

Laboratório Nacional de Computação Científica
Programa de Pós Graduação em Modelagem Computacional

**Biota-Rio: um banco de dados para a biodiversidade do
Estado do Rio de Janeiro**

Por

Vinicius Schmitz Pereira Nunes

PETRÓPOLIS, RJ - BRASIL

JULHO DE 2008

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

**BIOTA-RIO: UM BANCO DE DADOS PARA A BIODIVERSIDADE
DO ESTADO DO RIO DE JANEIRO**

Vinicius Schmitz Pereira Nunes

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO LABORATÓRIO
NACIONAL DE COMPUTAÇÃO CIENTÍFICA COMO PARTE DOS REQUI-
SITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM
MODELAGEM COMPUTACIONAL

Aprovada por:

Prof. Claudia Augusta de Moraes Russo, D.Sc.

(Presidente)

Prof. Jauvane Cavalcante de Oliveira, Ph.D.

Prof. Fabiano Lopes Thompson, Ph.D.

PETRÓPOLIS, RJ - BRASIL
JULHO DE 2008

Nunes, Vinicius Schmitz Pereira

N972b biota-rio: um banco de dados para a biodiversidade do estado do rio de janeiro / Vinicius Schmitz Pereira Nunes. Petrópolis, RJ. : Laboratório Nacional de Computação Científica, 2008.

xx, yy p. : il.; 29 cm

Orientador: Claudia Augusta de Moraes Russo

Dissertação (M.Sc.) – Laboratório Nacional de Computação Científica, 2008.

1. Taxonomia. 2. Banco de dados. I. Russo, Claudia Augusta de Moraes. II. LNCC/MCT. III. Título.

CDD 578.028 5

Agradecimentos

- À minha orientadora Prof^o Claudia Russo, por ter me dado a oportunidade de trabalhar neste bonito projeto que é o Biota-Rio.
- À Dra Ana Tereza Ribeiro de Vasconcelos, por ter me acolhido em seu laboratório, LABINFO, e por todo o apoio na realização do meu trabalho.
- À minha família: pais e irmãos, pessoas que sempre estiveram comigo me apoiando nos meus trabalhos.
- Aos meus amigos: Ale, Anselmo, Mônica, Oberdan, Reinaldo, Tarcísio.
- À Ana Neri e Ana Paula, funcionárias da pós que estão sempre dispostas a nos ajudar.
- Às instituições que apoiaram esse trabalho: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, e Laboratório Nacional de Computação Científica - LNCC/MCT.

Resumo da Dissertação apresentada ao LNCC/MCT como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

BIOTA-RIO: UM BANCO DE DADOS PARA A BIODIVERSIDADE DO ESTADO DO RIO DE JANEIRO

Vinicius Schmitz Pereira Nunes

Julho , 2008

Orientador: Claudia Augusta de Moraes Russo, D.Sc.

O uso da tecnologia da informação na implementação de bases de dados com informações sobre a biodiversidade, vem ganhando uma grande importância. Isso se deve a grande necessidade de se promover um acesso livre a dados e informações sobre as espécies. O acesso a essas informações é considerado como fundamental para o desenvolvimento de diferentes áreas do conhecimento. A presente dissertação teve como objetivo o desenvolvimento de um sistema computacional para acesso a informações sobre a biodiversidade do Estado do Rio de Janeiro. O sistema Biota-Rio deverá conter informações sobre as espécies de animais, vertebrados e invertebrados, que foram descritas para o Estado do Rio de Janeiro. Informações como descrições originais, localidades-tipo, ambiente e outras informações estarão disponíveis no sistema. Biota-Rio é um sistema de acesso livre via web, que foi implementado utilizando a linguagem de programação PERL. O MySQL foi o SGBD utilizado na implementação do banco de dados onde estarão depositadas as informações sobre as espécies. Até o momento, foram feitos levantamentos das referências para trinta e seis espécies de mamíferos utilizando a base de dados Zoological Record.

Abstract of Dissertation presented to LNCC/MCT as a partial fulfillment of the requirements for the degree of Master of Sciences (M.Sc.)

**BIOTA-RIO: A DATABASE OF BIODIVERSITY IN THE STATE
OF RIO DE JANEIRO**

Vinicius Schmitz Pereira Nunes

July, 2008

Advisor: Claudia Augusta de Moraes Russo, D.Sc.

The use of information technology in biodiversity related databases is increasing due to the open access to a large sum of data. Access to this information may be regarded as crucial to different areas of knowledge. The present work aims at the development of a open access computational system on biodiversity of the State of Rio de Janeiro. The Biota-Rio system focuses on the animal species, vertebrates and invertebrates, which were originally described for the State of Rio de Janeiro. In this sense, key information such as original descriptions, type localities, habitats and other are directly available on the system. Biota-Rio is a system in which the user may freely access the information via web that was implemented using the PERL language. The MySQL was the selected DBMS in the implementation of the database in which the informations are deposited. To date, surveys have been made of references to thirty-six species of mammals using the Zoological Record database.

Sumário

1	Introdução	1
1.1	Bancos de dados	1
1.2	Projeto de um banco de dados	3
1.3	Os modelos de dados	5
1.3.1	Modelo Entidade-Relacionamento	5
1.3.2	Modelo Relacional	10
1.4	Normalização de tabelas	13
1.5	SQL	20
1.5.1	Linguagem de Definição dos Dados	21
1.5.2	Linguagem de Manipulação dos Dados	23
1.6	Diversidade Biológica	24
1.6.1	Classificação Biológica	24
1.6.2	A Teoria da Evolução	26
1.6.3	Sistemática Filogenética	28
1.7	Sistemas de informação e diversidade biológica	29
1.8	Objetivos	32
2	Metodologia	33
2.1	Modelo ER do banco de dados Biota-Rio	33
2.2	O banco de dados Biota-Rio	36
2.3	Ferramentas utilizadas na implementação do sistema Biota-Rio	45
2.4	Levantamento bibliográfico	47

3	Resultados e Discussão	49
3.1	Consultando a partir do nome de uma espécie ou subespécie	50
3.2	Consultando a partir do nome de um táxon supra-específico	54
3.3	Levantamento das referências	59
3.4	Discussão	59
4	Conclusão e Perspectivas	61
	Referências Bibliográficas	63

Lista de Figuras

Figura

1.1	Uma visão simplificada do ambiente de um sistema de banco de dados. (Adaptado de (Elmasri e Navathe, 2002)).	2
1.2	Etapa 1 - Análise dos requisitos.	3
1.3	Etapa 2 - Esquema do banco de dados Universidade.	4
1.4	Etapa 3 - Implementação das tabelas a partir do esquema conceitual. . .	4
1.5	Exemplo de atributos.	6
1.6	Exemplo de um relacionamento do banco de dados Universidade.	6
1.7	A tabela acima mostra, de forma sucinta, os símbolos e seus respectivos significados usados em um modelo ER.	7
1.8	Tabela com os tipos de entidade e os respectivos atributos, do banco de dados Universidade.	7
1.9	Tabela com os relacionamentos e os respectivos tipos de entidades par- ticipantes do banco de dados Universidade.	8
1.10	Razões de cardinalidade.	9
1.11	Exemplo de cardinalidade 1:1.	9
1.12	Exemplo de cardinalidade 1:N.	9
1.13	Exemplo de relacionamento com razão de cardinalidade N:N.	10
1.14	Os atributos e tuplas (linhas) da relação <i>Departamento</i>	10
1.15	Tabelas do banco de dados Universidade. O nome da chave primária encontra-se sublinhado.	13
1.16	Exemplo de dependência funcional.	14

1.17	Exemplo de dependência funcional parcial.	14
1.18	Exemplo de dependência funcional transitiva.	15
1.19	Exemplo de dependência funcional multivalorada.	15
1.20	Exemplo de dependência funcional cíclica.	16
1.21	O atributo <i>Endereço</i> é multivalorado.	16
1.22	Exemplo de tabela normalizada pela 1FN.	17
1.23	Exemplo de aplicação da 2FN.	17
1.24	Exemplo de aplicação da 3FN.	18
1.25	Exemplo de aplicação da 4FN.	19
1.26	Exemplo de aplicação da 5FN.	20
1.27	Exemplo de esquema de um banco de dados relacional.	21
1.28	Alguns dos tipos de dados.	22
1.29	Exemplo de uma consulta.	23
1.30	Exemplo de inserção.	24
1.31	Exemplo de deleção.	24
1.32	Exemplo de atualização.	24
2.1	Esquema do banco de dados Biota-Rio na forma de um diagrama ER.	33
2.2	Tabela com a relação dos atributos para cada tipo de entidade do modelo ER do banco de dados Biota-Rio.	35
2.3	Tabela com as treze tabelas que compõe o banco de dados Biota-Rio e seus respectivos campos.	37
2.4	Continuação da tabela da Figura 35. Os tópicos que se encontram mar- cados com um asterisco (*), serão abordados mais adiante.	38
2.5	Modelo de dados do banco Biota-Rio.	42
2.6	Representação da tabela <i>Taxon</i> e três de seus campos, <i>Id_Taxon</i> , <i>NoPai</i> e <i>NomeTaxon</i>	43
2.7	Representação da tabela <i>Organismo</i> e três de seus campos, <i>Id_Organismo</i> , <i>Id_Taxon</i> , <i>OrgPai</i> e <i>NomeOrganismo</i>	44

2.8	Diagrama representando o processo de hierarquização nas tabelas <i>Taxon</i> e <i>Organismo</i>	45
2.9	Arquitetura do sistema Biota-Rio.	46
3.1	Página principal do sistema Biota-Rio e em destaque o campo para consulta.	50
3.2	Tabela com as informações do organismo.	52
3.3	Caixa de texto para uma nova consulta e o link para a página com a relação dos genes mitocondriais do organismo no GenBank.	53
3.4	Tabela com a relação dos genes mitocondriais de um organismo. (A tabela acima é apenas um exemplo fictício, pois as informações mostradas ainda não estão cadastradas na base).	53
3.5	Árvore de táxons.	54
3.6	Continuidade da árvore taxonômica.	55
3.7	Árvore taxonômica dentro de Mammalia.	56
3.8	Árvore taxonômica dentro de Didelphimorphia.	57
3.9	Árvore taxonômica para o gênero <i>Monodelphis</i>	57
3.10	Árvore taxonômica para o táxon Craniata.	58
3.11	Exemplo de táxons diferentes, mas que possuem o mesmo nome.	58

Lista de Tabelas

Tabela

Capítulo 1

Introdução

1.1 Bancos de dados

Os bancos de dados tornaram-se uma das principais aplicações da tecnologia de informação no cotidiano do mundo moderno. Durante um dia de nosso cotidiano, deparamos com muitas atividades que de alguma forma envolvem alguma interação com banco de dados. Exemplos disso não faltam, quando vamos ao banco fazer algum tipo de movimentação bancária, quando compramos algum produto pela internet, quando fazemos uma consulta à base de dados de uma biblioteca, ou, então, quando solicitamos o CPF à Receita Federal. Em tudo isso, e em muitos outros exemplos, existe um banco de dados envolvido no processo.

Um banco de dados pode ser definido como sendo uma coleção de dados relacionados e um conjunto de programas que permitem ao usuário criar e manter uma coleção de dados (Silberschatz et al., 2006). Esse conjunto de programas é também chamado de Sistema de Gerenciamento de Banco de Dados ou simplesmente SGBD. Na definição de banco de dados dois conceitos se destacam, os conceitos de dados e SGBD.

Segundo Elmasri (Elmasri e Navathe, 2002), dados podem ser definidos como sendo fatos conhecidos que podem ser registrados e que apresentam algum significado. O nome de um aluno de uma universidade, o seu número de matrícula, as disciplinas que cursa, são exemplos de dados que se encontram relacionados. Isso quer dizer que esses dados possuem um significado, ou seja, são informações que

se referem a uma pessoa que é aluna de uma universidade.

Já o SGBD, por outro lado, permite ao usuário definir, construir e manipular uma base de dados. Definir um banco de dados envolve a especificação dos tipos de dados a serem armazenados na base. Construir é o processo de armazenar os dados em algum meio de armazenamento controlado pelo SGBD. E manipular um banco de dados envolve a utilização de funções para consulta, recuperação e atualização dos dados e, geração de relatórios (Elmasri e Navathe, 2002).

O SGBD e a base de dados são, conjuntamente, denominados de Sistema de Banco de Dados. Pela Figura 1.1 , o usuário constrói uma consulta, a consulta é então processada por um conjunto de softwares do SGBD que irão acessar a base de dados e retornar, para o usuário, as informações requisitadas na consulta.

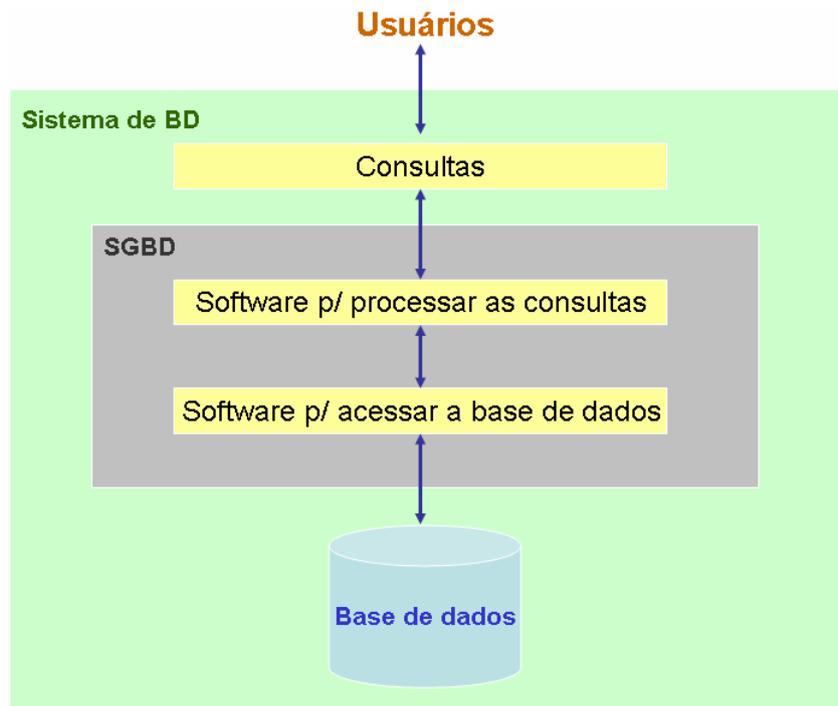


Figura 1.1: Uma visão simplificada do ambiente de um sistema de banco de dados. (Adaptado de (Elmasri e Navathe, 2002)).

Na seção seguinte será apresentada as etapas que constituem o projeto de um banco de dados, ou seja, desde a análise dos requisitos até a implementação das tabelas.

