

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

SAMANTHA DE CASTRO CAÑETE

**Interface de Gerenciamento e Consultas
Visuais em Banco de Dados de
Biodiversidade**

Dissertação apresentada como requisito parcial
para a obtenção do grau de
Mestre em Ciência da Computação

Profa. Dra. Carla Maria Dal Sasso Freitas
Orientador

Porto Alegre, abril de 2011

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Cañete, Samantha De Castro

Interface de Gerenciamento e Consultas Visuais em Banco de Dados de Biodiversidade / Samantha De Castro Cañete. – Porto Alegre: PPGC da UFRGS, 2011.

73 p.: il.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR-RS, 2011. Orientador: Carla Maria Dal Sasso Freitas.

1. Sistemas de informação de biodiversidade. 2. Visualização de informações. I. Freitas, Carla Maria Dal Sasso. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos Alexandre Netto

Vice-Reitor: Prof. Rui Vicente Oppermann

Pró-Reitor de Pós-Graduação: Prof. Aldo Bolten Lucion

Diretor do Instituto de Informática: Prof. Flávio Rech Wagner

Coordenador do PPGC: Prof. Álvaro Freitas Moreira

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

*“There is no way to happiness,
happiness is the way.”*
— BUDDHA

AGRADECIMENTOS

A professora Carla Freitas, que me guiou durante todo mestrado. Pela sua atenção, paciência, ensinamentos e críticas que tanto contribuíram para o meu crescimento pessoal. Foi um grande privilégio tê-la como orientadora. Aos professores do grupo de CG, em especial, João Comba, Luciana Nedel e Manuel Oliveira, pelos ensinamentos e críticas construtivas.

Aos biólogos do Laboratório de Evolução de Vertebrados, do Departamento de Genética da UFRGS, parceiros de projeto, que contribuíram com informações e sugestões no decorrer deste trabalho. Ao Rafael Henkin, Pedro Estrela, Thales Freitas e Renata Galante, pelo apoio.

A toda minha família, por sempre terem me oferecido todo o seu apoio e carinho. Agradeço em especial a minha "abuela", minha mãe e irmãos por todo amor, força e por sempre terem acreditado em mim. Ao meu Aurélio (Pablo) pelo incentivo, apoio, paciência e companheirismo durante todos estes anos.

Aos amigos e companheiros do laboratório de CG (Grupo de Computação Gráfica, Processamento de Imagens e Interação da UFRGS), que contribuíram direta ou indiretamente na realização deste trabalho. Em especial, ao Denison pela motivação, ajuda, paciência, conselhos e festas. A Lenna e todos amigos do lab pelas inúmeras ajudas durante todo mestrado. Admiro muito a todos e tenho muito orgulho de pertencer a este grupo.

Aos meus amigos que estão distantes, em especial a Rafa, Grasi, Bina e Bruno, que mesmo à distancia souberam me motivar e ofereceram sempre o melhor dos seus conselhos.

A todos meus amigos, que sempre me acolheram com carinho e compartilharam comigo momentos inesquecíveis. Por toda bondade, amizade e paciência, muito obrigada!.

Finalmente, agradeço ao CNPQ, pelo importante apoio financeiro na forma de bolsa e auxílio ao projeto de pesquisa ao qual a dissertação está vinculada.

SUMÁRIO

LISTA DE ABREVIATURAS E SIGLAS	9
LISTA DE FIGURAS	11
RESUMO	13
ABSTRACT	15
1 INTRODUÇÃO	17
1.1 Motivação	17
1.2 Objetivos e contribuições	18
1.3 Organização do texto	19
2 CONCEITOS BÁSICOS E TRABALHOS CORRELATOS	21
2.1 Sistemática e Árvore Taxonômica	21
2.2 Mecanismo para compartilhamento de informação de biodiversidade	23
2.3 Trabalhos Relacionados	25
2.4 Comentários finais	34
3 MODELO PROPOSTO PARA O ARMAZENAMENTO DE DADOS DE BIODIVERSIDADE	37
3.1 Visão Geral	37
3.1.1 Cenário de atividades de pesquisa dos biólogos	37
3.1.2 Cenário de uso dos usuários não cadastrados	39
3.2 Modelo de Dados da Coleção	39
3.3 Projeto do banco de dados	41
4 ARQUITETURA DO SISTEMA	47
4.1 Visão Geral do Sistema TaxonomyBrowser	47
4.2 Arquitetura da Aplicação	48
5 INTERFACE VISUAL	51
5.1 Interface administrativa	51
5.2 Interface de consulta	52
5.3 Interface sobre mapa	55
5.4 Interface de análise	58
5.4.1 Análise com sistema R	59
5.4.2 Visualizações alternativas	61
5.4.3 Exemplo de cenário de uso	64

6 CONCLUSÕES E TRABALHOS FUTUROS	69
REFERÊNCIAS	71

LISTA DE ABREVIATURAS E SIGLAS

ABCD	Access to Biological Collection Data
IABIN	Inter American Biodiversity Information Network
API	Application Programming Interface
CSS	Cascading Style Sheets
GBIF	Global Biodiversity Information Facility
GPL	General Public License
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
IPT	Integrated Publishing Toolkit
MaNIS	Mammal Networked Information System
MVC	Model View Controller
NGO	Non-Governmental Organization
ONG	Organização Não-Governamental
PHP	PHP Hypertext Processor
SGBD	Sistema de Gerência de Bancos de Dados
SI	Sistema Internacional de Medidas
SQL	Structured Query Language
SSTN	Species and Specimen Thematic Network
TDWG	Taxonomic Database Working Group
UMN	University of Minnesota
XHTML	Extensible HyperText Markup Language
XML	Extensible Markup Language
WeBios	Web Service Multimodal Tools for Biodiversity Research

LISTA DE FIGURAS

Figura 2.1	Árvore da vida	22
Figura 2.2	Exemplo de parte da árvore taxonômica das espécies. O binome é formado por gênero e espécie.	23
Figura 2.3	Atlas/Biota vinculado ao portal Sinbiota	25
Figura 2.4	Seleção da coleção a ser consultada no speciesLink	26
Figura 2.5	Resultado de uma consulta no specieslink	27
Figura 2.6	Resultados da consulta a registros com nome científico começando com A, numa coleção específica exibido em mapa no speciesLink	27
Figura 2.7	Resultado de outra consulta exibido com Google Maps	28
Figura 2.8	Arquitetura de sistema de informações de biodiversidade proposto no projeto WeBios	29
Figura 2.9	Resultado de busca por uma espécie no portal IABIN	30
Figura 2.10	Portal Species 2000: navegação na árvore taxonômica	30
Figura 2.11	Portal Species 2000: consulta a partir dos níveis taxonômicos	31
Figura 2.12	Tree of Life: exemplo de página descritiva de um grupo taxonômico	31
Figura 2.13	Portal AIAMAZONIA: descrição simplificada das espécies	32
Figura 2.14	Portal MANIS: arquitetura do sistema	33
Figura 2.15	Portal MANIS: consulta exibida sobre o mapa.	33
Figura 2.16	Portal ARCTOS: consulta a informações de coleção	34
Figura 2.17	Quadro comparativo dos trabalhos revisados	35
Figura 3.1	Diagrama de atividades de pesquisa dos usuários biólogos	38
Figura 3.2	Diagrama Entidade-Relacionamento para nodo taxonômico	39
Figura 3.3	Diagrama E-R do Modelo Implementado no TaxonomyBrowser	40
Figura 3.4	Tabelas. (a) Tabela de usuários do sistema. (b) Tabela de papéis do sistema.	41
Figura 3.5	Tabela de características.	42
Figura 3.6	Tabelas. (a) Tabela Grupo de características. (b) Tabela de Tipos de características.	42
Figura 3.7	Tabelas. (a) Tabela de unidades física. (b) Tabela de Quantidade de medidas suportadas.	43
Figura 3.8	Tabela de medidas das características.	43
Figura 3.9	Tabela que armazena os dados básicos dos espécimes coletados.	44
Figura 3.10	Tabelas que armazenam dados sobre os níveis taxonômicos (a) Armazena os nodos taxonômicos hierarquicamente. (b) Identifica os níveis taxonômicos.	44
Figura 3.11	Tabela que armazena os dados das publicações de referência.	45

Figura 3.12	Modelo relacional do banco de dados da aplicação, tabelas e ligações lógicas entre tabelas.	46
Figura 4.1	Arquitetura do Sistema TaxonomyBrowser. (A)Interface de Gerência. (B) Interface Visual. (C) Interface de Consulta Otimizada. (D) Base de Dados da Coleção.	47
Figura 4.2	Arquitetura MVC da aplicação.	48
Figura 5.1	Janela de autenticação de usuário no TaxonomyBrowser.	51
Figura 5.2	Layout padrão do sistema para todas as entidades.	52
Figura 5.3	Visualização de um nodo da árvore taxonômica.	53
Figura 5.4	Inserção de um exemplar da espécie <i>calomys venustus</i>	54
Figura 5.5	Navegação na coleção mostrando lista de exemplares e dados de um exemplar.	55
Figura 5.6	Diagrama de sequência de consulta.	56
Figura 5.7	Interface visual do editor de buscas.	57
Figura 5.8	Interface visual mostrando o resultado de busca sobre o mapa.	58
Figura 5.9	Visualização do resultado da busca sobre o mapa. (a) Com os espécimes agrupados. (b) Após selecionar um espécime específico.	59
Figura 5.10	Interface de análise mostrando a tabela resultado de uma busca que pode ser exportada para a execução de um script selecionado da lista apresentada ou para visualização em uma das formas também exibida nessa interface.	60
Figura 5.11	Resultado do processamento de script R.	61
Figura 5.12	Visualização utilizando coordenadas paralelas de uma amostra de <i>Ctenomys flamarioni</i>	62
Figura 5.13	Visualização utilizando <i>heat map</i> de uma amostra de <i>Ctenomys minutus</i>	63
Figura 5.14	Visualização utilizando gráfico de barras de uma mostra de <i>Ctenomys lami</i> . (a) Gráfico com três características sendo comparadas. (b) Gráfico com duas características.	63
Figura 5.15	Tabela com algumas características de uma amostra de <i>Ctenomys flamarioni</i>	64
Figura 5.16	Resultado da busca pelos <i>Ctenomys minutus</i> com número cromossômico de 46 a 48.	65
Figura 5.17	Coordenadas paralelas exibindo 114 exemplares de <i>Ctenomys minutus</i> com número cromossômico de 46 a 48.	66
Figura 5.18	Resultados das buscas por <i>Ctenomys minutus</i> com 46 (a) e 48 (b) cromossomos.	67
Figura 5.19	Resultados das buscas por <i>Ctenomys minutus</i> com 47 cromossomos.	68

RESUMO

Com a ampliação da quantidade de informação disponível, gerada pelas diversas atividades humanas, surgiu a necessidade de sistemas computacionais capazes de indexar e analisar estas informações. A manipulação de dados, a derivação de novas informações e a extração de características desses dados propiciam descoberta de padrões e geração de conhecimento, tornando tais sistemas uma importante ferramenta científica. Sistemas com essas funcionalidades são criados com as mais variadas finalidades.

No presente trabalho, focamos os sistemas para gerenciamento de informações na área da Biologia, mais especificamente, nos estudos da evolução de mamíferos. Tais sistemas podem ser enquadrados nos chamados sistemas de informação de biodiversidade. Tem-se verificado um número crescente de iniciativas direcionadas para o compartilhamento de informações de biodiversidade através da web. Tais sistemas visam disponibilizar, através da rede mundial de computadores, informações sobre espécimes catalogados em coleções nos mais variados locais, tais como museus de ciências naturais e instituições de pesquisa.

Este trabalho descreve os componentes principais de um sistema de informações de biodiversidade, cujo objetivo é integrar recursos de visualização e de análise de dados e funções convencionais de gerenciamento encontradas em bancos de dados. O sistema, denominado TaxonomyBrowser, visa auxiliar os biólogos no gerenciamento de informações a respeito de espécimes por eles coletados, assim como exemplares já catalogados em coleções de museus. A integração com visualização de mapas através da API do Google Maps e a utilização de ferramentas de análise de dados através do Google Charts e do sistema R amplia a aplicabilidade do sistema, bem como, facilita a interação com os usuários. A validação, assim como todo projeto do sistema, foi realizada com a ajuda de biólogos da UFRGS, responsáveis pela coleta e catalogação de espécimes de mamíferos.

Palavras-chave: Sistemas de informação de biodiversidade, visualização de informações.

Visual Interface for Managing and Querying a Biodiversity Data Base System

ABSTRACT

The increasing amount of information available due to the several, different human activities leads to the need for computing systems able to index and analyze these large data volume. Data manipulation, derivation of new data from the available ones and the extraction of characteristics from these data allow finding patterns and generating knowledge, turning such computing systems important scientific tools. Thus, systems with these features are created with a variety of purposes.

In this work we focus on systems for the management of biological data, more specifically on those supporting the study of mammals' evolution. Such systems are classified as biodiversity information systems. It has been verified an increasing number of approaches to share biodiversity information through the web. Such systems aim at providing, through the World Wide Web, information about catalogued specimens in collections at various locations, like natural sciences museums and research institutions.

This work describes the main components of a biodiversity information system developed to integrate visualization and data analysis tools and conventional data management functions provided by relational databases. The system, named TaxonomyBrowser, aims at helping biologists to manage the information about the specimens they collect in field-work, as well as specimens that have already been catalogued in museums' collections. The integration of map visualization through the Google Maps API and the application of Google Charts and R System as data analysis tools improve the system applicability, as well as facilitate user interaction. All phases of the system design were performed together with UFRGS biologists, in a participatory design experience, and currently they are carrying on a thorough validation of the system.

Keywords: Biodiversity information systems, information visualization.

1 INTRODUÇÃO

Sistemas computacionais para auxiliar o gerenciamento de grandes volumes de dados são frequentemente utilizados localmente, por um conjunto relativamente pequeno de usuários. Entretanto, tem-se observado a expansão desses sistemas na web, com o objetivo principal de agregar informações oriundas de diversos grupos de pesquisa ao redor do mundo, e permitir trabalho colaborativo e compartilhamento de dados.

Com essa ampliação da quantidade de informação disponível, tanto local como remotamente, nas mais variadas atividades humanas, criou-se a necessidade de sistemas capazes de indexar e analisar estas informações. A manipulação de dados, a derivação de novas informações e a extração de características desses dados propiciam descoberta de padrões e geração de conhecimento, tornando tais sistemas uma importante ferramenta científica. Sistemas com essas funcionalidades são criados com as mais variadas finalidades.

Neste trabalho, focamos sistemas que dão suporte aos dados que descrevem espécimes coletados por biólogos, os quais fazem parte do que se convencionou chamar de sistemas de informação de biodiversidade (SOBERÓN; PETERSON, 2004; GURALNICK; NEUFELD, 2005). Uma boa parcela desses sistemas é destinada a manter as informações sobre exemplares depositados em coleções de museus, de instituições ou grupos de pesquisa. Alguns desses sistemas estão disponíveis na web, mas, a grande maioria, contudo, é baseada em bancos de dados convencionais e é restrita a usuários locais, conforme já mencionado.

Tem-se verificado um número crescente de iniciativas direcionadas para o compartilhamento de informações de biodiversidade através da web, o que demanda algum tipo de padronização para facilitar a comunicação de dados entre diferentes sistemas. Para tal, o padrão mais utilizado é o Darwin Core (TDWG, 2010) através do protocolo DiGir. Um exemplo de iniciativa desse tipo é a IABIN (Rede Interamericana de informação sobre a Biodiversidade) (IABIN, 2010), que referencia (como um portal) um conjunto de bancos de dados localizados nos mais variados locais.

Constata-se, assim, que há interesse de grupos de pesquisa em criar conhecimento de forma distribuída, e existe a carência de um sistema com flexibilidade suficiente para ser usado por diferentes pesquisadores, com diferentes graus de conhecimento sobre os dados.

1.1 Motivação

O avanço do homem sobre o espaço, que antes era ocupado por espécies nativas, implica na diminuição e, algumas vezes, até na extinção de algumas espécies. Uma política ou ação que vise à conservação da biodiversidade requer dados de espécies que habitam a

região alvo. Este é um dos papéis fundamentais dos sistemas de informação de biodiversidade: conhecer para conservar.

Não é suficiente armazenar uma grande quantidade de dados. Também é preciso ter a capacidade de processá-los de forma eficaz e expor ao usuário especialista o resultado desse processamento de forma eficiente e compreensível. Além disso, tal sistema deve, também, fornecer um mecanismo de acesso às informações por usuários que não fazem parte do meio científico, possibilitando a disseminação da informação.

Frequentemente, também, mais do que conhecer para decidir como e porquê conservar uma determinada espécie, é a própria evolução de um conjunto de espécies que motiva o estudo sobre um conjunto de dados. O estudo pode abranger desde várias medidas morfológicas e características anatômicas, até sequências de DNA obtidas de exemplares coletados. Além disso, durante o processo de análise de dados, novas informações são geradas, as quais podem vir a ser importantes para pesquisas futuras sendo, portanto, necessário armazená-las.

Um usuário típico de coleções de dados de espécimes realiza muito esforço manual para coletar, catalogar e armazenar as informações que serão usadas em estudos futuros. Primeiramente, ele cataloga informações básicas sobre exemplares coletados. Uma vez catalogadas as informações, estas podem ser recuperadas em mapas ou textualmente. Esses dados, então, podem ser processados de acordo com o objetivo do estudo. Os resultados gerados em laboratório ou computacionalmente revelam novas informações, as quais podem ser agregadas à base de dados. Todas essas atividades são realizadas com o auxílio das mais diversas ferramentas, com pouca ou nenhuma integração automática, cabendo aos usuários realizar o fluxo de dados entre elas. Neste contexto, surge a necessidade de desenvolver sistemas específicos, com o propósito de auxiliar pesquisadores a aprimorar ou complementar seu conhecimento e entendimento sobre os seres vivos, suas interações e seu habitat.

1.2 Objetivos e contribuições

O objetivo desse trabalho é o desenvolvimento de um sistema de fácil acesso, flexível e com diversas ferramentas para agregação de informações e análise das mesmas, a fim de facilitar o estudo de conjuntos de dados sob diferentes pontos de vista. Apesar deste trabalho inicial estar focado em exploração de dados sobre espécies animais, o sistema desenvolvido é flexível e pode ser usado para outros grupos de seres vivos.

O sistema descrito neste trabalho, nomeado *TaxonomyBrowser*, tem como objetivo o armazenamento de dados de coleções de espécimes animais, além de prover ferramentas para consulta e análise dos dados armazenados. O nome é derivado do fato de ser a classificação taxonômica (FUTUYMA, 2009) dos espécimes armazenados a informação primordial a eles associada e esta classificação ser a forma básica de acesso aos dados dos elementos da coleção. O *TaxonomyBrowser* provê uma interface de análise de dados que possibilita o acesso tanto de usuários experientes, quanto de usuários leigos. Isso é feito através de ferramentas visuais de consulta e exibição de dados. Com o intuito de facilitar o entendimento dos dados, é possível navegar nos registros do banco de dados através da organização provida pela árvore taxonômica, visualizar os dados através de Google Maps (GOOGLE, 2010a), ou através de gráficos da API Google Charts (GOOGLE, 2010b), ou ainda analisá-los com o sistema R (R-PROJECT, 2000).

As principais características e funcionalidades do *TaxonomyBrowser* são:

- armazenamento e navegação na coleção de dados de acordo com a árvore taxonô-

mica

- definição dos níveis da árvore taxonômica de acordo com as necessidades do usuário
- definição das características (e suas medidas) dos espécimes por nível taxonômico
- busca por espécimes através de consultas por interface visual
- representação em mapas do resultado da consulta
- análise de resultados de busca através do sistema R
- análise gráfica dos resultados de busca ou análise.

O TaxonomyBrowser visa ter impactos positivos para disponibilidade de dados de coleções de espécimes, beneficiando a geração, visualização, qualificação e disseminação de dados para pesquisa em biodiversidade. Para a especificação do modelo de dados, contou-se com a participação de biólogos parceiros do projeto, e a validação do sistema está sendo feita por eles, usando dados reais.

O resultado obtido neste trabalho é o sistema aqui descrito, já disponível no endereço eletrônico <http://darwin.inf.ufrgs.br/taxonomybrowser/>. Sua utilização já vem auxiliando biólogos do Departamento de Genética da UFRGS, no tratamento e visualização de dados sobre espécimes coletadas no Rio Grande do Sul e no Centro-Oeste do Brasil, e da FIOCRUZ, no armazenamento de dados de roedores. O sistema também já teve uma primeira descrição publicada (CAÑETE et al., 2010).

1.3 Organização do texto

O trabalho está organizado da forma como segue. O capítulo 2 apresenta conceitos sobre biodiversidade, juntamente com os trabalhos correlatos visando traçar um comparativo entre os sistemas já estabelecidos e o aqui proposto. No capítulo 3 é descrito o modelo e a base de dados, definidos a partir das necessidades dos biólogos envolvidos no projeto. O capítulo 4 apresenta a arquitetura do TaxonomyBrowser, enquanto o capítulo 5 aborda as interfaces de consulta e análise, através de exemplos. E, finalmente, o capítulo 6 conclui e elenca os próximos desenvolvimentos.

2 CONCEITOS BÁSICOS E TRABALHOS CORRELATOS

Este capítulo apresenta conceitos básicos relacionados à biodiversidade e são descritos os trabalhos e projetos mais relacionados com o foco deste trabalho.

2.1 Sistemática e Árvore Taxonômica

A Sistemática é o campo da Biologia que se dedica ao estudo do relacionamento entre as espécies, ou seja, à compreensão da chamada "árvore da vida"(Figura 2.1). Duas sub-áreas abrigam as tarefas associadas à sistemática: a Taxonomia e a Análise Filogenética (FUTUYMA, 2009).

Enquanto a Taxonomia corresponde à classificação e à descrição, e à consequente denominação dos organismos, a Análise Filogenética preocupa-se com os relacionamentos de ancestralidade, ou seja, com as relações de origem entre os diferentes organismos. Observando a Figura 2.1, vemos a localização dos "Animais" como um ramo dessa árvore. A análise filogenética estuda justamente a formação desses ramos de descendência, cujos resultados, em termos de características, acabam se refletindo na classificação dos organismos, ou seja, na taxonomia.

Dentro de cada uma das folhas dessa árvore existe, na realidade, uma outra árvore que indica como os organismos evoluíram do ponto de vista biológico, ou seja, como foram se diferenciando ao longo de milhões de anos até atingirem sua forma atual, com a classificação que conhecemos.

