

Modelagem e consultas de dados relativos a taxas de rendimento escolar na educação básica brasileira

Estevan Gladstone¹, João Luis Guio¹, Matheus Andrade¹, Tiago Montalvão¹

¹Departamento de Ciência da Computação
Universidade Federal do Rio de Janeiro (UFRJ)

Abstract. *This paper describes the process of modeling a database found in a government website, since the creation of an ER model, going through the logical model up to the creation of tables in a SQL physical system through the MySQL Database Management System. A set of queries is presented together with the Web application developed to access the database.*

Resumo. *Este artigo descreve o processo de modelagem de uma base de dados encontrada em site do governo, desde a criação de um modelo ER, passando pelo modelo lógico e por fim a criação de tabelas em um sistema físico SQL, através do Sistema de Gerenciamento de Banco de Dados MySQL. Um conjunto de consultas é apresentado, juntamente com a descrição da aplicação Web desenvolvida para acessar o banco de dados.*

Introdução

Este artigo começa descrevendo brevemente a base de dados apresentada em <http://dados.gov.br/dataset/taxas-de-rendimento-escolar-na-educacao-basica>. Esta base apresenta dados indicativos do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) sobre diferentes tipos de taxas escolares da educação básica brasileira: aprovação, reprovação e abandono. Estas taxas são separadas por escolas, por ano e por ano escolar na escola.

A fim de deixar a modelagem mais rica em relações, o nosso grupo criou mais uma entidade, representando empresas terceirizadas que prestam serviços às escolas.

As modelagens das entidades no modelo ER e no lógico, e a subsequente tradução para modelo físico, foram realizadas com o uso do software brModelo.

Modelagem ER

A modelagem Entidade-Relacionamento (Fig.1) leva em consideração os objetos mais importantes a serem modelados e os transforma em entidades. Sendo assim, temos 5 entidades:

- Escola
- Municipio
- Estado
- Taxa
- Terceirizada

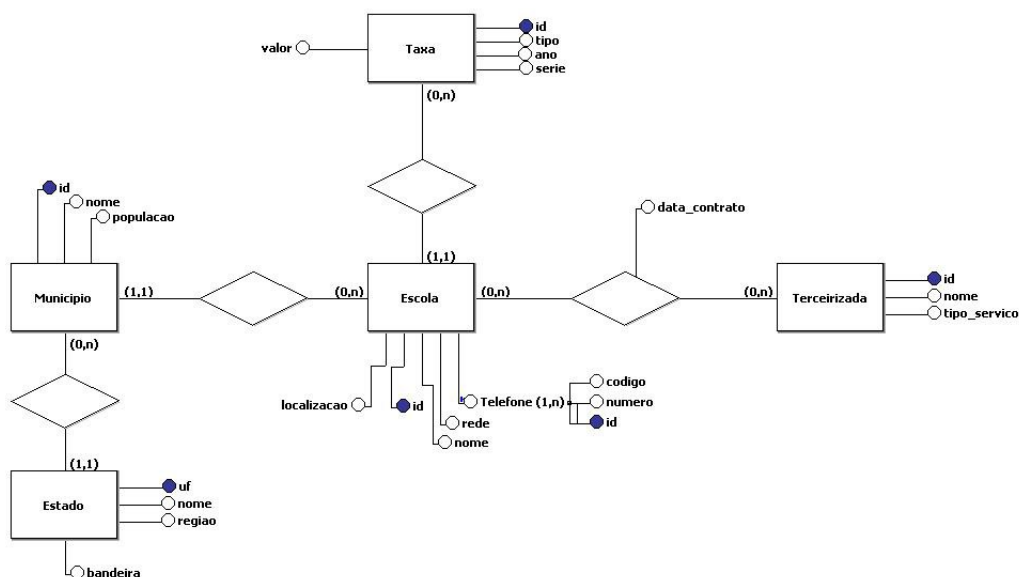


Figura 1. Modelagem ER da base de dados

Todos eles possuem um atributo identificador e pelo menos mais dois outros atributos, além da entidade Escola possuir um atributo (Telefone) multivalorado.

As cardinalidades das relações presentes são 1:N e N:N.

A relação N:N entre Escola e Terceirizada é uma relação que possui um atributo, representando a data do contrato dos serviços prestados pela empresa terceirizada.