1.2 Projeto de um banco de dados

O projeto de um banco de dados envolve, basicamente, três etapas: a análise de requisitos, o projeto conceitual e a implementação do projeto (Teorey et al., 2006). Antes de abordarmos cada uma das etapas na elaboração do projeto de um banco de dados é necessário definir que parcela do mundo real será representado pelo banco de dados. Essa parcela da realidade é chamada de **minimundo** ou **Universo de Discurso** (Elmasri e Navathe, 2002). Para ilustrar o processo de desenvolvimento do projeto de um banco de dados, será usado como exemplo a implementação de um banco de dados *Universidade*.

Durante a análise dos requisitos são definidas as características que o sistema deve ter para que ele possa melhor atender as necessidades do usuário. Por exemplo, quais informações o usuário necessita acessar no sistema e quais recursos o sistema deverá ter para que o usuário possa acessar essas informações (Figura 1.2).



Figura 1.2: Etapa 1 - Análise dos requisitos.

A fase seguinte é o desenvolvimento do esquema conceitual do banco de dados a partir dos requisitos que foram definidos na etapa anterior. O objetivo do projeto do esquema conceitual é representar, de forma diagramática, os requisitos de dados do mundo real de tal maneira que seja inteligível ao projetista do banco de dados e ao usuário final (Teorey et al., 2006). A Figura 1.3 mostra o esquema conceitual de um banco de dados.

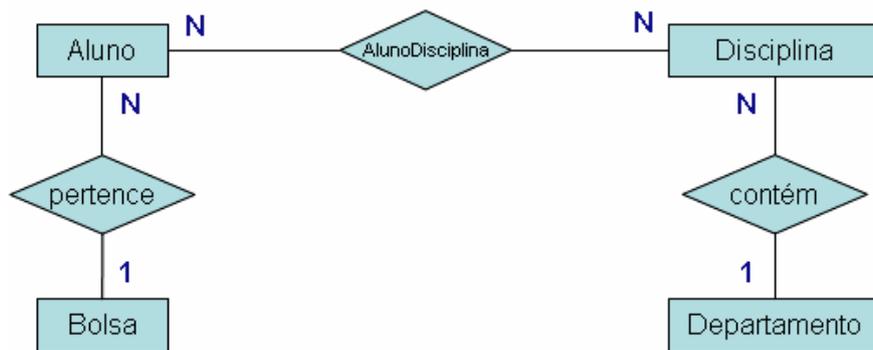


Figura 1.3: Etapa 2 - Esquema do banco de dados Universidade.

A última etapa do projeto de um banco de dados é a implementação das tabelas que irão compor o banco. Nessa última etapa, é feita a transformação do modelo conceitual em tabelas. Assim, o projetista deverá definir, a partir das características do banco, qual o melhor modelo de implementação a ser utilizado (Figura 1.4).

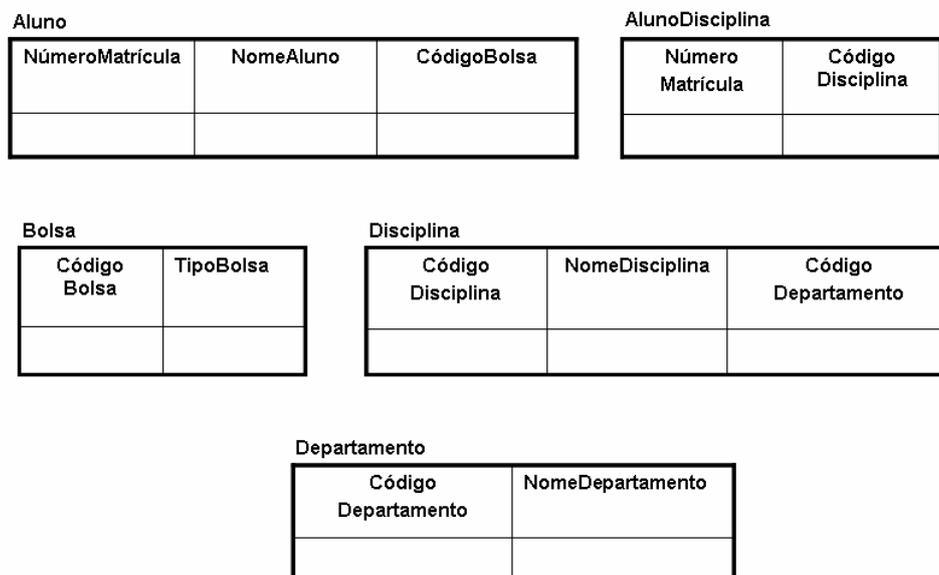


Figura 1.4: Etapa 3 - Implementação das tabelas a partir do esquema conceitual.

Na figura acima estão representadas as tabelas que compõem o banco de dados Universidade e os seus respectivos campos. Por exemplo, a tabela Aluno é formada pelos campos *NúmeroMatrícula*, *NomeAluno* e *CódigoBolsa*.

Na próxima seção, será apresentada uma breve introdução sobre modelos de

dados. Posteriormente será discutido o modelo Entidade-Relacionamento, utilizado na segunda etapa do projeto de um banco de dados. E, para finalizar a seção 1.3, será apresentado o modelo Relacional utilizado na última etapa do projeto de um banco de dados.

1.3 Os modelos de dados

Um modelo de dados pode ser entendido como sendo um conjunto de conceitos usados para descrever o projeto de um banco de dados. A descrição de um banco de dados, a partir de um determinado modelo, é denominada de ESQUEMA. A maioria dos modelos de dados possui certas regras para exibir os esquemas como diagramas. Esses diagramas são denominados de DIAGRAMA DO ESQUEMA (Elmasri e Navathe, 2002). A Figura 1.3 é um exemplo de diagrama do esquema do modelo ER.

Segundo Elmasri (Elmasri e Navathe, 2002) existem dois grupos de modelos. O modelo conceitual e os modelos de implementação. O modelo conceitual mais usado no projeto de um banco de dados é o modelo Entidade-Relacionamento. Já o modelo de implementação mais usado é o modelo Relacional.

Na próxima subseção será abordado o modelo ER e, em seguida, o modelo relacional, usado no banco de dados desta dissertação.

1.3.1 Modelo Entidade-Relacionamento

Em 1976, Peter Chen publicou um trabalho intitulado *The Entity-Relationship Model: Toward the unified view of data*. (Chen, 1976), no qual definia uma abordagem para o processo de modelagem dos dados. Após a sua publicação, essa metodologia teve ampla aceitação e passou a ser um referencial definitivo no processo de modelagem conceitual (Cougo, 1997). Essa abordagem se baseia em uma técnica de diagramação e de conceitos que são importantes no processo de modelagem dos dados. Os principais conceitos no modelo ER são: entidade, relacionamento e atributos.

Uma *entidade* pode ser entendida como sendo um objeto do mundo real que é distinguível de todos os outros objetos (Silberschatz et al., 2006). Um aluno de uma universidade, por exemplo, é uma entidade.

Um conjunto de entidades é formado por todas as entidades que compartilham dos mesmos atributos. Esse conjunto é chamado de TIPO DE ENTIDADE (Elmasri e Navathe, 2002); (Silberschatz et al., 2006). O conjunto de todos os alunos de uma universidade pode ser definido como um conjunto de entidades do tipo *Aluno*. Um conjunto de entidades é representado, em um diagrama ER, por um retângulo.

Os *atributos* podem ser entendidos como sendo as propriedades que descrevem uma entidade. (Elmasri e Navathe, 2002). Os atributos de uma entidade são representados, no diagrama ER, por elipses. O *Nome do aluno* e *Número de matrícula* são exemplos de atributos de cada entidade *aluno* matriculado em uma universidade (ver Figura 1.5).

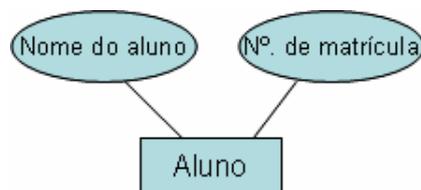


Figura 1.5: Exemplo de atributos.

Relacionamentos representam associações no mundo real entre entidades ou de uma entidade com ela mesma (Teorey et al., 2006). No modelo ER os relacionamentos são representados por losangos. Na Figura 1.6, *AlunoDisciplina* é, por exemplo, o relacionamento entre os tipos de entidades *Aluno* e *Disciplina*. Isso significa que as entidades *aluno* relacionam-se com as entidades *disciplina* por meio da matrícula.



Figura 1.6: Exemplo de um relacionamento do banco de dados Universidade.

A Figura 1.7 apresenta, de forma resumida, os principais símbolos e seus respectivos significados.

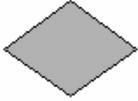
Símbolos	Significado
	Conjunto de entidades
	Relacionamento
	Atributo

Figura 1.7: A tabela acima mostra, de forma sucinta, os símbolos e seus respectivos significados usados em um modelo ER.

Com base no diagrama do esquema ER (veja Figura 1.3), o banco de dados é formado por quatro tipos de entidades: *Departamento*, *Aluno*, *Bolsa* e *Disciplina*. Cada tipo de entidade possui os seus próprios atributos. A Figura 1.8 mostra uma tabela com os atributos de cada um dos tipos de entidades.

Tipo de Entidade	Atributos
Aluno	- Número de matrícula - Nome do aluno
Disciplina	- Código da disciplina - Nome da disciplina
Departamento	- Código do departamento - Nome do departamento
Bolsa	- Código da bolsa - Tipo de bolsa

Figura 1.8: Tabela com os tipos de entidade e os respectivos atributos, do banco de dados Universidade.

Os relacionamentos entre as entidades também podem ser vistos no esquema. São três os relacionamentos. A Figura 1.9 apresenta, de forma sucinta, os relacionamentos e os tipos de entidades participantes.

Relacionamento	Tipos de Entidades participantes
AlunoDisciplina	Aluno e Disciplina
pertence	Bolsa e Aluno
contém	Departamento e Disciplina

Figura 1.9: Tabela com os relacionamentos e os respectivos tipos de entidades participantes do banco de dados Universidade.

Um outro importante conceito no modelo ER é o conceito de chave. Uma chave é um atributo de um tipo de entidade cujo conteúdo é único para cada entidade, ou seja, cada entidade é identificada univocamente pelo conteúdo deste atributo. Por exemplo, no tipo de entidade Aluno, cada entidade *aluno* possui o seu próprio número de matrícula. O número de matrícula é único e exclusivo do aluno. Portanto, o atributo *Número de matrícula* é uma chave ou atributo chave. Quando uma entidade não possui uma chave ela é chamada de ENTIDADE FRACA (Elmasri e Navathe, 2002); (Silberschatz et al., 2006).

Pode ocorrer que um tipo de entidade tenha mais de um atributo chave. Quando isso ocorre tem-se uma SUPERCHAVE. Cabe ao projetista do banco definir qual dos atributos da superchave será a chave. O atributo que for escolhido como sendo a chave, ou seja, o atributo que irá identificar cada entidade dentro do conjunto de entidades, é denominado de chave primária. Suponha que o tipo de entidade *Aluno* tenha um terceiro atributo, *Número do CPF*, e esse atributo fosse o escolhido pelo projetista para identificar cada entidade no conjunto das entidades, então, o atributo *Número do CPF* seria a chave primária.

Uma outra característica do modelo ER é a razão de cardinalidade, ou simplesmente cardinalidade. Essa razão expressa o número de entidades ao qual uma outra entidade pode estar associada (Silberschatz et al., 2006). A Figura 1.10

mostra uma tabela com as razões de cardinalidade.

1:1	Um para um
1:N	Um para muitos
N:N	Muitos para muitos

Figura 1.10: Razões de cardinalidade.

Para exemplificar o relacionamento 1:1, suponha o tipo de entidade *Professor* no modelo da Figura 1.3. Na Figura 1.11, observa-se o relacionamento *alocado* entre os tipos de entidade *Professor* e *Departamento*. Nesse exemplo, cada entidade *professor* se relaciona com apenas uma entidade *departamento*. Isso significa que cada professor está alocado em um único departamento.



Figura 1.11: Exemplo de cardinalidade 1:1.

A Figura 1.12 ilustra um exemplo de relacionamento com cardinalidade 1:N, encontrado no esquema da Figura 1.3. O exemplo mostra o relacionamento *contém* entre os tipos de entidade *Departamento* e *Disciplina*. A cardinalidade indica que cada entidade *departamento* se relaciona com várias entidades *disciplina*, ou seja, um único departamento pode ser responsável por várias disciplinas (ver Figura 1.12).



Figura 1.12: Exemplo de cardinalidade 1:N.

Na Figura 1.13 há um exemplo de relacionamento com razão de cardinalidade N:N entre os tipos de entidade *Aluno* e *Disciplina*. Isso significa que cada aluno

pode estar matriculado em muitas disciplinas, e também, que em cada disciplina podem estar matriculados muitos alunos.



Figura 1.13: Exemplo de relacionamento com razão de cardinalidade N:N.

1.3.2 Modelo Relacional

O modelo relacional foi introduzido pela primeira vez por Tedd Codd, da IBM Research, em 1970. Na época ele atraiu a atenção devido a sua simplicidade. Segundo Elmasri (Elmasri e Navathe, 2002), a base teórica do modelo relacional está na teoria dos conjuntos. Atualmente é o modelo de implementação mais usado em banco de dados. Para Teorey (Teorey et al., 2006), isso se deve ao grau de independência dos dados. Isso significa que em um mesmo modelo são permitidos relacionamentos com grau de cardinalidade 1:1, 1:N e N:N.

Um banco de dados relacional pode ser definido como sendo um conjunto de tabelas e cada tabela é formada por um conjunto de linhas. Cada tabela no modelo é também chamada de *relação* e cada campo, ou coluna, de uma tabela é também chamado de *atributo*. As linhas de uma tabela são chamadas de *tuplas* (Elmasri e Navathe, 2002). A Figura 1.14 exemplifica o que foi dito anteriormente.

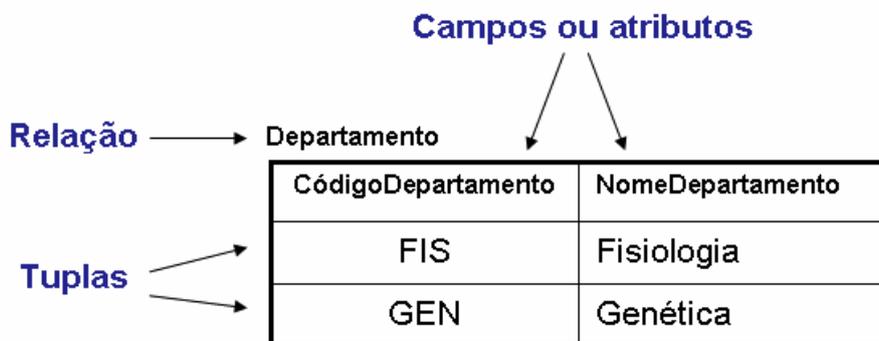


Figura 1.14: Os atributos e tuplas (linhas) da relação *Departamento*.

A criação do modelo relacional a partir do modelo Entidade-Relacionamento

se faz a partir de um conjunto de etapas. Essas etapas compõem aquilo que é chamado de MAPEAMENTO ER-para-RELACIONAL (Elmasri e Navathe, 2002).

