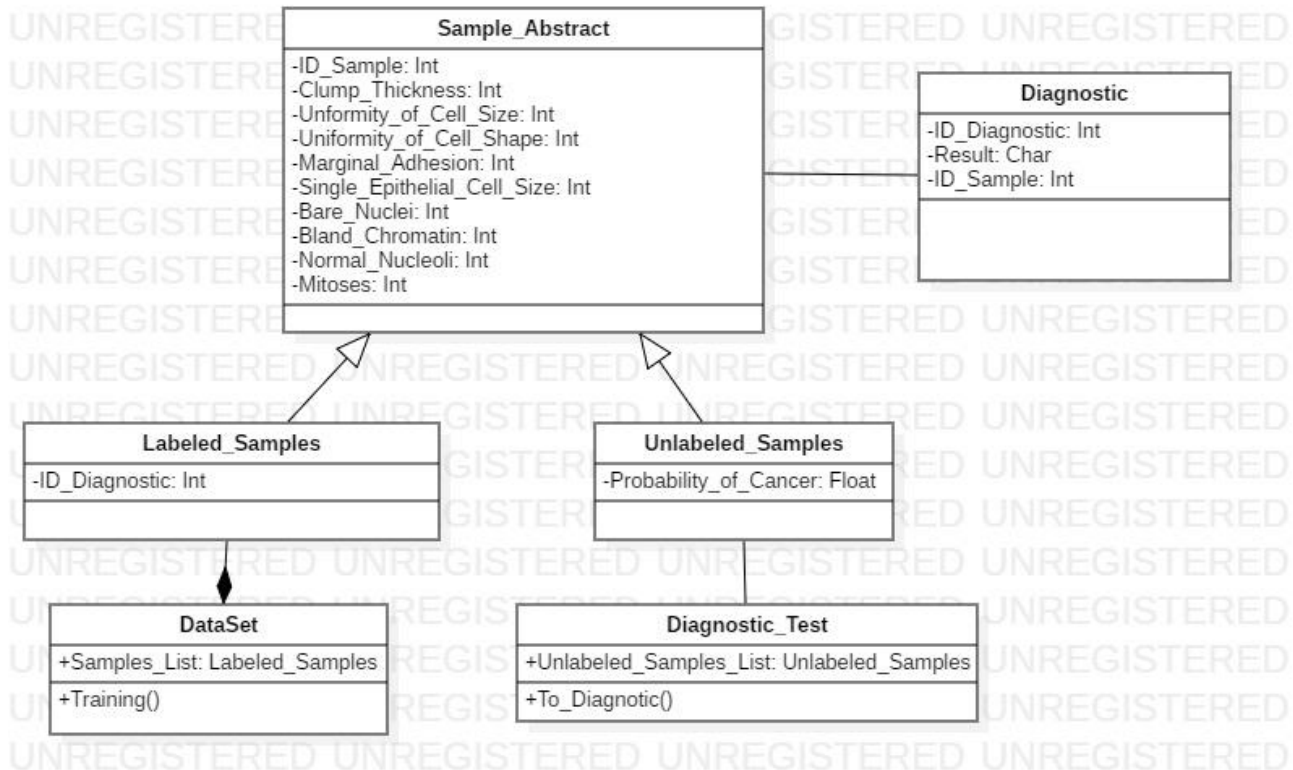
BANCO DE DADOS

Nome: Murilo Guidetti Andrietta
Nome: Naomi Takemoto

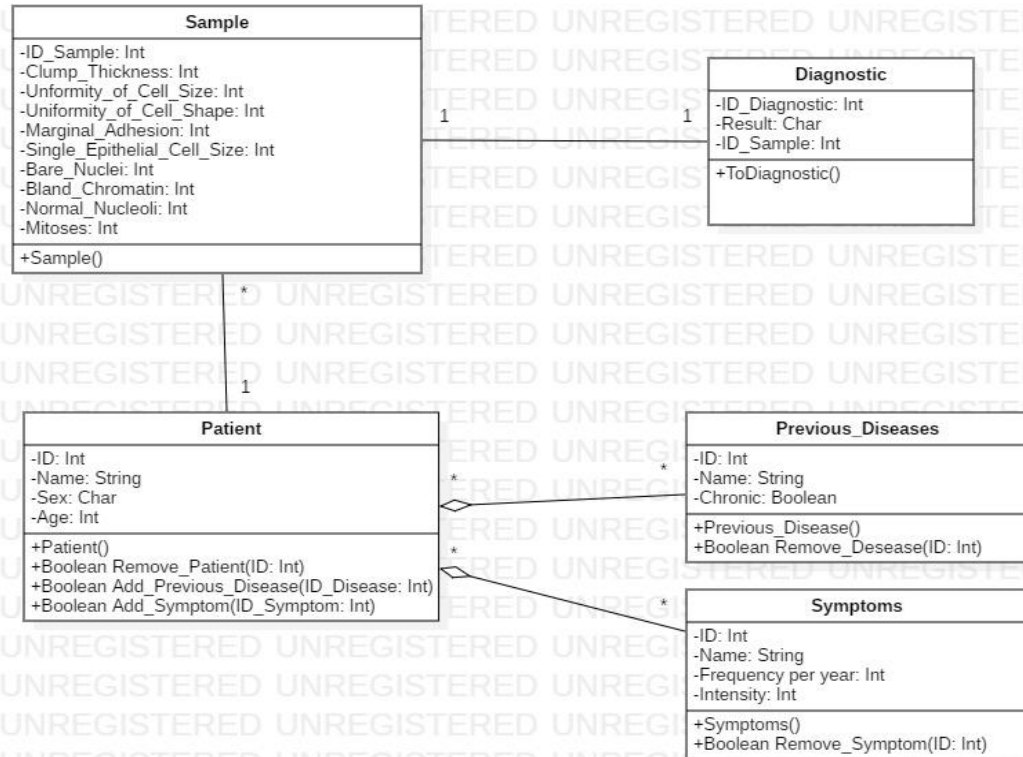
RA:147472
RA:184849



Modelo Conceitual Etapa 1



Modelo conceitual da etapa 2





Consultas Interessantes

Número de amostras Malígnas - 239/683 (35%)

Número de Amostras Benignas - 444/683 (65%)

Número de amostras malignas que apresentam anomalias no tamanho e forma das células - 98/239 (41%)

Número de amostras malignas que apresentam anomalias no núcleo - 70/239 (30%)



Consultas Interessantes

Número de amostras benignas que possuem baixo valor de anomalia na forma e tamanho da célula - 424/444 (95%)

Número de amostras que possuem baixo valor de anomalia no núcleo - 415/683 (61%)

Número de amostras malignas que possuem média de todos os atributos maiores que 5 - 126/239 (52%)

Modelo Relacional: SQL



Detalhes - Consultas interessantes

Número de amostras malignas

```
In [6]: DROP TABLE IF EXISTS AmostrasMalignas;  
  
        CREATE VIEW AmostrasMalignas AS  
        SELECT COUNT(*) Malignos  
        FROM Sample s  
        WHERE s.Id_Diagnostic = 4;  
  
        SELECT * FROM AmostrasMalignas;