Modelo Lógico

A transformação de modelo ER para lógico (Fig.2) dá-se de maneira semi-automática pelo software brModelo. Algumas configurações manuais são necessárias, assim como alguns ajustes no modelo gerado:

- **Atributo multivalorado Telefone:** foi criada uma tabela separada para tal, com chave estrangeira escola_id para a respectiva escola. Havia a possibilidade de incluir os atributos de Telefone na tabela Escola, mas isto causaria grandes redundâncias.
- **Chave primária da tabela TerceirizadaEscola:** esta é uma tabela relacional, que guarda informações de todas as relações entre uma empresa e uma escola. Ela contém chaves estrangeiras referenciando as chaves primárias das respectivas tabelas do relacionamento. Estas chaves devem constituir a chave primária da nova tabela, o que não acontecia inicialmente no modelo gerado.

Podem ser informados os tipos de cada atributo de cada tabela na hora da criação do modelo lógico no brModelo. Isto facilita a futura tradução em modelo físico. É válido mencionar que a tabela Estado possui um campo que será uma imagem, e assim é do tipo BLOB (Binary Large Object), que representa no caso uma imagem, mas poderia representar arquivos de mídia, como áudio ou vídeo, ou até mesmo grandes documentos.

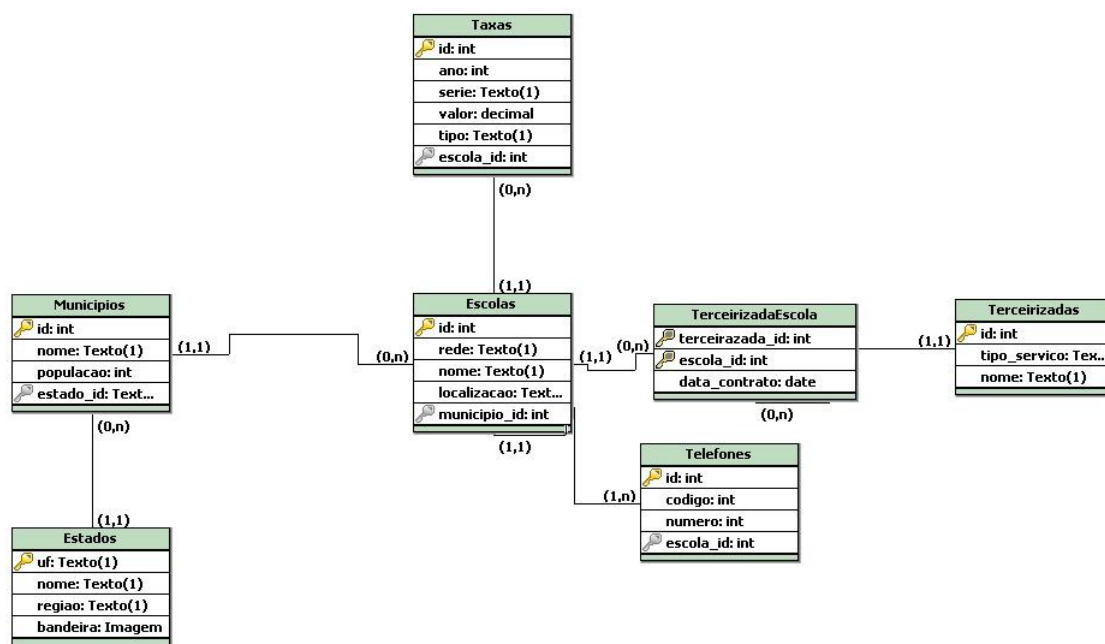


Figura 2. Modelo lógico da base de dados

Análise da forma normal

Os dados nas planilhas da base de dados da qual extraímos os dados encontrava-se em nenhuma forma normal, sendo assim classificada como pertencente à 0FN.

A modelagem apresentada encontra-se pelo menos na 3FN, pois não é encontrada dependência funcional nem parcial nem transitiva de chave. Isto foi garantido na hora de fazer a modelagem ER e a respectiva tradução para o modelo lógico. Um exemplo de situação que não permitiria a modelagem de estar na 2FN seria se as informações de escolas, municípios e estados estivessem todos na mesma tabela. Estado depende da escola, mas também depende de município que, por sua vez, depende do atributo escola.