O estudo da sistemática engloba etapas de coleta, preservação e estudo de espécimes. A nomenclatura científica, atribuída aos organismos e às categorias nas quais são classificados, provê um padrão aceito internacionalmente, e cada nome aplica-se apenas a uma espécie.

As características em comum entre os diversos organismos vivos dão origem a agrupamentos taxonômicos. Essa classificação já foi baseada no tipo de locomoção dos animais, método substituído pela nomenclatura binomial, a qual foi primeiramente proposta pelo naturalista suíço Gaspard Bauhin, no século XVII, e formalizada por Carl Linnaeus, naturalista sueco, no século seguinte. Essa nomenclatura é constituída por binomes, formados a partir de identificação de um gênero e uma espécie, permitindo a rápida e única identificação de cada grupo de espécimes. A partir desse conceito básico, são formadas hierarquias de classificações, onde os níveis mais altos englobam as várias classificações dos níveis inferiores, agregando informações de forma incremental, com diferentes níveis de detalhamento, o que facilita a pesquisa e análise dos dados.

À medida que os estudos se aprofundaram, essa classificação passou a englobar as relações evolutivas entre os organismos (a análise filogenética). Neste texto, damos particular ênfase à taxonomia, pois ela está na base da organização o sistema desenvolvido.

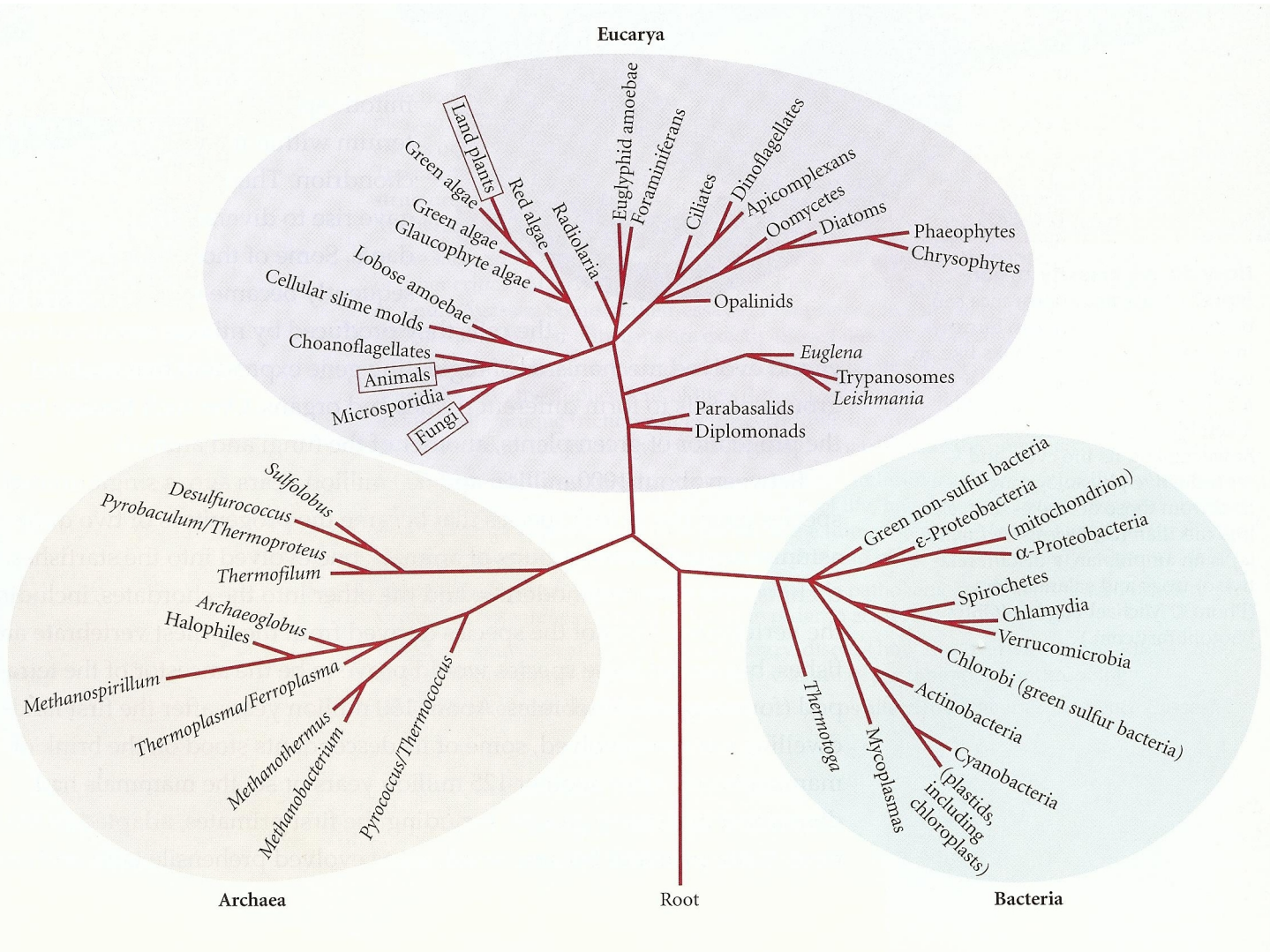


Figura 2.1: Árvore da vida (FUTUYMA, 2009).

Linnaeus iniciou seu trabalho de classificação com apenas alguns níveis hierárquicos básicos. Entretanto, com a descoberta de novas informações sobre os espécimes, níveis intermediários tiveram de ser implementados (DAYRAT, 2005). Modificações na classificação taxonômica ocorrem até hoje, a cada nova descoberta que justifique uma alteração. É possível a ocorrência da troca de nome de determinadas classificações, modificações nas relações parentais entre os nodos da árvore, criação de novos nodos ao surgirem novas classificações.

A representação de uma parte da árvore taxonômica pode ser vista na Figura 2.2. Um caminho específico dentro dessa árvore gera uma classificação científica de um espécime. Por exemplo, a classificação científica de um roedor conhecido como tuco-tuco, comumente encontrado nas planícies da América do Sul, é dada por: 1- *Animalia* (reino), 2- *Chordata* (filo), 3- *Mammalia* (classe), 4- *Rodentia* (ordem), 5- *Hystricomorpha* (subordem), 6- *Ctenomyidae* (família), 7- *Ctenomys* (gênero), e *Ctenomys minutus* (espécie).

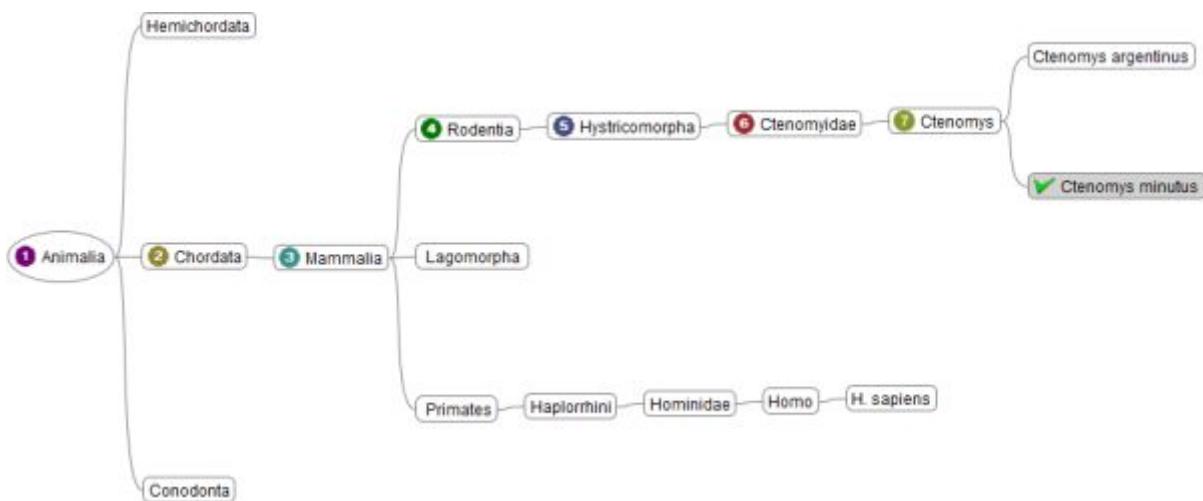


Figura 2.2: Exemplo de parte da árvore taxonômica das espécies. O binome é formado por gênero e espécie.

Dessa forma, através da árvore taxonômica, é possível visualizar a classificação de um espécime de forma hierárquica. É importante ressaltar que cada um dos nodos da árvore representa um conjunto de espécimes com uma série de atributos semelhantes. Por exemplo, todos os espécimes pertencentes a uma determinada classe possuirão atributos em comum (talvez não presentes em outras classes). Tudo que pertencer a essa classe (ordens, famílias, gêneros, espécies, etc) possuirá as mesmas características da classe por herança, além de possuir as suas características próprias.

Assim, ao analisar um caminho na árvore taxonômica, partindo da raiz e chegando a um nodo folha, temos um conjunto de atributos que caracterizam aquele nodo de forma única. Ele pode possuir atributos próprios e também outros herdados dos níveis superiores. O tratamento dado neste trabalho à árvore taxonômica será visto com mais detalhes no capítulo 3.

2.2 Mecanismo para compartilhamento de informação de biodiversidade

Com o objetivo de garantir a interoperabilidade na hora de consultar diferentes bases de dados de biodiversidade espalhadas por todo o mundo é necessário contar com

Tabela 2.1: Campos obrigatórios do Darwin Core.

Campo	Descrição
ScientificName	Táxon de mais baixo nível no qual o organismo foi identificado.
InstitutionCode	Um código que identifica a instituição ou centro ao qual a coleção pertence.
CollectionCode	Um valor alfanumérico único que identifica a coleção dentro da instituição.
DateLastModified	A data e a hora da última modificação do registro.
CatalogNumber	Um texto alfanumérico único que identifica um registro individual dentro de uma coleção.

um conjunto de informações padronizadas para as bases de dados de coleções museus e instituições de pesquisa e observações e coletas no campo. Isto pode ser obtido com a utilização de metadados, ou seja, é um conjunto de atributos ou elementos, utilizados para descrever uma fonte de dados (MATTHEW et al., 2001).

Existem vários padrões de metadados que são usados para classificar informação na web. Entre eles, se encontra o padrão Darwin Core (TDWG, 2010), que descreve o grupo mínimo de dados padrões para busca e recuperação das bases de dados de biodiversidade. Darwin Core apenas inclui um conjunto básico de dados, provavelmente disponível para a maioria dos registros sobre espécimes. É um padrão aprovado e usado pela Agência de Informação da Biodiversidade Global GBIF (Global Biodiversity Information Facility), para dados de coleções de espécimes.

O Darwin Core funciona como um glossário de termos que fornece definições semânticas estáveis, de forma que possam ser reutilizadas em inúmeros contextos. O Darwin Core serve como base para a construção de formatos mais complexos de intercâmbio de informação, garantindo a interoperabilidade de um conjunto comum de termos. Os arquivos nesse padrão dispõem os dados em um arranjo lógico com uma geometria semelhante à de uma estrela, em que cada arquivo-núcleo está cercado por um grande número de extensões. Cada registro de extensão remete para um arquivo-núcleo. Isso permite compartilhar conjuntos de dados inteiros.

Darwin Core conta com uma série de campos que permitem descrever os mais diversos espécimes, dos quais na tabela 2.1, são apresentados os cinco campos obrigatórios.

Embora o Darwin Core seja insuficiente para as necessidades de todas as disciplinas biológicas, pode ser adaptado para servir a novos propósitos. O núcleo de Darwin pode ser estendido pela adição de novos termos para compartilhar informação adicional. Para fazer isso é necessário estar familiarizado com as recomendações e os procedimentos definidos no núcleo de Darwin Core Namespace Policy. Basicamente, antes de propor um novo termo, deve-se considerar as condições existentes neste e em outros padrões compatíveis para determinar se o novo conceito pode ser acomodado por uma simples revisão da descrição e comentários de um termo existente, sem perder o sentido atual do

termo.

2.3 Trabalhos Relacionados

Dentre os sistemas de gerenciamento de informações de biodiversidade pesquisados no decorrer do desenvolvimento deste trabalho, encontramos um conjunto de portais, que tem como objetivo, em sua maioria, disponibilizar informações sobre espécimes coletados ou registrados em museus ou instituições de pesquisa e monitoramento. Há iniciativas de âmbito mundial assim como de âmbito nacional, das quais destacamos as iniciativas no Brasil.

Os sistemas brasileiros que destacamos são: Sinbiota (Sistema de Informação Ambiental do Biota) (SINBIOTA, 2010), WeBios (Web Service Multimodal Tools for Biodiversity Research, Assessment and Monitoring) (WEBIOS, 2010) e BioCORE (Biodiversity and Computing Research) (BIOCORE, 2010).

O sistema Sinbiota visa o gerenciamento de dados de biodiversidade no estado de São Paulo. Foi criado com o objetivo de sistematizar as informações de coleta de dados a respeito da biodiversidade do estado de São Paulo, de tal forma a integrar as informações geradas pelos pesquisadores vinculados ao Programa Biota/Fapesp, e relacioná-las a uma base cartográfica digital (SINBIOTA, 2010). Foi desenvolvido o Atlas/Biota, um portal onde é possível consultar dados de registros de espécimes coletados. Os resultados dessas consultas podem ser exibidos em um mapa do estado de São Paulo. O Atlas/Biota conta com várias camadas que podem ser exibidas sobre o mapa, como podemos ver na Figura 2.3, onde a camada correspondente a "Registros do speciesLink no estado de São Paulo" foi exibido sobre o mapa. É possível interação com o mapa no sentido de selecionar uma área e buscar os espécimes coletados nessa área. É possível, também, realizar consultas simples para plotagem no mapa das coletas relacionadas a um nome científico, autor, município, grupo taxonômico, identificador ou ecossistema.

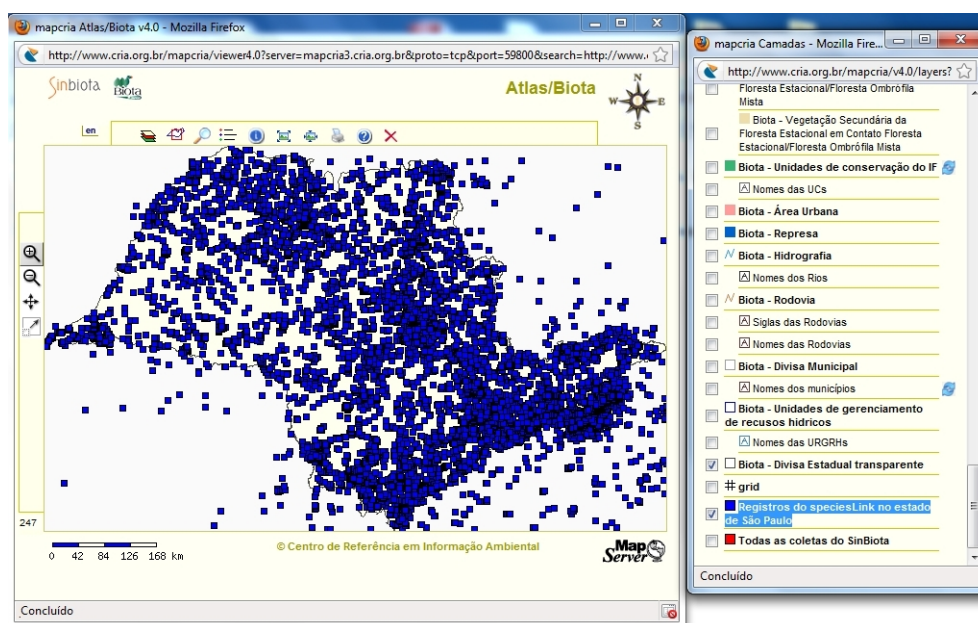


Figura 2.3: Atlas/Biota vinculado ao portal Sinbiota

O sistema SpeciesLink (SPECIESLINK, 2010) é um sistema distribuído de informação que integra, em tempo real, dados primários de coleções científicas. A arquitetura

da rede speciesLink permite que cada coleção mantenha o seu próprio sistema (software, sistema operacional, rotina) e o total controle sobre os dados a serem compartilhados, permitindo filtrar dados sensíveis, atualizar, modificar ou até retirar os dados disponíveis em rede. Esta arquitetura de rede viabilizou o desenvolvimento de um sistema onde as coleções associadas têm total autonomia e controle para reter ou compartilhar os seus dados. Permite, portanto, que cada instituição pratique a sua própria política de acesso a dados. Além disso, foram desenvolvidas várias ferramentas para auxiliar na melhoria da qualidade dos dados, na avaliação do perfil dos acervos das coleções associadas à rede e indicadores, além de ferramentas de busca, recuperação e exibição dos dados em formato de texto, integrados a serviços de georreferenciamento automático e visualização em bases cartográficas. As consultas podem ser feitas através da utilização de 3 passos: primeiro, a escolha de coleção (Figura 2.4); logo a aplicação de alguns filtros (Figura 2.5); e, finalmente, a exibição de resultados referentes a algumas coleções disponibilizadas pelo sistema. O resultado pode ser visualizado através do sistema de mapas como ilustrado na Figura 2.6 e em alguns casos também podem ser visualizados no sistema de mapas do Google Maps como podemos ver na Figura 2.7.

The image shows the 'speciesLink' search interface. At the top, there are logos for 'speciesLink' and several institutions. Below the logos, there is a navigation bar with 'Busca simples' and '1 | Coleções'. A yellow box highlights the search options section, which includes a header: 'Use as opções abaixo para selecionar as coleções a serem pesquisadas'. The options are organized into four categories: 'Tipo de Acervo', 'Localização da coleção', 'Rede', and 'Fonte dos dados'. Each category has a dropdown menu. To the right of these categories is a list of collection names with checkboxes next to them.

speciesLink
english | ajuda

Busca simples **1 | Coleções**

Use as opções abaixo para selecionar as coleções a serem pesquisadas

Tipo de Acervo

- Animais
- Vertebrados
- Mamíferos

Localização da coleção

- América do Sul
- Brasil
- todos

Rede

- Brasil
- todos

Fonte dos dados

- Voucher

CM - Coleção de Mamíferos da UNEMAT - Campus Nova Xavantina
 CRMRO - Coleção de Referência da Mastofauna de Rondônia
 DZSJRP-Chiroptera - Coleção de Quirópteros DZSJRP
 DZUP-Mammalia - Coleção Mastozoológica DZUP
 IAL-roedores - Coleção Zoológica de Referência da Seção de Vírus Transmitidos por Artrópodos
 INPA-Mamíferos-tipos - Coleção de tipos de Mamíferos INPA
 MBML-Mamíferos - Coleção de Mamíferos
 MHNCI-Mamíferos - Coleção de Mamíferos MHNCI
 MZUEL-Mamíferos - Coleção de Mamíferos do Museu de Zoologia
 UFES-MAM - Coleção de Mamíferos da Universidade Federal do Espírito Santo
 ZEE_MAM - Zoneamento Ecológico Econômico do Acre - Mammalia
 ZUEC-MAM - Coleção de Mamíferos do Museu de Zoologia da UNICAMP

Figura 2.4: Seleção da coleção a ser consultada no speciesLink

No projeto WeBios (WEBIOS, 2010) foi proposto um sistema que permite a especificação de consultas multimodais sobre fontes heterogêneas de dados de biodiversidade. Alguns exemplos desses dados são fotos, dados geográficos, ontologias e metadados específicos de domínio. Seu objetivo é fornecer um mecanismo unificado de consultas sobre diversos bancos de dados, para pesquisadores que trabalham com questões ambientais e de biodiversidade. A ideia é que as diversas fontes de dados sejam acessadas através de

The screenshot shows the specieslink web interface. At the top, there are logos for various institutions including UNICAMP, ETDA, and OBIS. Below the logos, there are navigation tabs: 'Busca simples', 'Coleções', 'Filtros', and 'Resultados'. The main content area displays search results for two collections: 'BUTIP-Mammalia' and 'ZUEC-MAM'. A table summarizes the results, including the number of records and georeferenced records for each collection. Below the table, there is a detailed explanation of the georeferencing process and a note about 'suspeitos' (suspect records). At the bottom, there are links for 'Indicadores', 'data cleaning', 'networkManager', and 'Busca avançada'.

Coleção	Georeferenciados		Conteúdo	Formato	Saída
	Registros Original	Automático			
BUTIP-Mammalia	2	2	0 Resumido	HTML	ver
ZUEC-MAM	2	2	0 Resumido	HTML	ver
Total	4	4	0 Resumido	HTML	ver

Figura 2.5: Resultado de uma consulta no specieslink

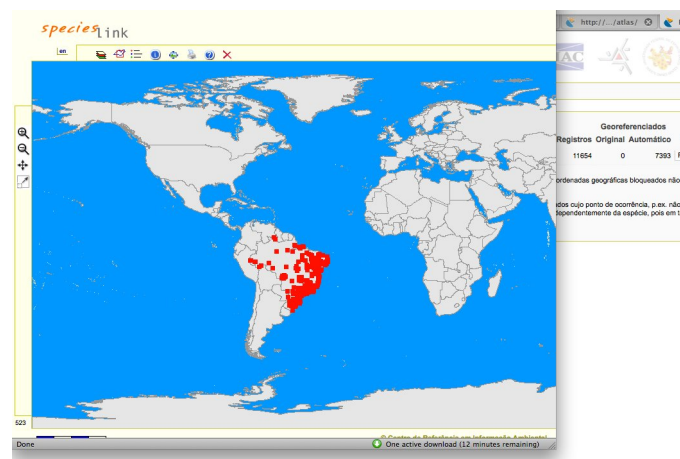


Figura 2.6: Resultados da consulta a registros com nome científico começando com A, numa coleção específica exibido em mapa no speciesLink

um portal único na WEB, enquanto a formulação das consultas, o pré-processamento e a visualização dos dados resultantes sejam executados como uma aplicação cliente remota (Figura 2.8). Desta forma, os pesquisadores podem realizar um trabalho exploratório extensivo, o qual considera informações oriundas de diferentes fontes. Com isso, é capaz de prover dados diferenciados sobre espécies, suas interações, seus habitats e correlações entre eles. O principal diferencial do Webios, em relação aos demais sistemas de biodiversidade é permitir a combinação de predicados baseados em conteúdos de imagem, predicados espaciais e textuais em uma mesma consulta (TORRES, 2007). Os sistemas de biodiversidade disponíveis não atacam essas questões. O projeto WeBios evoluiu para o projeto BioCore (BIOCORE, 2010), descrito a seguir.