```

Out[6]: 239

Número de Amostras Benígnas

```
In [7]: DROP TABLE IF EXISTS AmostrasBenignas;  
  
        CREATE VIEW AmostrasBenignas AS  
        SELECT COUNT(*) Benignos  
        FROM Sample s  
        WHERE s.Id_Diagnostic = 2;  
  
        SELECT * FROM AmostrasBenignas;
```

Out[7]: 444



Detalhes - Consultas interessantes

Número de amostras malignas que apresentam anormalias no tamanho e forma das células.

```
In [8]: SELECT COUNT(*) M
        FROM Sample, AmostrasMalignas
        WHERE (Uniformity_of_Cell_Size >= 7) AND (Uniformity_of_Cell_Shape >= 7) AND (Sample.Id_Diagnostic = 4);
```

Out[8]: 98

Número de amostras malignas que apresentam anormalias no núcleo.

```
In [9]: SELECT COUNT(*) M
        FROM Sample, AmostrasMalignas
        WHERE (Normal_Nucleoli >= 7) AND (Bare_Nuclei >= 7) AND (Sample.Id_Diagnostic = 4);
```

Out[9]: 70

Número de amostras benígnas que possuem baixo valor de anomalia na forma e tamanho da célula.

```
In [10]: SELECT COUNT(*) M
         FROM Sample, AmostrasMalignas
         WHERE (Uniformity_of_Cell_Size <= 3) AND (Uniformity_of_Cell_Shape <= 3) AND (Sample.Id_Diagnostic = 2);
```