Na primeira etapa, cada tipo de entidade E , no esquema ER, dá origem a uma tabela que inclui todos os atributos do tipo de entidade E . É necessário que um dos atributos chaves seja escolhido como chave primária (Elmasri e Navathe, 2002). Por exemplo, a tabela *Departamento* foi criada com os atributos, *CódigoDepartamento* e *NomeDepartamento*, que foram definidos no modelo ER (ver figuras 1.3 e 1.4). A chave primária é o atributo *CódigoDepartamento*.

Na segunda etapa do processo, para cada relacionamento binário R com cardinalidade 1:N, identifique a relação S que representa o tipo de entidade participante no lado N do relacionamento. Inclua como chave estrangeira em S a chave primária da relação T que representa o outro tipo de entidade participante em R . Isso porque cada entidade do tipo de entidade que possui cardinalidade 1, se relaciona com várias entidades do tipo de entidade com cardinalidade N (Elmasri e Navathe, 2002). Um exemplo disso é o relacionamento entre as tabelas *Departamento* e *Disciplina*, onde o lado N do relacionamento está na tabela *Disciplina*. Portanto, é ela que recebe, como chave estrangeira, o atributo *CódigoDepartamento*, que é a chave primária da tabela *Departamento* (ver figuras 1.3 e 1.4).

Na terceira etapa, para cada relacionamento R com cardinalidade N:N é necessário criar uma nova relação S , para representar R , e incluir em S , como chaves estrangeiras, as chaves primárias das relações que participam do relacionamento (Elmasri e Navathe, 2002). Por exemplo, o relacionamento entre as tabelas *Aluno* e *Disciplina* apresenta cardinalidade N:N. Esse relacionamento deu origem a uma tabela, *AlunoDisciplina*, formada pelas chaves primárias de ambas as tabelas.

Assim como no modelo Entidade-Relacionamento, as chaves estão presentes no modelo relacional. Uma chave é um campo que possui valores únicos e exclusivos em cada linha. Na tabela *Departamento*, cada departamento terá o seu próprio código de identificação. Logo, o atributo *CódigoDepartamento* será uma chave na tabela *Departamento*.

Um aspecto importante na utilização de chaves primárias, é que elas são usadas na indexação de tabelas. Ou seja, cada chave primária funciona como um índice que referencia uma única linha da tabela. A indexação é considerada como a principal técnica de otimização de consultas em banco de dados relacionais (Dubois, 2003).

Pode ocorrer que em uma relação exista mais de um atributo chave. Nesse caso cada uma das chaves é chamada de chave candidata. O projetista terá que escolher, entre as chaves candidatas, aquela que irá identificar cada uma das tuplas em uma relação. A chave escolhida é denominada de chave primária (Primary Key - PK), como mencionado anteriormente.

Uma outra importante chave presente no modelo relacional é a chave estrangeira. Uma chave estrangeira pode ser definida como sendo o atributo de uma tabela que guarda valores de uma chave primária que pertence a uma outra tabela (Milani, 2006). É por meio das chaves estrangeiras que se estabelecem os relacionamentos entre as tabelas.

Na Figura 1.15, com exceção da tabela *AlunoDisciplina*, as demais tabelas apresentam chave primária. Os campos *NúmeroMatrícula*, *CódigoBolsa*, *CódigoDisciplina* e *CódigoDepartamento* são, respectivamente, chaves primárias das tabelas *Aluno*, *Bolsa*, *Disciplina* e *Departamento*. Isso porque em cada uma dessas colunas os valores guardados são exclusivos em cada linha. Por exemplo, na tabela *Departamento* o departamento de Genética possui o seu próprio código de identificação e esse é exclusivo dele, ou seja, o departamento de Genética só pode ter um único código de identificação e nenhum outro departamento poderá ter esse mesmo código.

Já os campos *CódigoBolsa*, da tabela *Aluno*, e *CódigoDepartamento*, da tabela *Disciplina*, são chaves estrangeiras. Eles guardam os valores referentes aos campos *CódigoBolsa*, da tabela *Bolsa*, e *CódigoDepartamento*, da tabela *Departamento*, que são chaves primárias. O campo *CódigoDepartamento* da tabela *Disciplina* possui, na primeira linha, o registro “GEN”, que é o código do departa-

mento “Genética” que se encontra no campo *CódigoDepartamento*, chave primária da tabela *Departamento* (ver Figura 1.15).

Aluno			AlunoDisciplina	
<u>NúmeroMatrícula</u>	NomeAluno	CódigoBolsa	Número Matrícula	Código Disciplina
2007230	Marcos Oliveira	MEST	2007230	GNB

Bolsa		Disciplina		
<u>Código Bolsa</u>	TipoBolsa	<u>Código Disciplina</u>	NomeDisciplina	Código Departamento
MEST	Mestrado	GNB	Genética Básica	GEN

Departamento	
<u>Código Departamento</u>	NomeDepartamento
GEN	Genética

Figura 1.15: Tabelas do banco de dados Universidade. O nome da chave primária encontra-se sublinhado.

O uso de chaves também contribui para que se evite a redundância de dados em um banco relacional. A redundância ocorre quando um mesmo conjunto de dados é armazenado em tabelas diferentes (Silberschatz et al., 2006). Isso leva a um desperdício de espaço de armazenamento. Por exemplo, suponha que as tabelas *Aluno* e *Bolsa*, do banco de dados Universidade (ver Figura 1.4), possuíssem o campo *NomeAluno*, isso seria redundância, pois a mesma informação estaria sendo armazenada em duas tabelas diferentes.

No próximo item será abordado um conjunto de regras que, quando aplicados em um projeto, permitem que os dados sejam armazenados sem redundância. Estas regras recebem o nome de FORMAS NORMAIS.

1.4 Normalização de tabelas

Para se compreender o que são as Formas Normais, é antes necessário abordar alguns conceitos que são necessários para melhor compreensão do tema.

O primeiro destes conceitos é a Dependência Funcional. Uma dependência funcional acontece quando cada valor de um ou mais atributos são identificados pelo

valor de um outro atributo, ou seja, um ou mais atributos estão relacionados a um outro atributo. Se $A \rightarrow B$, diz-se que o atributo A é dependente funcionalmente de B (Rocha-Júnior, 2007); (Silberschatz et al., 2006).

Um exemplo disso é o campo *NumeroMatricula* da tabela *Aluno*. Os valores “Marcos Oliveira”, no campo *NomeAluno*, e “MEST”, no campo *CódigoBolsa*, são referenciados pelo valor “2007230”, do campo *NúmeroMatricula*. Portanto, os campos *NomeAluno* e *CódigoBolsa* dependem funcionalmente do campo *NúmeroMatricula* (ver Figura 1.16).

Aluno

<u>NúmeroMatricula</u>	NomeAluno	CódigoBolsa
2007230	Marcos Oliveira	MEST

Figura 1.16: Exemplo de dependência funcional.

Um outro importante conceito é a Dependência Funcional Parcial. Uma dependência funcional parcial acontece quando a chave primária de uma tabela é composta por mais de um atributo e, pelo menos um dos atributos não chave da tabela não depende funcionalmente de toda a chave primária (Rocha-Júnior, 2007); (Elmasri e Navathe, 2002). Por exemplo, suponha que no banco de dados haja uma tabela *Nota* e essa tenha como chave primária os campos *NúmeroMatricula* e *CódigoDisciplina*, porém o campo *NomeDisciplina* depende apenas do campo *CódigoDisciplina*. A Figura 1.17 ilustra a tabela *Nota*.

Nota

<u>NúmeroMatricula</u>	<u>CódigoDisciplina</u>	NomeDisciplina	Nota
2007230	GNB	GenéticaBásica	85

Figura 1.17: Exemplo de dependência funcional parcial.

A Dependência Funcional Transitiva é outro conceito fundamental para se compreender as Formas Normais. Uma dependência funcional transitiva ocorre quando pelo menos um atributo da tabela depende de um outro atributo que não

seja a chave primária (Rocha-Júnior, 2007). Um exemplo disso seria uma tabela *Funcionário* (*MatriculaFuncionário*, *NomeFuncionário*, *CódigoCargo*, *NomeCargo*, *SalárioCargo*), onde *MatriculaFuncionário* seria a chave primária da tabela. Nesse exemplo, somente os atributos *NomeFuncionário* e *CódigoCargo* estariam na dependência da chave primária. Já os atributos *NomeCargo* e *SalárioCargo* estariam na dependência do atributo *CódigoCargo*. A Figura 1.18 ilustra a tabela *Funcionário*.

Funcionário

<u>MatriculaFuncionário</u>	NomeFuncionário	CódigoCargo	NomeCargo	SalárioCargo
U370	Pedro	1	Professor	R\$ 6.000,00

Figura 1.18: Exemplo de dependência funcional transitiva.

Um outro importante conceito no processo de normalização é a Dependência Funcional Multivalorada. Essa dependência ocorre quando para cada valor de um atributo **A**, chave primária de uma tabela, há um conjunto de valores para outros atributos **B** e **C** e que estão associados ao atributo **A**. Os atributos **B** e **C** estão na dependência do atributo **A**, mas são independentes entre si (Rocha-Júnior, 2007). Suponha uma tabela *Livro* (*CódigoLivro*, *Autor*, *TítuloLivro*, *Assunto*, *Ano*), onde *CódigoLivro* é a chave primária da tabela. Para cada livro, pode haver um conjunto de valores tanto no campo *Autor* como no campo *Assunto* e ambos os campos são independentes entre si (ver Figura 1.19).

Livro

<u>CódigoLivro</u>	Autor	TítuloLivro	Assunto	Ano
1	Korth	Introdução a Sistemas de Banco de dados	Banco de dados	2003
2	Silberchatz	Introdução a Sistemas de Banco de dados	Teste funcional	2003
3	Mario Jino	Introdução ao teste de software	Teste de carga	2006
4	Eduardo Delamaro	Introdução ao teste de software	Teste de aceitação	2006
5	Carlos Maldonado	Introdução ao teste de software	Teste de regressão	2006

Figura 1.19: Exemplo de dependência funcional multivalorada.

Uma outra dependência, é a Funcional Cíclica que ocorre quando há dependências como: $A \rightarrow B$, $B \rightarrow C$, $C \rightarrow A$. Por exemplo, a dependência entre $professor \rightarrow disciplina$, $disciplina \rightarrow apostila$, $apostila \rightarrow professor$ é um caso de dependência cíclica (Rocha-Júnior, 2007). A Figura 1.20 mostra uma tabela que apresenta dependência funcional cíclica.

Curso

<u>Professor</u>	<u>Disciplina</u>	<u>Apostila</u>
Rodrigo	Programação I	Tutorial de Programação I
Carlos	Algoritmos Genéticos	Tutorial de Algoritmos Genéticos
Paulo	Engenharia de software	Tutorial de Engenharia de software

Figura 1.20: Exemplo de dependência funcional cíclica.

O processo de normalização se inicia com a Primeira Forma Normal. A PRIMEIRA FORMA NORMAL, ou simplesmente 1FN, acontece quando todos os atributos de uma tabela apresentam apenas valores atômicos, ou seja, a tabela não possui atributos multivalorados (Rocha-Júnior, 2007); (Silberschatz et al., 2006). Valores atômicos são aqueles que não podem ser subdivididos em valores menores, por exemplo, o CPF. Um exemplo de atributo multivalorado é o atributo *Endereço* da tabela *Aluno*. Na Figura 1.21 o atributo *Endereço* não apresenta valores atômicos, sendo, portanto, um atributo composto. Logo, a tabela *Aluno* não encontra-se na 1FN.

Aluno

<u>NúmeroMatrícula</u>	<u>NomeAluno</u>	<u>Endereço</u>
2007230	Marcos Oliveira	Av. Brasil, 977 – Centro – Belo Horizonte-MG

Figura 1.21: O atributo *Endereço* é multivalorado.

Para que a tabela *Aluno* esteja na 1FN, é necessário que o atributo *Endereço* seja dividido em outros atributos. A Figura 1.22 mostra a tabela *Aluno* em acordo com a 1FN.

Aluno

<u>NúmeroMatricula</u>	<u>NomeAluno</u>	<u>TipoLogradouro</u>	<u>Logradouro</u>	<u>Número</u>	<u>Bairro</u>	<u>Cidade</u>	<u>UF</u>
2007230	Marcos Oliveira	Avenida	Brasil	977	Centro	Belo Horizonte	MG

Figura 1.22: Exemplo de tabela normalizada pela 1FN.

Pela SEGUNDA FORMA NORMAL, ou 2FN, quando uma tabela estiver na 1FN e não apresentar dependência funcional parcial, essa tabela se encontra na 2FN. Caso a tabela apresente dependência funcional parcial, é necessário separar o atributo que apresenta dependência parcial em uma outra tabela. A Figura 1.23 mostra as tabelas *Nota* e *Disciplina* depois de aplicada a 2FN na tabela *Nota* da Figura 1.17.

Nota			Disciplina	
<u>NúmeroMatricula</u>	<u>CódigoDisciplina</u>	<u>Nota</u>	<u>CódigoDisciplina</u>	<u>NomeDisciplina</u>
2007230	GNB	85	GNB	Genética Básica

Figura 1.23: Exemplo de aplicação da 2FN.

Uma tabela se encontra na TERCEIRA FORMA NORMAL, ou 3FN, se estiver na 2FN e não possuir dependência funcional transitiva. Retomando o exemplo da Figura 1.18, observa-se que aquela tabela apresenta dependência funcional transitiva, pois os atributos *NomeCargo* e *SalárioCargo* estão na dependência do atributo *CódigoCargo*, que não é a chave primária da tabela.

Para resolver tal situação é necessário dividir a tabela *Funcionário* em duas tabelas: *Cargo* e *Funcionário*. A tabela *Cargo* seria formada pelos atributos que estão na dependência do atributo *CódigoCargo*. Já a tabela *Funcionário* seria formada pelos atributos *MatriculaFuncionário*, chave primária da tabela, *NomeFuncionário* e *CódigoCargo*. Esse último atributo seria chave primária na tabela *Cargo* e chave estrangeira na tabela *Funcionário*. A próxima figura mostra as

tabelas *Cargo* e *Funcionário*.

Cargo			Funcionário		
CódigoCargo	NomeCargo	SalárioCargo	MatriculaFuncionário	NomeFuncionário	CódigoCargo
1	Professor	R\$ 6.000,00	U370	Pedro	1

Figura 1.24: Exemplo de aplicação da 3FN.

Pela QUARTA FORMA NORMAL, ou 4FN, uma tabela se encontra nessa forma, se estiver na 3FN e não existirem dependências funcionais multivaloradas (Rocha-Júnior, 2007). Para exemplificar a 4FN, será necessário retomar o exemplo da Figura 1.19.