Algumas das ferramentas resultantes do projeto WeBios foram um serviço de ontologia ((DALCIO, 2007), (DALCIO et al., 2008)) e um serviço ecologicamente ciente (GOMES, 2007) baseado também em ontologia e repositórios distribuídos. Basicamente, o sistema BioCORE possui a mesma arquitetura do WeBios (Figura 2.8), mas, além de permitir gerenciar dados de coletas na natureza, também incorpora informações do Museu de Zoologia da UNICAMP (MALAVERRI; VILAR; MEDEIROS, 2009). Uma característica que se destaca no projeto BioCORE é a possibilidade de usá-lo como um ambiente para ensino e compartilhamento de informações entre pesquisadores. Para permitir o



Figura 2.7: Resultado de outra consulta exibido com Google Maps

acesso e a recuperação de informação de forma transparente, o banco de dados foi encapsulado em um serviço web, combinando registros de ocorrências de espécies e registros de catálogos (MALAVERRI, 2009).

Dentre os sistemas de biodiversidade propostos no Brasil, apenas o speciesLink e o SinBiota compartilham dados com sistemas de mesma finalidade, de âmbito mundial. Em particular, esses dois sistemas trocam informações com o Global Biodiversity Information Facility (GBIF), comentado a seguir.

O GBIF é um portal agregador de dados, concebido a partir de uma iniciativa multinacional. Seu objetivo é disponibilizar informações sobre biodiversidade por meio de uma rede global distribuída de forma gratuita (GBIF, 2010). Os participantes devem fornecer os dados nos formatos esperados pelo sistema, os quais são o Darwin Core ou o ABCD (Access to Biological Collection Data) (ABCD, 2010), que são padrões para compartilhamento de coleções completas de dados biológicos. Um software chamado Integrated Publishing Toolkit (IPT, 2010) foi desenvolvido com o objetivo de facilitar a publicação de dados pelos participantes e é oferecido gratuitamente às instituições registradas. Os dados contidos no portal de dados do GBIF incluem registro de espécimes coletados e registrados em coleções. Entretanto, como o número de fontes varia bastante, a disponibilidade dos dados também varia. As consultas referentes a espécies e espécimes são feitas através do portal IABIN, o qual será descrito a seguir.

IABIN - Inter American Biodiversity Information Network (Rede Interamericana de informação sobre a Biodiversidade) (IABIN, 2010) é um fórum que propicia a colaboração técnica e coordenação entre os países das Américas para coletar, compartilhar e utilizar informações sobre biodiversidade que sejam relevantes para os processos de tomada de decisões sobre conservação e gerenciamento da biodiversidade, assim como para a educação nas diferentes regiões. Os principais objetivos desta iniciativa são: construir

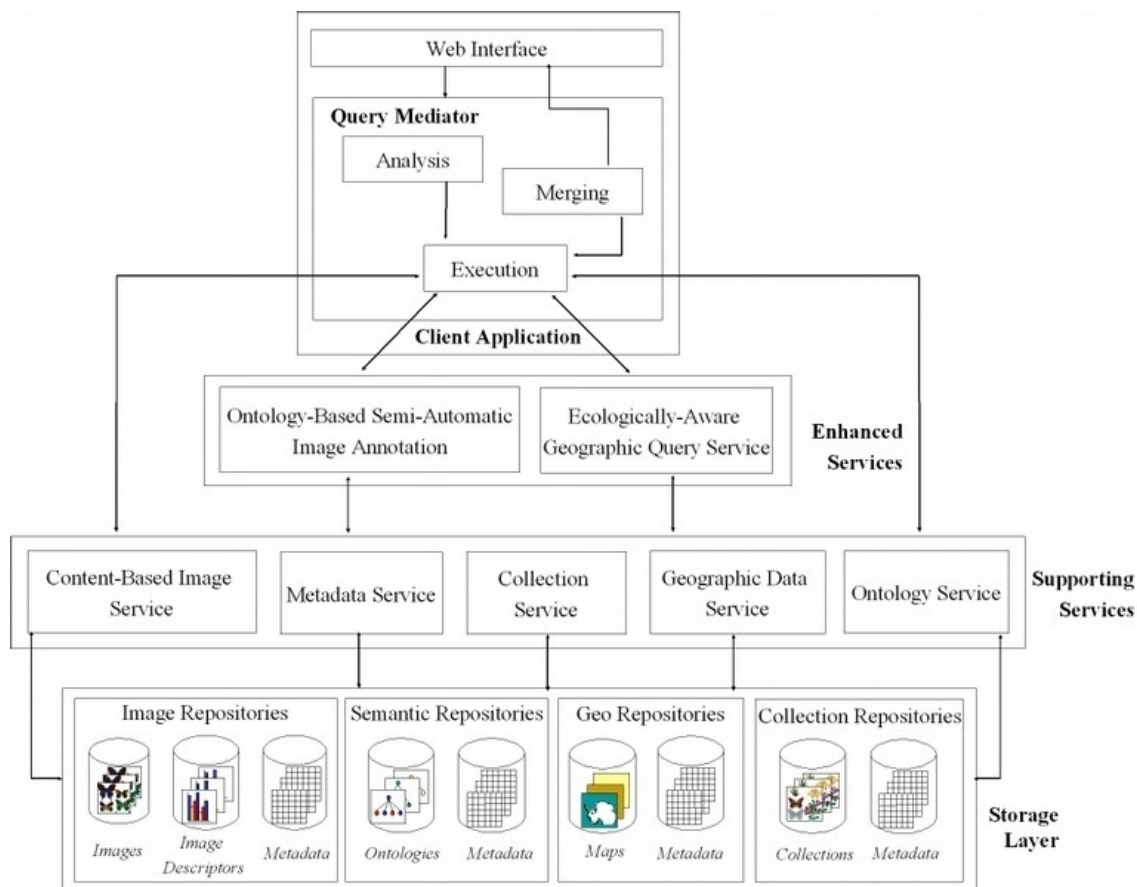


Figura 2.8: Arquitetura de sistema de informações de biodiversidade proposto no projeto WeBios (WEBIOS, 2010)

uma infra-estrutura para o intercâmbio de informações sobre a biodiversidade; fortalecer a capacidade técnica para trocar informações entre os países das Américas sem limites de fronteira políticas, lingüísticas e institucionais; prover acesso a informações que sejam úteis para as tomadas de decisões a fim de melhorar os esforços de conservação da biodiversidade; e melhorar o armazenamento, a utilização e distribuição de informações sobre biodiversidade que sejam cientificamente confiáveis e atualizadas.

A IABIN conta com 6 redes temáticas, sendo uma delas a SSTN - Species and Specimen Thematic Network (Rede temática de espécies e espécimes) (SSTN, 2010) cujo objetivo geral é facilitar a integração e divulgação de dados na forma digital sobre espécies e espécimes que formam a biodiversidade das Américas. Como podemos ver na Figura 2.9, ao se realizar uma consulta por um determinado grupo taxonômico (no exemplo, uma espécie de morcego), o portal retorna informações sobre a espécie. Mas, também é possível realizar buscas específicas por espécimes de coleções de diversos bancos de dados.

O portal IABIN é uma das instituições que utiliza o sistema GeoSur, que oferece acesso a dados e mapas da América Latina. Entretanto, o serviço do GeoSur não está disponível para ser utilizado em outro navegador que não seja o Internet Explorer, diferente dos mapas da Google, por exemplo, que podem ser acessados de qualquer navegador.

Assim como os anteriores, o projeto Species 2000 (SPECIES2000, 2010) é um programa que reúne diversas instituições do mundo e possui como objetivo a criação de um catálogo de todas as espécies de animais, plantas, fungos e microrganismos. Cada instituição executa um conjunto de softwares denominado SPICE (XU et al., 2001), fornecido

The screenshot shows the IABIN website interface. At the top, there is a header with the IABIN logo and the text "Red Interamericana de Información sobre Biodiversidad". Below the header, there is a navigation bar with links for Home, Browse by Vernacular Names, Advanced Search, Datasets, and About. The main content area is titled "Species record details: Thyroptera discifera (INBio Especies - INBio: Especies de Costa Rica)". It features a table of contents on the left, a central image of the species, and detailed information on the right. The taxonomic classification is listed as follows:

- Scientific name: *Thyroptera discifera* (Lichtenstein and Peters, 1855) 1855 (Interpreted as *Thyroptera discifera*)
- Kingdom: Animalia
- Phylum: Chordata
- Class: Mammalia
- Order: Chiroptera
- Family: Thyropteridae
- Genus: *Thyroptera*
- Species Publication Reference: Moneteb. K. Preuss. Akad. Wiss. Berlin, 1855: 335
- Common Names: Murci@lago
- Typification:

Figura 2.9: Resultado de busca por uma espécie no portal IABIN

pelo próprio Species 2000, os quais permitem que o servidor central do Species 2000 colete os dados. Desta forma, mesmo que a origem dos dados seja diferente, os dados coletados estão no formato correto. As consultas podem ser feitas por nomes, científicos ou comuns, e o usuário também pode navegar na árvore taxonômica como podemos ver na Figura 2.10 e realizar consultas através da classificação taxonômica, como ilustrado na Figura 2.11.

The screenshot shows the Species 2000 Catalogue of Life website. At the top, there is a header with the Species 2000 logo and the text "Catalogue of Life, 2nd October 2010 indexing the world's known species". Below the header, there is a navigation bar with links for Browse, Search, and Info. The main content area is titled "Browse taxonomic tree" and displays a list of taxonomic levels:

- Kingdom: Animalia
 - Phylum Acanthocephala
 - Phylum Acoelomorpha
 - Phylum Annelida
 - Phylum Arthropoda
 - Phylum Brachiopoda
 - Phylum Bryozoa
 - Phylum Cephalorhyncha
 - Phylum Chaetognatha
 - Phylum Chordata
 - Class Actinopterygii
 - Class Amphibia
 - Class Appendicularia
 - Class Ascidiacea
 - Class Aves
 - Class Cephalaspidomorphi
 - Class Cephalochordata
 - Class Elasmobranchii
 - Class Holocephali

Figura 2.10: Portal Species 2000: navegação na árvore taxonômica

Com caráter mais informativo, educacional e de divulgação, o portal Tree of Life (LIFE-PROJECT, 2010) é uma coleção de informações sobre biodiversidade compilada colaborativamente por profissionais e estudantes da área. O objetivo é fornecer um conjunto de páginas web com imagens, vídeos, textos e outras informações para cada espécie e grupo de organismos vivos ou extintos. Conexões entre páginas web do Tree of Life seguem links filogenéticos entre grupos de organismos. Assim, os visitantes podem seguir

Species 2000 Catalogue of Life, 2nd October 2010
indexing the world's known species

ITIS

Browse Search Info

Browse taxonomic classification

Top level group: **Animalia**

Phylum: []

Class: **Mammalia**

Order: **Rodentia**

Superfamily: []

Family: **Ctenodactylidae**

Genus: **Ctenodactylus**

Species: []

Intraspecific taxon: **gundi wai**

Match whole words only

Clear form Search

Annual Checklist Interface v1.5.1 developed by ETI Bioinformatics.

Figura 2.11: Portal Species 2000: consulta a partir dos níveis taxonômicos

a hierarquia dos grupos taxonômicos e aprender sobre filogenia e evolução, bem como sobre as características desses grupos. As consultas à base de dados são feitas de forma manual e exploratória como ilustrado na Figura 2.12, ou seja, na forma de navegação na árvore filogenética geral. Além da localização na árvore, a página relativa a um determinado grupo taxonômico mostra imagens, links para vídeo e referências bibliográficas.

Tree of Life web project

home browse help features learning contribute about

Search advanced

Animals

Metazoa

Animals

page content
articles & notes
treehouses
collections
Animals Mollusca
people
options

Explore Other Groups

other Eukaryotes
containing groups
subgroups
random page

Animals

Metazoa
Animalia

References

Adouta, A., G. Balavoine, N. Lefillat, O. Laspinet, B. Prod'homme, and R. de Rosa. 2000. The new animal phylogeny: Reliability and implications. *Proceedings of the National Academy of Sciences (USA)* 97:4453-4456.

Anderson, C. L. 1998. Phylogenetic relationships of the Myxozoa. Pages 341-350 in *Evolutionary Relationships among Protozoa* (G.H. Coombs, K. Vickerman, M.A. Sleight, and A. Warren, eds.) Chapman & Hall, London.

Anderson, C. L., E. H. Canning, and B. Okamura. 1998. A trophoblast origin for Myxozoa? *Nature* 392:244-247.

Ayala, F. J., A. Zhetsky, and F. J. Ayala. 1998. Origin of the metazoan phyla: Molecular clocks confirm paleontological estimates. *Proceedings of the National Academy of Sciences (USA)* 95:640-642.

Bauhoff, J., P. Martinez, J. Pass, and M. Ruidor. 2008. Back in time: a new systematic proposal for the Bilateria. *Philosophical Transactions of the Royal Society Series B*

Figura 2.12: Tree of Life: exemplo de página descritiva de um grupo taxonômico

O SIAMAZONIA (Sistema de informação da Diversidade Biológica e Ambiental da Amazônia Peruana) (SIAMAZONIA, 2010) é gerenciado por uma organização descentralizada, não contando com instituição líder, mas com uma instituição facilitadora, que

administra a rede. A rede é formada por instituições e pessoas especializadas na diversidade biológica e ambiental da Amazônia Peruana. Esta organização disponibiliza dados e informações relevantes a esta temática, as quais são sistematizadas e apresentadas através de uma página web, como podemos ver na Figura 2.13. O SIAMAZONIA tem como objetivo principal desenvolver políticas de conservação e proteção ambiental na Amazônia Peruana e, para tal, estão sendo inventariados os seres vivos da Amazônia, de todos os reinos. Assim como o Tree of Life, que não conta com dados de características das espécies, o SIAMAZONIA contém apenas dados simples da árvore taxonômica, nome comum e uma imagem ilustrativa (Figura 2.13).

The screenshot shows the SIAMAZONIA website interface. At the top, there is a header with the logo and navigation links: INICIO, BUSCAR, GLOSARIO, GUIA DE USO, MAPA WEB, CONTACTOS, and INGRESAR. Below the header is a search form titled 'BUSCAR FICHA' with the following fields: Reino (Animalia), Familia, Género, Especie, Nombre Científico, and Nombre Común. A 'Buscar' button is located below the form. To the left of the search form is a vertical menu with various categories such as 'Diversidad Amazónica', 'Biodiversidad y Sociedad', 'Conservación', 'Investigación', 'Biotecnología', 'Zonificación Ecológica', 'Económica', 'Legislación y Política', 'Recursos Informativos', 'Información Básica', 'Literatura en Línea', 'Material Educativo', 'Imágenes', 'Mapas', 'Estadísticas', 'Herramientas', 'Entorno Regional', 'Áreas Naturales Protegidas', 'Normas', 'Archivos del Sitio', 'Comunicación', 'Directorio de Contactos', 'Foro de Discusión', 'Vínculos de Interés', 'Red SIAMAZONIA', 'Presentación', 'Nodos', 'Comité Directivo', and 'Normas de Redacción'. Below the search form is a table with the following columns: Nº, Nombre Científico, Nombre Común, Familia, and ver. The table contains the following data:

Nº	Nombre Científico	Nombre Común	Familia	ver
	<i>Epirates carchia</i>	Boc arcoiris Inglés: Rainbow bow	Boidae	
1-				
	<i>Felis concolor</i>	Yuma; León Americano	Felidae	
2-				
	<i>Iguana iguana</i>	Iguana verde Inglés: Green Iguana	Iguanidae	
3-				
	<i>Lutrea longicauda</i>	Mutria, labllo de río, pisua, perari, mayayuma.	Mutellidae	
4-				

Figura 2.13: Portal AIAMAZONIA: descrição simplificada das espécies

Já o MaNIS (Mammal Networked Information System) é um portal para compartilhamento de bases de dados de espécimes, oriundos principalmente de dados de museus. Entre os objetivos do projeto do MaNIS, listados em (MANIS, 2010), está facilitar o acesso a dados de espécimes via web, realçar o valor científico das coleções de espécimes, catalogação de itens de museus a fim de auxiliar o trabalho de seu curador, e ser genérico o suficiente para poder ser usado com outras famílias. Na Figura 2.14, temos a estrutura do sistema do MaNIS mostrando o portal acessado pelo usuário através das bases de dados das coleções que são apresentadas. As bases de dados das coleções estão localizadas fisicamente em cada museu, porém, os dados são exportados e o servidor do portal possui uma cópia de cada base no formato específico do MaNIS. O portal conta também com o sistema BerkeleyMapper (BERKELEYMAPPER, 2005) que foi criado e construído usando localidades de espécimes de museus de história natural. Porém, é adaptável para mapear qualquer coleção de pontos. O BerkeleyMapper conta com uma coleção de scripts e serviços ligados para proporcionar funções de mapeamento integrado usando o GoogleMaps, TerraServer (TERRASERVER, 2001) e outros componentes de software livre. Um exemplo de utilização é ilustrado na Figura 2.15

O Arctos (ARCTOS, 2010) reúne dados de espécimes de coleções de museus. Foi desenvolvido como um projeto no Museu de Zoologia de Vertebrados da Universidade de Berkeley, Califórnia (EUA). O sistema permite uma grande variedade de critérios de busca, como localização geográfica, características do espécime ou elemento taxonômico. Uma das principais funcionalidades do sistema é permitir que o usuário salve as consultas

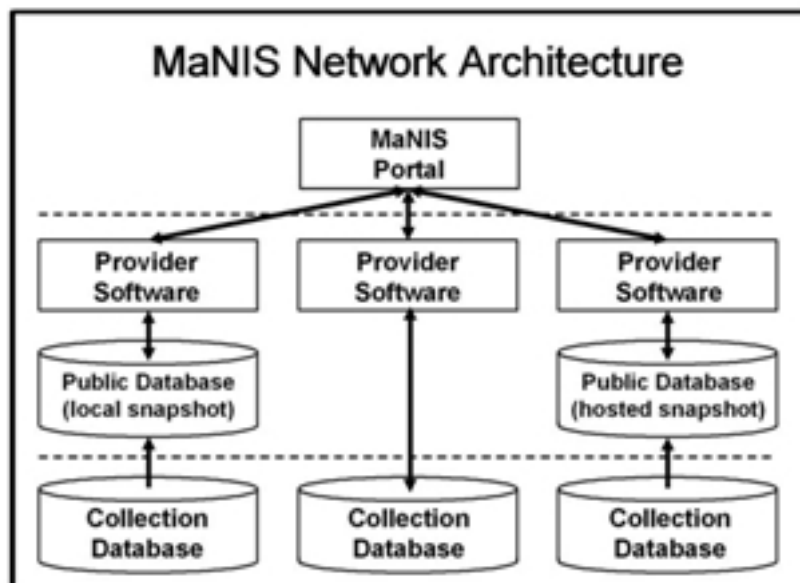


Figura 2.14: Portal MANIS: arquitetura do sistema



Figura 2.15: Portal MANIS: consulta exibida sobre o mapa.

feitas no site para uso posterior, além de permitir que os resultados das consultas sejam exportados de diversas maneiras, como texto normal ou XML. O sistema oferece um funcionamento personalizado para cada coleção, permitindo que alguns critérios de busca sejam utilizados somente naquela coleção. Assim como o MANIS também oferece a utilização do sistema de mapa BerkeleyMapper para a exibição dos resultados das consultas. Contudo, é mais um sistema que não provê meios de visualização analítica dos dados gerados nas consultas.

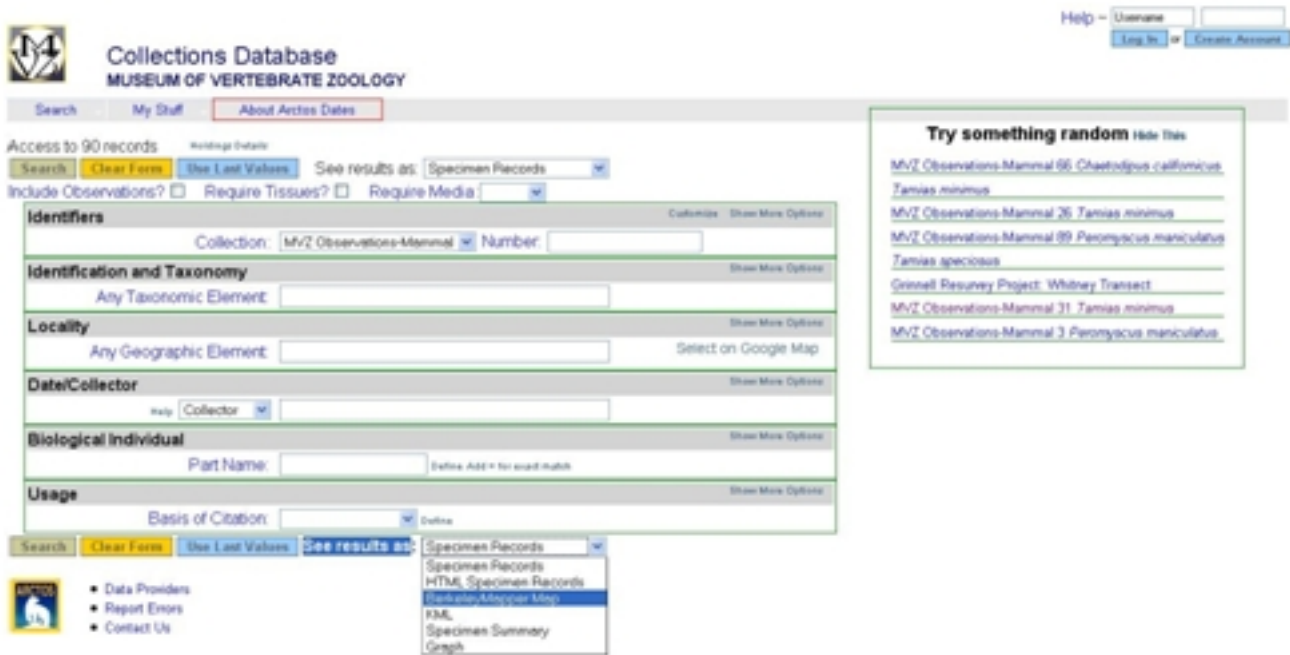


Figura 2.16: Portal ARCTOS: consulta a informações de coleção

2.4 Comentários finais

Para traçar o quadro comparativo dos diversos sistemas encontrados Figura 2.17, foram definidos alguns critérios relevantes, os quais foram levantados na análise de requisitos para o sistema aqui desenvolvido:

- Navegação na árvore taxonômica: permite acessar níveis taxonômicos diretamente na árvore, facilitando a busca de algum nível específico pelo usuário.
- Definição das características dos exemplares: possibilita ao usuário um estudo detalhado e específico sobre cada exemplar.
- Bibliografia relacionada aos registros: apresenta bibliografia relacionada nos resultados da busca.
- Dados de exemplares: contem informações de espécimes não apenas de espécies.
- Darwin Core: possibilita o compartilhamento de informações com outros sistemas.
- Exportação de dados: permite a exportação dos resultados das consultas para uso em outra ferramenta.
- Georeferenciamento: apresenta o resultado das consultas através de mapas.
- Integração com o R: permite a exportação dos dados e a utilização do sistema R para análise das consultas.
- Geração de gráficos: apresenta os resultados de consultas através de gráficos.