Out[10]: 424



Detalhes - Consultas interessantes

Número de amostras que possuem baixo valor de anomalia no núcleo.

```
In [11]: SELECT COUNT(*) M
FROM Sample, AmostrasMalignas
WHERE (Normal_Nucleoli <= 3) AND (Bare_Nuclei <= 3) AND (Sample.Id_Diagnostic = 2);
```

Out[11]: 415

Número de amostras malignas que possuem média de todos os atributos maiores que 5.

```
In [12]: DROP TABLE IF EXISTS Attributes_Average;

CREATE VIEW Attributes_Average AS
SELECT Id_Sample, ((Clump_Thickness + Uniformity_of_Cell_Size + Uniformity_of_Cell_Shape + Marginal_Adhesion + Single_Epithelial_Cell_Size +
    Bare_Nuclei + Bland_Chromatin + Normal_Nucleoli + Mitoses)/9) AS Media FROM Sample;

SELECT COUNT(*) FROM Attributes_Average AA, Sample S WHERE (AA.Id_Sample = S.Id_Sample) AND (AA.Media > 5) AND (S.Id_Diagnostic = 4);
```

Out[12]: 126



Consultas: Algumas Estatísticas

Média dos atributos das amostras malignas

```
match (s:Sample)-[h:Has]-(n)
match (s)-[h]-(d:Diagnostic)
where d.code = 4
return avg(n.code)
```

4.0

Média dos atributos das amostras benígnas

```
match (s:Sample)-[h:Has]-(n)
match (s)-[h]-(d:Diagnostic)
where d.code = 2
return avg(n.code)
```

2.0



Consultas: Algunas Estadísticas

Busca por atributos redundantes

```
match (m_a:Marginal_Adhesion)
match(m:Bare_Nuclei)
match(s1:Sample)
where (s1)-[:Has]->(m_a) and (s1)-[:Has]->(m_a) and m.code = m_a.code
return count(s1)
```

1454



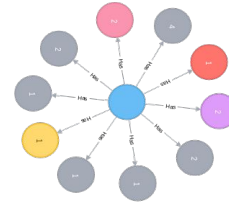
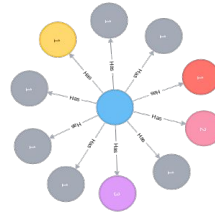
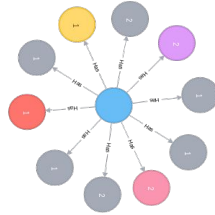
Modelo Lógico Adaptado para Cypher

- O modelo lógico consiste na descrição de nós e arestas
- Exemplos:
 - No: Sample(name, code)
 - Aresta: Sample -> Clump_Thickness

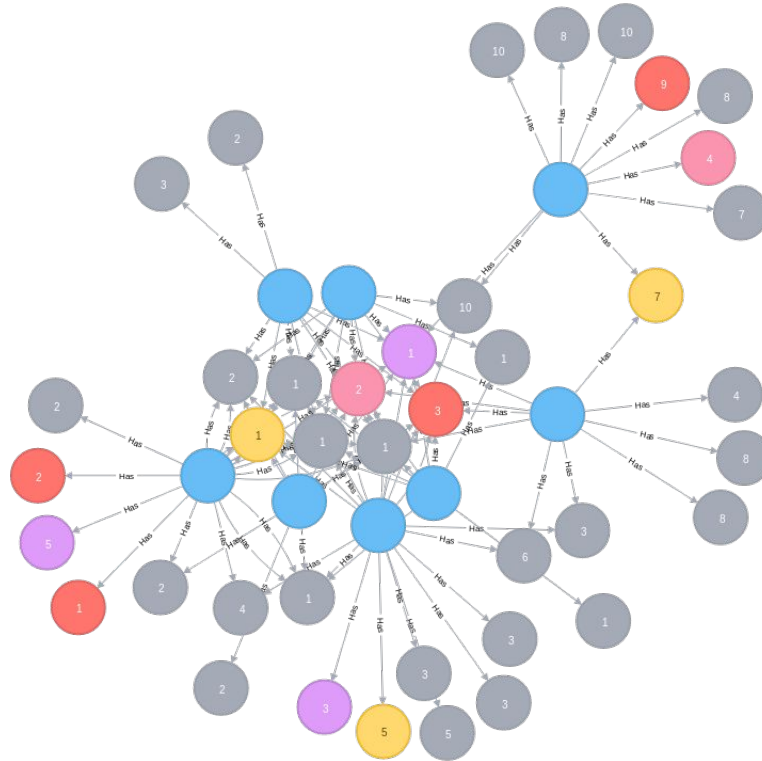
Neste modelo cada amostra (Sample) possui 11 atributos.