Na tabela da Figura 1.19, observa-se que há dependência funcional multivalorada nos atributos *Autor* e *Assunto*. Assim, para que essa tabela esteja em acordo com 4FN é necessário separar os atributos multivalorados em tabelas diferentes. As tabelas e seus respectivos campos seriam: *Livro* (*CódigoLivro*, *Título* e *Ano*); *Assunto* (*CódigoAssunto* e *Assunto*); *Autor* (*CódigoAutor* e *Nome*), *AutorLivro* (*CódigoAutor* e *CódigoLivro*) e *LivroAssunto* (*CódigoLivro* e *CódigoAssunto*).

O relacionamento *Autor* e *Livro* deu origem a tabela *AutorLivro*. Isso se deve ao fato de que esse relacionamento apresenta razão de cardinalidade N:N, pois cada livro pode ter vários autores, assim como uma pessoa pode ser autora de vários livros. O mesmo ocorre com a tabela *LivroAssunto*. Ela é o relacionamento entre as tabelas *Livro* e *Assunto*. Um livro pode abordar assuntos diferentes, ou então, um mesmo assunto pode ser abordado por livros diferentes. A Figura 1.25 mostra as tabelas *Livro*, *Assunto*, *Autor*, *LivroAutor* e *LivroAssunto*.

Assunto	
<u>CódigoAssunto</u>	Assunto
1	Banco de dados
2	Teste funcional
3	Teste de carga
4	Teste de aceitação
5	Teste de regressão

Autor	
<u>CódigoAutor</u>	Nome
1	Korth
2	Silberchatz
3	Mario Jino
4	José Maldonado
5	Márcio Delamaro

AssuntoLivro	
<u>CódigoAssunto</u>	CódigoLivro
1	1
1	2
2	3
3	3
4	3
5	3

AutorLivro	
<u>CódigoAutor</u>	CódigoLivro
1	1
1	2
2	2
3	3
4	3
5	3

Livro		
<u>CódigoLivro</u>	Título	Ano
1	Introdução a Banco de dados	2007
2	Introdução a Sistemas de Banco de dados	2003
3	Introdução ao teste de software	2006

Figura 1.25: Exemplo de aplicação da 4FN.

A última das regras de normalização é a QUINTA FORMA NORMAL (5FN) ou FORMA NORMAL DE PROJEÇÃO DE JUNÇÃO (FNPJ). Quando uma tabela apresenta dependência funcional cíclica, é necessário separar o ciclo por meio de relacionamentos N:N (Rocha-Júnior, 2007).

Para exemplificar o processo da 5FN, será necessário retornar ao exemplo da Figura 1.20. Nesse exemplo, há dependência funcional cíclica entre os atributos *Professor*, *Disciplina* e *Apostila* da tabela *Curso*. Para resolver esse problema criam-se três tabelas: *ProfessorDisciplina*, *DisciplinaApostila* e *ApostilaProfessor*. Dessa forma soluciona-se o problema da dependência funcional cíclica, pois cada uma dessas tabelas tornam-se relacionamentos com cardinalidade N:N. A Figura 1.26 mostra as três tabelas em acordo com a 5FN.

Professor	Disciplina
Rodrigo	Programação I
Carlos	Algoritmos Genéticos
Paulo	Engenharia de software

Apostila	Professor
Tutorial de Programação I	Rodrigo
Tutorial de Algoritmos Genéticos	Carlos
Tutorial de Engenharia de software	Paulo

Disciplina	Apostila
Programação I	Tutorial de Programação I
Algoritmos Genéticos	Tutorial de Algoritmos Genéticos
Engenharia de software	Tutorial de Engenharia de software

Figura 1.26: Exemplo de aplicação da 5FN.

Alguns autores afirmam que a aplicação das formas 1FN, 2FN e 3FN elimina a maioria das redundâncias encontradas em banco de dados. As poucas redundâncias restantes podem ser eliminadas com o uso das formas FNBC, 4FN e 5FN (Teorey et al., 2006); (Rocha-Júnior, 2007).

A próxima seção irei apresentar a linguagem SQL, hoje a linguagem de consulta padrão para SGBDs relacionais (Elmasri e Navathe, 2002).

1.5 SQL

O nome SQL, derivado de Structured Query Language, se refere a uma linguagem que foi desenvolvida pela IBM para servir como interface para SGBDs relacionais. Segundo Elmasri ((Elmasri e Navathe, 2002), a linguagem SQL pode ser considerada uma das principais razões para o sucesso dos bancos de dados relacionais.

SQL é uma linguagem que possui um conjunto de instruções para definição e para manipulação de dados. Portanto, SQL fornece uma Linguagem de Definição dos Dados (DDL), para especificar o esquema de um banco de dados, e uma Linguagem de Manipulação de Dados (DML), para expressar consultas, inserções, exclusões e atualizações em banco de dados relacionais (Silberschatz et al., 2006).

1.5.1 Linguagem de Definição dos Dados

No parágrafo anterior foi dito que a DDL é utilizada para especificar o ESQUEMA de um banco de dados. Um esquema em SQL é identificado através de um NOME DE ESQUEMA. Esse inclui um conjunto de informações que descrevem cada elemento do esquema. Os ELEMENTOS de um esquema são as tabelas e suas restrições e domínios (Elmasri e Navathe, 2002). A Figura 1.27 mostra um exemplo do esquema de um banco de dados.

```
create table Aluno
  (NomeAluno char(20),           Not Null
  NumeroMatricula int(7),       Not Null
 CodigoBolsa char(4),           Not Null
  primary key (NumeroMatricula),
  foreign key (CodigoBolsa))

create table Bolsa
  (CodigoBolsa char(4),         Not Null
  TipoBolsa char(10),          Not Null
  primary key (CodigoBolsa))

create table Disciplina
  (CodigoDisciplina char(3),    Not Null
  NomeDisciplina char(15),     Not Null
  CodigoDepartamento char(3), Not Null
  primary key (CodigoDisciplina),
  foreign key (CodigoDepartamento))

create table Departamento
  (CodigoDepartamento char(4), Not Null
  NomeDepartamento char(10),  Not Null
  primary key (CodigoDepartamento))
```

Figura 1.27: Exemplo de esquema de um banco de dados relacional.

No esquema da Figura 1.27, em cada declaração **create table** uma tabela está sendo criada no banco de dados. Entre parênteses encontram-se os nomes dos campos que compõem a tabela, a definição da chave primária, as chaves estrangeiras e os tipos de dados em cada campo.

Por TIPO DE DADO entende-se como sendo um conjunto de valores que são definidos para cada campo de uma tabela. Os tipos são também chamados de DOMÍNIOS (Date, 2004). No exemplo da Figura 1.27, o campo CódigoDepartamento, da tabela Departamento, recebe somente valores do tipo CHAR, tipo para caracteres, e com tamanho máximo de quatro caracteres.

O padrão SQL aceita vários tipos de dados. A Figura 1.28 mostra uma tabela com alguns dos tipos de dados.

Tipo de dado	Descrição do tipo
INT	É um tipo numérico somente para inteiros. Tamanho: 4 bytes.
TINYINT	É também um tipo numérico para inteiros. Tamanho: 1 byte.
SMALLINT	Tipo para números inteiros. Tamanho: 2 bytes.
CHAR	Seqüência de caracteres (string) de tamanho fixo. O número de caracteres é definido pelo usuário. Número máximo de caracteres: 255
VARCHAR	String de tamanho variável. Número máximo de caracteres: 65535
TEXT	Para textos com mais de 65535 caracteres.
BLOB	Tipo para imagens, vídeo e áudio. Tamanho: 65535 bytes.

Figura 1.28: Alguns dos tipos de dados.

Na Figura 1.27, observa-se nos campos das tabelas a cláusula **NOT NULL**. Isso significa que esses campos não aceitam o valor nulo. Além do comando **create table** outros comandos podem ser utilizados para alterar o esquema de um banco de dados. O primeiro destes comandos é o comando **drop schema**. Ele é usado

na remoção do esquema de um banco de dados. Um outro comando é o **drop table**, sendo utilizado na remoção de uma tabela do banco de dados (Elmasri e Navathe, 2002). O comando **alter table** serve para alterar alguma definição de uma tabela. Por exemplo, quando se quer alterar o tipo de dado em um campo (Elmasri e Navathe, 2002).

1.5.2 Linguagem de Manipulação dos Dados

SQL também possui comandos para manipulação dos dados. Essa manipulação dos dados envolve os processos de inserção, deleção, atualização e de consulta dos dados. O processo de consulta envolve três cláusulas: **select**, **from** e **where**. A cláusula **select** é usada para listar os atributos desejados no resultado de uma consulta. Para listar as tabelas a serem consultadas, utiliza-se a cláusula **from**. Já a cláusula **where** é uma expressão condicional que identifica as tuplas (linhas) a serem recuperadas na consulta (Elmasri e Navathe, 2002); (Silberschatz et al., 2006). A Figura 1.29 mostra um exemplo de consulta.

```
SELECT NomeAluno  
FROM Aluno  
WHERE NomeAluno = 'Marcos'
```

Figura 1.29: Exemplo de uma consulta.

O exemplo da Figura 1.29 mostra que o atributo selecionado foi *NomeAluno* (cláusula **select**) e a tabela selecionada foi *Aluno* (cláusula **from**). Já a condição é que sejam selecionadas somente as tuplas que contenham o nome “Marcos” no atributo *NomeAluno*.

Em SQL, além dos comandos para consulta, há três comandos que são utilizados na modificação do banco de dados: **insert**, **delete** e **update**. O comando **insert** é utilizado para inserir um novo registro em uma tabela. Para tal, deve-se especificar a tabela e em seguida os valores que serão inseridos (Elmasri e Navathe, 2002); (Silberschatz et al., 2006). A Figura 1.30 mostra um exemplo de inserção.

```
INSERT INTO Departamento (CodigoDepartamento, NomeDepartamento)
VALUES ('EST', Estatística')
```

Figura 1.30: Exemplo de inserção.

O comando **delete** é utilizado para remover tuplas de uma tabela. Ele possui a cláusula **where** para selecionar quais as tuplas que serão excluídas da tabela (Elmasri e Navathe, 2002); (Silberschatz et al., 2006). Um exemplo do uso do comando **delete** se encontra na Figura 1.31.

```
DELETE FROM Aluno
WHERE NomeAluno = 'Carlos'
```

Figura 1.31: Exemplo de deleção.

Por fim, o comando **update** é utilizado para modificar os valores de atributos de uma ou mais tuplas selecionadas. Também no comando **update**, existe a cláusula **where** para selecionar as tuplas que serão modificadas (Elmasri e Navathe, 2002); (Silberschatz et al., 2006). A Figura 1.32 mostra um exemplo de declaração para o comando **update**.

```
UPDATE Aluno
SET NomeAluno = 'Carlos Torrado'
WHERE NomeAluno = 'Carlos Torres'
```

Figura 1.32: Exemplo de atualização.

1.6 Diversidade Biológica

1.6.1 Classificação Biológica

É fascinante a diversidade de formas diferentes de vida que existem no planeta Terra. Entretanto, conhecer essa diversidade tem sido, até hoje, um grande desafio para a ciência. Isso acontece por duas razões. Primeiro é a necessidade de se nomear os vários organismos componentes da diversidade de uma forma pouco

arbitrária, e segundo é a importância de se conhecer os processos que deram origem a essa diversidade que irão auxiliar no desenvolvimento do primeiro.

Nomear e classificar são algumas das formas de seres humanos se comunicarem e estruturarem o conhecimento. Essas têm sido, já faz muito tempo, maneiras de se conhecer a diversidade biológica. É, portanto, da necessidade de se comunicar e conhecer que passamos a dar nomes e a classificar os organismos vivos. É assim que surge a taxonomia, o ramo da biologia que se preocupa em nomear e classificar os organismos vivos (Papavero, 1994).

A primeira tentativa de classificar os organismos vivos foi feita por Aristóteles no século IV a.C. Aristóteles tinha um interesse particular pelos animais. Ele chegou a escrever alguns trabalhos sobre os animais tais como *A História dos Animais* e *A Origem dos Animais*. A classificação feita por ele, se baseou em alguns caracteres morfológicos. Mais tarde, já no século XVIII, o trabalho de Aristóteles teria grande influência sobre a taxonomia proposta por Lineu (Amorim, 2002).

Carolus Linnaeus, ou simplesmente Lineu, foi quem deu início à moderna taxonomia. Em 1735, ele publicou um trabalho intitulado de *Systema Nature* onde apresentava o seu sistema de classificação e nomeação binomial, que usamos até hoje. Lineu propôs um sistema onde os organismos seriam classificados em cinco categorias: Reino, Classe, Ordem, Gênero e Espécie (Amorim, 2002). Mais tarde foram incluídas novas categorias, tais como: Domínio, Filo e Família. Dessa forma cada organismo recebe um nome binomial específico, que o inclui em um gênero, em uma família, em uma ordem, em uma classe, em um filo, em um reino e em um domínio.

Além disso, os organismos vivos seriam agrupados em classes chamadas de táxons. Segundo Amorim (Amorim, 2002), Táxon é qualquer agrupamento de organismos baseado em uma definição. Os grupos Animalia, Chordata (vertebrados e outros cordados), Mammalia, Perissodactyla (cavalos e antas), são exemplos de táxons. Uma outra importante característica do sistema lineano é o método binomial, ou seja, o uso de duas palavras para nomear as espécies. Antes desse sistema,

o nome das espécies trazia as principais características do organismo. Assim, uma espécie de abelha era conhecida por uns 8 ou 10 nomes, dependendo do pesquisador responsável pela classificação desta.

Com a adoção oficial da 10^a edição de *Systema Nature*, de 1758, pelo Código Internacional de Nomenclatura Zoológica ou ICZN, a descrição de uma nova espécie passou a seguir um conjunto de regras. Pelo código, a primeira palavra do nome de uma espécie se refere ao gênero ao qual ela pertence, enquanto que a segunda palavra é chamada de epíteto. Ambas as palavras devem vir destacadas no texto, ou em itálico ou sublinhadas. Além disso, a primeira palavra deve sempre começar com letra maiúscula e a segunda sempre com letra minúscula.

Pelo código, um mesmo táxon não pode ter nomes científicos diferentes. Isso é chamado de SINONÍMIA. Casos de sinonímia são freqüentes e podem ocorrer em todos os níveis taxonômicos. Quando ocorre sinonímia, o sinônimo válido para o táxon é o mais antigo. Esse sinônimo é chamado de SINÔNIMO SÊNIOR.

1.6.2 A Teoria da Evolução

Durante muito tempo perdurou entre os naturalistas a idéia de que as espécies seriam entidades fixas. O argumento era que as espécies foram criadas por Deus de forma perfeitamente adaptada ao seu ambiente e por isso não haveria razões para que elas passassem por mudanças. Por volta do século XVII começaram a surgir, entre alguns naturalistas, as primeiras idéias sobre a transformação das espécies. Erasmus Darwin, avô de Charles Darwin, foi um desses naturalistas que começaram a especular a respeito da transformação dos organismos vivos. Porém, nenhum desses naturalistas estava interessado em propor uma teoria que explicasse o porque as espécies mudavam (Ridley, 2006).

