

Identificação de Doenças Mentais

Trabalho MC 536 - Etapa 3
171828 - Gabriel Alves Tabchoury
185441 - Paulo Afonso Martins Januário

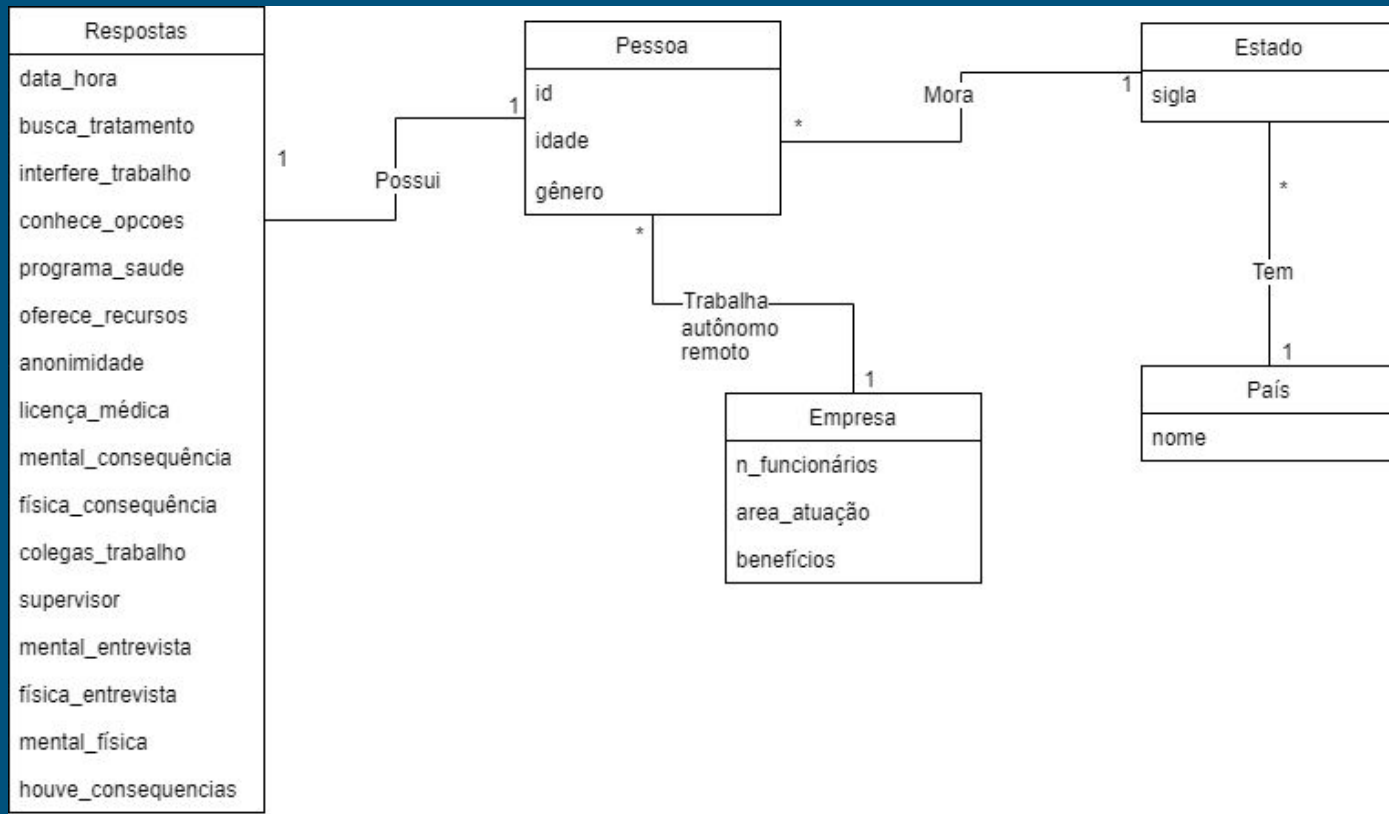
UNICAMP - 2s 2018



Etapas 2 e 3...



Modelo Conceitual



Modelo Lógico

- **PESSOA** (id, idade, genero, estado, pais)
- **PESSOA_RESPOSTA** (idPessoa, data_hora, trabalho proprio, historico_familiar, tratamento, interferencia_trabalho, trabalho_remoto, empresa_ti, conhece_opcoes, programa_saude, busca_tratamento, anonimidade, licenca_medica, mental_consequencia, fisica_consequencia, mental_entrevista, fisica_entrevista, mental_fisica, houve_consequencias)
 - idPessoa -> Chave Estrangeira: Pessoa(id)

Análise de Dados

Fonte: <https://www.kaggle.com/osmi/mental-health-in-tech-survey>

- Dados insuficientes (Empresa e Região)
 - 1256 entrevistas
 - Mais da metade sem estado
 - Maior parte dos dados -> EUA e Inglaterra
- Conceito inicial
 - Gênero, Idade e Região
- Análise baseada em Gênero e Idade