Essa relação é representada pelas arestas entre a amostra e o atributo.

Possível representação gráfica do modelo lógico



Para evitar repetição de nós



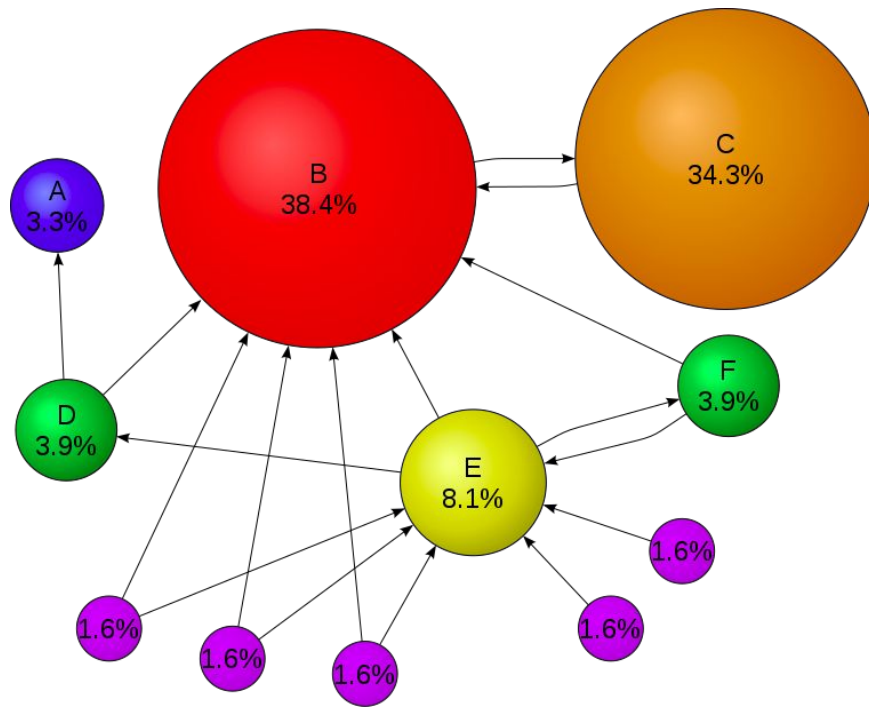


Consultas: Page Rank

- Verifica, dentre outros fatores, quais são os nós com mais links, este algoritmo é comum em rank de páginas por sites de pesquisa.
- Usado pelo Google para posicionar os websites pesquisados.
 - *A métrica PageRank de uma página representa a probabilidade de uma pessoa chegar a essa página, clicando aleatoriamente em hiperligações*



Consultas: Page Rank





Consultas: Page Rank

```
//Algoritmo Page Rank
```

```
CALL algo.pageRank.stream('Page', 'LINKS', {iterations:20, dampingFactor:0.85})
```

```
YIELD nodeId, score
```

```
RETURN algo.getNodeById(nodeId).name AS page, algo.getNodeById(nodeId).code AS code , score
```

```
ORDER BY score DESC
```



Consultas: Page Rank

page	code	score
"Mitoses"	1	6.721027
"Diagnostic"	2	5.242027
"Normal_Nucleoli"	1	5.145552
"Bare_Nuclei"	1	4.7444455
"Marginal_Adhesion"	1	4.6988769999999995
"Single_Epithelial_Cell_Size"	2	4.4953020000000001
"Uniformity_of_Cell_Size"	1	4.407276
"Uniformity_of_Cell_Shape"	1	4.09173900000000005
"Diagnostic"	4	3.093788
"Bland_Chromatin"	2	2.034263
"Bland_Chromatin"	3	1.9970075



Consultas: Closeness Centrality

- A centralidade de proximidade de um nó mede sua distância média (distância inversa) a todos os outros nós.
- Uma alta pontuação de proximidade têm as menores distâncias para todos os outros nós