Foi Jean-Baptiste Lamarck quem pela primeira vez propôs uma teoria que explicasse a transformação das espécies. Em 1809 ele publicou uma obra, *Philosophie Zoologique*, na qual expôs a sua teoria evolutiva. Para Lamarck a evolução das espécies acontecia por meio da transmissão, para as gerações seguintes, dos

caracteres adquiridos pelo uso e desuso dos seus órgãos.

Depois de percorrer o mundo por cinco anos em um navio chamado *Beagle*, observando e coletando plantas e animais, Charles Robert Darwin começou a questionar sobre a variação entre os pássaros que ele havia coletado nas ilhas Galápagos. Darwin não queria somente explicar a mudança das espécies, mas também o fato de elas serem bem adaptadas ao seu ambiente.

Em 1838, quando estava lendo o livro *Essay on the principle of population*, de Thomas Malthus, Darwin se deparou com uma afirmação de Malthus que foi fundamental na formulação da idéia da luta pela vida (*Struggle of existence*). Segundo Malthus a taxa de reprodução das espécies seria maior do que o número de descendentes que conseguem sobreviver e que, por isso, o tamanho das populações se mantinha mais ou menos constante. Darwin, então, concluiu que a maioria dos descendentes era eliminado a cada geração por conta da dura competição por alimento, pela predação ou pelas adversidades ambientais. Combinando a idéia da luta pela vida com as suas observações sobre a variação, Darwin percebeu que algumas variações aumentam as chances de sobrevivência e por isso os indivíduos que as possuem deixam um maior número de descendentes.

Em 1858, depois de 20 anos de trabalho, Darwin resolveu publicar a sua teoria sobre a evolução das espécies. O fato que levou Darwin a publicá-la foi uma carta que ele havia recebido de Alfred Russel Wallace. Independentemente, Wallace havia chegado a uma teoria muito semelhante à de Darwin (Ridley, 2006). Eles então acertaram de anunciá-la na Sociedade Lineana de Londres. Em 1859, Darwin finalmente publica a sua principal obra sobre a evolução das espécies *On the Origin of Species* (A Origem das Espécies).

Uma das críticas que foram feitas à teoria de Darwin foi o fato de ela não incluir uma teoria consistente sobre a hereditariedade. Infelizmente, o trabalho de Gregor Mendel *Versuche über Pflanzen-Hybriden* (Experimentos em hibridação de plantas), publicado em 1865, foi pouco conhecido até o início do século XX. Essa situação começou a mudar em 1900 quando os trabalhos de Mendel foram

redescobertos. Em 1920 o mendelismo já era uma teoria da hereditariedade bem aceita (Ridley, 2006). Finalmente, na década de 1930, com os trabalhos teóricos de Ronald A. Fisher, Sewall Wright e John B. S. Haldane, surge a teoria sintética da evolução a partir da síntese das teorias de Mendel e Darwin.

1.6.3 Sistemática Filogenética

Quando Darwin publicou *A Origem das Espécies*, ele já havia percebido que a evolução ocorria em um padrão de árvore. Para Darwin as ramificações em uma árvore representavam os eventos onde duas novas espécies surgiam a partir de um ancestral comum, ou seja, um evento de especiação era uma ramificação de uma árvore que continha vários eventos de especiação ilustrados.

Depois de Darwin, um outro naturalista que concebeu a evolução como uma árvore foi Ernest Haeckel. Foi Haeckel quem primeiro usou a expressão FILOGENIA para nomear a árvore da vida. Para ele, a filogenia tinha como objetivo o estudo da origem e a história evolutiva das diversas linhagens de organismos.

Durante muito tempo, a classificação dos organismos se baseou em critérios de semelhança e diferenças entre os organismos de uma maneira arbitrária e pouco criteriosa. Conceitos como ancestral, descendente, relacionamento filogenético, embora presentes na teoria evolutiva, não eram usados na prática da classificação dos organismos. Foi Willi Hennig, um entomólogo alemão, quem pela primeira vez propôs uma nova metodologia de classificação dos diversos grupos de organismos.

Em 1950, Hennig publicou o trabalho *Grundzüge einer Theorie der Phylogenetischen Systematik* onde apresentava a sua metodologia. Porém, durante 16 anos as suas idéias ficaram desconhecidas da comunidade científica. Foi somente em 1966, com a publicação, em inglês, de uma versão do seu trabalho, é que as suas idéias puderam ser divulgadas. O seu método de reconstrução filogenética ficou conhecido como SISTEMÁTICA FILOGENÉTICA ou CLADISMO.

A sistemática filogenética busca classificar a diversidade biológica com base nas relações de parentesco (filogenéticas) entre os grupos de organismos que com-

partilham um mesmo ancestral comum e exclusivo. Esses grupos de organismos que compartilham de um único e exclusivo ancestral comum, Hennig denominou de MONOFILÉTICOS. Os grupos monofiléticos são também chamados de CLADOS. Um clado pode pertencer a qualquer nível taxonômico.

1.7 Sistemas de informação e diversidade biológica

Em 1992 aconteceu na cidade do Rio de Janeiro a Conferência das Nações Unidas para o Meio Ambiente e o Desenvolvimento. Um dos principais resultados da conferência foi a Convenção sobre Diversidade Biológica (CDB), que hoje se encontra ratificada por mais de 180 países. O CDB prevê, em seu artigo 17, o intercâmbio de informações de todas as fontes disponíveis do público, pertinentes à conservação e à utilização sustentável da diversidade biológica. Indica também que esse intercâmbio de informações deve incluir a troca de resultados de pesquisas técnicas, científicas e sócio-econômicas e que deve também, quando possível, incluir a repatriação das informações.

O artigo do CDB mencionado anteriormente vem confirmar uma importante necessidade, o uso da tecnologia da informação na implementação de bases de dados para informações sobre biodiversidade. Estudos vêm confirmando a necessidade de se abordar de forma multidisciplinar a problemática sobre o acesso às informações referentes à biodiversidade (Canhos, 2006); (Canhos et al., 2006).

No cenário internacional três projetos, entre outros, se destacam: o GBIF (Global Biodiversity Information Facility), ITIS (Integrated Taxonomic Information System) e Species 2000. O GBIF (<http://www.gbif.org/>) é uma organização intergovernamental que tem como missão tornar o acesso às informações sobre a biodiversidade mundial de forma gratuita e universalmente disponível através da internet (Canhos et al., 2006); (Yesson et al., 2007). Ele é hoje o maior portal com esse tipo informação (Yesson et al., 2007). O GBIF também procura desenvolver novas ferramentas em tecnologia da informação para melhor divulgação das informações sobre a biodiversidade mundial (Edwards, 2004); (Yesson et al., 2007).

O ITIS (<http://www.itis.gov>) é uma parceria de agências federais e outras organizações dos Estados Unidos, Canadá e México, com informações de especialistas de todo o mundo. O sistema também é uma base de dados com os nomes científicos e comuns da biota de interesse para a América do Norte. Ele contém mais de meio milhão de nomes de espécies de todos os reinos. O ITIS é parte da National Biological Information Infrastructure (USNBII) e membro associado do GBIF.

Já o *Species 2000* (<http://www.sp2000.org>) é uma organização que envolve taxonomistas de vários países. O seu objetivo é compilar um índice com os nomes válidos de espécies conhecidas em todo o mundo. É uma organização associada ao GBIF. Existem dois programas regionais: Species 2000 Europa (<http://www.sp2000europa.org>), financiado pela Comissão Européia para trabalhar com bases de dados globais e regionais baseadas na Europa, e Species 2000 Ásia-Oceania (www.sp2000asiaoceania.org) programa que tem como objetivo a promoção de uma base de dados taxonômica naquela região.

O *Catalogue of Life* (<http://www.catalogueoflife.org/annual-checklist/>) é formado pelo consórcio entre o Species 2000 e o ITIS, e tem como objetivo a construção, até 2011, de um amplo catálogo com os nomes das espécies conhecidas no mundo inteiro. Todos os anos esse catálogo é atualizado e atualmente se encontra na versão 2007 (<http://www.catalogueoflife.org/annual-checklist/2007/>).

Um outro importante projeto é o *Tree of Life* (<http://www.tolweb.org/>). Ele reúne um conjunto de informações sobre a biodiversidade, elaboradas em colaboração entre centenas de especialistas de todo o mundo. Cada página contém fotos, textos e outras informações para cada espécie e para cada grupo de organismos, vivos ou extintos. As conexões entre as páginas seguem padrões de ramificações filogenéticas entre os grupos de organismos, de tal maneira que o usuário possa navegar na hierarquia da vida e aprender sobre filogenia e evolução, bem como as características de cada um dos grupos (Maddison et al., 2007).

Já no âmbito nacional, em 1999 foi lançado pela FAPESP um programa

abrangente de estudo da diversidade biológica do Estado de São Paulo. O programa intitulado “BIOTA/FAPESP: O INSTITUTO VIRTUAL DA BIODIVERSIDADE” tem entre os seus objetivos: compreender os processos que geram e mantêm a biodiversidade, assim como aqueles que resultam na sua redução; tornar as informações importantes para o aperfeiçoamento das políticas públicas de conservação e uso sustentável da biodiversidade, disponíveis para os órgãos responsáveis por sua definição e implementação; e assegurar o acesso público, livre e gratuito a estas informações (Joly e Speglich, 2003).

Ainda no âmbito nacional, foi desenvolvido pelo LABINFO do Laboratório Nacional de Computação Científica (LNCC/MCT) juntamente com o Laboratório de Biodiversidade Molecular do Departamento de Genética da UFRJ o MamMiBase, um banco de dados de genoma mitocondrial completo de espécies de mamíferos. Esse banco reúne um conjunto de informações, como: tamanho dos genes mitocondriais, distância p , conteúdo de cada uma das bases nos genes, taxa de transição/transversão, parâmetro gama e outras informações, que são importantes em estudos filogenéticos (Vasconcelos et al., 2005).

Em 2006, foi publicado um trabalho realizado pelo Ministério do Meio Ambiente e coordenado por Thomas Lewinsohn (Lewinsohn, 2005) sobre o estado do conhecimento da biodiversidade brasileira. Uma das propostas desse estudo foi o uso de tecnologias de bioinformática na organização, processamento e difusão de informações sobre biodiversidade. Uma dessas tecnologias é o desenvolvimento de banco de dados, que possam ser acessados pela internet, com informações sobre a biodiversidade.

O principal bioma do Estado do Rio de Janeiro é a Mata Atlântica. Por ser considerado um dos biomas mais biodiversos do planeta, foi incluído na lista dos *hotspots* (Myers et al., 2000). Por isso, o Estado do Rio de Janeiro é considerado possuidor de uma das mais ricas e ao mesmo tempo ameaçadas faunas do Brasil e do mundo (Rodrigues et al., 2005). Diante de tais evidências torna-se relevante a implementação de uma base de dados com informações sobre as espécies de animais

descritas para o Estado do Rio de Janeiro.

1.8 Objetivos

Portanto, esta dissertação teve como objetivo a construção do sistema Biota-Rio, uma base de dados de acesso livre pela internet com informações sobre a fauna descrita para o Estado do Rio de Janeiro. Informações como: as descrições originais, as fotos, as localidades-tipo, o ambiente em que as espécies vivem e outras informações referentes a todas as espécies de vertebrados e invertebrados, serão cadastradas no Biota-Rio.

Os objetivos específicos do presente trabalho foram:

- A construção do sistema Biota-Rio, ou seja, o desenvolvimento do modelo de dados, a implementação das tabelas usando um SGBD, e a implementação dos módulos de consulta.
- E o levantamento das espécies de mamíferos descritas para o Estado do Rio de Janeiro.

A continuidade do levantamento, para os outros grupos de animais, e a inclusão das informações na base de dados serão feitas em momento posterior a esse trabalho de mestrado.

Capítulo 2

Metodologia

Neste capítulo será apresentado o modelo de dados do banco Biota-Rio, em seguida as ferramentas que foram utilizadas na implementação do sistema e, finalizando, o levantamento bibliográfico.

2.1 Modelo ER do banco de dados Biota-Rio

O modelo ER do banco de dados Biota-Rio está baseado em treze tipos de entidades, são elas: *Ambiente*, *Artigo*, *Autor*, *Categorias*, *Gene*, *Instituição*, *Localidade-tipo*, *Nome comum*, *Nome do Gene*, *Organismo*, *Revista*, *Situação* e *Táxon*. A Figura 2.1 mostra o diagrama Entidade-Relacionamento do banco de dados Biota-Rio.

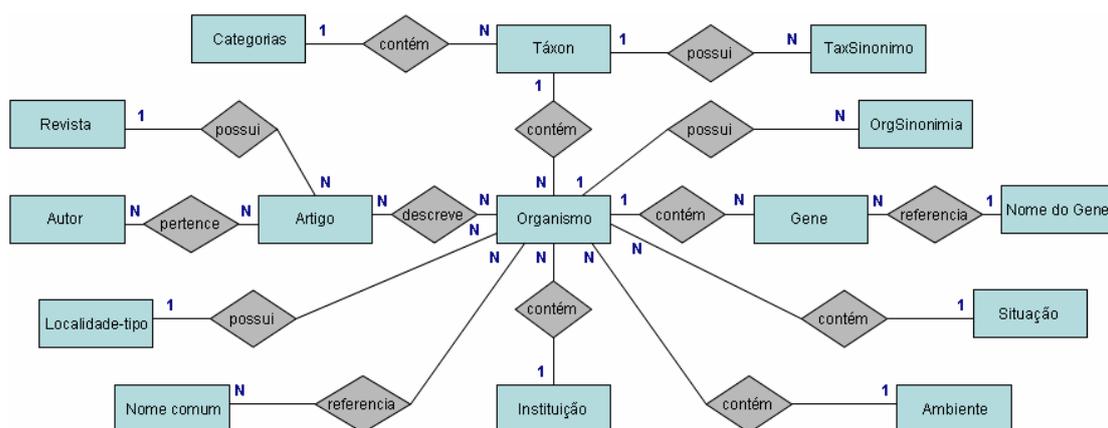


Figura 2.1: Esquema do banco de dados Biota-Rio na forma de um diagrama ER.

A Figura 2.1 também mostra os relacionamentos que compõe o modelo ER do banco de dados Biota-Rio. São ao todo doze relacionamentos. Cada um dos relacionamentos apresenta a sua própria razão de cardinalidade. Por exemplo, o relacionamento entre os tipos de entidades *Categorias* e *Táxon* apresenta cardinalidade 1:N. Isso significa que uma entidade *categoria* se relaciona com várias entidades *táxon*. Os táxons *Mammalia* e *Aves* são exemplos de táxons que pertencem à categoria *Classe*.

A definição dos atributos de cada um dos tipos de entidades foi feita considerando as informações biológicas que são fundamentais na sistemática de uma espécie. A Figura 2.2 mostra uma tabela com cada um dos tipos de entidades do banco de dados Biota-Rio, e seus respectivos atributos.