Modelo físico

A tradução para o modelo físico também foi realizado pelo software brModelo. Isto é feito de maneira automática, mas uma pequena modificação (inserção do caractere ';' após cada comando SQL) também foi necessária.

Após a criação de um arquivo com os comandos SQL para a criação das tabelas, este foi executado em um banco de dados, através do SGBD MySQL, e as tabelas foram criadas, ainda vazias.

Inserção de valores no banco de dados

A extração de dados das planilhas e a respectiva inserção no banco foram feitas em três passos:

1. Tradução do arquivo XLS das planilhas para o formato CSV

2. Criação de scripts (em PHP e em Lua) que manipulam os dados dos arquivos .CSV e geram um outro .CSV, agora pronto para ser inserido no BD
3. Inserção no BD, dos arquivos .CSV já tratados, através do phpMyAdmin, configurado em um servidor local.

Consultas

A seguir, apresentamos uma lista de consultas, com sua descrição (o que envolve e seu enunciado):

– Id, nome, população e id do estado de todas os municípios do estado do Rio de Janeiro (envolve apenas seleção e projeção)

```
SELECT id , nome , populacao , estado_id AS uf FROM
municipios WHERE uf = 'RJ';
```

– Nome, localização, rede e município de todas escolas do estado do RJ ou do RS (envolve junção de apenas duas relações)

```
SELECT escola.nome , localizacao , rede , municipio.nome
AS municipio , municipio.estado_id AS uf
FROM escolas INNER JOIN municipios ON escolas .
municipio_id=municipio.id
WHERE uf = 'RJ' OR uf = 'RS';
```

– Todas as escolas federais que não tem taxas associadas apesar de cadastradas no sistema (envolve junção externa de apenas duas relações)

```
SELECT nome , rede , localizacao
FROM escolas LEFT JOIN taxas ON escolas.id=taxas .
escola_id
WHERE rede = 'Federal' AND taxas.id = null;
```

– Nomes, localização e taxas de todas as escolas da rede Federal e suas respectivas taxas de aprovação do ensino medio caso existam (envolve junção externa de apenas duas relações)

```
SELECT nome , localizacao , (.colocar taxas aqui.)
FROM escolas LEFT JOIN taxas ON escolas.id=taxas .
escola_id
WHERE rede = 'Federal' AND ( (taxas.tipo = null OR
taxas.tipo = 'Aprovação') AND (taxas.serie = null
or taxas.serie LIKE '%Médio'));
```

– Obtém o nome, rede, localização e uf de todas as escolas da região Sudeste (envolve junção de três relações)

```
SELECT escolas.nome , rede , localizacao , uf
FROM escolas INNER JOIN municipios ON escolas .
municipio_id=municipio.id INNER JOIN estados ON
estado.id=municipio.estado_id
```

```
WHERE estados.regiao = 'Sudeste';
```

– Obtém todos os telefones da escola de nome CEFET CELSO SUCKOW DA FONSECA (envolve junção de duas relações)

```
SELECT codigo , numero  
FROM escolas INNER JOIN telefones ON telefones.  
    escola_id=escolas.id  
WHERE nome = 'CEFET_CELSO_SUCKOW_DA_FONSECA';
```

– Obtém o número de escolas cadastradas no sistema por estado

```
SELECT count(distinct escolas.id), estados.nome, uf,  
    regiao, (bandeira?)  
FROM escolas INNER JOIN municipios ON escolas.  
    municipio_id=municipio.id INNER JOIN estados ON  
    estado.id=municipio.estado_id
```

References

Bibliographic references must be unambiguous and uniform. We recommend giving the author names references in brackets, e.g. [Knuth 1984], [Boulic and Renault 1991], and [Smith and Jones 1999].

The references must be listed using 12 point font size, with 6 points of space before each reference. The first line of each reference should not be indented, while the subsequent should be indented by 0.5 cm.

Referências

Boulic, R. and Renault, O. (1991). 3d hierarchies for animation. In Magnenat-Thalmann, N. and Thalmann, D., editors, *New Trends in Animation and Visualization*. John Wiley & Sons Ltd.

Knuth, D. E. (1984). *The T_EX Book*. Addison-Wesley, 15th edition.

Smith, A. and Jones, B. (1999). On the complexity of computing. In Smith-Jones, A. B., editor, *Advances in Computer Science*, pages 555–566. Publishing Press.