Consultas: Closeness Centrality

```
//The Closeness Centrality algorithm
```

```
CALL algo.closeness.stream('Node', 'LINK')
```

```
YIELD nodeId, centrality
```

```
RETURN algo.getNodeById(nodeId).name AS name, algo.getNodeById(nodeId).code AS code, centrality
```

```
ORDER BY centrality DESC
```

```
LIMIT 20;
```

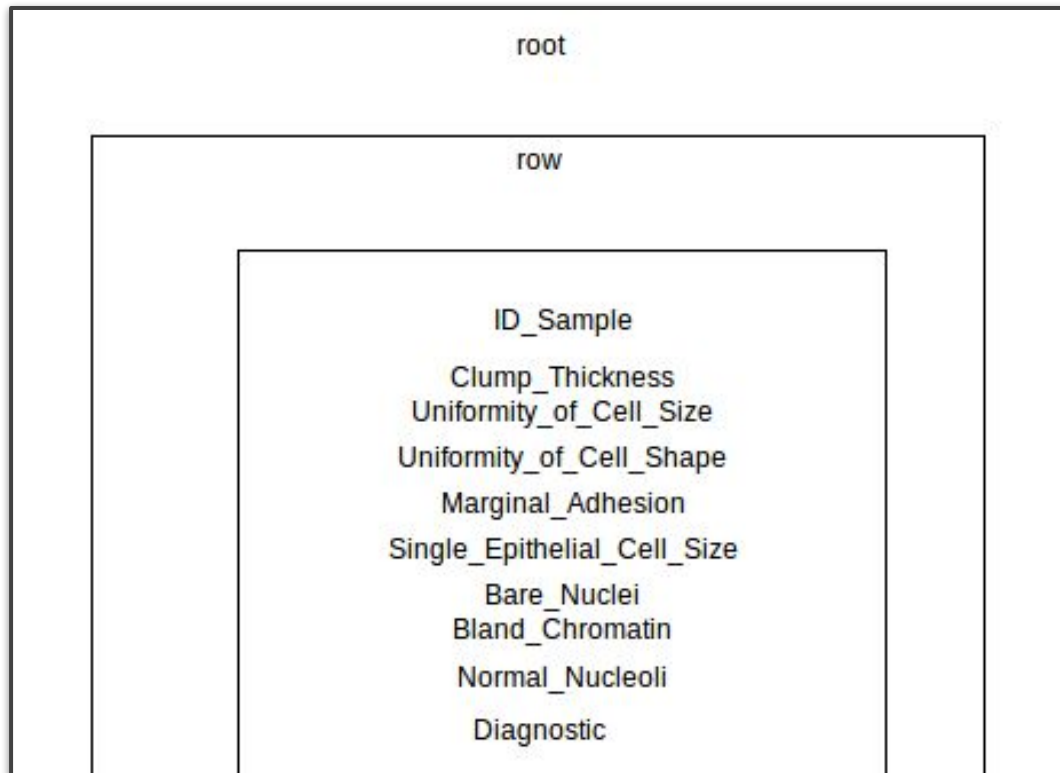


Consultas: Closeness Centrality

name	code	centrality
"Mitoses"	1	0.6923076923076923
"Normal_Nucleoli"	1	0.5555555555555556
"Diagnostic"	2	0.5546995377503852
"Bare_Nuclei"	1	0.5263157894736842
"Marginal_Adhesion"	1	0.5247813411078717
"Single_Epithelial_Cell_Size"	2	0.5084745762711864
"Uniformity_of_Cell_Size"	1	0.5006954102920723
"Sample"	1299596	0.482897384305835
"Sample"	1171710	0.482897384305835
"Sample"	1198641	0.482897384305835
"Sample"	695091	0.4822505023442733
"Sample"	1168736	0.4816053511705686



Modelo Lógico XML





Consultas Interessantes

- Número de amostras benígnas

```
count(//row[Diagnostic=2])
```

- Número de amostras malignas com anomalias na uniformidade das células

```
count(//row[Uniformity_of_Cell_Size >= 7][Uniformity_of_Cell_Shape >= 7][Diagnostic=4])
```

- Número de amostras malignas com Normal Nucleoli maior que 7

```
count(//row[Normal_Nucleoli >= 7][Bare_Nuclei >= 7][Diagnostic=4])
```