Tipo de Entidade	Atributos
Ambiente	1- Tipo de bioma. 2- Informações.
Artigo	1-Título. 2-Resumo. 3-Páginas.
Autor	1-Nome do autor.
Categorias	1-Nome da categoria.
Gene	1-AcessNumber. 2-Gi. 3-Endereço URL no NCBI.
Instituição	1-Nome da instituição. 2-Seção. 3-Cidade. 4-País.
Localidade-tipo	1-Nome da localidade. 2-Informações.
Nome comum	1-Nome.
Nome do Gene	1-Nome.
Organismo	1-Nome do organismo. 2-Descrição original. 3-Descrição alternativa. 4-Diagnose. 5-Genoma Mitocondrial. 6-Informações. 7-Foto.
OrgSinonimia	1-Sinônimo
Revista	1-Nome da revista. 2-Volume. 3-Edição. 4-Número. 5-Ano.
Situação	1-Tipo de situação.
Táxon	1-Nome do táxon.
TaxSinonimo	1-Sinônimo

Figura 2.2: Tabela com a relação dos atributos para cada tipo de entidade do modelo ER do banco de dados Biota-Rio.

Como pode ser observado na Figura 2.2, é no tipo de entidade *Nome do Gene* que se encontra o atributo *Nome*, que referencia o nome do gene, e não no tipo de entidade *Gene*. Isso é explicado pelo fato de que cada gene cadastrado na base de dados é identificado pelo seu *AccessNumber*. Não é o nome que identifica o gene, pois um mesmo nome pode referenciar vários genes com *AccessNumber* diferentes. Por exemplo, o gene mitocondrial COI (Citocromo oxidase I) está presente em todos os organismos com respiração aeróbica, entretanto, o gene COI de cada organismo é identificado, no GenBank, pelo seu *AccessNumber*.

2.2 O banco de dados Biota-Rio

O banco de dados Biota-Rio é um banco relacional, onde cada um dos tipos de entidades definidos no modelo ER deu origem a uma tabela. As Figuras 2.3 e 2.4 apresentam as treze tabelas do banco, seus respectivos campos e uma descrição sucinta de cada campo.

Tabela	Campo(s)	Descrição de cada campo
Ambiente	1-Id_Ambiente	- Chave primária.
	2-TipoAmbiente	- Tipo de bioma da localidade-tipo.
	3-InfoAmbiente	- Informações relevantes sobre o bioma.
Artigo	1-Id_Artigo	- Chave primária.
	2-Id_Revista	- Chave estrangeira.
	3-Titulo	- Título do artigo.
	4-Resumo	- Resumo do artigo.
	5-Paginas	- Páginas do artigo.
Autor	1-Id_Autor	- Chave primária.
	2-NomeAutor	- Nome do Autor de um artigo.
Categorias	1-Id_Categoria	- Chave primária.
	2-NomeCategoria	- Nome de uma categoria taxonômica.
Gene	1-Id_Gene	- Chave primária.
	2-Id_NomeGene	- Chave estrangeira.
	3-Id_Organismo	- Chave estrangeira.
	4-AccessNumber	- AccessNumber de gene no NCBI.
	5-Gi	- Gi de gene no NCBI.
	6-EnderecoWeb	- Endereço URL de um gene no NCBI.
	7-Variante	- Informa qual o variante cadastrado.
Instituição	1-Id_Instituição	- Chave primária.
	2-NomeDaInstituição	- Nome da instituição onde se encontra depositado os tipos.
	3-Seção	- Setor da instituição.
	4-Cidade	- Cidade onde se encontra a instituição.
	5-Pais	- País.
LocalidadeTipo	1-Id_LocalidadeTipo	- Chave primária.
	2-NomeLocalidade	- Nome da cidade, distrito.
	3-InfoLocalidade	- Informações sobre a localidade-tipo.
NomeComum	1-Id_NomeComum	- Chave primária.
	2-Nome	- Nome(s) comum(ns) dos organismos.
NomeDoGene	1-Id_NomeGene	- Chave primária.
	2-NomeDoGene	- Nome do gene.

Figura 2.3: Tabela com as treze tabelas que compõe o banco de dados Biota-Rio e seus respectivos campos.

Na Figura 2.4 será apresentada uma descrição de cada um dos campos das tabelas que compõem o banco de dados Biota-Rio:

Organismo	1-Id_Organismo	- Chave primária.
	2-Id_Ambiente	- Chave estrangeira.
	3-Id_Instituição	Chave estrangeira.
	4-Id_LocalidadeTipo	- Chave estrangeira.
	5-Id_Situação	- Chave estrangeira.
	6-Id_Taxon	- Chave estrangeira.
	7-OrgPai	- Guarda o id do organismo-pai, caso o organismo que está sendo cadastrado seja subespécie*.
	8-NomeOrganismo	- Nome do organismo, que pode ser uma espécie ou subespécie.
	9-DescOriginal	- Descrição original do organismo.
	10-DescAlternativa	- Descrição alternativa.
	11-Diagnose	- Diagnose.
	12-NumTombo	- Número de tombo do tipo.
	13-InfoOrganismo	-Outras informações sobre o organismo.
	14-Foto	- Foto.
OrgSinonimia	1-Id_Sinonimia	- Chave primária.
	2-Id_Organismo	- Chave estrangeira.
	3-Sinonimo	- Guarda o sinônimo de uma espécie.
Situação	1-Id_Situação	- Chave primária.
	2-TipoSituação	- Situação do organismo na lista vermelha da IUCN.
Taxon	1-Id_Taxon	- Chave primária.
	2-Id_Categoria	- Chave estrangeira.
	3-NoPai	- Guarda o id do táxon-pai*.
	5-NomeTaxon	- Nome do táxon.
	6-GenusFlag	- Flag para gênero ou subgênero.
	7-InfoTaxon	- Informações relevantes sobre o táxon.
TaxSinonimo	1-Id_Sinonimo	- Chave primária.
	2-Id_Taxon	- Chave estrangeira.
	3-Sinonimo	- Guarda o sinônimo de um táxon.

Figura 2.4: Continuação da tabela da Figura 35. Os tópicos que se encontram marcados com um asterisco (*), serão abordados mais adiante.

- Tabela *Ambiente*: O campo *Id_Ambiente* é a chave primária da tabela. Já o campo *TipoAmbiente* se refere ao tipo de bioma encontrado na localidade-

tipo. O campo *InfoAmbiente* guardará informações mais detalhadas sobre a situação do bioma, como descrição, estado de conservação do bioma na localidade-tipo, ecossistemas encontrados na localidade (restinga, mangue, ilhas oceânicas, campos de altitude).

- Tabela *Artigo*: O campo *Id_Artigo* é a chave primária da tabela. *Id_Revista* é chave estrangeira. Os campos *Titulo*, *Resumo* e *Paginas* referenciam, respectivamente, o título de um artigo, o resumo e as páginas de um artigo.
- Tabela *Autor*: O campo *Id_Autor* é a chave primária da tabela e o campo *NomeAutor* guarda o nome dos autores de um artigo.
- Tabela *Categoria*: O campo *Id_Categoria* é a chave primária da tabela, enquanto o campo *TipoCategoria* guarda as categorias taxonômicas (Reino, Filo, Subfilo...).
- Tabela *Gene*: *Id_Gene* é a chave primária da tabela. *Id_NomeGene* e *Id_Organismo* são chaves estrangeiras. *AccessNumber* é o campo que guarda o número de acesso no NCBI para cada gene. *Gi* é o campo que guarda o GI do gene no NCBI. Já o campo *EnderecoWeb* guarda o endereço URL da página do gene no NCBI.
- Tabela *Instituição*: *Id_Instituição* é a chave primária da tabela. *NomeDaInstituição* é o campo que recebe o nome da instituição (museu, universidade) onde se encontra depositado a série-tipo (holótipo e parátipos). *Seção* é o campo que se refere à seção (departamento, laboratório) da instituição onde se encontra a série-tipo. *Cidade* é o campo que guarda o nome da cidade onde se localiza a instituição, e *País* é o país onde se encontra a instituição na qual o holótipo foi tombado.
- Tabela *Localidade-tipo*: *Id_Localidade-tipo* é a chave primária da tabela. O campo *NomeLocalidade* recebe o nome da localidade (município) onde foi feita a descoberta da espécie. O campo *InfoLocalidade* recebe outras

informações (região do Estado do Rio de Janeiro, principais atividades econômicas) sobre o município.

- Tabela *NomeComum*: *Id_NomeComum* é a chave primária da tabela. O campo *Nome* recebe o nome comum de cada espécie.
- Tabela *NomeDoGene*: O campo *Id_NomeGene* é a chave primária da tabela. O campo *NomeDoGene* guarda o nome dos genes mitocondriais.
- Tabela *Organismo*: O campo *Id_Organismo* é chave primária da tabela. Os campos *Id_Ambiente*, *Is_Instituição*, *Id_LocalidadeTipo*, *Id_Situação*, *Id_Taxon*, são todos chaves estrangeiras. O campo *OrgPai* guarda o id do “organismo-pai”, caso o organismo que está sendo cadastrado seja subespécie (esse item será abordado mais adiante). *NomeOrganismo* é o campo que guarda o nome científico do organismo. *DescOriginal* é o campo que recebe a descrição original da espécie, ou seja, descrição feita pelo autor da descoberta da espécie. Em alguns casos, a descrição original é sumária devido ao veículo em que a descrição foi publicada, ou então, aos recursos tecnológicos disponíveis na época da publicação. Nestes casos, uma descrição alternativa e mais completa que esteja disponível na literatura será também depositada em um campo específico denominado *DescAlternativa*. O campo *Diagnose* recebe a diagnose de uma espécie, sendo ela uma outra importante informação no processo de identificação da mesma. *NumTombo* é o campo que guarda o número de tombo do holótipo. *InfoOrganismo* é um campo onde outras informações relevantes (se a espécie é endêmica ao Estado do Rio de Janeiro, em quais outros estados brasileiros a espécie é encontrada) são depositadas. No campo *Foto*, serão depositadas, quando disponíveis, as fotos das espécies ou de algumas de suas estruturas. Esse procedimento pode facilitar na identificação de uma espécie.
- Tabela *OrgSinonimia*: O campo *Id_Sinonimo* é a chave primária da tabela, enquanto que o campo *Id_Organismo* é uma chave estrangeira que relaciona

a sinônimo de uma espécie, em caso de sinonímia.

- Tabela *Revista*: O campo *Id_Revista* é a chave primária da tabela. Os campos *NomeRevista*, *Numero*, *Volume*, *Edição* e *Ano*, guardam respectivamente o nome, o número, o volume, a edição e o ano de uma revista.
- Tabela *Situação*: O campo *Id_Situação* é a chave primária da tabela. O campo *TipoSituação* se refere a situação da espécie na lista vermelha da IUCN.
- Tabela *Taxon*: O campo *Id_Taxon* é a chave primária da tabela. *Id_Categoria* é uma chave estrangeira que relaciona a tabela *Taxon* com a tabela *Categoria*. O campo *NoPai* guarda o id do “táxon-pai” (esse item será abordado com mais adiante). *NomeTaxon* é o campo onde são depositados os nomes dos táxons. *GenusFlag* é o campo que guarda uma FLAG para identificar os táxons que pertencem às categorias gênero ou subgênero. Por exemplo, o táxon Metazoa também pode ser identificado pelo nome Animalia. O campo *InfoTaxon* é onde são cadastradas informações relevantes sobre o táxon.
- Tabela *TaxSinonimo*: *Id_Sinonimo* é a chave primária da tabela. *Id_Taxon* é uma chave estrangeira e, *Sinonimo* é o campo que guarda o sinônimo de um táxon.

Além das treze tabelas, o modelo ER (ver Figura 2.1) mostra que existem três relacionamentos do tipo N:N. Nesses casos, é exigido que o relacionamento se torne uma tabela (Date, 2004). São elas: *ArtigoAutor*, *OrgArtigo* e *OrgNome-Comum*. Portanto, o banco de dados Biota-Rio é constituído por 16 tabelas. A Figura 2.5 apresenta o modelo de dados do banco Biota-Rio com todas as suas tabelas. O modelo foi feito usando a ferramenta de modelagem DBDesigner 4.0 (<http://fabforce.net/dbdesigner4/>).

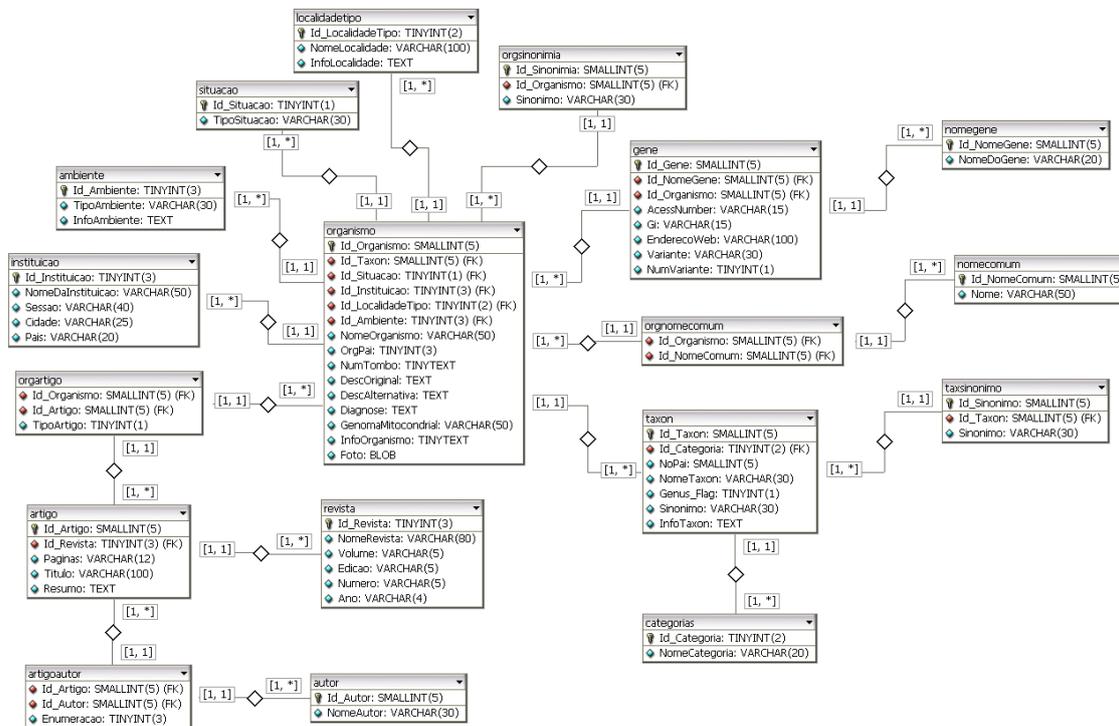


Figura 2.5: Modelo de dados do banco Biota-Rio.

Na Figura 2.5, pode-se observar o tipo de dado e o tamanho máximo de cada campo. O campo *TipoAmbiente* da tabela *Ambiente*, por exemplo, recebe valores do tipo *VARCHAR* e com tamanho máximo de trinta caracteres.