Nosso trabalho tem por objetivo prover um sistema unificado de agregação de conhecimento de forma colaborativa, fornecendo ferramentas de suporte à construção de uma base de dados sólida e consistente. Os serviços são disponibilizados através de uma interface que viabiliza a utilização do sistema tanto por profissionais experientes da área, os quais fazem consultas avançadas e necessitam de ferramentas para análise dos resultados, quanto por alunos e público em geral, que queiram aprender sobre espécimes. Além disso, nosso sistema provê diferentes formas de visualização dos dados, as quais vão desde loca-

	Navegação na árvore taxonômica	Consulta sobre características de exemplares	Bibliografia associada aos registros	Dados de Exemplares	Darwin Core	Exportação de dados	Georeferenciamento	Integração com o R	Geração de gráficos
Sinbiota	não	sim	sim	sim	não	não	sim	não	não
SpeciesLink	não	sim	sim	sim	não	sim	sim	não	não
WeBios	*	*	*	*	*	*	*	*	*
BioCORE	*	*	*	*	*	*	*	*	*
Species 2000	sim	não	sim	não	sim	não	não	não	não
GBIF	sim	não	não	não	sim	sim	não	não	não
MaNIS	não	não	não	não	sim	sim	não	não	não
Arctos	não	sim	sim	sim	sim	sim	sim	não	não
SIAMAZONIA	sim	não	não	não	sim	não	sim	não	não
IABIN	sim	sim	sim	sim	sim	sim	sim	não	não
Tree of Life	sim	não	sim	não		não	não	não	não
Taxonomy Br	sim	sim	sim	sim	sim	sim	sim	sim	sim

Figura 2.17: Quadro comparativo dos trabalhos revisados. As características não disponíveis são representadas por '*'.

lização espacial de espécimes sobre mapas, até geração de diversos gráficos para análise preliminar dos dados.

A integração de recursos de visualização e de análise de dados às funções comumente encontradas nos sistemas de biodiversidade, constitui uma característica diferenciadora do nosso sistema. A integração com visualização de mapas, através da API do Google Maps/Google Earth, e a adição das ferramentas de análise de dados ao ambiente R e à API do Google Charts, ampliam a aplicabilidade e usabilidade do sistema. Ou seja, usando essas ferramentas externas, não restringimos o sistema a uma área ou região específicas. Dessa maneira, o mesmo sistema pode ser aplicado no gerenciamento, disponibilização, visualização e análise de diversos tipos de dados sobre qualquer região do globo.

3 MODELO PROPOSTO PARA O ARMAZENAMENTO DE DADOS DE BIODIVERSIDADE

Este capítulo descreve o modelo de dados proposto para o TaxonomyBrowser. Para a especificação do modelo de dados contamos com a participação de biólogos parceiros do projeto, dos quais investigamos as atividades de pesquisa. A validação do sistema vem sendo feita, pelos próprios usuários, através da entrada de dados reais.

3.1 Visão Geral

A base de dados foi modelada para armazenar informações de espécimes coletadas por biólogos, processar consultas para análise de dados bem como disponibilizar os dados da coleção para o público em geral.

Assim, os usuários previstos para o TaxonomyBrowser são divididos em dois grupos: os biólogos que mantêm o banco de dados e tem acesso pleno ao sistema, podendo atualizar os dados e processar qualquer tipo de consulta, e os usuários não cadastrados que tem acesso apenas às consultas georeferenciadas.

No caso dos usuários cadastrados, os biólogos que atualizam os dados e/ou utilizam todos os recursos do TaxonomyBrowser, realizam essa atualização ou utilização como parte de suas atividades de pesquisa. Os usuários não cadastrados podem ter metas de pesquisa ou apenas de busca de informações para aprendizado.

A seguir, estão descritos os cenários de uso e as atividades de ambas as categorias de usuários. Tais cenários guiaram tanto a concepção do modelo de dados e operações sobre o banco de dados como o projeto da interface visual.

3.1.1 Cenário de atividades de pesquisa dos biólogos

Inicialmente, os biólogos iniciam suas atividades de pesquisa a partir de uma (ou mais) situações dentre as seguintes: (a) estabelecimento de um plano de coletas de amostras (espécimes) em campo; (b) uso de amostras já coletadas e armazenadas fisicamente na coleção do laboratório; (c) uso de amostras de espécimes registradas em coleções de museus; (d) uso de dados já armazenados em banco de dados.

Dependendo da situação, as atividades iniciais são diferentes, culminando num conjunto de dados que são analisados resultando na confirmação ou não da hipótese de pesquisa.

O diagrama da Figura 3.1 mostra as atividades dos biólogos de forma combinada, ou seja, agregando os 4 pontos de partida mencionados acima.

A primeira alternativa de início do trabalho de pesquisa é aquela em que os objetivos da pesquisa foram estabelecidos mas não há material biológico disponível, ou seja,

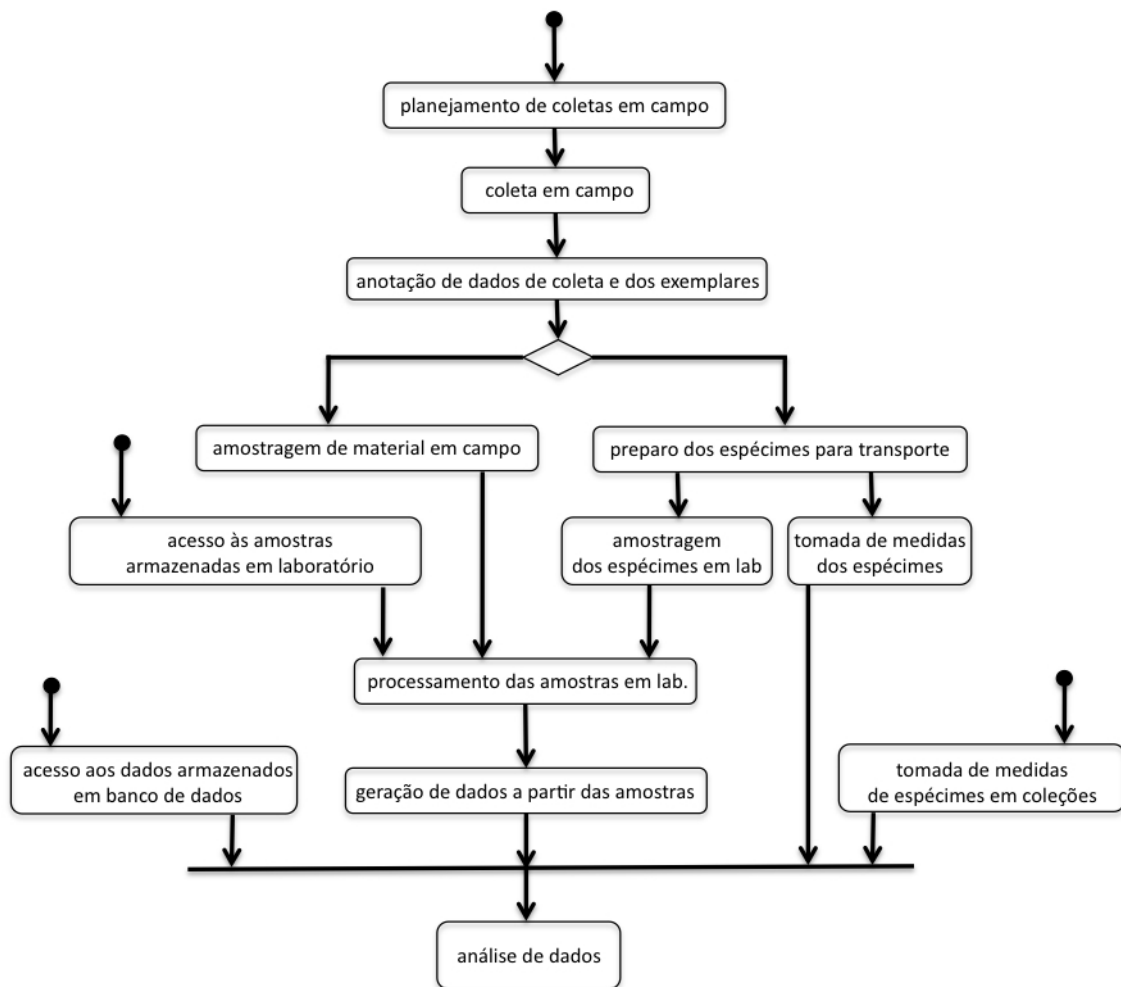


Figura 3.1: Diagrama de atividades de pesquisa dos usuários biólogos

é necessário coletar amostras em campo. A primeira etapa é, então, o planejamento das coletas. As coletas seguem metodologias diferentes dependendo do objeto da pesquisa. Tanto o procedimento como o período e o que coletar são dependentes do espécime e do objetivo. Entretanto, é comum a todas as coletas, registro de informações básicas: data, localização geográfica, coletor(es), identificação do espécime (número de campo), observações eventuais (habitat, condições do ambiente, fêmeas grávidas, etc.). No caso de mamíferos, em geral, são tomadas medidas anatômicas, peso, cor da pelagem, sexo e outras características relevantes. Dependendo do estudo, o espécime pode ser simplesmente amostrado (retirada de pelo, sangue, medula óssea ou amostra de tecidos diversos), marcado e solto na natureza, ou pode ser retirado definitivamente do campo para incorporação à coleção. Numa segunda etapa, em função do que foi coletado, os biólogos retornam ao laboratório para e armazenagem definitiva do exemplar ou das amostras, e posterior processamento. O processamento das amostras também segue procedimentos diferentes, dependendo do estudo:

- se o estudo for a nível cromossômico, o sangue (ou a medula óssea) é processado para observação em lâminas em microscópio.
- se o estudo for morfológico, são tomadas medidas de ossos específicos para complementação das medidas anatômicas tomadas no momento da coleta.

- se o estudo for evolutivo, com marcadores moleculares, o DNA é extraído do tecido amostrado e processado para posterior sequenciamento.

No caso de estudos morfológicos, podem ser usados exemplares de museus, dos quais são tomadas as medidas. Em todos os três casos, os biólogos podem utilizar amostras já armazenadas fisicamente no laboratório ou dados registrados em banco de dados próprios ou compartilhados.

Atualmente, os primeiros registros feitos pelos biólogos em campo são realizados em cadernos próprios, com a localização obtida a partir de GPS. Posteriormente, os dados são registrados em planilhas ou arquivos texto, conforme as ferramentas computacionais que serão utilizadas. Com a implantação do TaxonomyBrowser como uma aplicação web, este pode ser utilizado para registro dos dados dos exemplares ainda em campo. Posteriormente, os dados podem ser atualizados com as informações geradas pelo processamento das amostras em laboratório, e os dados necessários para análise podem ser selecionados.

3.1.2 Cenário de uso dos usuários não cadastrados

Este cenário foi concebido para usuários que desejam acessar informações básicas de ocorrência dos espécimes. Neste caso, o usuário acessa o TaxonomyBrowser e tem possibilidade de:

- navegar na árvore taxonômica, buscando informações sobre os diversos níveis da árvore taxonômica até chegar aos exemplares de uma espécie específica.
- acessar consultas preparadas pelos biólogos para visualização em mapas gerados através da API do Google Maps, o que permite também acessar informações sobre exemplares específicos pela interação com o mapa.

3.2 Modelo de Dados da Coleção

Os exemplares depositados na coleção e seus dados descritivos são representados por nodos taxonômicos. Esses nodos representam cada nível dentro do sistema taxonômico de classificação dos organismos vivos. Como podemos ver na Figura 3.2, nodos taxonômicos (entidade Nodo Taxonômico) se relacionam para permitir a representação da árvore taxonômica.



Figura 3.2: Diagrama Entidade-Relacionamento para nodo taxonômico

Um nodo taxonômico pode conter uma série de características que podem ser mensuráveis. Todos os nodos filhos deste nodo taxonômico herdam as características do nodo pai. Cada nodo filho pode conter também características especializadas, as quais somente apareceram neste nodo. Desta forma, ao analisar um caminho na árvore taxonômica, partindo da raiz e chegando a um nodo taxonômico folha, temos um conjunto de atributos que caracterizam aquele nodo de forma única. Ele pode possuir características próprias e, também, outras herdadas dos nodos superiores. Por exemplo: um espécime da espécie *Tadarida brasiliensis* (morcego) e um *Sus scrofa* (javali) possuem características comuns herdadas do nodo taxonômico da classe *Mammalia*, como a cor do pelo e a presença

de glândulas mamárias são características comuns aos dois. Contudo, o morcego possui características específicas da ordem *Chiroptera*, como o comprimento do antebraço. A organização de vários nodos é feita de forma hierárquica, gerando a árvore taxonômica como foi visto no capítulo 2.

A Figura 3.3 mostra o diagrama Entidade-Relacionamento reunindo as entidades do modelo que descrevem os nodos taxonômicos.

Cabe ressaltar que o projeto do modelo do banco de dados incorpora os elementos que são parte do padrão Darwin Core (TDWG, 2010), acrescentando outras características de interesse de nossos usuários alvo. O Darwin Core foi escolhido tendo em vista principalmente a interoperabilidade futura do trabalho com projetos semelhantes.

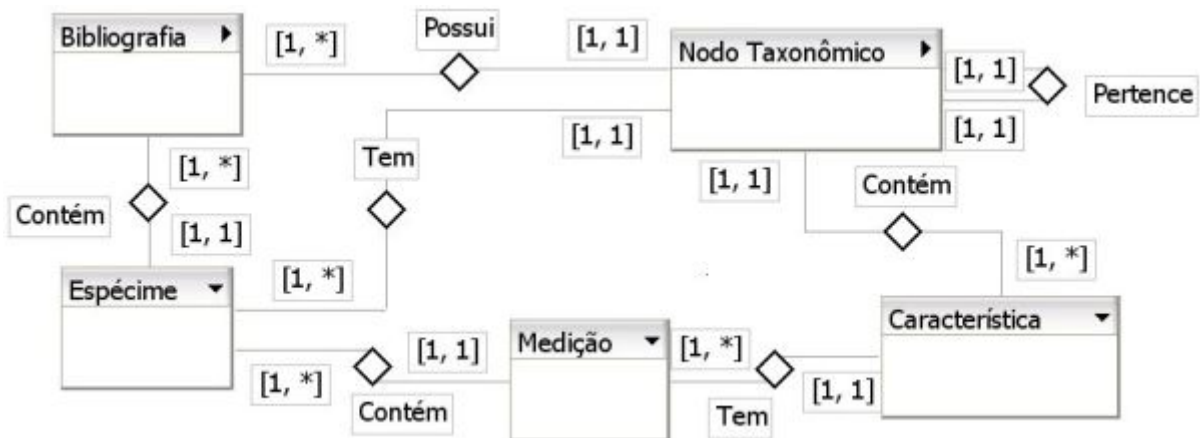


Figura 3.3: Diagrama E-R do Modelo Implementado no TaxonomyBrowser

Cada nodo taxonômico armazena uma referência a um nodo pai, que representa um agrupamento maior na classificação. Para construir a hierarquia, são necessários o nome do nível taxonômico (Reino, Filo, Classe, Ordem, Subordem, Família, Gênero ou Espécie), o nome científico deste nível taxonômico (por exemplo, um nome científico para o Gênero pode ser *Ctenomys*), além de informações extras que podem ser adicionadas quando houver necessidade. O nodo taxonômico pode ainda conter uma lista de referências bibliográficas (entidade Bibliografia) que identifica as obras que dizem respeito ao nível taxonômico que está sendo descrito. A entidade Bibliográfica armazena o título da obra, o autor, o editor e o ano de publicação, ou seja, as informações necessárias para identificação única da publicação, para que esta possa ser encontrada e referenciada pelo usuário.

As características estão agrupadas de acordo com a natureza dos caracteres que representam: morfométricos, taxonômicos e filogenéticos. Uma característica (entidade Característica) possui também um tipo específico para armazenagem, que pode ser: unidade física, nome literal, número inteiro, número real, arquivo de imagem, arquivo de som, arquivo de texto e arquivo de vídeo. Uma unidade física está ligada a uma unidade de medida (entidade Medição) no sistema internacional de medidas (SI). As unidades físicas facilitam conversões de medidas entre buscas feitas no sistema. A entidade Espécime designa um exemplar coletado ou uma amostra de um organismo. Esta entidade também contém informações de identificação do exemplar, seu coletor e data de coleta, e informações geográficas sobre o local de coleta. O sistema é flexível e permite a inserção de novas características, de acordo com as necessidades do estudo a ser desenvolvido pelo usuário.

Para gerenciar o sistema, quatro operações comuns a todas as entidades são suficientes: visualização da entidade, inserção de nova entidade, edição de entidade existente e remoção. A visualização é uma operação que requer apenas leitura no banco de dados, onde as informações referentes à entidade são exibidas para o usuário de maneira clara e legível. Na adição, uma nova entidade é criada no sistema, respeitando a não duplicação de entidades e no caso do nodo taxonômico, respeita-se a classificação científica possível para a hierarquia da árvore. Por exemplo, um nodo taxonômico de Gênero não pode vir antes de um nodo Família. A edição de entidade permite a atualização das informações da entidade, bem como permite trocar a hierarquia da classificação científica do nodo na árvore taxonômica.

A remoção de uma entidade facilita a correção de erros na entrada de dados no sistema, removendo completamente todas as informações referentes a uma entidade, bem como a entidade em si e seus filhos. Nesta operação é feita a deleção da entidade no sistema de modo que a consistência das outras entidades se mantenha, ou seja, outras entidades que estão ligadas à entidade que foi removida ficam com informações incompletas, gerando inconsistência no sistema. Logo não tem mais sentido de existirem sozinhas e, portanto, também devem ser deletadas. Um exemplo disso é quando um nodo taxonômico é removido do sistema, logo toda a subárvore deste nodo deve ser removida também. Na sequência, todos os espécimes com referência a esses nodos removidos devem ser removidos, pois as suas classificações científicas não existem mais no sistema.

3.3 Projeto do banco de dados

A partir da modelagem conceitual do sistema foi implementada uma base de dados utilizando o modelo de dados relacional para a aplicação de gerenciamento de espécimes. O sistema gerenciador de bancos de dados (SGBD) escolhido foi o MySQL. As justificativas de sua escolha são principalmente a fácil configuração da instalação do banco, portabilidade para diferentes sistemas operacionais, um grande conjunto de ferramentas para facilitar o desenvolvimento e administração de aplicações que necessitem de banco de dados e também sua licença GNU General Public License (GPL).

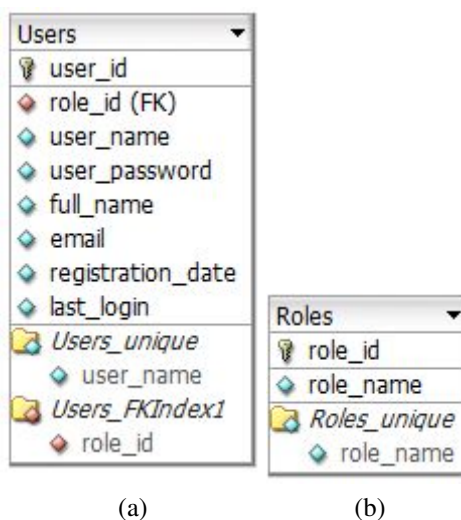
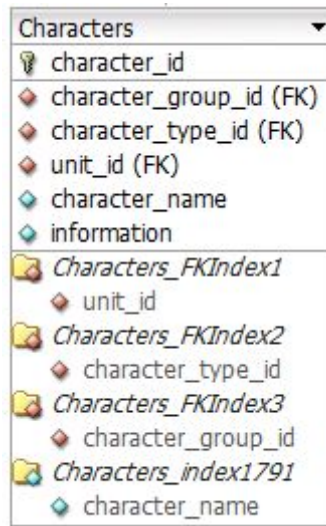


Figura 3.4: Tabelas. (a) Tabela de usuários do sistema. (b) Tabela de papéis do sistema.

Como existe a necessidade de diferentes níveis de acesso, foi necessário implementar um sistema de autenticação que utiliza papéis de usuários para permitir certas operações

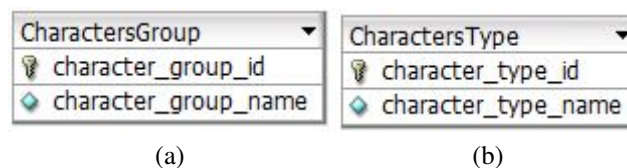
dentro do TaxonomyBrowser. A tabela *Users*, na Figura 3.4a, mostra as informações que são armazenadas sobre um usuário do sistema: identificador único do usuário, nome do usuário, senha, nome completo do usuário, endereço de email, data de criação do usuário e a data que o usuário entrou pela última vez no sistema e o papel atribuído a ele no sistema.



Characters	
🔑	character_id
🔗	character_group_id (FK)
🔗	character_type_id (FK)
🔗	unit_id (FK)
🔹	character_name
🔹	information
📁	Characters_FKIndex1
🔗	unit_id
📁	Characters_FKIndex2
🔗	character_type_id
📁	Characters_FKIndex3
🔗	character_group_id
📁	Characters_index1791
🔹	character_name

Figura 3.5: Tabela de características.

A tabela *Roles* na Figura 3.4b armazena os papéis reconhecidos pelo sistema, contendo identificador único do papel e nome completo do papel. Os possíveis papéis de usuários encontrados no sistema são: usuário anônimo e usuário administrador. O usuário anônimo não faz autenticação no sistema; logo, somente poderá visualizar informações sobre os espécimes e a árvore taxonômica. O usuário administrador pode, além de visualizar as informações, criar entidades, editar entidades e remover entidades do sistema. Ele tem controle completo para gerenciar as informações dentro do sistema. O sistema também prevê a criação de usuários administradores com diferentes níveis de acesso.



CharactersGroup	
🔑	character_group_id
🔹	character_group_name

(a)

CharactersType	
🔑	character_type_id
🔹	character_type_name

(b)

Figura 3.6: Tabelas. (a) Tabela Grupo de características. (b) Tabela de Tipos de características.

