

## Normalização: microdados 2000 e Logradouros MSP 2002

### Caso 1 - Normalização da tabela de micro dados de pessoas do Censo 2000

Está na 1ª FN – os domínios de todas as suas colunas são atômicos.

Chave(s) candidata(s):

- (CONTROLE, FAM, PESSOA): chave primária (PK).

Verificação da 2ª FN - dependências funcionais (não triviais) parciais da chave primária:

- CONTROLE → (UF, MESO, MICRO, MUN, DIST, SUBDIST, RM, AREAP, REG, SITSET, SITDOM, TIPSET) – não está na 2ª FN;
- PK → demais atributos.
- Observação: neste caso particular observa-se que: PK → PESO, (CONTROLE, FAM) → PESO e CONTROLE → PESO

Obtenção das tabelas de domicílios e de famílias a partir da tabela original de microdados de pessoas:

- **Select distinct** CONTROLE, UF, MESO, MICRO, MUN, DIST, SUBDIST, RM, AREAP, REG, SITSET, SITDOM, TIPSET, PESO  
**into** DOMICILIOS  
**from** MICRODADOS\_PESSOAS;
- **Select distinct** CONTROLE, FAM, PESO  
**into** FAMILIAS  
**from** MICRODADOS\_PESSOAS

#### Observações:

1. As duas novas tabelas facilitam as consultas referentes aos domicílios e às famílias. CONTROLE é chave primária da tabela DOMICILIOS. A chave primária de FAMILIAS é a composição de CONTROLE e FAMÍLIA. Na tabela original, elas passam ser chaves estrangeiras.
2. Como o banco de dados é estático (não atualizado no dia-a-dia, após sua construção) não há preocupação com o controle da redundância de dados da tabela original, a qual deve ser mantida. No caso de um banco de dados dinâmico, a redução da tabela original (normalização) deve ser analisada do ponto de vista dos recursos computacionais requeridos para lidar com as redundâncias (das formas não normalizadas) versus os requeridos para a recuperação de dados por junção das tabelas (das formas normalizadas).
3. Há DF transitivas na tabela DOMICILIOS, por exemplo, CONTROLE → MUN e MUN → UF, indicando que ela não está na 3ª FN. Não parece haver vantagem em sua redução. Pode-se verificar que a tabela de micro dados, se reduzida à 2ª FN, pela exclusão dos atributos DF de CONTROLE, exceto PESO, estaria na 3ª FN.

## Caso 2 - Normalização da tabela de logradouros do Município de São Paulo, usada nos exercícios 3 e 4.

Verificação da 1ª FN:

A tabela original e a resultante após o exercício 4, possui redundância óbvia quanto aos atributos: RUA, TIPO, T\_TULO, PREPOSI\_O, LOGRADOURO, TITULO, PREPOSICAO e LOG1.

A composição mais apropriada dependerá da estrutura do geocodificador a ser utilizado. Do ponto de vista do exercício 4, RUA não está na 1ª FN, mas sua remoção não traz perda de informação, portanto será considerada, assim como T\_TULO, PREPOSI\_O, LOGRADOURO, TITULO, PREPOSICAO.

Por não terem utilidade, ID e ID2 também podem ser ignorados (inclusive removidos). FID e Shape são de uso do aplicativo ArcGIS. MUNIC\_PIO, no caso, é sempre São Paulo, e pode ser desconsiderado. Os atributos a considerar na análise são:

X2002 ( X2002\_ID, FROMLEFT, TOLEFT, FROMRIGHT, TORIGHT, LENGTH, DIRECTION, CODLOG, CEP, TIPO, BAIRRO, LOG1 ).

Chave(s) candidata(s):

- (X2002\_ID): chave primária (PK);
- (CODLOG, FROMLEFT): chave candidata;
- (CODLOG, FROMRIGHT): chave candidata;
- (CODLOG, TOLEFT): chave candidata;
- (CODLOG, TORIGHT): chave candidata;
- (TIPO, LOG1, FROMLEFT, CEP) chave candidata;
- (TIPO, LOG1, FROMRIGHT, CEP) chave candidata;
- (TIPO, LOG1, TOLEFT, CEP) chave candidata;
- (TIPO, LOG1, TORIGHT, CEP) chave candidata;

Observações:

1. A rigor, apenas a primeira chave é candidata, pois se observam valores nulos nas colunas FROMLEFT, TOLEFT, FROMRIGHT, TORIGHT e CEP, impossibilitando a criação de chaves únicas (unique index). Vamos considerar o caso teórico em que esses campos não aceitassem valores nulos;
2. Se admitirmos que não haja dois logradouros com mesmo CEP podemos substituir as últimas 4 chaves candidatas (as quais passariam ser super-chaves) pelas correspondentes:
  - (CEP, FROMLEFT): chave candidata;
  - (CEP, FROMRIGHT): chave candidata;
  - (CEP, TOLEFT): chave candidata;
  - (CEP, TORIGHT): chave candidata;

Verificação da FNBC – há determinante que não seja chave candidata:

- CODLOG → (TIPO, LOG1) – não está na FNBC;

Obtenção da tabelas de logradouros a partir da tabela original X2002:

- **Select distinct** CODLOG, TIPO, LOG1  
**into** LOGRADOURO  
**from** X2002;

Ou mesmo:

- **Select** CODLOG, TIPO, LOG1, **sum**(LENGTH) **as** LENGTH  
**into** LOGRADOURO  
**from** X2002  
**group by** CODLOG, TIPO, LOG1;

Observação: se a tabela X2002 fosse, de fato, decomposta, eliminando-se as colunas TIPO e LOG1, a geocodificação não iria funcionar.