Uma das formas de consulta ao banco de dados Biota-Rio será através de uma árvore com a hierarquização dos táxons. Para implementar essa árvore foi necessário representá-la no modelo de dados, mais precisamente nas tabelas *Taxon* e *Organismo*. Ambas as tabelas apresentam campos que são fundamentais no processo de representação dos dados de forma hierárquica. Na tabela *Taxon* dois campos se destacam, são eles: *Id_Taxon* e *NoPai*. A Figura 2.6 mostra a tabela *Taxon* e os dois campos que são utilizados no processo de hierarquização dos táxons.

Id_Taxon	NoPai	NomeTaxon
1	0	A
2	1	B
3	1	C
4	1	D
5	2	E
6	3	F
7	3	G
8	4	H

Figura 2.6: Representação da tabela *Taxon* e três de seus campos, *Id_Taxon*, *NoPai* e *NomeTaxon*.

Cada linha na tabela da Figura 2.6 se refere a um único táxon, ou seja, cada táxon é referenciado por um único valor no campo *Id_Taxon*, que é chave primária. O valor único para cada linha na chave primária é também chamado de id. O campo *NoPai* é usado para guardar o valor id do “táxon-pai” quando um novo táxon é cadastrado na tabela. Em outras palavras, o campo *NoPai* funciona como um ponteiro para o “táxon-pai”. Por exemplo, o táxon **G**, que tem id igual a 7, tem como “táxon-pai” o táxon **C** com id igual 3. Já o táxon **C**, que possui id igual a 3, tem como “táxon-pai” o táxon **A** com id igual a 1.

A representação que ocorre na tabela *Organismo* é semelhante a que ocorre na tabela *Taxon*. A Figura 2.7 é uma representação do que se passa na tabela *Organismo*.

Id_Organismo	Id_Taxon	OrgPai	NomeOrganismo
1	5	0	EI
2	5	0	EJ
3	6	0	FK
4	7	0	GL
5	8	0	HM
6	8	0	HN
7	0	2	EJO
8	0	5	HMP

Figura 2.7: Representação da tabela *Organismo* e três de seus campos, *Id_Organismo*, *Id_Taxon*, *OrgPai* e *NomeOrganismo*.

Na tabela *Organismo*, assim como na tabela *Taxon*, cada organismo cadastrado recebe um único id, ou seja, é referenciado por um único valor na chave primária *Id_Organismo*. O campo *Id_Taxon* é uma chave estrangeira que guarda o id dos táxons que pertencem ou a categoria gênero ou a categoria subgênero. O campo *OrgPai* guarda o id da espécie quando uma subespécie está sendo cadastrada na tabela. Esse campo, assim como o campo *NoPai* na tabela *Taxon*, funciona como um ponteiro referenciando a espécie ao qual uma subespécie pertence. Quando o organismo é uma espécie o valor atribuído ao campo *OrgPai* é zero. Por exemplo, o organismo **FK**, que possui id igual a 3, pertence ao táxon **F**, pois o valor no campo *Id_Taxon* é 6. Como o valor no campo *OrgPai* é igual a zero, então se trata de uma espécie. Por outro lado, o organismo **EJO**, com id igual a 7, é uma subespécie que pertence a espécie **EJ**, pois o valor no campo *OrgPai* é 2. Note que nesse caso o valor no campo **Id_Taxon** é zero. A Figura 2.8 mostra um diagrama, em forma de árvore, que resume o que se passa nas tabelas *Taxon* e *Organismo*.

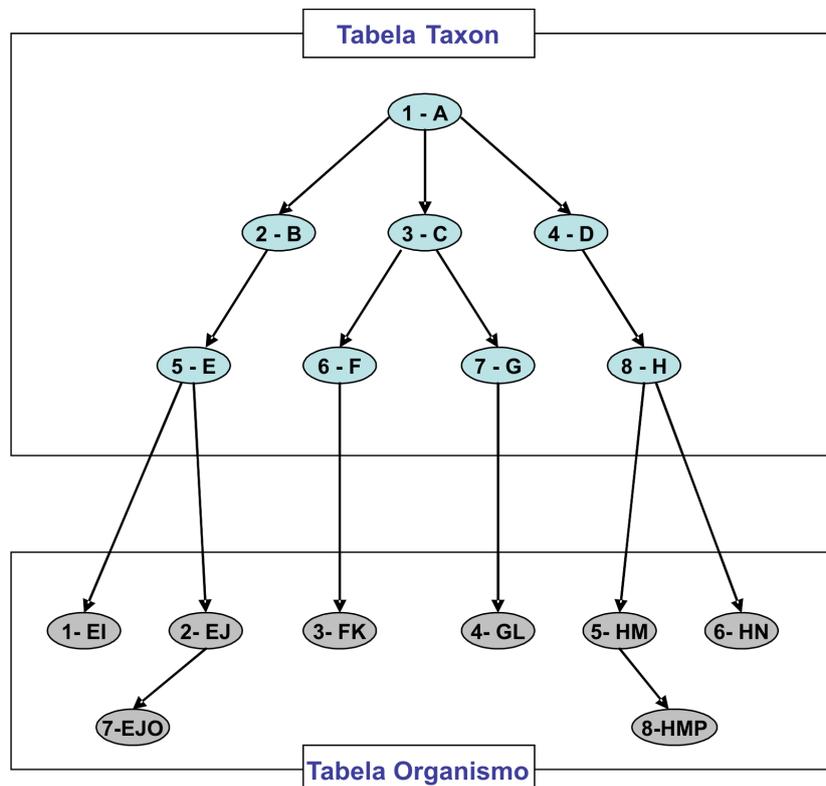


Figura 2.8: Diagrama representando o processo de hierarquização nas tabelas *Taxon* e *Organismo*.

Na Figura 2.8, os táxons e os organismos são representados pelas elipses e os retângulos representam as tabelas. As elipses da tabela *Taxon* representam os táxons e as da tabela *Organismo* os organismos, que podem ser espécie ou subespécie. Cada elipse na tabela *Taxon* é identificada por um número, o id do táxon, e por uma letra, o nome do táxon. Já na tabela *Organismo*, cada elipse é identificada por um número, o id do organismo, e por duas letras, se o organismo for espécie, e por três letras se organismo for subespécie.

2.3 Ferramentas utilizadas na implementação do sistema Biota-Rio

No desenvolvimento do sistema Biota-Rio, foi utilizada a linguagem de programação PERL versão 5.6 (<http://www.perl.org>), por ser uma linguagem compatível com várias plataformas de desenvolvimento e por ter uma grande quantidade de bibliotecas disponíveis gratuitamente na rede. Duas dessas bibliotecas se destacam, o DBI e o CGI. Para que os scripts PERL possam se conectar e

interagir com o banco de dados é preciso uma interface que estabeleça a comunicação. Quem faz essa interface é o DBI (DataBase Interface) (<http://dbi.perl.org>). Já o CGI (Common Gateway Interface) (<http://hoohoo.ncsa.uiuc.edu/cgi>) é um conjunto de protocolos que permitem a troca de informações entre o navegador (Browser) e os processos executados em um servidor, que são, neste caso, escritos com a linguagem PERL.

O SGBD utilizado foi o MySQL versão 5.1 (<http://www.mysql.com>). O MySQL é um dos SGBDs de código fonte aberto mais utilizados no mundo. Isso se deve a sua portabilidade, pois é um SGBD compatível com vários sistemas operacionais, sua velocidade e sua estabilidade (Dubois, 2003); (Milani, 2006). Muitas corporações utilizam o MySQL, alguns exemplos são a Sony, Cisco, Telemar, Nokia, entre outras (<http://www.mysqlbrasil.com.br/referencias-cases>). Mesmo entre os bancos de dados com aplicação biológica são exemplos, o projeto Tree of Life (www.tolweb.org) e o Fishbase (www.fishbase.org). Todo o sistema Biota-Rio é executado em um servidor SunOS 5.8. com o servidor para páginas na Web, Apache 2.0 (www.apache.org). O diagrama da Figura 2.9 representa a arquitetura do sistema Biota-Rio.

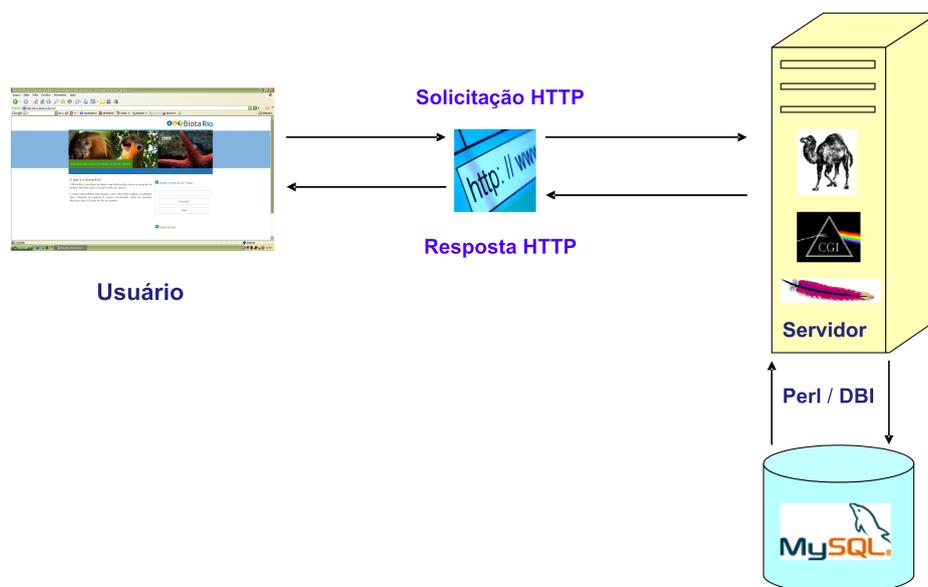


Figura 2.9: Arquitetura do sistema Biota-Rio.

2.4 Levantamento bibliográfico

Para recuperar a diversidade das espécies de animais descritas para o Estado do Rio de Janeiro, foram feitos levantamentos das referências dos artigos que contém as descrições originais usando a base de dados Zoological Record (ZR). O Zoological Record é uma publicação conjunta da Zoological Society of London e BIOSIS, sendo considerada a base mais completa de informação zoológica (Hughes, 2001). Disponível desde 1864, o ZR aborda hoje cerca de 6000 revistas científicas e outros veículos de informação nas mais diversas áreas da zoologia.

Cada livro do Zoological Record é dividido em cinco partes (Index), que são: *Author Index*, *Subject Index*, *Geographical Index*, *Paleontological Index* e *Systematic Index*.

O *Author Index* contém os nomes dos autores, o título do trabalho e as referências. O índice está organizado em ordem alfabética pelo sobrenome do autor principal e para cada trabalho há um número específico que o identifica.

No *Subject Index* os trabalhos estão separados por temas, tais como: biofísica, biometria, bioquímica, citologia, genética, histologia, metabolismo, morfologia, nutrição, reprodução, veterinária e outros temas. Portanto, caso o usuário esteja interessado em pesquisar os trabalhos por temas específicos, poderá buscar nesse índice.

Já no *Geographical Index* os trabalhos estão separados por regiões geográficas, tais como: região Australasia, Neártico, Neotropical, Palearctica, Ilhas Oceânicas, Oceano Antártico, Oceano Ártico, Oceano Atlântico, Oceano Índico, Oceano Pacífico e outras regiões. Cada uma dessas regiões agrupa um determinado número de países e dependendo do país, os trabalhos estarão agrupados por estados. Por exemplo, a região Neotropical inclui o Brasil e esse, por sua vez, está dividido nos estados. Essa organização permite, portanto, a busca dos trabalhos que estejam relacionados com o Estado do Rio de Janeiro.

O *Paleontological Index* está organizado por períodos geológicos permitindo, assim, a busca dos trabalhos sobre fósseis. No último índice, o *Systematic Index*,

os trabalhos estão divididos por grupo taxonômico. Nele o usuário pode encontrar novos nomes de gêneros, subgêneros, espécies e subespécies.

Para o levantamento das referências foram feitas consultas no *Geographical Index* e no *Systematic Index*, desde o ano de 1864 até o ano de 2006. O levantamento das referências para mamíferos foi completado e foi feito na Biblioteca Central do campus da Fundação Oswaldo Cruz, no Rio de Janeiro e na Biblioteca do Museu Nacional.

Capítulo 3

Resultados e Discussão

No capítulo anterior foi apresentado como o sistema Biota-Rio foi construído e como foi feito o levantamento bibliográfico referente às espécies de mamíferos descritas para o Estado do Rio de Janeiro. Neste próximo capítulo, serão apresentados os módulos de consulta e como as informações estarão disponíveis para a consulta.

As consultas ao sistema Biota-Rio estão centralizadas em seu web site (www.biota-rio.lncc.br). Na página principal do sistema, há uma caixa de texto para o usuário digitar o nome de um táxon, que pode ser o nome de uma espécie, subespécie ou ainda de um táxon supra-específico (táxon que pertence a alguma categoria acima de espécie). A Figura 3.1 mostra a página principal do sistema Biota-Rio e em destaque a caixa de texto para digitar a palavra-chave.



Figura 3.1: Página principal do sistema Biota-Rio e em destaque o campo para consulta.

3.1 Consultando a partir do nome de uma espécie ou subespécie

Caso o usuário digite o nome de uma espécie ou subespécie, uma outra janela, com informações sobre o organismo, será aberta. As informações são tabeladas na seguinte ordem:

- Rank taxonômico.
- Descrição original da espécie ou subespécie.
- Descrição alternativa.
- Diagnose.
- Número de tomo.
- Instituição de depósito do holótipo.
- Localidade-tipo.

- Ambiente.
- Descrição do ambiente.
- Status na Lista Vermelha da IUCN.
- Sinônimos.
- Referência principal.
- Outras referências.

Cada item relacionado acima é uma linha da tabela. A descrição de cada um dos itens se encontra no capítulo anterior. Na primeira linha da tabela se encontra o nome científico do organismo. A Figura 3.2 mostra a tabela com as informações do organismo.

Nome Científico: <i>Monodelphis theresa</i>	
Informações sobre a espécie	
Rank Taxonomico:	Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda; Amniota; Mammalia; Theria; Metatheria; Didelphimorphia; Didelphidae; Didelphinae; <i>Monodelphis</i> .
Descrição Original:	.
Descrição Alternativa:	.
Diagnose:	.
Número de Tombo:	Holotype: NHM 1921.8.6.2.
Inst. de Depósito:	Natural History Museum, London, United Kingdom.
Localidade-tipo:	Teresópolis.
Ambiente (Bioma):	Mata Atlântica.
Descrição do Ambiente:	.
Status na Lista Vermelha:	.
Sinônimos:	.
Ref. Principal:	Thomas, O.; A new short tailed Opossum from Brazil; Ann. Mag. Nat. Hist. London, (9), 8; pg. 441.
Outras Referências:	.

Figura 3.2: Tabela com as informações do organismo.