A tabela *Characters* armazena características dos espécimes, como ilustrado na Figura 3.5. Ela guarda um identificador único da característica, o nome e informações extras, caso haja necessidade. Ela também está ligada a um grupo de características que é armazenado pela tabela *CharactersGroup*, a qual pode ser vista na Figura 3.6a, como um identificador único do grupo de características e um nome do grupo. No sistema, os grupos de características possíveis são o grupo de características Morfométricas, Taxonômicas e Filogenéticas. As características também estão classificadas por tipos, os quais são armazenados na tabela *CharactersType*, descrita na Figura 3.6b.

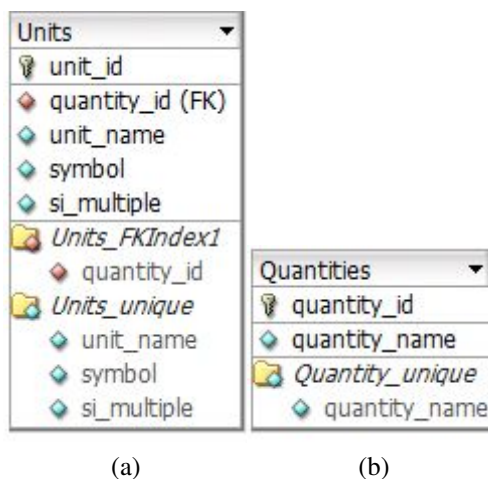


Figura 3.7: Tabelas. (a) Tabela de unidades física. (b) Tabela de Quantidade de medidas suportadas.

Os tipos de características suportados pelo sistema são: unidade física, nome (literal), número inteiro, número real, arquivo de imagem, arquivo de som, arquivo de texto e arquivo de vídeo. Qualquer nova característica adicionada ao sistema deve pertencer a um desses tipos previstos.

Caso a característica seja do tipo unidade física, a mesma estará ligada a tabela *Units*, ilustrada na Figura 3.7a. Esta tabela armazena os tipos de unidades físicas suportadas pelo sistema. Para estas unidades são guardadas um identificador único, nome, símbolo e um multiplicador de conversão para o Sistema Internacional de medidas. Essas unidades também estão ligadas a uma quantidade de medida no SI, e as quantidades de medida suportadas no sistema são armazenadas na tabela *Quantities* da Figura 3.7b. Algumas das quantidades de medidas suportadas pelo sistema são: tamanho, tempo, massa, temperatura, quantidade de substância, intensidade luminosa, aceleração. Novas unidades de medida também podem ser adicionadas ao sistema de acordo com as necessidades do usuário.

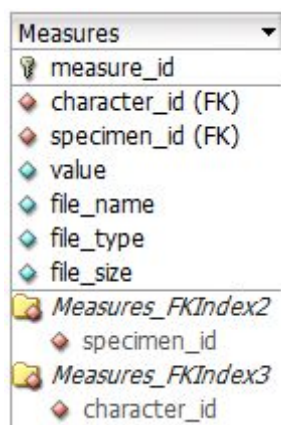


Figura 3.8: Tabela de medidas das características.

A tabela *Measures* na Figura 3.8 armazena as medições das características para os espécimes adicionados ao sistema. Cada medida guarda o valor da medição, caso esse valor seja de um tipo numérico, ou um valor literal. Se a medida for um tipo arquivo, o nome do arquivo é armazenado juntamente com suas propriedades de tipo e tamanho

do arquivo. O arquivo é, então, armazenado no sistema operacional com um nome único gerado automaticamente. A tabela de espécimes está ligada à tabela de medições como visto na Figura 3.3.

Specimens	
🔑	specimen_id
🔗	taxonomy_id (FK)
🔗	collection_id
🔗	collected_by
🔗	collected_data
🔗	latitude
🔗	longitude
🔗	information
📁	Specimens_FKIndex1
🔗	taxonomy_id

Figura 3.9: Tabela que armazena os dados básicos dos espécimes coletados.

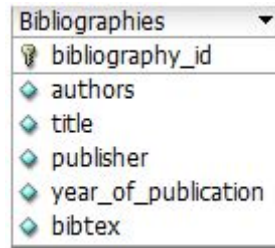
A tabela *Specimens*, ilustrada na Figura 3.9, armazena informações sobre os espécimes coletados pelos biólogos. Esta tabela contém um identificador único de espécime, um identificador de coleção do espécime, o nome do coletor, a data da coleta, informações geográficas de onde o espécime foi coletado (latitude e longitude) e informações adicionais, caso seja necessário. O espécime está ligado a um nodo taxonômico da tabela *Taxonomy* da Figura 3.10a, o qual representa a classificação científica do espécime coletado. Esta tabela contém informações do nome científico do nodo taxonômico, uma referência ao nodo taxonômico pai, informações extras e ainda uma referência ao nível taxonômico deste nodo. Os níveis taxonômicos são armazenados na tabela *TaxonomyRank*, vista na Figura 3.10b. Os níveis taxonômicos suportados pelo sistema são: Reino, Filo, Classe, Ordem, Subordem, Família, Gênero e Espécie.

(a) Taxonomy	
🔑	taxonomy_id
🔗	taxonomy_rank_id (FK)
🔗	parent_id (FK)
🔗	scientific_name
🔗	information
📁	Taxonomy_FKIndex1
🔗	parent_id
📁	Taxonomy_FKIndex2
🔗	taxonomy_rank_id

(b) TaxonomyRank	
🔑	taxonomy_rank_id
🔗	taxonomy_rank_name

Figura 3.10: Tabelas que armazenam dados sobre os níveis taxonômicos (a) Armazena os nodos taxonômicos hierarquicamente. (b) Identifica os níveis taxonômicos.

Quando adicionado um exemplar como um filho de um nodo taxonômico Espécie, toda a classificação científica do animal já é previamente subentendida, pois o nodo Espécie selecionado pelo usuário já contém essas informações. Caso não se saiba a que espécie o exemplar coletado pertence, pode-se atribuí-lo a um nodo taxonômico de um nível superior ao de Espécie (por exemplo, ao nível de um nodo Gênero), ou até mesmo a um nodo de espécie indefinida.









Bibliographies	
	bibliography_id
	authors
	title
	publisher
	year_of_publication
	bibtex

Figura 3.11: Tabela que armazena os dados das publicações de referência.

Por último, mas não menos importante, a tabela *Bibliographies*, ilustrada na Figura 3.11, armazena as referências bibliográficas utilizadas pelos biólogos quando adicionam espécimes e nodos na árvore taxonômica. Esta tabela guarda as obras bibliográficas utilizando um identificador único por cada obra, nome dos autores, editor, ano de publicação e ainda mais o código bibtex da mesma bibliografia, para utilização futura ou mesmo manter compatibilidade com outros sistemas bibliográficos.

A Figura 3.12 ilustra o modelo completo do banco de dados relacional implementado para a aplicação, mostrando as tabelas de entidades e de relacionamentos. Sobre esse banco de dados foram criadas as operações de visualização, inserção, remoção e edição das entidades do sistema, na aplicação detalhada no próximo capítulo.

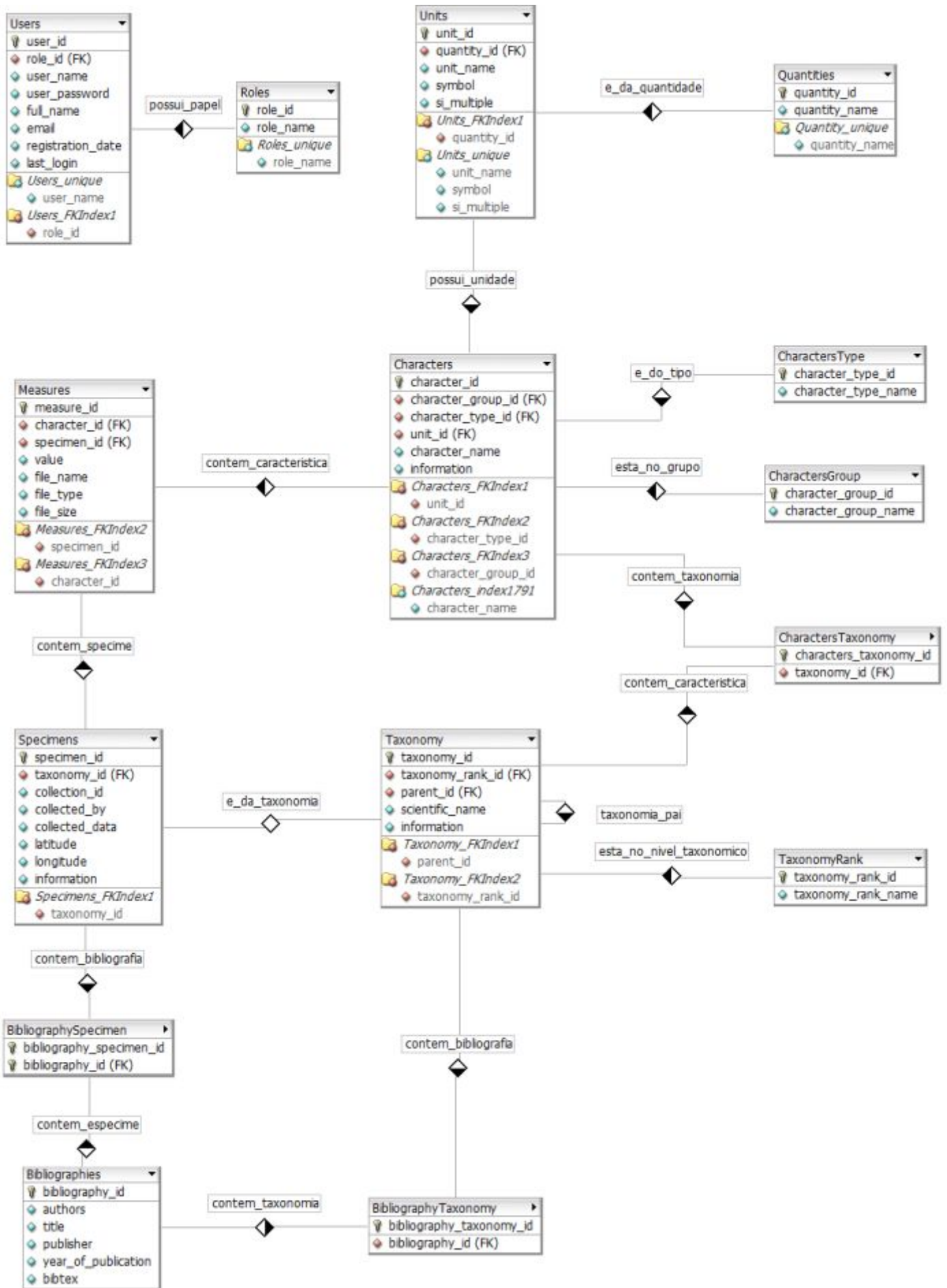


Figura 3.12: Modelo relacional do banco de dados da aplicação, tabelas e ligações lógicas entre tabelas.

4 ARQUITETURA DO SISTEMA

Este capítulo descreve a arquitetura do sistema desenvolvido de acordo com o modelo de dados proposto para o TaxonomyBrowser.

4.1 Visão Geral do Sistema TaxonomyBrowser

A arquitetura do sistema TaxonomyBrowser pode ser separada em quatro componentes distintos. A Interface de Gerência da Coleção (A) permite a manutenção de todos os dados de espécimes através das quatro operações básicas: visualização, inserção, alteração e remoção. A Interface Visual, baseada no Google Maps/Earth, (B) permite consultas georeferenciadas aos dados da coleção. A Interface de Consulta Otimizada (C) permite a execução de consultas avançadas para análise de dados por parte dos biólogos. As consultas otimizadas são processadas através de aplicações externas, neste caso, scripts no sistema R e gráficos na API Google Charts, as quais são integradas no sistema TaxonomyBrowser de forma transparente ao usuário. A base de dados da coleção (D) armazena todos os dados das coleções, sendo que a camada de mapeamento faz a tradução dos dados da coleção para cada uma das interfaces de consulta. Uma visão em alto nível dessa arquitetura pode ser vista na Figura 4.1.

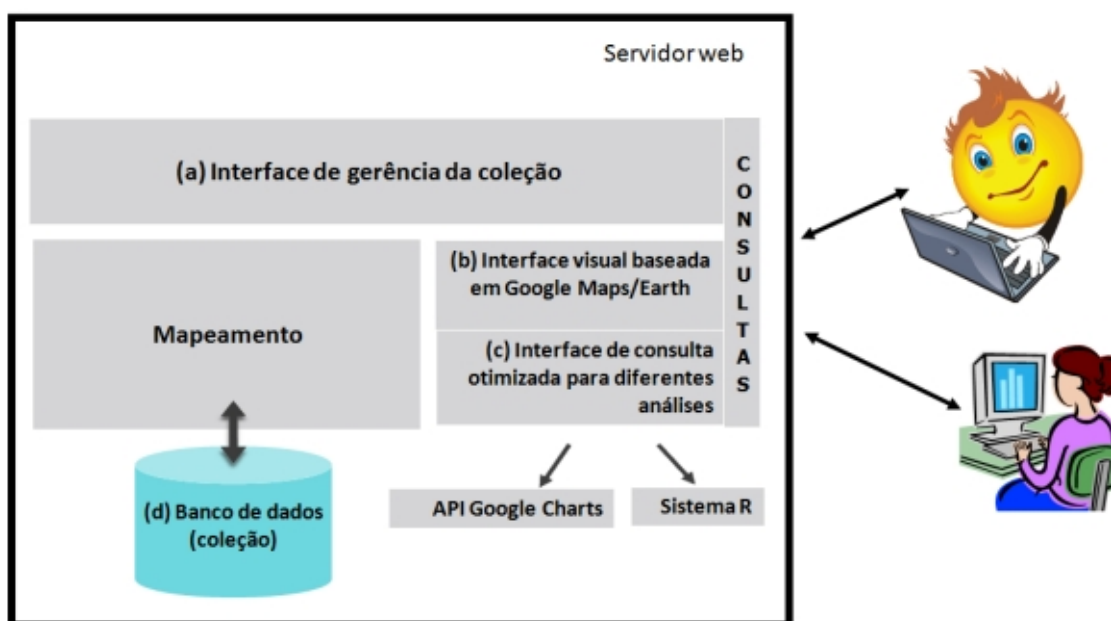


Figura 4.1: Arquitetura do Sistema TaxonomyBrowser. (A)Interface de Gerência. (B) Interface Visual. (C) Interface de Consulta Otimizada. (D) Base de Dados da Coleção.

Conforme já mencionado, o sistema prevê dois tipos de usuários com diferentes graus de acesso. Por exemplo, um biólogo que tenha acesso pleno ao sistema, pode atualizar os dados e processar qualquer tipo de consulta, podendo acionar scripts de análise no sistema R (R-PROJECT, 2000) e gerar gráficos através da API do Google Charts (GOOGLE, 2010b). Um usuário não cadastrado no sistema tem acesso às consultas georeferenciadas e aos dados da coleção.

O sistema TaxonomyBrowser foi implementado sobre o sistema SGBD MySQL (ORACLE, 2000), tendo em vista sua portabilidade para diferentes sistemas operacionais, além de prover um vasto conjunto de ferramentas que facilitam o desenvolvimento e administração de sistemas, e também por sua licença GNU (FREE-SOFTWARE-FOUNDATION, 2000) General Public License. O sistema R está instalado no mesmo servidor e os scripts escritos pelos usuários são também armazenados no banco de dados.

4.2 Arquitetura da Aplicação

Como este software foi desenvolvido para a plataforma *web*, ele é utilizado através de um navegador, podendo ser acessado por múltiplos usuários simultaneamente, incluindo visitantes não cadastrados. Estes somente poderão fazer operações restritas aos dados, como por exemplo, visualizá-los. Esse sistema de gerenciamento foi escrito em linguagem PHP versão 5.3.1, utilizando orientação a objetos e o padrão de arquitetura de software MVC (Model-View-Controller). Esse padrão de desenvolvimento de software separa de forma explícita, em camadas, as operações de acesso aos dados, a lógica da aplicação e a apresentação dos dados, deixando o sistema mais compreensível para futuras atualizações e correções feitas por outros desenvolvedores.

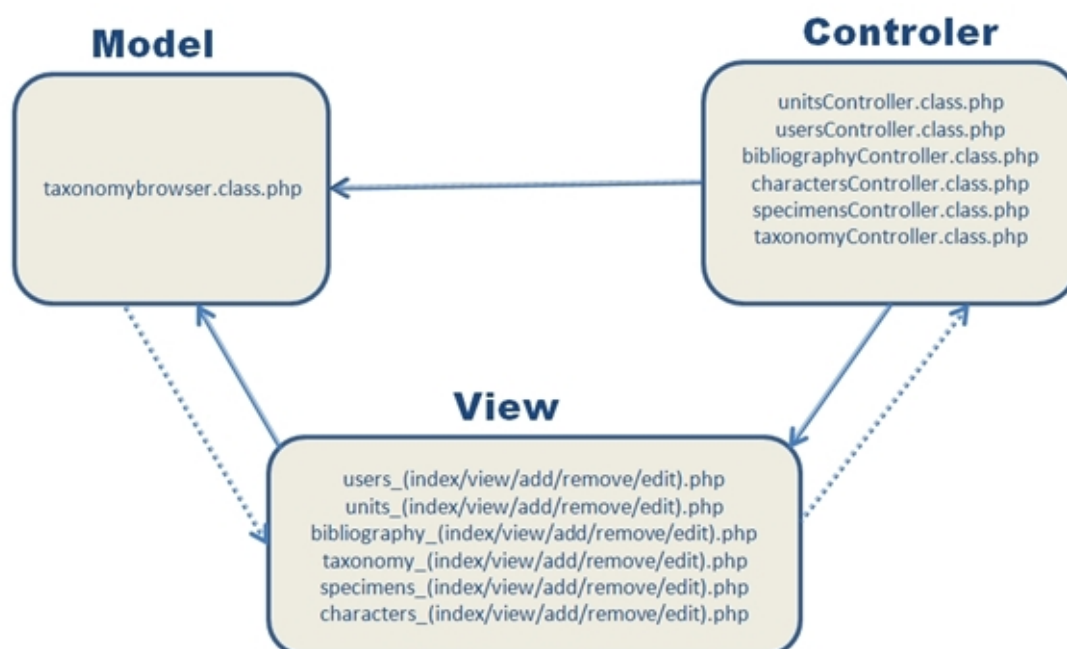


Figura 4.2: Arquitetura MVC da aplicação.

A Figura 4.2 ilustra um diagrama simples exemplificando a relação entre Model, View e Controller. As linhas sólidas indicam associação direta e as tracejadas indicam associação indireta. A camada que controla o acesso aos dados (Model) foi implementada utilizando funções nativas da biblioteca da linguagem PHP para conexão e manipulação

Tabela 4.1: Classes Controladoras, cujo desenvolvimento reflete diretamente na existência de uma entidade real na camada de dados

unitsController.class.php	Controla unidades de medições
unitsController.class.php	Controla usuários do sistema
bibliographyController.class.php	Controla obras bibliográficas
charactersController.class.php	Controla características de espécimes
specimensController.class.php	Controla espécimes e medições das características
taxonomyController.class.php	Controla a árvore taxonômica

Tabela 4.2: Classes Controladoras de Gerenciamento.

loginController.class.php	Controla autenticação de usuários
logoutController.class.php	Controla saída de usuários do sistema
errorController.class.php	Trata erros do sistema
indexController.class.php	Responde como entrada principal do sistema

do bando MySQL. Uma quantidade abrangente de métodos para visualizar, adicionar, remover e atualizar entidades do sistema foram implementados na classe `taxonomybrowser.class.php`.

A camada que gerencia a lógica da aplicação foi implementada utilizando a abstração `Controller`. Para cada entidade do sistema, uma classe `Controller` foi implementada. Estas classes são responsáveis por responder às ações dos usuários vindas da camada de visualização, alterar a camada de dados baseado na entrada dos usuários e, ainda, validar dados gerados pelos usuários do sistema. Na tabela 4.1 são apresentados as classes `Controllers`, cujo desenvolvimento refletem diretamente na existência de uma entidade real na camada de dados.

Também existem classes `Controllers` que não são entidades físicas do sistema. Elas são apenas classes de gerenciamento de funções específicas do sistema e são apresentadas na tabela 4.2.

A camada de apresentação do sistema é baseada nas linguagens XHTML, CSS e Javascript. XHTML e CSS são as linguagens usadas para apresentação e layout de documentos padronizados para navegadores de Internet. Javascript é utilizado para tornar algumas ações dinâmicas dentro da interface do usuário, sem necessidade de redirecionamento HTTP, deixando a interface livre dos detalhes de implementação, ou seja, transparente para o usuário. A tabela 4.3 apresenta os scripts PHP que compõem a camada de apresentação das entidades existentes no `Taxonomy Browser`.

Ainda, na camada de visualização, existem alguns scripts em linguagem PHP que não operam sobre entidades do sistema, mas sim fazendo renderizações de operações específicas, são apresentados na tabela 4.4.

Tabela 4.3: Scripts PHP que compõem a camada de apresentação das entidades existentes no Taxonomy Browser.

users_index.php users_view.php users_add.php users_remove.php users_edit.php	Mostra os dados da entidade Usuário e permite disponibiliza as operações possíveis sob esta entidade.
units_index.php units_view.php units_add.php units_remove.php units_edit.php	Mostra os dados da entidade Unidade de medida e permite ao usuário operar sob a mesma.
bibliography_index.php bibliography_view.php bibliography_add.php bibliography_remove.php bibliography_edit.php	Renderiza e operada sob os dados das obras bibliográficas.
taxonomy_index.php taxonomy_view.php taxonomy_add.php taxonomy_remove.php taxonomy_edit.php	Mostra os dados de um nodo taxonômico e permite ao usuário operar sob a árvore taxonômica.
specimens_index.php specimens_view.php specimens_add.php specimens_remove.php specimens_edit.php	Renderiza informações sobre um espécime e permite fazer operações sob o mesmo.
characters_index.php characters_view.php characters_add.php characters_remove.php characters_edit.php	Permite ao usuário visualizar características e operar sob as mesmas.

Tabela 4.4: Scripts PHP da camada de visualização.

Index.php	Script de visualização da página inicial do sistema.
navigation_bar.php	Gera a visualização do menu de navegação das entidades.
login.php	Visualiza a interface de autenticação do sistema.
error.php	Visualiza erros gerados pelos sistema.
header.php footer.php	Scripts que são utilizados como template para visualização.

5 INTERFACE VISUAL

O sistema Taxonomy Browser possui uma interface visual que permite aos usuários realizar todas as operações mencionadas anteriormente, dependendo de suas permissões de acesso.

Para um melhor gerenciamento das funcionalidades do sistema, a interface visual foi subdividida em quatro interfaces principais, sendo elas a interface administrativa, a interface de consulta, a interface sobre mapa e a interface de análise.

