Projeto: TabNetCells

- Recuperar os arquivos CSVs do site www.tabnet.datasus.gov.br/idb
- Para cada elemento de cada tabela, gerar um arquivo com o valor do elemento juntamento com título, notas e links para as correspondentes tabela e Ficha de Qualificação.
- Promover a indexação dos arquivos gerados pelo Google.
- Acompanhar o possível aumento do acesso das tabelas.

Recuperar os arquivos CSVs do site www.tabnet.datasus.gov.br/idb

Endereço redireciona para a versão mais atualizada das tabelas.

Algoritmo:

Se a página for .html e contiver link para arquivo csv - recupera arquivo.

Se a página for .html e não contiver link para arquivo csv - pega todos os links para páginas .html e .def

Se página for .def - monta combinações de opções gerando links para páginas .html

Geração de arquivo para cada elemento da tabela





C.17 Taxa de mortalidade específica por doenças transmissíveis

Óbitos p/infec.respirat.agudas por Região Metropolitana Período: 2010

Óbitos_p/infec.respirat.agudas 590

Nota(s)

Nas tabulações por faixa etária ou sexo, estão suprimidos os casos com idade ou sexo ignorados, respectivamente.

TME - Taxa de mortalidade específica: óbitos por 100.000 habitantes.

Taxas calculadas a partir dos óbitos informados ao SIM. As análises devem considerar as limitações de cobertura e qualidade da informação da causa de óbito.

A taxa de mortalidade específica não padronizada por idade está sujeitas à influência de variações na composição etária da população, o que exige cautela nas comparações entre áreas geográficas e para períodos distintos.

Tabela de dados do TabNet gerada com os seguintes filtros: Capital: Todas as categorias Coluna: Não Ativa Conteúdo: Óbitos p/infec.respirat.agudas Faixa etária: Todas as categorias Linha: Região Metropolitana Períodos Disponíveis: 2010 Região: Todas as categorias Região Metropolitana: Todas as categorias Sexo: Todas as categorias Unidade da Federação: Todas as categorias URL TabNet URL Ficha de Qualificação da RIPSA

DATASUS - Departamento de Informática do SUS / Ministério da Saúde Espl. dos Ministérios Bloco G Anexo A, 1o andar - Sala 107 | 70058-900 | Brasília/DF | Brasíl

Promover a indexação dos arquivos gerados pelo Google

- TabNetCells

index.html

- img
- CSS
- ed6

- a.1

a01 tb1 ce1.html

Tabelas

Grupo A (8 def, 44 html)

Grupo B (17 def, 11 html)

Grupo C (12 def, 15 html)

Grupo D (31 def, 36 html)

Grupo G (8 def, 11 html)

Grupo E (12 def, 21 html)

Grupo F (23 def, 1 html)

Total: 111 def e 139 html

Aproximações

Total – 111 def, 139 html

139 arquivos x 20 linhas = 2780 células

111 arquivos x 12 combinações x 20 linhas = 26400 células

Total aproximado de células = 29420

Saída do programa 1 ano gerado no arquivo .def

```
loading page (GET): [http://tabnet.datasus.gov.br/csv/A090537200_6_42_254.csv] - OK loading page (GET): [http://tabnet.datasus.gov.br/csv/A090537200_6_42_254.csv] - OK loading page (GET): [http://tabnet.datasus.gov.br/csv/A090538200_6_42_254.csv] - OK loading page (GET): [http://tabnet.datasus.gov.br/csv/A090539200_6_42_254.csv] - OK loading page (GET): [http://tabnet.datasus.gov.br/csv/A090539200_6_42_254.csv] - OK loading page (GET): [http://tabnet.datasus.gov.br/csv/A090540200_6_42_254.csv] - OK loading page (GET): [http://tabnet.datasus.gov.br/csv/A090540200_6_42_254.csv] - OK loading page (GET): [http://tabnet.datasus.gov.br/csv/A090541200_6_42_254.csv] - OK loading page (GET): [http://tabnet.datasus.gov.br/csv/A090541200_6_42_254.csv] - OK loading page (GET): [http://tabnet.datasus.gov.br/csv/A090541200_6_42_254.csv] - OK
```

Total cell files created: 33827

Files saved at: /home/heitor/Projetos/TabNetCells/TabNetCells

CONSTRUÍDO COM SUCESSO (tempo total: 39 minutos 2 segundos)

Desafios a serem encarados

Filtragem de determinadas células cujos valores em si não apresentam informação relevante (mecanismo de filtragem).

133 páginas não processadas (requerem mais um campo selecionado (d23) ou "Nenhum registro selecionado" (d1.9).

Escolha de número ótimos de opções nos arquivos .def objetivando-se produzir um número aceitável de arquivos finais a serem indexados.

Fragilidade a ser considerada

O programa depende fortemente da estrutura e de certos rótulos dos arquivos csv atuais (campos não são obrigatórios).