Análise dos Dados

Número de entrevistados por gênero

```
SELECT P.genero, COUNT(*)  
FROM Pessoa_Resposta R, Pessoa P  
WHERE R.idPessoa=P.id  
GROUP BY P.genero
```

index	GENERO	COUNT(*)
0	Female	247
1	Male	991
2	Non-binary	17

Análise dos Dados

Interferência da saúde mental no trabalho por gênero

```
SELECT T.genero, T.total, I.quantidade, ROUND((CAST(I.quantidade as
double)/CAST(T.total as double))*100,1) porcentagem
FROM
  (SELECT genero, COUNT(*) total
   FROM Pessoa_Resposta, Pessoa
   WHERE idPessoa=id
   GROUP BY genero) T,
  (SELECT genero, COUNT(*) quantidade
   FROM Pessoa, Pessoa_Resposta
   WHERE idPessoa=id
   AND (interferencia_trabalho='Sometimes' OR interferencia_trabalho='Often')
   GROUP BY genero) I
WHERE T.genero = I.genero
```

index	GENERO	TOTAL	QUANTIDADE	PORCENTAGEM
0	Non-binary	17	11	64.7
1	Female	247	142	57.5
2	Male	991	452	45.6

Análise dos Dados

Interferência da saúde mental no trabalho por idade

```
SELECT P.idade, COUNT(*)  
FROM Pessoa_Resposta R, Pessoa P  
WHERE R.idPessoa=P.id  
AND (R.interferencia_trabalho='Sometimes'  
OR R.interferencia_trabalho='Often')  
GROUP BY P.idade
```

index	IDADE	COUNT(*)
0	33	34
1	34	34
2	35	28
3	36	18
4	37	21
5	38	22
6	39	19
7	40	19
8	41	10
9	42	11
10	43	14
11	44	6
12	45	5
13	46	8
14	47	1
15	48	2
16	49	3
17	50	2
18	51	1
19	54	2
20	55	2
21	57	1
22	60	1

23	61	1
24	-1726	1
25	329	1
26	18	3
27	19	5
28	20	2
29	21	8
30	22	12
31	23	26
32	24	22
33	25	34
34	26	37
35	27	25
36	28	33
37	29	30
38	30	36
39	31	28
40	32	38

Análise dos Dados

Porcentagem de pessoas que trabalham em TI as quais acreditam que o seus chefes dão importância para a saúde mental

```
SELECT ROUND((
  CAST((SELECT COUNT(*)
  FROM Pessoa_Resposta
  WHERE mental_fisica='Yes'
  AND empresa_ti='Yes') as double) /
  CAST(( SELECT COUNT(*)
  FROM Pessoa_Resposta
  WHERE empresa_ti='Yes') as double))*100,1) Sim,
ROUND((
  CAST((SELECT COUNT(*)
  FROM Pessoa_Resposta
  WHERE mental_fisica='No'
  AND empresa_ti='Yes') as double) /
  CAST(( SELECT COUNT(*)
  FROM Pessoa_Resposta
  WHERE empresa_ti='Yes') as double))*100,1) Nao,
ROUND((
  CAST((SELECT COUNT(*)
  FROM Pessoa_Resposta
  WHERE mental_fisica='Don't know'
  AND empresa_ti='Yes') as double) /
  CAST(( SELECT COUNT(*)
  FROM Pessoa_Resposta
  WHERE empresa_ti='Yes') as double))*100,1) Nao_sei
```

index	Key	Value
0	SIM	28.9
1	NAO	24.5
2	NAO_SEI	46.5

Análise dos Dados

Descobrir se pessoas que cuidam do próprio negócio buscam mais tratamento que as outras

```
SELECT trabalho proprio, tratamento, count(*)  
FROM Pessoa_Resposta  
WHERE trabalho proprio <> 'NA'  
GROUP BY trabalho proprio, tratamento  
ORDER BY trabalho proprio
```

index	TRABALHO_PROPRIO	TRATAMENTO	COUNT(*)
0	No	No	545
1	No	Yes	550
2	Yes	Yes	74
3	Yes	No	68



Etapa 4



Fonte de Dados para as análises

- **SPARQL**
 - **MeSH RDF** : <https://id.nlm.nih.gov/mesh/>
- **XQuery**
 - **Mental Health in Tech Survey**:
<https://www.kaggle.com/osmi/mental-health-in-tech-survey>

SPARQL

Tipos de transtornos de ansiedade

```
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/>
PREFIX mesh2018: <http://id.nlm.nih.gov/mesh/2018/>

SELECT DISTINCT ?label
FROM <http://id.nlm.nih.gov/mesh>
WHERE {
    mesh:D001008 meshv:treeNumber ?treeNum .
    ?childTreeNum meshv:parentTreeNumber+ ?treeNum .
    ?descriptor meshv:treeNumber ?childTreeNum .
    ?descriptor rdfs:label ?label .
}

LIMIT 50
```

SPARQL

Todos os
conceitos
relacionados a
uma busca

```
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/>
PREFIX mesh2018: <http://id.nlm.nih.gov/mesh/2018/>
```

```
SELECT DISTINCT ?descriptor ?label
FROM <http://id.nlm.nih.gov/mesh>
WHERE {
    ?concept rdfs:label ?s .
    FILTER regex(str(?s), "Depression")
    ?descriptor meshv:concept ?concept .
    ?descriptor meshv:concept ?new .
    ?descriptor meshv:treeNumber ?t .
    FILTER regex(str(?t), "F03")
    ?new rdfs:label ?label
}
```

```
LIMIT 50
```