Na mesma página onde se encontra a tabela com as informações sobre o organismo, há uma caixa de texto onde o usuário poderá digitar o nome de um outro táxon e realizar uma nova consulta na base de dados. Também nessa mesma página há um link para uma outra página que mostra a relação de genes mitocondriais disponíveis no GenBank. O link se encontra ao lado da caixa de texto onde o usuário poderá fazer uma nova consulta na base de dados. A Figura 3.3 mostra a caixa de texto e o link.



Figura 3.3: Caixa de texto para uma nova consulta e o link para a página com a relação dos genes mitocondriais do organismo no GenBank.

A relação dos genes mitocondriais do organismo no GenBank, se encontra tabelada. Em cada linha da tabela estão o nome do gene e o link para a página com as informações do gene no GenBank. O nome do gene se encontra na coluna da esquerda e na coluna da direita o link é o AccessNumber do gene. A Figura 3.4 mostra um exemplo de como os dados estarão organizados na tabela.

Nome Científico: <i>Uroderma bilobatum</i>	
Genes mitocondriais no GenBank	
COI	EF080791
CytB	AF435172
NAD3	DQ312378

Figura 3.4: Tabela com a relação dos genes mitocondriais de um organismo. (A tabela acima é apenas um exemplo fictício, pois as informações mostradas ainda não estão cadastradas na base).

3.2 Consultando a partir do nome de um táxon supra-específico

Quando o usuário faz uma consulta usando o nome de um táxon que seja supra-específico, uma nova página se abre e nela se encontra uma árvore com os táxons que estão incluídos no táxon consultado. Cada táxon é um link, o que permite ao usuário navegar pela árvore. A Figura 3.5 mostra a árvore tomando o táxon Teleostomi como exemplo.

Linhagem: [Metazoa](#); [Eumetazoa](#); [Bilateria](#); [Coelomata](#); [Deuterostomia](#); [Chordata](#); [Craniata](#); [Vertebrata](#); [Gnathostomata](#).

Teleostomi

- ◊ **Euteleostomi**
 - **Actinopterygii**
 - **Actinopteri**
 - **Polypteriformes**
 - **Sarcopterygii**
 - **Coelacanthimorpha**
 - **Dipnoi**
 - **Tetrapoda**

Figura 3.5: Árvore de táxons.

A seqüência de figuras abaixo mostra as mudanças que ocorrem na árvore quando algum táxon é selecionado na árvore. Na Figura 3.5, ao ser selecionado o táxon Tetrapoda, uma nova página é aberta resultando na árvore da Figura 3.6.

Tetrapoda

- ◊ Amniota
 - Mammalia
 - Prototheria
 - Theria
 - Sauropsida
 - Sauria
 - Testudines
- ◊ Amphibia
 - Batrachia
 - Anura
 - Caudata
 - Gymnophiona
 - Caeciliidae
 - Ichthyophiidae
 - Rhinatremitidae
 - Scolecophoridae
 - Typhlonectidae
 - Uraeotyphlidae

Figura 3.6: Continuidade da árvore taxonômica.

Selecionando o táxon Mammalia na árvore da Figura 3.6, o resultado será uma árvore conforme a Figura 3.7.

Mammalia

- ◊ Prototheria
 - Monotremata
 - Ornithorhynchidae
 - Tachyglossidae
- ◊ Theria
 - Eutheria
 - Afrotheria
 - Euarctoglires
 - Laurasiatheria
 - Xenarthra
 - Metatheria
 - Dasyuromorphia
 - Didelphimorphia
 - Diprotodontia
 - Microbiotheria
 - Notoryctemorphia
 - Paucituberculata
 - Peramelemorphia

Figura 3.7: Árvore taxonômica dentro de Mammalia.

Selecionando o táxon *Didelphimorphia* na árvore da Figura 3.7, o resultado será a árvore mostrada na Figura 3.8.

Didelphimorphia

- ◊ Didelphidae
 - Caluromyinae
 - Caluromys
 - Caluromysiops
 - Glironia
 - Didelphinae
 - Chironectes
 - Cryptonanus
 - Didelphis
 - Gracilmanus
 - Lestodelphys
 - Lutreolina
 - Marmosa
 - Marmosops
 - Metachirus
 - Micoureus
 - Monodelphis
 - Phlander
 - Thylamys
 - Tlacuatzin
 - Hyladelphys

Figura 3.8: Árvore taxonômica dentro de Didelphimorphia.

Selecionando o táxon *Monodelphis* na árvore da Figura 3.8, o resultado será a árvore da Figura 3.9.

Monodelphis

- ◊ Monodelphis theresa

Figura 3.9: Árvore taxonômica para o gênero *Monodelphis*.

Por fim, selecionando a espécie *Monodelphis theresa* uma nova página será aberta com as informações da espécie, conforme descrito no item anterior (Figura 3.2).

Acima da árvore (ver Figura 3.5) encontra-se uma seqüência de táxons na forma de linhagem. Nessa seqüência, encontram-se todos os táxons que pertencem às categorias anteriores à categoria do táxon consultado. Os nomes dos táxons, assim como na árvore, são links. Isso permite ao usuário poder “caminhar” para os

táxons que são anteriores ao táxon consultado. Na Figura 3.5 se o táxon Craniata for selecionado, isso resulta na página mostrada na figura 3.10.

Linhagem: [Metazoa](#); [Eumetazoa](#); [Bilateria](#); [Coelomata](#); [Deuterostomia](#); [Chordata](#).

Craniata

- ◊ [Hyperotreti](#)
 - [Myxiniiformes](#)
 - [Myxinidae](#)
 - ◊ [Vertebrata](#)
 - [Gnathostomata](#)
 - [Chondrichthyes](#)
 - [Teleostomi](#)
 - [Hyperoartia](#)
 - [Petromyzontiformes](#)
-

Figura 3.10: Árvore taxonômica para o táxon Craniata.

Em alguns casos, táxons diferentes podem ter o mesmo nome. Quando isso ocorre, uma página informa ao usuário qual o táxon ele quer consultar. Para isso, basta ao usuário selecionar o táxon que aparece na forma de um link. A Figura 3.11 exemplifica isso.

[Craniata -> Chordata](#)
[Craniata -> Craniiformea](#)

Figura 3.11: Exemplo de táxons diferentes, mas que possuem o mesmo nome.

A palavra consultada foi Craniata, porém existem dois táxons diferentes que possuem esse mesmo nome. Craniata se refere a um táxon que é classe do subfilo Craniiformea (filo Brachiopoda), mas também é o nome de um táxon que é subfilo do filo Chordata.

3.3 Levantamento das referências

Como já foi colocado na Introdução dessa dissertação, foram feitos levantamentos das referências para as espécies de mamíferos. Essas referências são dos artigos onde se encontram as descrições originais das espécies de mamíferos descritas para o Estado do Rio de Janeiro. Os levantamentos foram feitos na base de dados Zoological Record. Ao todo foram levantadas trinta e seis espécies de mamíferos.

3.4 Discussão

Como já foi apresentado, o sistema Biota-Rio irá reunir um conjunto de informações sobre as espécies de animais descritas para o Estado do Rio de Janeiro. O acesso a informações como descrição original e localidade-tipo, implica na busca do artigo original. Porém, essa é uma busca bastante trabalhosa, pois muitas das descrições foram feitas no início do século XX, ou durante o século XIX, ou mesmo no século XVIII. Portanto, trata-se de material antigo. Para dificultar ainda mais, muitas das descrições foram publicadas em revistas de baixa circulação.

Um outro fator que também dificulta o acesso às informações referentes às espécies de animais, é o fato de que, em muitos casos, a descrição de uma nova espécie foi feita por pesquisadores estrangeiros. A consequência disso é que os holótipos dessas espécies se encontram depositadas em museus do exterior. Portanto, ter acesso a informações como: número de tombo e a instituição de depósito do holótipo, acaba sendo difícil.

Mesmo o levantamento das referências dos artigos que trazem as descrições originais, é um processo lento. Na cidade do Rio de Janeiro, por exemplo, apenas duas instituições possuem a base de dados Zoological Record: a biblioteca do Museu Nacional e a biblioteca da Fundação Oswaldo Cruz. A coleção do Museu Nacional, principalmente os volumes mais antigos, encontra-se incompleta. Foram observados intervalos de até dez anos entre um volume e outro. Já a coleção da Fundação Oswaldo Cruz encontra-se mais completa e em melhor estado de conser-

vação. Portanto, reunir esse conjunto de informações é um processo trabalhoso e que dispende tempo. Paralelo ao processo de levantamento das informações, foram feitas a modelagem e a implementação do sistema Biota-Rio.

O sistema além de reunir informações como as descrições originais, fotos, localidades-tipo e outras, também terá uma base de dados com os nomes dos táxons. Atualmente existem 1662 nomes de táxons, de espécie até reino, cadastrados na base. É por meio dessa base que se constrói a árvore de táxons mostrada no item 3.2 do presente capítulo.

Uma outra importante característica do sistema Biota-Rio, é o fato de possuir uma base de dados com os links para o GenBank dos genes mitocondriais das espécies de animais descritas para o Estado do Rio de Janeiro. Isso faz com que o sistema reúna não somente importantes informações para sistemática clássica, como também o acesso a informações utilizadas na sistemática molecular.

É importante salientar que atualmente não há, nem no Estado do Rio de Janeiro e nem mesmo no Brasil, um banco de dados que reúna todas as informações que o Biota-Rio reunirá.

Capítulo 4

Conclusão e Perspectivas

O uso de ferramentas computacionais no estudo de problemas biológicos tem crescido de forma muito acelerada. A bioinformática e a biologia computacional têm sido as responsáveis por esse rápido desenvolvimento. Com o advento da genômica, o desenvolvimento de sistemas de informação se tornou uma das grandes aplicações da tecnologia da informação na biologia.

No início da atual década, começaram a surgir os primeiros sistemas de informação desenvolvidos como ferramentas na organização e no livre acesso às informações sobre biodiversidade. Isso se deve à grande importância que essas informações têm para diferentes áreas do conhecimento.

Nesse sentido, o sistema Biota-Rio é uma ferramenta que vem contribuir em estudos de sistemática animal, e mesmo em outras áreas do conhecimento. Isso se torna ainda mais relevante quando se leva em conta o fato de a Mata Atlântica ser o principal bioma do Estado do Rio de Janeiro.

No que se refere às perspectivas em relação ao sistema Biota-Rio destacam-se:

- Implementação de novos módulos de consulta. Serão implementadas consultas por localidade-tipo, por nome comum, por autor, por gene e outros módulos.
- Implementação de módulos que irão automatizar o processo de inserção e atualização do dados.

- O uso da tecnologia XML na descrição das informações depositadas no banco de dados.
- Cadastramento das descrições originais das espécies de vertebrados e invertebrados.
- Integração de uma nova base de dados com as seqüências dos genes mitocondriais das espécies descritas para o Estado do Rio de Janeiro.
- Desenvolvimento de um tutorial em sistemática biológica com o objetivo de fomentar associações entre sistematas clássicos, inseguros para promover trabalhos em sistemática molecular, e biólogos moleculares que carecem de uma base teórica em sistemática quando publicam suas filogenias com dados moleculares.

Referências Bibliográficas

- D.S Amorim. **Fundamentos de Sistemática Filogenética**. Holos, 2 edição, 2002.
- D. Canhos, S. Souza, e V. P. Canhos. Diretrizes e estratégias para modernização de coleções biológicas e a consolidação de sistemas integrados de informação sobre biodiversidade, 2006. <http://www.cria.org.br/cgee/documentos/informacao.doc> (acessado em 28/01/2008).
- V. P. Canhos. Diretrizes e estratégias para modernização de coleções biológicas e a consolidação de sistemas integrados de informação sobre biodiversidade, 2006. <http://www.cria.org.br/cgee/documentos/estrategia.doc> (acessado em 28/01/2008).
- P. P. Chen. The entity-relationship model: Toward the unified view of data. **ACM Transactions on Database Systems**, 1(1):9–36, 1976.
- P. Cougo. **Modelagem Conceitual e projeto de banco de dados**. Campus, 1 edição, 1997.
- C. J. Date. **Introdução a Sistemas de Banco de Dados**. Campus, 8 edição, 2004.
- P. Dubois. **Cookbook MySQL**. O'Reilly, 1 edição, 2003.
- J. L. Edwards. Research and societal benefits of the global biodiversity information facility. **Bioscience**, 54:485–486, 2004.

- E.R. Elmasri e S. Navathe. **Sistemas de banco de dados: fundamentos e aplicações**. LTC, 3 edição, 2002.
- J. Hughes. Characterization of unique serials indexed in the zoological records., 2001. <http://www.istl.org/01-spring/refereed.htm> (acessado em 28/01/2008).
- C. A. Joly e E. Speglich. Programa biota/fapesp: Um novo paradigma no estudo da conservação e do uso sustentável da biodiversidade. **Ciência e Cultura**, 55 (3):41–43, 2003.
- T. Lewinsohn. **Avaliação do estado do conhecimento da biodiversidade brasileira**. Brasília: Ministério do Meio Ambiente, 1 edição, 2005.
- D. R. Maddison, K.S. Schulz, e W. P. Maddison. The tree of life web project. **Zootaxa**, 1668:19–40, 2007.
- A. Milani. **MySQL: Guia do programador**. Novatec, 1 edição, 2006.
- N. Myers, R.A. Mittermeier, C. G. Mittermeier, G. A. B. Fonseca, e J. Kent. Biodiversity hotspots for conservation priorities. **Nature**, 403:852–858, 2000.
- N. Papavero. **Fundamentos práticos de taxonomia zoológica**. UNESP, 2 edição, 1994.
- M. Ridley. **Evolução**. Artmed, 3 edição, 2006.
- A.S. Rocha-Júnior. Normalização de dados. **SQL Magazine**, 47:22–35, 2007.
- A. S. L. Rodrigues, J. D. Pilgrim, e J. F. Lamoreux. The value of the iucn red list for conservation. **Trends in Ecology and Evolution**, 20(20):1–6, 2005.
- A. Silberschatz, H. F. Korth, e S. Sudarshan. **Sistema de banco de dados**. Campus, 5 edição, 2006.
- T. Teorey, S. Lightstone, e S. Nadeau. **Projeto e Modelagem de banco de dados**. Campus, 4 edição, 2006.

- A. T. R. Vasconcelos, C. H. Castelletti, C.S. Caruso, F. Yokaichiya, G. R. G Armoa, G. S. P. Pereira, I. T. da Silva, Schrago C. E. G, A. L. V. Fernandes, E. R. da Silveira, A. G. Carneiro, C. J. M. Viana, D. Gramkow, e C. A. M. Russo. Mammibase: a mitochondrial genome database for mammalian phylogenetic studies. **Bioinformatics**, 21(10):2566–2567, 2005.
- C. Yesson, P. W. Brewer, T. Sutton, Caithness N, e J.S. Pahwa. How global is the global biodiversity information facility? **PLoS ONE**, 2(11):1–10, 2007.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)