O usuário visitante pode navegar pela interface visualizando os espécimes e a árvore taxonômica registrada no sistema. Já os usuários cadastrados no sistema poderão se autenticar no mesmo e usar suas permissões específicas de edição, inserção ou remoção de entidades. Este processo é feito utilizando a janela de autenticação, ilustrada na Figura 5.1.



Figura 5.1: Janela de autenticação de usuário no TaxonomyBrowser.

5.1 Interface administrativa

A interface administrativa permite inserir nodos taxonômicos, incluindo a definição de suas características (nome de apresentação, tipo, unidade de medida) e a inclusão, alteração ou remoção de espécimes vinculados a um nodo taxonômico folha.

A interface gráfica do sistema possui um layout padronizado para todos os tipos de entidades do sistema. Existe uma aba fixa na parte superior do layout para navegar entre diferentes tipos de Entidades, como mostra a Figura 5.2(a), assim como selecionar . Na parte esquerda do layout existe uma barra de navegação para instâncias de uma entidade, como pode ser visto na Figura 5.2(b). A partir desse navegador é possível chegar às informações de uma determinada entidade e também adicionar novas informações para uma entidade, como ilustrado na Figura 5.2(d). Quando as informações de uma entidade são selecionadas, é possível operar nessas informações utilizando a barra de operações, mos-

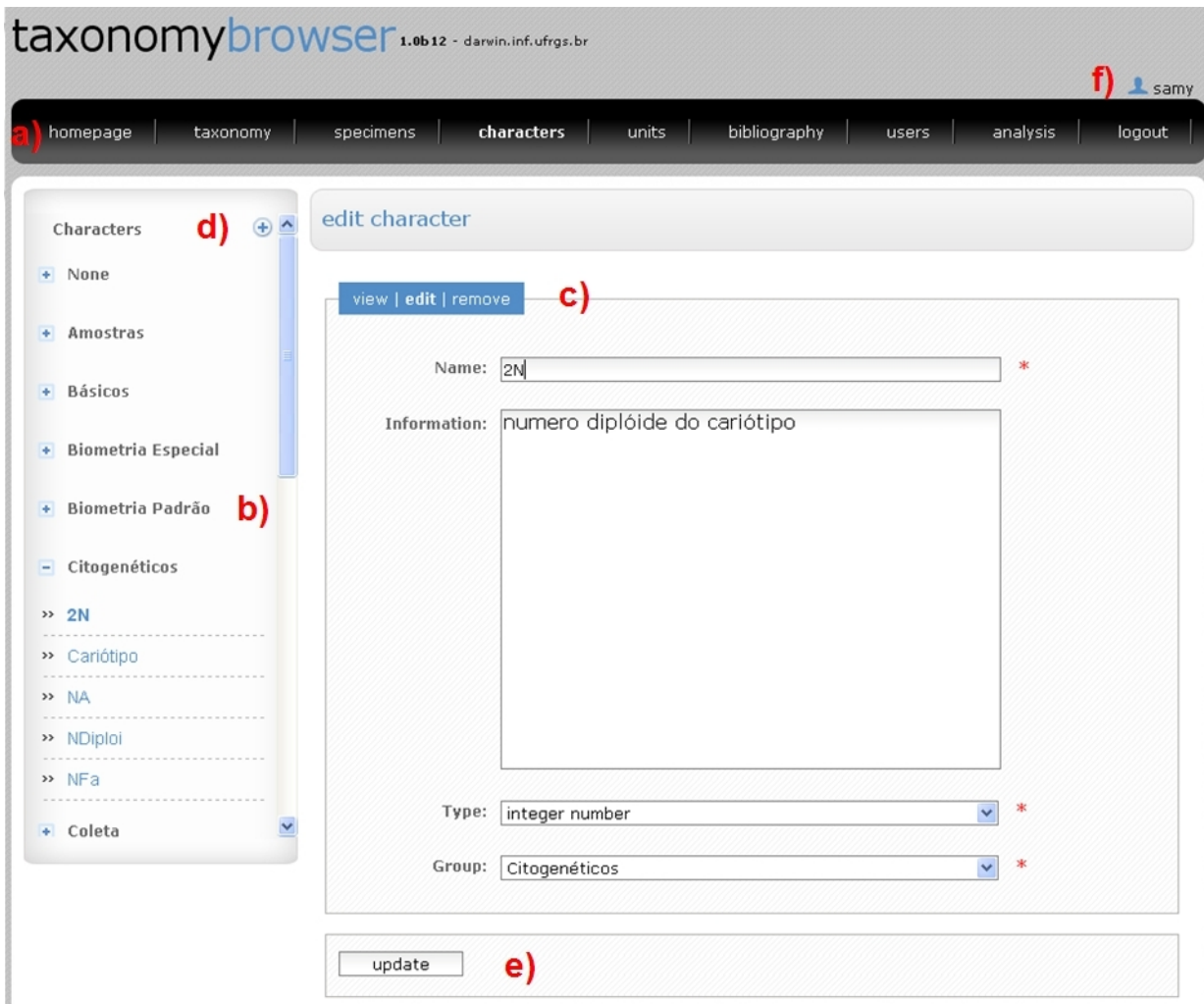


Figura 5.2: Layout padrão do sistema para todas as entidades.

trada na Figura 5.2(c). Após editar os dados sobre a entidade, deve-se aplicar a operação para refletir a mudança na camada de dados. Isso é feito através do botão de ação da operação, como mostra a Figura 5.2(e). A interface também mostra informações sobre o usuário que está autenticado no sistema (ver Figura 5.2(f)).

Na Figura 5.3 podem ser observadas as informações sobre um nodo da árvore taxonômica. No caso, está sendo visualizado um nodo que representa a subfamília *cavinaae*. Já na Figura 5.1, observa-se a inserção um novo exemplar da espécie *talomys venustus* no banco de dados. Na inserção de um exemplar de uma determinada espécie, o usuário deve informar os dados indicados como obrigatórios (marcados com *), os quais são, o nível taxonômico (*species*, no caso deste exemplo), o nome da espécie, a identificação do espécime (*Collection ID*) e o coletor. As demais informações que aparecem na figura são herdadas do nível taxonômico *species* e dos superiores na árvore taxonômica.

5.2 Interface de consulta

A interface de consulta para usuários não cadastrados está restrita à navegação na árvore taxonômica, acessando as informações dos nodos taxonômicos nos diversos níveis, ao acesso ao conjunto de exemplares a partir da aba specimens (Fig. 5.5), e ao uso de

taxonomybrowser 1.0b12 - darwin.inf.ufrgs.br

admin

homepage | taxonomy | specimens | characters | units | bibliography | users | analysis | logout

caviinae {subfamily} +

Genus

>> cavia

view taxonomy information

view | edit | remove

Scientific Name: caviinae

Rank Name: subfamily

Lineage: Chordata; Mammalia; Rodentia; Caviidae;

Information:

Characters

- none

Inherited:

- Amostras
- Básicos
- Biometria especial
- Biometria padrão
 - Rodentia {order} : Comprimento total - CT
- Citogenéticos
- Coleta
- Coletores
- Diagnósticos
- Estatus espécime
- Etários
- Filogenéticos
- Genéticos - DNA

Figura 5.3: Visualização de um nodo da árvore taxonômica.

Specimens **add specimen**

new specimen:

Rank Name: species *

Scientific Name: **C** calomys callosus
E calomys expulsus
L calomys laucha
P cavia porcellus
S cavia sp
calomys sp
T calomys tener
V calomys venustus *

Collection ID: xd3rf *

Collected By: Pedro C. *

Collected Date: 23 / 05 / 2008

Latitude: -22.906014

Longitude: -47.060571

Information: The Córdoba vesper mouse (*Calomys venustus*) is a rodent species in the Cricetidae family from South America . It is endemic to the area of Córdoba Province, central Argentina, where it is found in the espinal (dry lowland thorn brush and grassland).

References: **A** authors (title)

add

Figura 5.4: Inserção de um exemplar da espécie *calomys venustus* sistema.

consultas já programadas por usuários cadastrados.

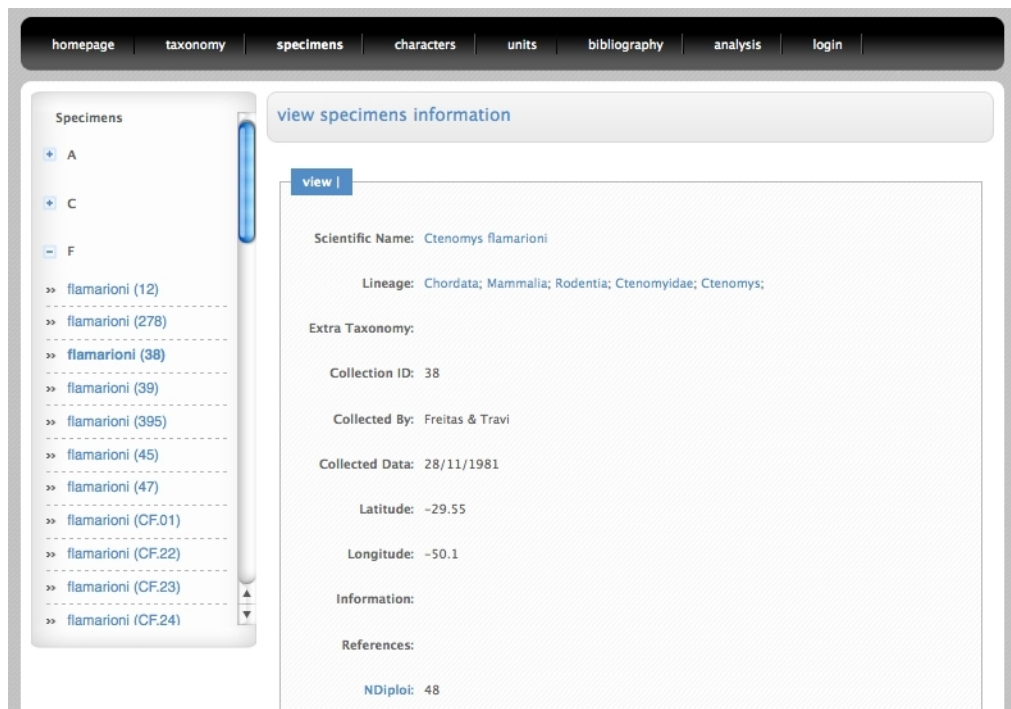


Figura 5.5: Navegação na coleção mostrando lista de exemplares e dados de um exemplar.

Os usuários cadastrados podem pesquisar por espécimes registrados no banco de dados, usando características relevantes, e apresentando-os em um mapa, se eles possuírem coordenadas geográficas de coleta. O usuário pode prosseguir, buscando dados específicos de um exemplar ou utilizando algum script com os dados recuperados na consulta. A Figura 5.6 ilustra as possibilidades de uso da interface visual.

Para especificar os critérios de busca de dados, o usuário utiliza um editor de buscas, combinado as características dos espécimes em expressões com uma operação (que pode ser $<$, $>$, $<=$, $>=$, $=$, $!=$, *between* e *like*) e dois valores, como podemos ver na Figura 5.7.

A operação *between* é a única que necessita de dois valores. As outras operações, entretanto, utilizam apenas o primeiro valor. Por exemplo, "comprimento da cauda" *between* 10 e 15 retornará todos os espécimes que tem o campo "comprimento de cauda" com valores entre 10 e 15. A operação *like* é utilizada como, por exemplo, *like* "%con", retornando todos os espécimes cujo nome termina em "con". Pode-se escolher a cor do ícone que representará os espécimes resultantes da busca sobre o mapa, permitindo diferenciar grupos de espécimes, facilitando a visualização e, por consequência, o estudo das interações de convivência entre espécies. A especificação de uma busca é armazenada para utilização posterior, sendo possível sua modificação ou a reexibição do resultado.

5.3 Interface sobre mapa

O tratamento de mapas é realizado através da API do Google Maps (GOOGLE, 2010a), a qual consiste basicamente de um conjunto de classes JavaScript que fornecem a interface necessária para construir aplicações sobre mapas, realizar consultas por endereços em coordenadas geográficas, acrescentar descrições a pontos no mapa, dentre outras funcionalidades.

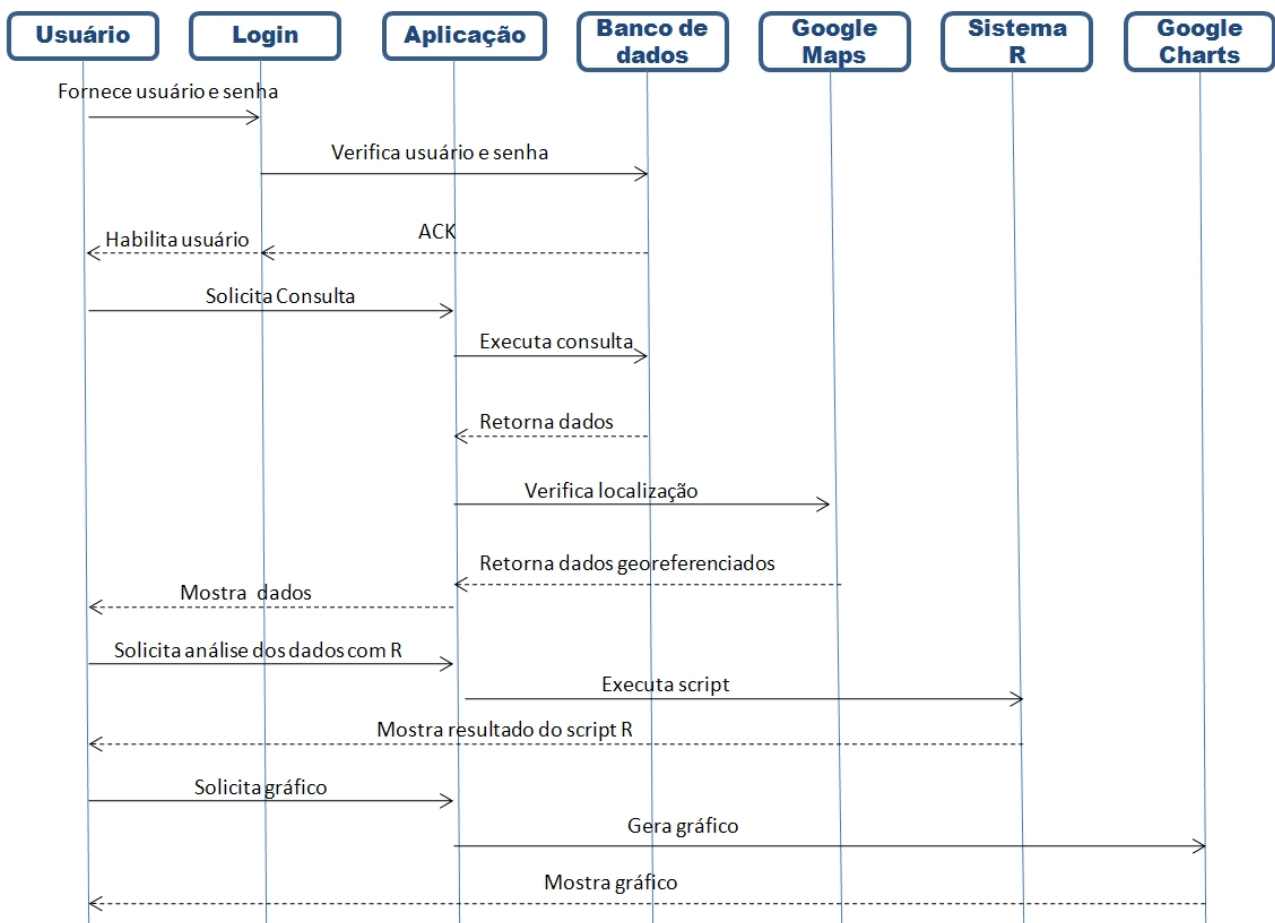


Figura 5.6: Diagrama de sequência de consulta.

A primeira versão da API Google Maps foi disponibilizada em junho de 2005 e, desde então, as funcionalidades estão em constante desenvolvimento. Qualquer pessoa responsável por um domínio da internet pode utilizar a API gratuitamente, desde que possua uma chave ("API Key"). Essa chave consiste de uma string que deve ser utilizada sempre que se incluir a API em uma página web. Ela é distribuída gratuitamente pelo Google, mediante aprovação da proposta de uso da página web onde a API será inserida. Os termos de uso não contêm maiores restrições, a não ser pela proibição em tentar alterar e/ou esconder o logotipo do Google embutido nos mapas. Uma outra restrição é quanto ao uso dos mapas em outros formatos: é proibida a utilização dos mapas em outros meios digitais que não as páginas autorizadas pelo Google (ou seja, que possuam uma chave).

Como pode-se ver na parte superior do mapa na Figura 5.8, existem cinco modos de visualização padrão: Mapa, Satélite, Híbrido, Terreno e *Earth*. O modo Mapa é apresentado na forma de um mapa político com estradas, ruas e seus respectivos nomes. No modo Satélite, a visualização dos mapas se dá através de fotos de satélites e, em certas áreas, por aerofotogrametria. O modo Híbrido é a sobreposição dos modos Mapa e Satélite de tal forma que o mapa político é sobreposto transparentemente ao mapa geográfico. O Earth faz a cartografia do planeta, agregando imagens obtidas de várias fontes, incluindo imagens de satélite, fotografia aérea e sistemas de informação geográfica sobre um globo em 3D. Para a visualização em modo Earth é necessária a instalação de um plugin JavaScript no navegador.

taxonomybrowser 1.0b5 - darwin.inf.ufrgs.br

admin

homepage taxonomy specimens characters units bibliography users searches logout

Searches

T

>> todos calomys sp

add search

new search:

Name: estudo calomys sp 1 *

Color: #429C72 *

Rank Name: species *

Scientific Name:

- C calomys callosus
- E calomys expulsus
- L calomys laucha
- P cavia porcellus *
- S cavia sp
- calomys sp
- T calomys tener
- V calomys venustus

Information: Todos calomys sp com peso entre 9 e 22 gramas.

2 n: integer number

n local: integer number

peso em gramas: physical unit BETWEEN 9.0 22.0

add

Figura 5.7: Interface visual do editor de buscas.

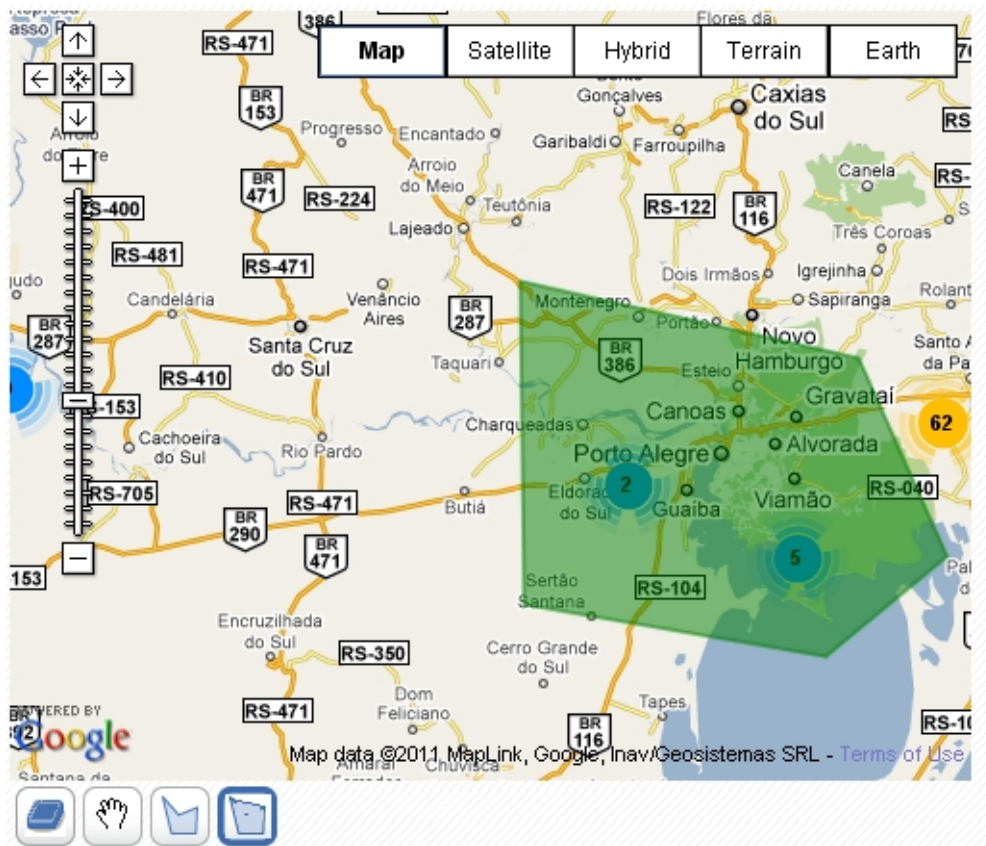


Figura 5.8: Interface visual mostrando o resultado de busca sobre o mapa.

A interface desenvolvida permite interagir com os dados apresentados na forma de ícones. Marcações na forma de polígonos permitem agrupar espécimes resultantes das buscas, selecionando-os para uma operação posterior. Também é possível gerar a envol-tória convexa de cada consulta, demarcando a menor área onde se encontram todos os espécimes recuperados com uma determinada busca.

A Figura 5.8 apresenta o resultado da busca especificada na Figura 5.7, com todos os exemplares de *calomys sp* com peso entre 9 e 22 gramas cadastrados até o momento. Estes espécimes estão representados no mapa como ícones que agrupam os exemplares encontrados em regiões próximas, indicando o número de exemplares registrados para aquela localização. Para realizar a análise de um subgrupo desses espécimes, podemos separá-los desenhando um polígono e, assim, considerar apenas os espécimes que estão incluídos na região demarcada.

Nas Figuras 5.9a e 5.9b podemos ver que a exibição com agrupamento permite um "zoom in", e os grupos de espécimes vão sendo desagrupados, até que estejam representadas as ocorrências individuais. Na Figura 5.9b podemos ver que um simples clique do mouse nos dá acesso às informações referentes ao indivíduo selecionado.