- Número de amostras benígnas que possuem grau de uniformidade do formato da célula menor ou igual a 3

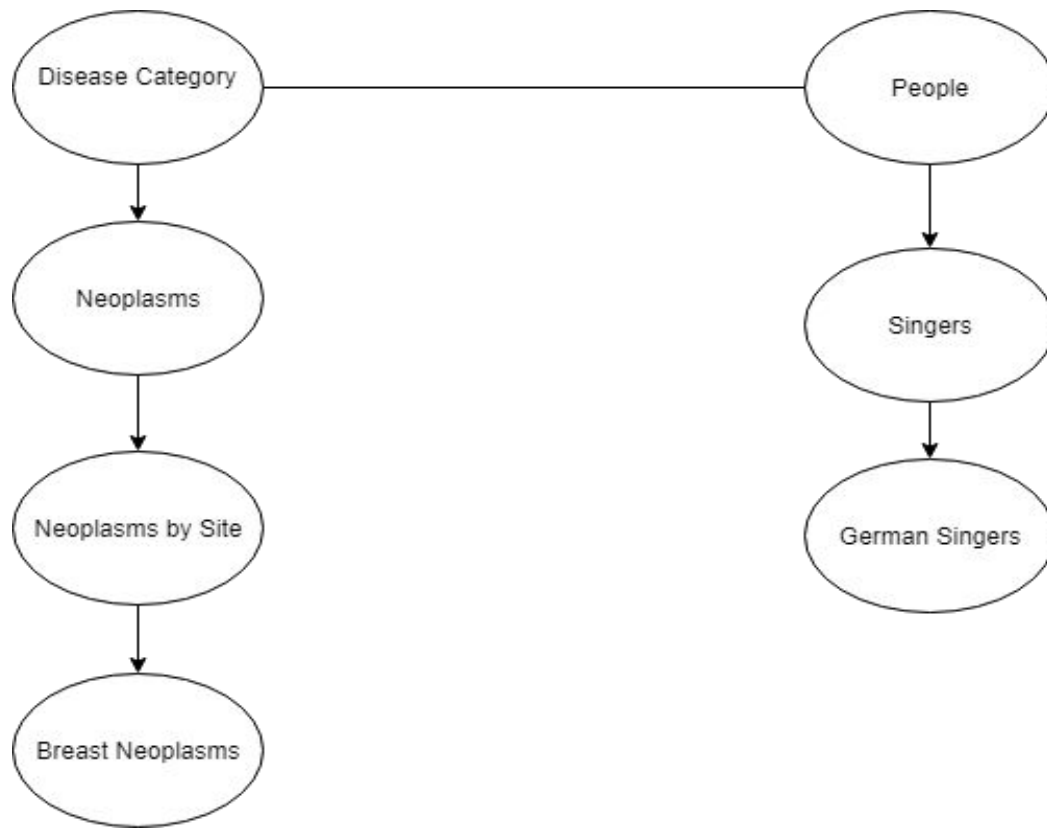
```
count(//row[Uniformity_of_Cell_Size <= 3][Uniformity_of_Cell_Shape <= 3][Diagnostic = 2])
```

- Número de amostras benígnas com Normal Nucleoli e Bare Nuclei menores que 3

```
count(//row[Normal_Nucleoli <= 3][Bare_Nuclei <= 3][Diagnostic = 2])
```

Modelo Grafos: RDF/SPARQL

Modelo Lógico RDF/SPARQL





Consultas Interessantes

Conceitos Relacionados à Câncer de Mama - MESH

```
SELECT *  
FROM <http://id.nlm.nih.gov/mesh>  
  
WHERE {  
  ?s ?p mesh:D001943  
}
```

s	p
http://id.nlm.nih.gov/mesh/C112748	http://id.nlm.nih.gov/mesh/vocab#indexerConsiderAlso
http://id.nlm.nih.gov/mesh/C586693	http://id.nlm.nih.gov/mesh/vocab#indexerConsiderAlso
http://id.nlm.nih.gov/mesh/C482680	http://id.nlm.nih.gov/mesh/vocab#indexerConsiderAlso
http://id.nlm.nih.gov/mesh/C485206	http://id.nlm.nih.gov/mesh/vocab#indexerConsiderAlso
http://id.nlm.nih.gov/mesh/C485207	http://id.nlm.nih.gov/mesh/vocab#indexerConsiderAlso
http://id.nlm.nih.gov/mesh/C485776	http://id.nlm.nih.gov/mesh/vocab#indexerConsiderAlso
http://id.nlm.nih.gov/mesh/C580973	http://id.nlm.nih.gov/mesh/vocab#indexerConsiderAlso
http://id.nlm.nih.gov/mesh/C583729	http://id.nlm.nih.gov/mesh/vocab#indexerConsiderAlso
http://id.nlm.nih.gov/mesh/C059917	http://id.nlm.nih.gov/mesh/vocab#indexerConsiderAlso
http://id.nlm.nih.gov/mesh/C407785	http://id.nlm.nih.gov/mesh/vocab#indexerConsiderAlso