SPARQL

Em que
categorias está
a Dispareunia?

```
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/>
PREFIX mesh2018: <http://id.nlm.nih.gov/mesh/2018/>
```

```
SELECT DISTINCT ?pai ?avo
FROM <http://id.nlm.nih.gov/mesh>
WHERE {
    mesh:D004414 meshv:treeNumber ?tree .
    ?tree meshv:parentTreeNumber ?parent .
    ?desc meshv:treeNumber ?parent .
    ?desc rdfs:label ?pai .
    ?desc meshv:broaderDescriptor ?broad .
    ?broad rdfs:label ?avo
}
```

```
LIMIT 50
```

SPARQL

Quais transtornos mentais foram estabelecidos no MeSH após 2001?

```
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 year(?dateEstablished)
FROM <http://id.nlm.nih.gov/mesh>

WHERE {
    mesh:D001523 meshv:treeNumber ?treeNum .
    ?childTreeNum meshv:parentTreeNumber+ ?treeNum .
    ?descriptor meshv:treeNumber ?childTreeNum .
    ?descriptor rdfs:label ?label .
    ?descriptor meshv:dateEstablished ?dateEstablished .
    FILTER (year(?dateEstablished) >= 2001)
}

ORDER BY DESC (?dateEstablished)
```


SPARQL

Contar quantos transtornos estão relacionados às subcategorias de transtornos mentais

```
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/>
PREFIX mesh2018: <http://id.nlm.nih.gov/mesh/2018/>
```

```
SELECT DISTINCT ?labelCat count(?descriptor)
FROM <http://id.nlm.nih.gov/mesh>
WHERE {
    ?subcat meshv:broaderDescriptor mesh:D001523 .
    ?subcat rdfs:label ?labelCat .
    ?subcat meshv:treeNumber ?treeNum .
    ?childTreeNum meshv:parentTreeNumber+ ?treeNum .
    ?descriptor meshv:treeNumber ?childTreeNum .
}
GROUP BY ?labelCat

LIMIT 50
```

XQuery

Porcentagem de pessoas que trabalham em TI as quais acreditam que os seus chefes dão importância para a saúde mental

```
let $pesquisadoc := doc('mydoc.xml')
let $total := count($pesquisadoc//PESQUISA[EMPRESA_TI="Yes"])
let $importancia := count($pesquisadoc//PESQUISA[MENTAL_FISICA="Yes"][EMPRESA_TI="Yes"])

return <result>{round($importancia div $total,2)}</result>
```

XQuery

Porcentagem de pessoas que buscam tratamento por gênero

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

let $numHT := count($pesquisadoc//PESQUISA[GENERO="Male"])
let $numMT := count($pesquisadoc//PESQUISA[GENERO="Female"])
let $numNBT := count($pesquisadoc//PESQUISA[GENERO!="Female"][GENERO!="Male"])

let $numH := count($pesquisadoc//PESQUISA[GENERO="Male"][BUSCA_TRATAMENTO="Yes"])
let $numM := count($pesquisadoc//PESQUISA[GENERO="Female"][BUSCA_TRATAMENTO="Yes"])
let $numNB := count($pesquisadoc//PESQUISA[GENERO!="Female"][GENERO!="Male"][BUSCA_TRATAMENTO="Yes"])

return <numeros>
  <Numero_Homens>{round($numH div $numHT, 2)}</Numero_Homens>
  <Numero_Mulheres>{round($numM div $numMT, 2)}</Numero_Mulheres>
  <Numero_Nao_Binario>{round($numNB div $numNBT, 2)}</Numero_Nao_Binario>
</numeros>
```

XQuery

Número de pessoas que já buscaram tratamento para saúde mental por idade

```
for $p in ($pesquisadoc//PESQUISA)
where $p[TRATAMENTO="Yes"]

let $idade := $p/IDADE

order by $idade
group by $idade

return
    <result>Idade:{$idade} | Qtd: {count($p)}</result>
```

XQuery

Descobrir se pessoas que cuidam do próprio negócio buscam mais tratamento que as outras

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

for $p in ($pesquisadoc//PESQUISA)
where $p[TRABALHO_PROPRIO!="NA"]

let $trabalho := $p/TRABALHO_PROPRIO
let $tratamento := $p/TRATAMENTO

order by $trabalho
group by $trabalho, $tratamento

return
    <result>Trabalho Proprio:{$trabalho} | Tratamento: {$tratamento} | Qtd: {count($p)}</result>
```

XQuery

Selecionar a interferencia do trabalho por genero

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

for $p in ($pesquisadoc//PESQUISA)
where $p[INTERFERENCIA_TRABALHO="Sometimes" or INTERFERENCIA_TRABALHO="Often"]

let $genero := $p/GENERO

group by $genero

return
    <result>Gênero:{$genero} | Qtd: {count($p)}</result>
```