5.4 Interface de análise

O sistema TaxonomyBrowser provê uma interface para geração de gráficos e para análise dos dados resultantes de uma busca. Essa interface de análise permite aos usuários submeter os dados de espécimes resultantes de uma busca a um script na linguagem R, a ser executado no sistema R instalado no mesmo servidor, e à API Google Charts. O

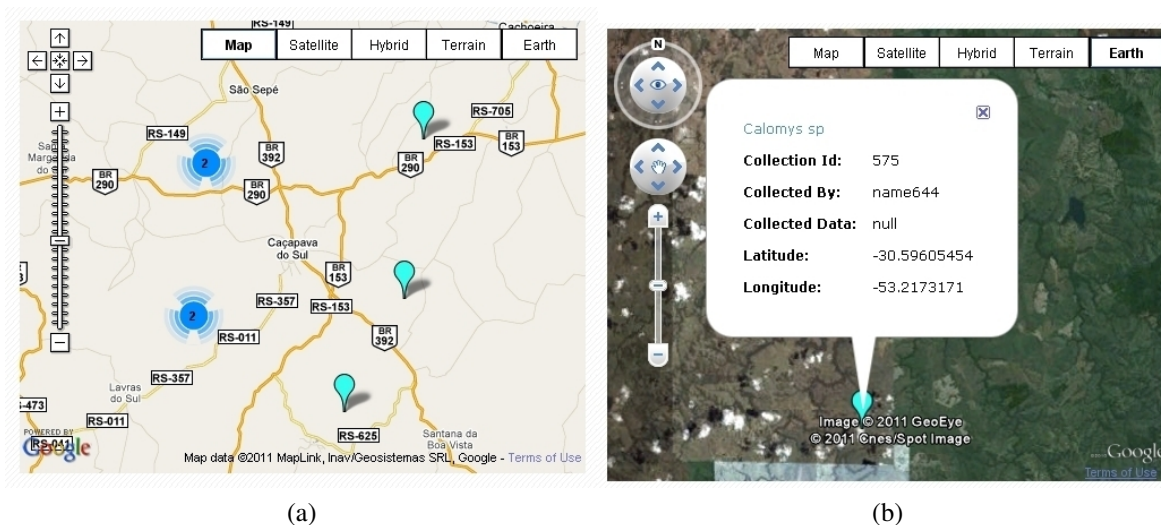


Figura 5.9: Visualização do resultado da busca sobre o mapa. (a) Com os espécimes agrupados. (b) Após selecionar um espécime específico.

o sistema também oferece a opção de salvar os dados de uma consulta para continuar a análise em um momento posterior, além de permitir que sejam registrados novos scripts.

5.4.1 Análise com sistema R

O ambiente R é ao mesmo tempo uma linguagem e um sistema interpretado que fornece uma ampla variedade de funções, como modelagem linear e não linear, testes estatísticos clássicos, análise de séries temporais, classificação, clustering, entre outros. A geração de gráficos está disponível também a partir de scripts.

O ambiente R está disponível gratuitamente como Software Livre sob os termos da GNU da Free Software Foundation General Public License. O fato de possuir licença livre fez com que vários grupos de pesquisa desenvolvessem centenas de módulos de diversas áreas do conhecimento: biologia, computação, matemática, física, economia, entre outras (R-PROJECT, 2000). Além dos benefícios já citados anteriormente, o ambiente R é, atualmente, utilizado pelos biólogos usuários de nosso sistema. Portanto, a integração do ambiente R ao TaxonomyBrowser foi um requisito do projeto.

Essa integração foi possível através de comunicação entre processos (IPC). Entre as soluções de IPC implementadas em sistemas UNIX, a que melhor atendia às necessidades da integração do TaxonomyBrowser com o ambiente R é a *named pipe*. A passagem de dados entre o TaxonomyBrowser e o sistema R é realizada através da exportação dos dados em formato .csv, pelo TaxonomyBrowser e sua conversão para XML, para importação pelo ambiente R (HENKIN, 2010). Posteriormente, o sistema retorna o resultado do script (normalmente uma imagem) ou um erro, caso do script contenha problemas de compatibilidade, por exemplo.

O usuário realiza uma busca, conforme explicado nas seções anteriores, e selecionando *export* na aba de operações, tem acesso aos dados, como pode ser observado na Figura 5.10. Nessa interface ele pode selecionar linhas e colunas da tabela exibida e exportar os dados para o script escolhido. O resultado da busca é exibido na tela, normalmente através de uma imagem, como podemos ver na Figura 5.11. Um usuário com experiência em R pode implementar scripts para análise de dados, enquanto outros usuários, menos familiarizados com a linguagem, podem simplesmente utilizar os scripts que estão dis-

view | edit | remove | export

Name: Lami grafico

Information:

Scientific Name: [Ctenomys lami](#)

Analysis Filters:

- Comprimento do Cranio: > 0.0
- Comprimento nasal: > 0.0
- Largura da mandibula: > 0.0
- Comprimento total - CT: > 0.0

Result:

specimen_id	taxonomy_id	scientific_name	taxonomy_rank_name	col
80	207	Ctenomys lami	species	TR.
81	207	Ctenomys lami	species	TR.
136	207	Ctenomys lami	species	TR.
137	207	Ctenomys lami	species	TR.
138	207	Ctenomys lami	species	79
139	207	Ctenomys lami	species	TR.

Export:

Processing:

```
script03.r
script02.r
script01.r
```

Graphical Visualizations:

```
Parallel Plot
Heat Map
Bar Chart
Table
Magic Table
```

Figura 5.10: Interface de análise mostrando a tabela resultado de uma busca que pode ser exportada para a execução de um script selecionado da lista apresentada ou para visualização em uma das formas também exibida nessa interface.

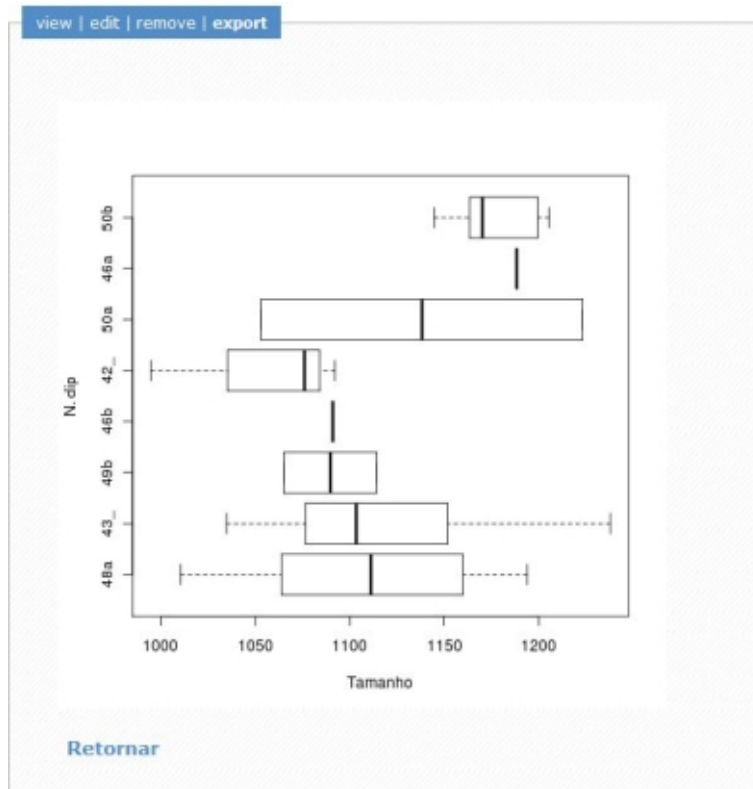


Figura 5.11: Resultado do processamento de script R.

poníveis no TaxonomyBrowser. Como os scripts manipulam as informações oriundas do banco de dados, os nomes de variáveis referenciadas no script devem ser compatíveis com os nomes retornados pela busca. Por exemplo, um script deve esperar que coordenadas geográficas sejam chamadas de Longitude e Latitude, com as letras iniciais em maiúsculo. O script deve referenciar o nome exato da tabela a ser analisada. Como os scripts são criados em um programa externo (o sistema R local ou um editor de textos) e enviados pelos próprios usuários para que sejam utilizados no TaxonomyBrowser, a responsabilidade pela correção de cada script é do respectivo autor, ou seja, o TaxonomyBrowser não realiza nenhuma verificação de correção dos scripts a ele submetidos.

5.4.2 Visualizações alternativas

A API Google Charts é uma ferramenta utilizada para gerar gráficos dinamicamente na Web. Não é necessária a instalação de qualquer software ou frameworks, portanto, não é necessário nem uma alteração no Taxonomy Browser. Esta API está disponível como Software Livre sob os termos da GNU.

Similar à utilização do ambiente R, como podemos ver na Figura 5.10, o usuário acessa o portal de dados, realiza uma busca, seleciona os dados da tabela sobre os quais deseja gerar um gráfico, escolhe, de acordo com a sua necessidade, qual visualização deseja utilizar, e depois recebe o resultado na tela, normalmente uma imagem ou uma mensagem de erro caso os dados da tabela não forem corretamente selecionados. A inclusão da API Google Charts no sistema se deu para facilitar o acesso de usuários leigos, ou seja, sem familiaridade com a linguagem R. A interface de geração de gráficos através da seleção de linhas e colunas de um resultado de busca é intuitiva, permitindo que qualquer usuário, mesmo sem nenhuma experiência, possa gerar gráficos com as informações

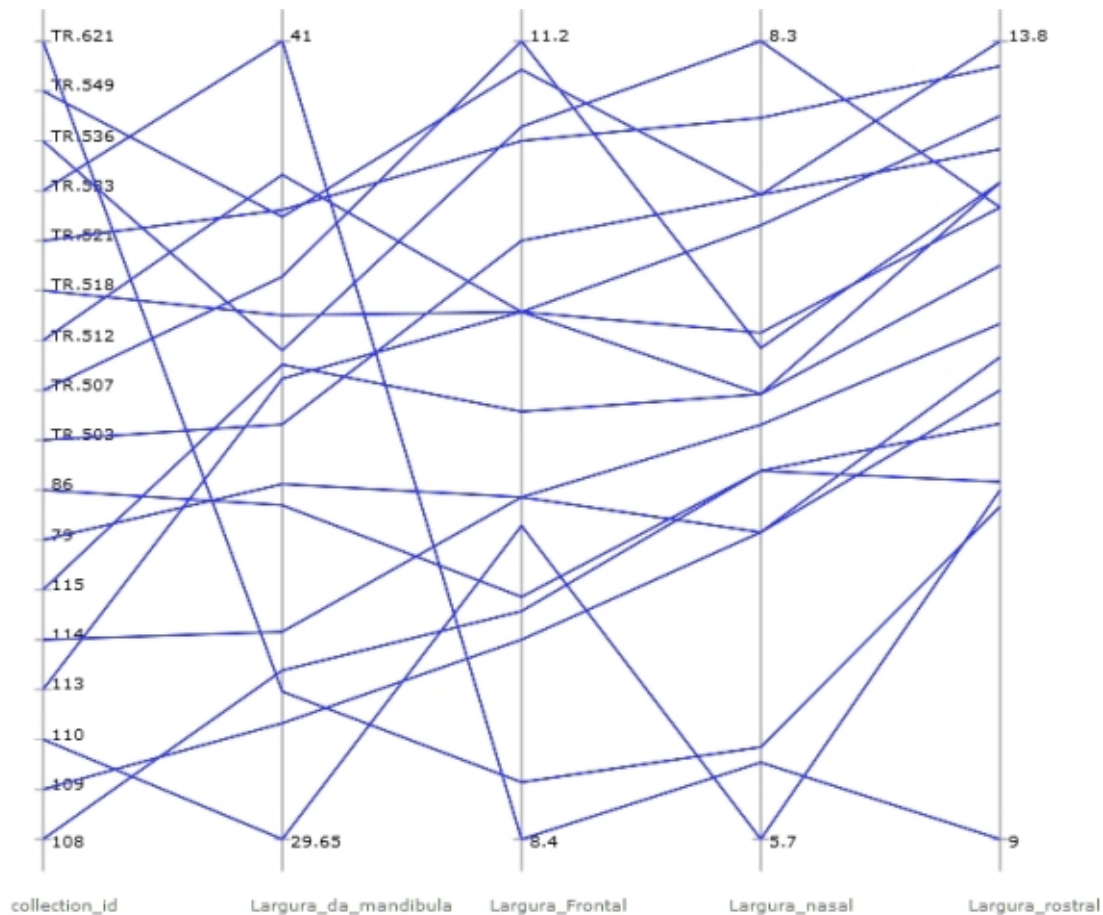


Figura 5.12: Visualização utilizando coordenadas paralelas de uma amostra de *Ctenomys flamarioni*.

de interesse de seu estudo.

Entre as várias opções de visualizações que oferece a API Google Charts, foram escolhidas as que mais satisfazem as necessidades do sistema. Entre elas estão:

- **Parallel Plot:** Coordenadas Paralelas é um método de visualização de dados multidimensionais (INSELBERG, 1985). Um espaço n -dimensional é representado por n linhas paralelas verticais, representando cada uma das dimensões, e os pontos nesse espaço são representados por linhas que cortam as verticais nas alturas relativas aos respectivos valores. No exemplo da Figura 5.12, a primeira linha vertical representa o identificador dos espécimes. As linhas seguintes representam as características que foram selecionadas para gerar o gráfico, lembrando que essas características devem ser numéricas. Cada exemplar é uma linha poligonal. A implementação da técnica na API é de (ALLAMRAJU, 2007).
- **Heat Map:** É um mapa de cores, comumente usado para visualizar matrizes de dados. Esta visualização suporta três cores básicas. Os valores mais baixos são representados pela cor verde, os intermediários por preto e os mais altos por vermelho. Assim como no gráfico *Parallel Plot*, a primeira coluna representa o identificador dos espécimes e as colunas seguintes representam as características que foram selecionadas para gerar o gráfico, como podemos ver na Figura 5.13. Neste exemplo, foram selecionados primeiramente os identificadores, depois as características correspondentes às medidas de crânio identificadas como M1 a M9.

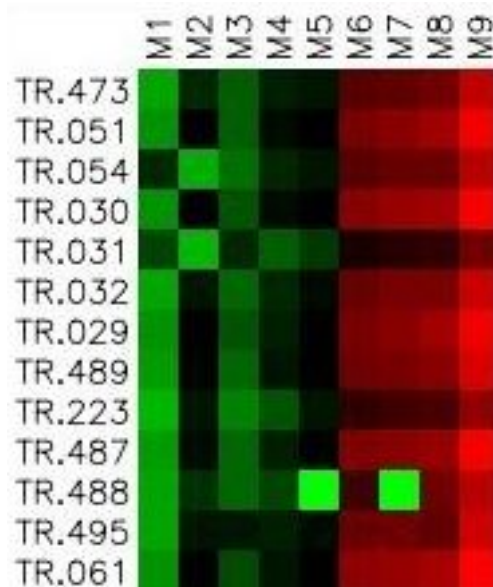


Figura 5.13: Visualização utilizando *heat map* de uma amostra de *Ctenomys minutus*.

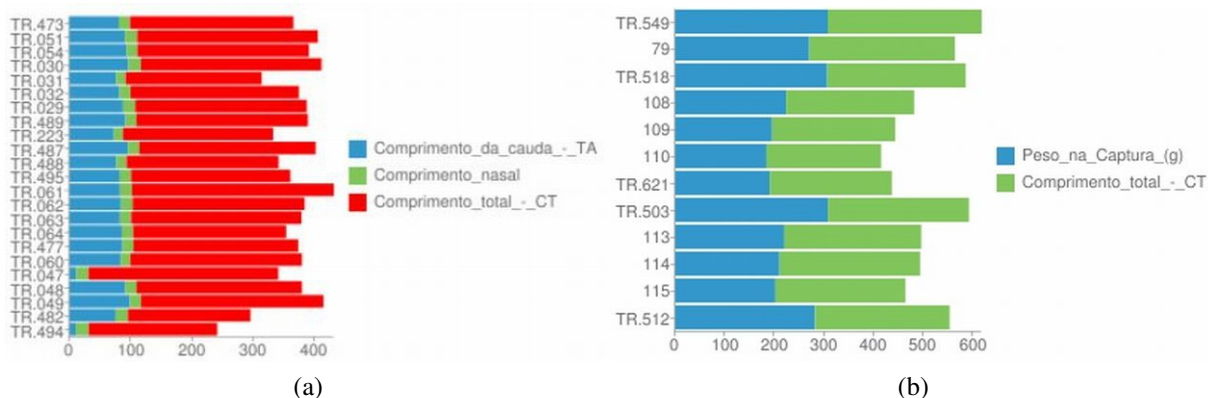


Figura 5.14: Visualização utilizando gráfico de barras de uma mostra de *Ctenomys lami*. (a) Gráfico com três características sendo comparadas. (b) Gráfico com duas características.

- *Bar Chart*: O gráfico de barras utiliza barras horizontais para representar os valores das características. Similar aos anteriores, a primeira coluna da tabela deve representar o identificador dos espécimes e as seguintes devem ser características numéricas, onde os valores serão representados através de barras coloridas. Cada característica será representada por uma cor, como podemos ver na Figura 5.14a, onde estão sendo comparados os comprimentos da cauda, os comprimentos nasais e os comprimentos totais de alguns espécimes da espécie *Ctenomys lami*.
- *Table*: Não é exatamente uma visualização, pois os dados selecionados são exibidos em formato de tabela, como podemos ver na Figura 5.15. Este formato permite mostrar os dados de forma explícita e é interessante, por exemplo, para a filtragem dos espécimes de interesse dentre aqueles resultantes de uma consulta. Outra motivação para a exibição em formato de tabela é a possibilidade de mostrar dados que não são numéricos, pois as técnicas providas pela API Google Charts não suportam o mapeamento de atributos simbólicos para gráficos.

	collection_id	latitude	longitude	Comprimento_da_cauda_TA	Comprimento_do_Cranio	Largura_da_mandibula	Peso_na_Captura_(g)
1	TR.473	-30.6602742	-51.4123021	80	45.5	35.1	187
2	45	-30.6602742	-51.4123021	89	44.6	34.6	178
3	47	-29.758801	-50.047575	87	44.3	34.15	219
4	TR.051	-29.758801	-50.047575	90	50	39.25	300
5	TR.054	-29.758801	-50.047575	93	45	34.4	180
6	TR.030	-30.174684	-50.207592	96	50.7	38.75	250
7	TR.031	-30.174684	-50.207592	76	37.4	28.3	85
8	TR.032	-30.174684	-50.207592	81	45.5	36	200
9	TR.029	-30.174684	-50.207592	87	49.4	38.7	270
10	TR.489	-30.174684	-50.207592	91	48	37.45	230
11	395	-30.7	-50.633333	85	41	31	150
12	TR.223	-30.7	-50.633333	72	40	31	115
13	TR.487	-30.542222	-52.576111	95	51.1	37.5	286
14	27B	-30.542222	-52.576111	88	46.4	36.4	236
15	TR.488	-30.542222	-52.576111	76	43.1	33.05	189
16	TR.495	-30.542222	-52.576111	82	45.7	36	210
17	TR.061	-30.542222	-52.576111	82	51.25	39.6	330
18	TR.062	-30.542222	-52.576111	84	46.1	35.33	208
19	TR.063	-30.542222	-52.576111	82	46.35	33.55	183
20	TR.064	-30.542222	-52.576111	85	47.5	31.4	147
21	TR.477	-30.542222	-52.576111	86	46.25	36	236
22	TR.060	-30.542222	-52.576111	83	45.9	36	215
23	TR.047	-29.5	-50.1	10.2	51	39.5	300
24	TR.048	-29.5	-50.1	91	46.5	36.3	200
25	TR.049	-29.5	-50.1	98	49.3	37.2	200

Figura 5.15: Tabela com algumas características de uma amostra de *Ctenomys flamarioni*.

5.4.3 Exemplo de cenário de uso

Nesta seção, resumidamente, apresentamos um cenário de uso do TaxonomyBrowser para consulta de dados sobre espécimes já cadastrados. Corresponde a uma das alternativas de início de atividade de pesquisa apresentadas na Figura 3.1.

Neste cenário, o banco de dados contém registros sobre espécimes de *Ctenomys minutus*, e o usuário quer investigar a relação existente entre populações dessa espécie, que apresentam números cromossômicos (número diplóide) entre 46 e 48, o que pode denotar o que se costuma chamar de zona híbrida, ou seja, animais com 46 cromossomos cruzam com animais com 48 cromossomos, o que resulta animais com 47 cromossomos.

Para avaliar essa hipótese, o usuário prepara uma busca pelos indivíduos com número diplóide entre 46 e 48 cromossomos (Figura 5.16), observando a existência de 114 exemplares nessas condições.

Os dados desses 114 exemplares podem ser visualizados na forma de coordenadas paralelas (Figura 5.17, o que permite analisar uma possível relação entre dados de biometria padrão (comprimentos e largura) e número diplóide. Observa-se, ainda, um indivíduo para o qual não estão registradas medidas e que, portanto, deve ser excluído de qualquer análise posterior e/ou anotado para correção.

Na sequência, sucessivamente, o usuário prepara buscas diferentes individualizando os grupos de indivíduos por número diplóide (Figuras 5.18a, 5.18b e 5.19).

O usuário pode, também, decidir por alguma análise estatística com todo o conjunto de dados ou com apenas parte dele. Pode, por exemplo, executar o script cujo resultado foi exibido na Figura 5.11.