Consultas Interessantes

Principais institutos/organizações relacionados ao tema de câncer de mama

```
: select ?search
  where{
    ?search dct:subject dbc:Breast_cancer
  }
```

person
http://dbpedia.org/resource/Clinical_Breast_Cancer
http://dbpedia.org/resource/Komen_Brinker_Award_for_Scientific_Distinction
http://dbpedia.org/resource/Male_breast_cancer
http://dbpedia.org/resource/Breast_cancer_management
http://dbpedia.org/resource/Breast_Cancer_Research_and_Treatment
http://dbpedia.org/resource/Abortion-breast_cancer_hypothesis
http://dbpedia.org/resource/Metastatic_breast_cancer
http://dbpedia.org/resource/Nottingham_Prognostic_Index
http://dbpedia.org/resource/Paget's_disease_of_the_breast



Consultas Interessantes

Pessoas que sobreviveram à doença

```
select ?search
where{
    ?search dct:subject dbc:Breast_cancer_survivors .
}
```

search
http://dbpedia.org/resource/Hildegard_Knef
http://dbpedia.org/resource/Olivia_Newton-John
http://dbpedia.org/resource/Sandra_Day_O'Connor
http://dbpedia.org/resource/Soong_Mei-ling
http://dbpedia.org/resource/Cheryl_Crane
http://dbpedia.org/resource/Drew_Gilpin_Faust
http://dbpedia.org/resource/Janet_Napolitano



Consultas Interessantes

Cantoras Alemãs que sobreviveram ao câncer de mama

```
: select ?search
  where{
    ?search dct:subject dbc:Breast_cancer_survivors .#search for survivors
    ?search dct:subject dbc:German_female_singers    #among the survivors, who was a female German Singer
  }
```

search
http://dbpedia.org/resource/Hildegard_Knef

Total: 1, Shown: 1





Consultas Interessantes

Pessoas que morreram de câncer de mama

```
: select ?person
  where{
    ?person dbo:deathCause dbr:Breast_cancer
  }
```

person
http://dbpedia.org/resource/Ingrid_Bergman
http://dbpedia.org/resource/Bibi_Besch
http://dbpedia.org/resource/Jeanne_Bates
http://dbpedia.org/resource/Roxie_Roker
http://dbpedia.org/resource/Laura_Ziskin
http://dbpedia.org/resource/Marcia_Strassman
http://dbpedia.org/resource/Norma_Crane



Consultas Interessantes

Topico relacionado ao câncer de mama que foi causa de morte de alguma pessoa

Esta consulta retorna dois resultados, o primeiro é o tópico (Male Breast Cancer), que foi responsável pela morte de uma pessoa. Esta por sua vez era um assassino serial.

```
: select ?search, ?person
where{
    ?search dct:subject dbc:Breast_cancer . #subjects related to breast_cancer
    ?person dbp:cause ?search              #person whose death cause is a subject related to breast cancer
}
```

search	person
http://dbpedia.org/resource/Male_breast_cancer	http://dbpedia.org/resource/Joseph_Christopher

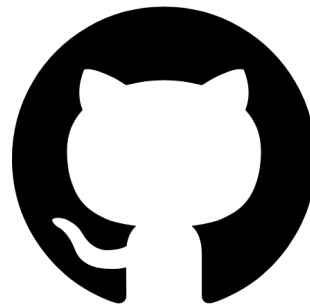
Total: 1, Shown: 1





Consultas Cypher

- [Sandbox](#)
- [GitHub](#)





Referências

Repositório gitHub:

<https://github.com/ntak1/404NotFound>

Page Rank:

<https://pt.wikipedia.org/wiki/PageRank>