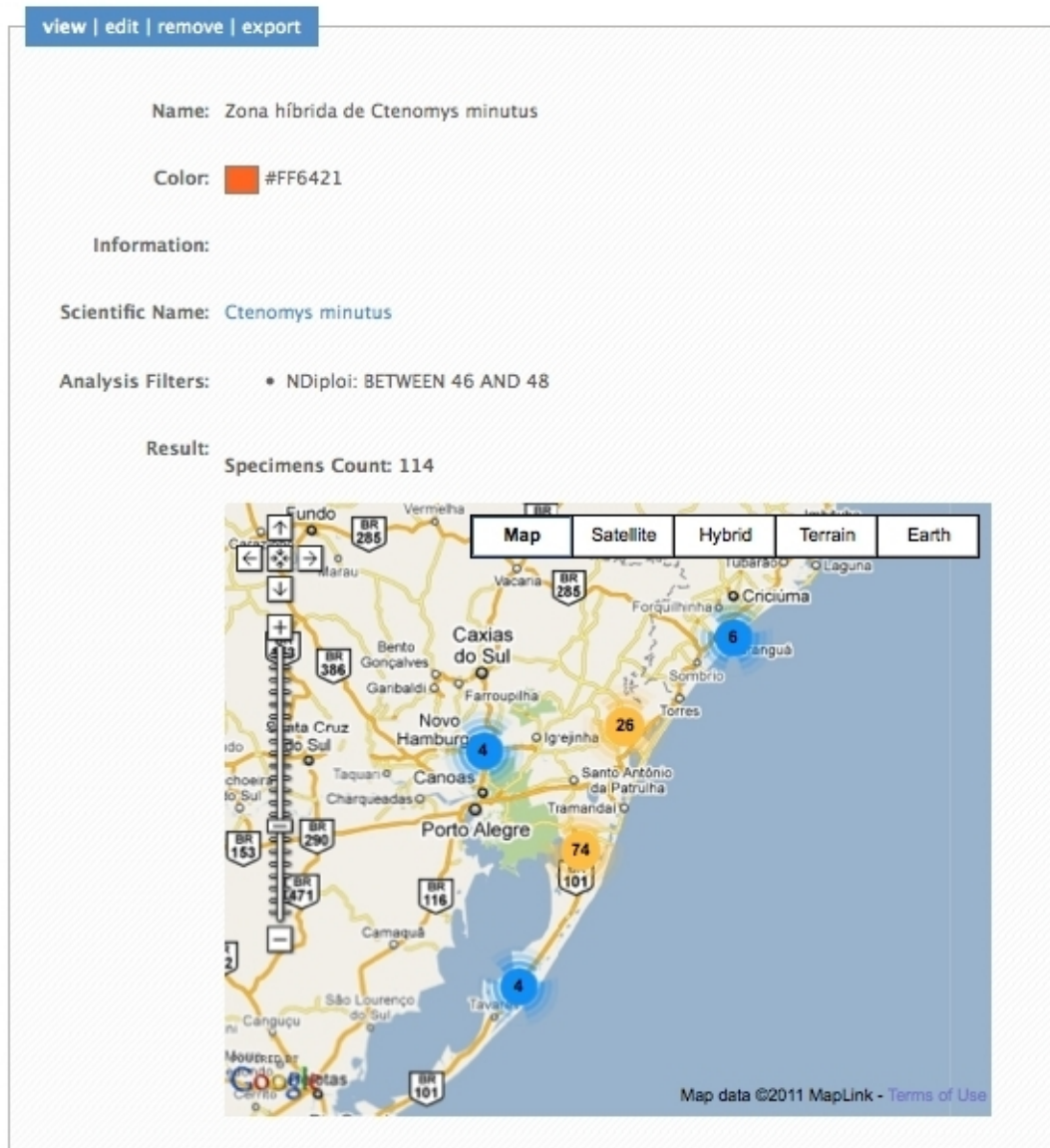


Figura 5.16: Resultado da busca pelos *Ctenomys minutus* com número cromossômico de 46 a 48.

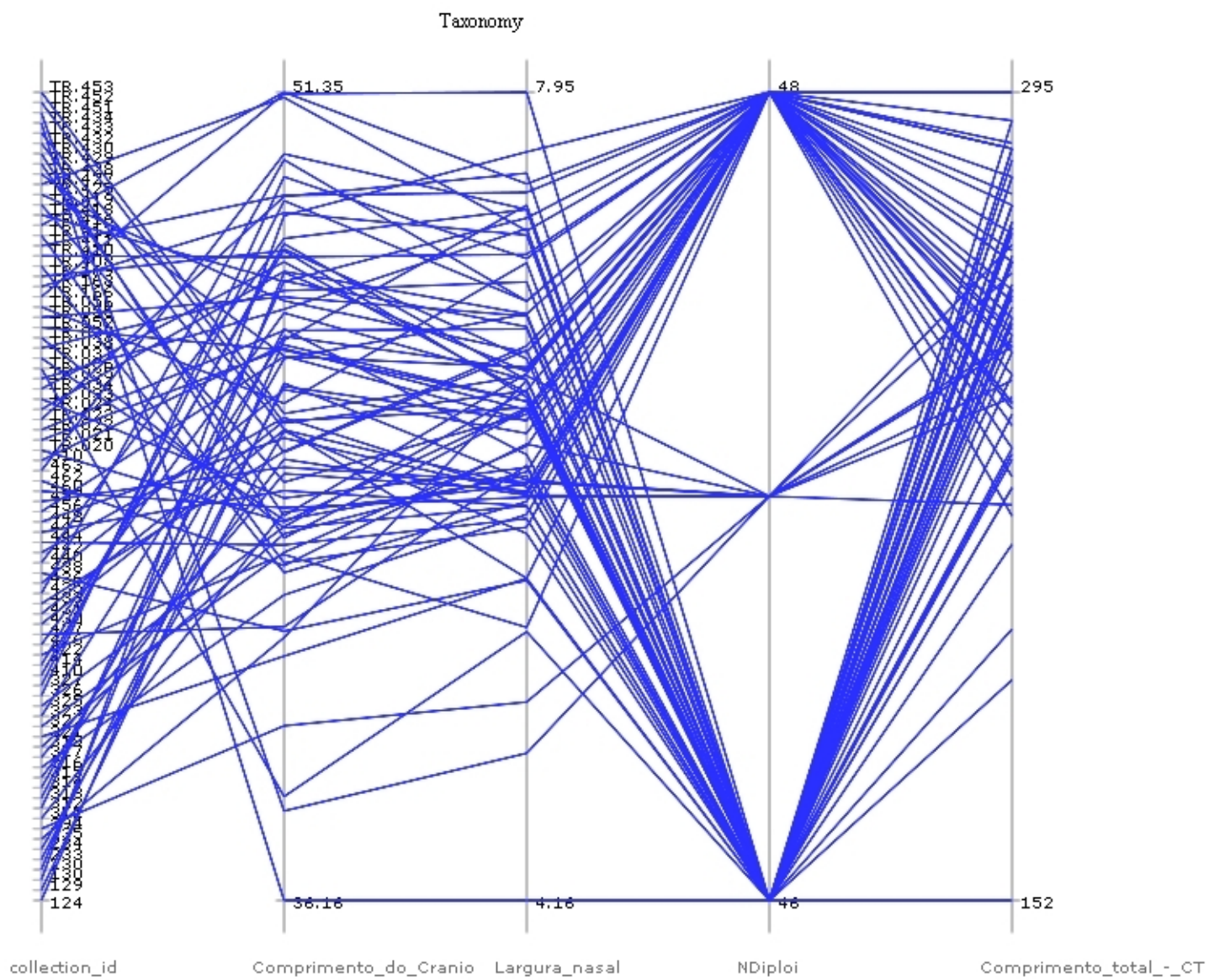
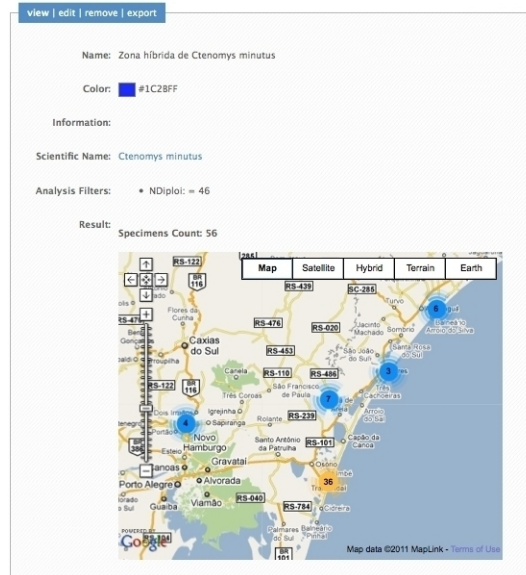
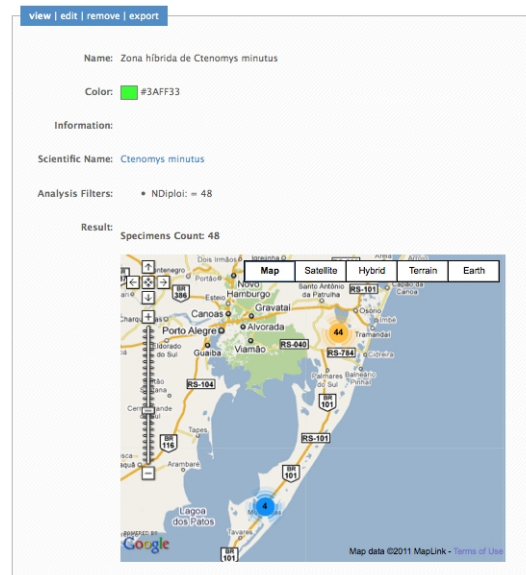


Figura 5.17: Coordenadas paralelas exibindo 114 exemplares de *Ctenomys minutus* com número cromossômico de 46 a 48.



(a)



(b)

Figura 5.18: Resultados das buscas por *Ctenomys minutus* com 46 (a) e 48 (b) cromossomos.

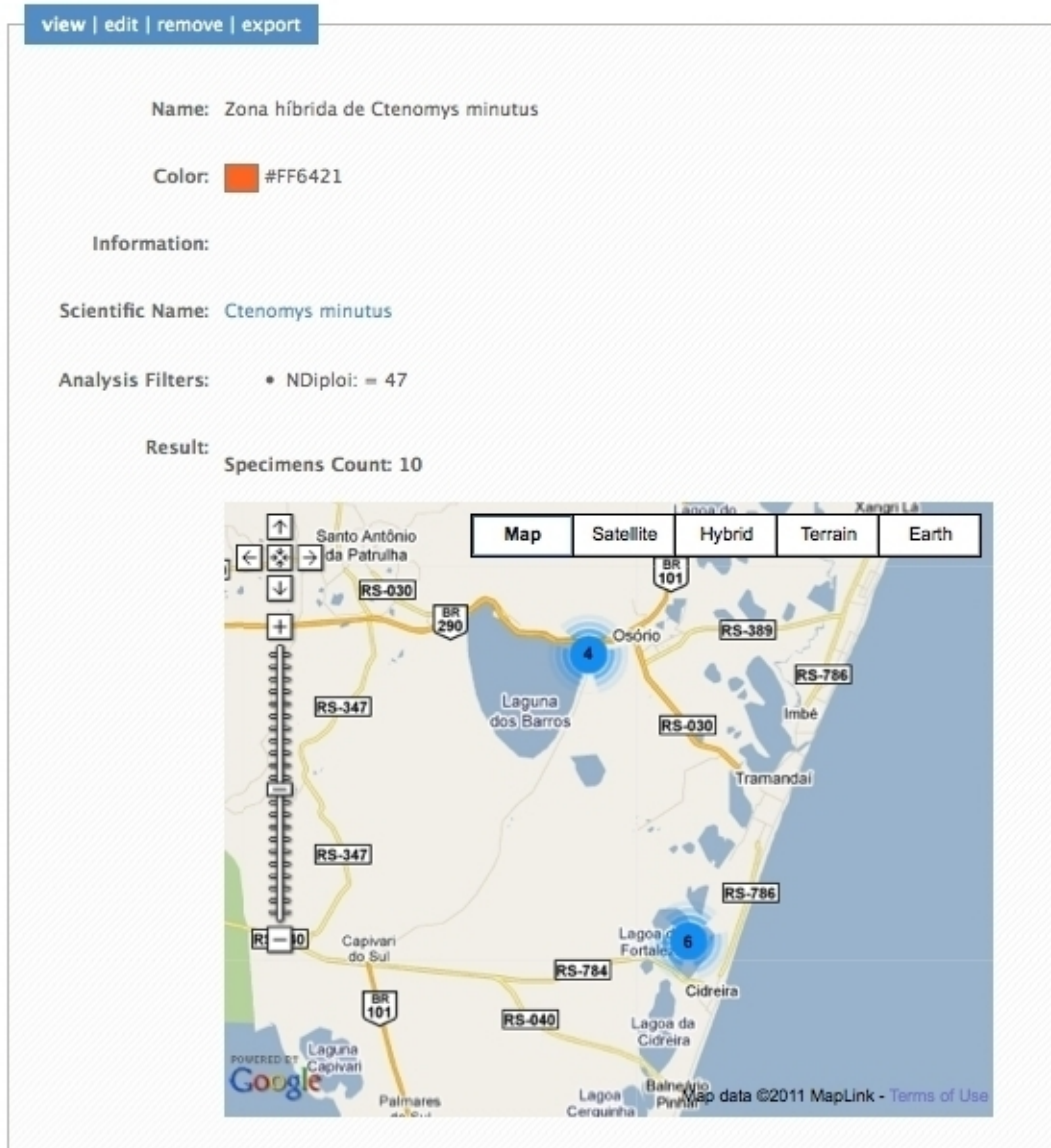


Figura 5.19: Resultados das buscas por *Ctenomys minutus* com 47 cromossomos.

6 CONCLUSÕES E TRABALHOS FUTUROS

O trabalho apresentado nesta dissertação se concentrou nos desafios relacionados à representação, integração e ao compartilhamento de dados típicos de pesquisas em biodiversidade. As dificuldades encontradas neste contexto estão relacionadas à possível grande quantidade de dados, a heterogeneidade desses dados e ao provimento de suporte de visualização e analítico para atividades de pesquisa que são bastante diversificadas. Este cenário demanda soluções flexíveis, que se adaptem a alterações e, ao mesmo tempo preservem a consistência dos dados.

Para tratar os problemas e necessidades levantados na fase de análise de requisitos do projeto, esta dissertação apresentou a especificação e a implementação de um banco de dados e de uma aplicação web, denominada TaxonomyBrowser, para gerenciamento, consulta e análise dos dados.

A utilização da API Google Maps permite a visualização das informações georreferenciadas através de mapas. E a integração com o sistema R e com a API Google Charts permite analisar graficamente os dados dentro do próprio sistema, sem a necessidade de mudança de ambiente. A generalidade do modelo permite também que seja utilizado para descrever coleções de museus ou de outros grupos de pesquisa, envolvendo outros organismos (outras classificações taxonômicas) e não somente mamíferos.

Uma outra característica é a possibilidade de integração do banco de dados com outros sistemas pertencentes a este domínio, através da adoção do padrão de metadados conhecido como Darwin Core, muito utilizado nas bases de dados e projetos de sistemas de biodiversidade.

A principal contribuição do TaxonomyBrowser é a integração de vários procedimentos em uma única plataforma que, além de servir como repositório de dados de coleção, é, também, uma plataforma de análise e divulgação de dados. Não há registro conhecido de ambiente integrado como o aqui descrito. Vale ressaltar que toda a concepção do sistema, incluindo o modelo do banco de dados, teve a participação direta de pesquisadores da área de aplicação, ou seja, de usuários do ambiente.

O TaxonomyBrowser está sendo amplamente avaliado pelos biólogos do Departamento de Genética da UFRGS e da ONG Mamíferos RS, assim como está em utilização experimental por pesquisadores da Fundação Oswaldo Cruz. O sistema está disponibilizado no endereço eletrônico <http://darwin.inf.ufrgs.br/taxonomybrowser>.

Dessas avaliações pelos biólogos já existem melhorias a serem feitas, que são listadas a seguir, juntamente com as extensões pretendidas como trabalhos futuros:

- Sobreposição de consultas diferentes no mesmo mapa.
- Possibilidade de anotação nos mapas e gráficos.
- Substituição da API Google Charts para obtenção de gráficos por implementações

próprias, interativas.

- Incorporação de outras técnicas de visualização sobre os mapas resultantes de consultas, como, por exemplo, a visualização de árvores filogenéticas (PARKS et al., 2010), que permitem estudos evolutivos.
- Integração de dados sobre o ambiente, como dados de sensores, por exemplo, para associação com os espécimes coletados.
- Melhoria das consultas implementadas utilizando técnicas de otimização de consultas.
- Aprimoramento da integração com o ambiente R, permitindo a manipulação de scripts dentro do TaxonomyBrowser assim como a edição dos arquivos de resultados de consultas a serem repassados para os scripts.

Finalmente, vale ressaltar que o Taxonomy Browser foi desenvolvido para dar suporte a todas as tarefas de coleta de dados de exemplares explicitados na Figura 3.1 e apresenta todas as características essenciais para sistemas de informações de biodiversidade, como mostrado na Figura 2.17.

REFERÊNCIAS

- ABCD. **Access to Biological Collection Data**. Disponível em:<<http://www.bgbm.org/tdwg/codata/>>, Acesso em: 10 jul. 2010.
- ALLAMRAJU, H. **Parallel Coordinates**. Disponível em:<<http://sites.google.com/site/sriharsha451/>>, Acesso em: 10 jul. 2010.
- ARCTOS. **Multi-Institution, Multi-Collection Museum Database** . Disponível em:<<http://arctos.database.museum/home.cfm>>, Acesso em: 10 jul. 2010.
- BERKELEYMAPPER. **BerkeleyMapper**. Disponível em:<<http://berkeleymapper.berkeley.edu/>>, Acesso em: 10 jul. 2010.
- BIOCORE. **Biodiversity and Computing Research**. Disponível em:<<http://www.lis.ic.unicamp.br/projects/biocore>>, Acesso em: 10 jul. 2010.
- CAÑETE, S. C. et al. **Integrando Visualização e análise de dados em sistema de gerenciamento de dados de biodiversidade**. Belo Horizonte, MG: Congress of the Brazilian Computer Science, SBC, 2010.
- DALTIO, J. **Aondê: um serviço web de ontologias para interoperabilidade em sistemas de biodiversidade**. Campinas: UNICAMP, 2007. 82 f. Dissertação (Mestrado) - Programa de Pós-Graduação em Computação, Universidade Estadual de Campinas, Campinas, 2007.
- DALTIO, J. et al. A framework to process complex biodiversity queries. In: ACM SYMPOSIUM ON APPLIED COMPUTING, 2008., New York, NY, USA. **Proceedings...** ACM, 2008. p.2293–2297. (SAC '08).
- DAYRAT, B. Towards integrative taxonomy. **Biol. J. Linn. Soc.**, [S.l.], v.85, n.3, p.407–415, 2005.
- FREE-SOFTWARE-FOUNDATION. **GNU Operating System**. Disponível em:<<http://www.gnu.org/>>, Acesso em: 10 jul. 2010.
- FUTUYMA, D. J. **Evolution**. Sunderland, MA 01375, USA: Sinauer Associates, Inc., 2009.
- GBIF. **Global Biodiversity Information Facility**. Disponível em:<<http://www.gbif.org/>>, Acesso em: 10 jul. 2010.

GOMES, L. C. **Uma Arquitetura para Consulta a Repositórios de Biodiversidade na Web**. Campinas: UNICAMP, 2007. 65 f. Dissertação (Mestrado) - Programa de Pós-Graduação em Computação, Universidade Estadual de Campinas, Campinas, 2007.

GOOGLE. **Google Maps API**. Disponível em:<<http://code.google.com/apis/maps/index.html>>, Acesso em: 10 jul. 2010.

GOOGLE. **Google Chart API**. Disponível em:<<http://code.google.com/apis/chart/>>, Acesso em: 10 jul. 2010.

GURALNICK, R.; NEUFELD, D. Challenges Building Online GIS Services to Support Global Biodiversity Mapping and Analysis: lessons from the mountain and plains database and informatics project. **Biodiversity Informatics**, [S.l.], v.2, n.0, 2005.

HENKIN, R. **Interface de Consultas Analíticas para Bases de Dados de Biodiversidade**. Porto Alegre: UFRGS, 2010. 39 f. Trabalho de Conclusão de Curso (Graduação) - Bacharelado em Ciência da Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010.

IABIN. **Inter-American Biodiversity Information Network**. Disponível em:<<http://sinbiota.cria.org.br/>>, Acesso em: 10 jul. 2010.

INSELBERG, A. The Plane with Parallel Coordinates. **The Visual Computer. Special Issue on Computational Geometry**, [S.l.], v.1, p.69–91, 1985.

IPT. **Integrated Publishing Toolkit**. Disponível em:<<http://ipt.gbif.org/>>, Acesso em: 10 jul. 2010.

LIFE-PROJECT, T.-T. of. **Tree of Life**. Disponível em:<<http://tolweb.org/tree/>>, Acesso em: 10 jul. 2010.

MALAVERRI, J. G. **Um Serviço de Gerenciamento de Coletas para Sistemas Biodiversidade**. Campinas: UNICAMP, 2007. 76 f. Dissertação (Mestrado) - Programa de Pós-Graduação em Computação, Universidade Estadual de Campinas, Campinas, 2009.

MALAVERRI, J. G.; VILAR, B. S. C. M.; MEDEIROS, C. B. A Tool based on Web Services to Query Biodiversity Information. In: WEBIST. **Anais...** [S.l.: s.n.], 2009. p.305–310.

MANIS. **Mammal Networked Information System**. Disponível em:<<http://manisnet.org/>>, Acesso em: 10 jul. 2010.

MATTHEW, J. B. et al. Managing Scientific Metadata. In: Piscataway, NJ, USA. **Anais...** IEEE Educational Activities Department, 2001. v.5, p.59–68.

ORACLE. **MySQL**. Disponível em:<<http://dev.mysql.com/>>, Acesso em: 10 jul. 2010.

PARKS, D. et al. GenGIS: a geospatial information system for genomic data. **Cold Spring Harbor Laboratory Press**, [S.l.], v.19, n.3, p.1896–1904, 2010.

R-PROJECT. **The R Project for Statistical Computing**. Disponível em:<<http://www.r-project.org/>>, Acesso em: 10 jul. 2010.

SIAMAZONIA. **Sistema de información de la Diversidad Biológica y Ambiental de la Amazonia Peruana**. Disponível em:<<http://www.siamazonia.org.pe/>>, Acesso em: 10 jul. 2010.

SINBIOTA. **Sistema de Informação Ambiental do Biota**. Disponível em:<<http://sinbiota.cria.org.br/>>, Acesso em: 10 jul. 2010.

SOBERÓN, J.; PETERSON, A. T. Biodiversity informatics: managing and applying primary biodiversity data. **Philosophical Transactions of the Royal Society of London B**, [S.l.], v.359, p.689–698, 2004.

SPECIES2000. **Species 2000**. Disponível em:<<http://www.sp2000.org/>>, Acesso em: 10 jul. 2010.

SPECIESLINK. **SpeciesLink**. Disponível em:<<http://splink.cria.org.br/>>, Acesso em: 10 jul. 2010.

SSTN. **Species and Specimen Thematic Network**. Disponível em:<<http://specimens.iabin.net/en/index.html>>, Acesso em: 10 jul. 2010.

TDWG. **Darwin Core**. Disponível em:<<http://www.tdwg.org/standards/450/>>, Acesso em: 10 jul. 2010.

TERRASERVER. **Terra Server**. Disponível em:<<http://www.terraserwer.com/>>, Acesso em: 10 jul. 2010.

TORRES, R. S. **Ambiente de Gerenciamento de Imagens e Dados Espaciais para Desenvolvimento de Aplicações em Biodiversidade**. Campinas: UNICAMP, 2007. 123 f. Tese (Doutorado) - Programa de Pós-Graduação em Computação, Universidade Estadual de Campinas, Campinas, 2007.

WEBIOS. **Web Service Multimodal Tools for Biodiversity Research, Assessment and Monitoring**. Disponível em:<<http://www.lis.ic.unicamp.br/projects/webios/>>, Acesso em: 10 jul. 2010.

XU, X. et al. **Experiences with a Hybrid Implementation of a Globally Distributed Federated Database System**. London, UK: Springer-Verlag, 2001. 212–224p.