

UNICAMP
BIBLIOTECA CENTRAL
SEÇÃO CIRCULANTE

Este exemplar corresponde à redação final da
Tese/Dissertação devidamente corrigida e defendida
por: Sandro de Paula
Matias
e aprovada pela Banca Examinadora.
Campinas, 28 de março de 2001
Mauro
COORDENADOR DE PÓS-GRADUAÇÃO
CPG-IC

Processamento de consultas ao banco de
dados de biodiversidade do BIOTA

Sandro de Paula Matias
Dissertação de Mestrado

Processamento de consultas ao banco de dados de biodiversidade do BIOTA

Sandro de Paula Matias¹

Dezembro de 2000

Banca Examinadora:

- Profa. Dra. Claudia M. Bauzer Medeiros
Instituto de Computação - Universidade Estadual de Campinas (Orientadora)
- Profa. Dra. Ana Maria de Carvalho Moura
Instituto Militar de Engenharia - Rio de Janeiro
- Prof. Dr. Geovane Cayres Magalhães
Instituto de Computação - Universidade Estadual de Campinas
- Profa. Dra. Maria Beatriz Felgar de Toledo
Instituto de Computação - Universidade Estadual de Campinas (Suplente)

¹O autor bacharel em Ciência da Computação pela Universidade Federal de Uberlândia.

UNIDADE BC
N.º CHAMADA:
T/ UNICAMP
M427p
V. Ex
TOMBO BC/ 44247
PROC. 16-392101
C D
PREÇO R\$ 11,00
DATA 03/05/01
N.º CPD

iv

CM-00154690-0

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP**

Matias, Sandro de Paula

M427p Processamento de consultas ao banco de dados de biodiversidade
do BIOTA / Sandro de Paula Matias -- Campinas, [S.P. :s.n.], 2000.

Orientadora : Claudia Maria Bauzer Medeiros

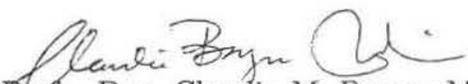
Dissertação (mestrado) - Universidade Estadual de Campinas,
Instituto de Computação.

I. Sistemas de informação geográfica. 2. Banco de dados. I.
Medeiros, Claudia Maria Bauzer. II. Universidade Estadual de
Campinas. Instituto de Computação. III. Título.

Processamento de consultas ao banco de dados de biodiversidade do BIOTA

Este exemplar corresponde à redação final da
Dissertação devidamente corrigida e defendida
por Sandro de Paula Matias e aprovada pela
Banca Examinadora.

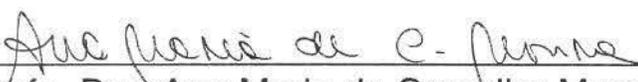
Campinas, 21 de Dezembro de 2000.


•Prof. Dra. Claudia M. Bauzer Medeiros
Instituto de Computação - Universidade
Estadual de Campinas (Orientadora)

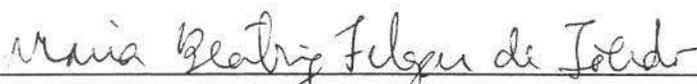
Dissertação apresentada ao Instituto de Com-
putação, UNICAMP, como requisito parcial para
a obtenção do título de Mestre em Ciência da
Computação.

TERMO DE APROVAÇÃO

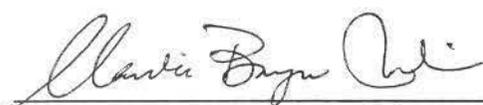
Tese defendida e aprovada em 15 de dezembro de 2000, pela
Banca Examinadora composta pelos Professores Doutores:



Profa. Dra. Ana Maria de Carvalho Moura
IME/URJ



Profa. Dra. Maria Beatriz Felgar de Toledo
IC – UNICAMP



Profa. Dra. Claudia Maria Bauzer Medeiros
IC – UNICAMP

© Sandro de Paula Matias, 2000.
Todos os direitos reservados.

Aos meus pais, Sancho e Luci,
com carinho.

... "vê mais longe,
a águia que voa mais alto."
Richard Back

Resumo

O SINBIOTASP é o sistema de informação para biodiversidade atualmente sendo desenvolvido como parte do programa BIOTA/FAPESP. Esta dissertação está voltada ao problema da implementação adequada de mecanismos para processamento de consultas no SINBIOTASP. Este tópico apresenta vários desafios tanto do ponto de vista de formulação das consultas quanto de seu processamento, pela variedade e volume dos dados considerados e pela gama de perfis de usuários do sistema.

As principais contribuições deste trabalho são: o levantamento de características de processamento de consultas de diversos sistemas de informação ambiental disponíveis na Web; uma proposta de sistematização dos tipos de consultas comuns em projetos de biodiversidade, sob critérios de processamento e interface; e a especificação de um conjunto básico de operadores espaciais, bem como interfaces para as consultas, envolvendo mapas e dados textuais, no contexto da informação ambiental para a biodiversidade. Como contribuição final, esta análise foi validada pelo desenvolvimento do módulo *Species Mapper* do SINBIOTASP, que permite o processamento via Web de consultas sobre coletas e distribuição de espécies.

Abstract

SINBIOTASP is the biodiversity information system being developed as part of the BIOTA/FAPESP program. This thesis is focused in the implementation issues of the query processing of the SINBIOTASP system. This subject presents many challenges in the formulation and the processing of the queries, due to the variety and the volume of the data and to the wide range of system user profiles.

The main contributions of this work are: a survey of the query processing features of many environmental information systems on the Web; a systematization of the query types which are typical of biodiversity application, considering processing and interface criteria; and specification of a basic set of spatial operators, as well general query interfaces, involving maps and textual data, in the context of biodiversity environmental information systems. As a final contribution, this analysis was validated by the development of the *Species Mapper* module of SINBIOTASP, which allows Web query processing on collection and distribution of species.

Agradecimentos

À Deus, a quem recorri em tantas ocasiões e por inúmeras razões, agradeço pelas inspirações, pela força, pela disposição, pela perseverança, pela saúde e, principalmente, por ter me iluminado quando me achava perdido entre dúvidas e incertezas. Obrigado pelos desafios e pela coragem de enfrentá-los.

Aos meus pais, Sancho e Luci, especialmente, pela dedicação, pelo carinho e pelo *enorme* apoio e eterno incentivo que sempre me ofereceram e, sobretudo, por estarem sempre dispostos e prontos a ajudar-me e terem aberto mão dos seus sonhos em função dos meus.

Aos meus irmãos, Flávio e Liliane, pelo apoio, pelo pensamento positivo e pelos estímulos constantes.

À minha namorada, pelo apoio irrestrito, paciência, compreensão, companheirismo e momentos de alegria e por me fazer acreditar ser possível alcançar mais este objetivo.

À minha orientadora Profa. Claudia Bauzer Medeiros, agradeço pela dedicação, pela paciência, pelos conselhos, pelas cobranças e pelos inúmeros ensinamentos, fundamentais para a conclusão deste trabalho e para a minha atual formação profissional. Pessoa pela qual jamais escondi minha sincera admiração e enorme orgulho de poder chamá-la de AMIGA.

Aos companheiros do grupo de Banco de Dados do IC, pela ajuda incondicional recebida e pela efetiva participação durante a elaboração deste trabalho através de sugestões de melhoria e de críticas construtivas. Um agradecimento especial àqueles que mais diretamente colaboraram para tornar concreto este trabalho: Renato Fileto, Luiz Mariano, Daniel Kaster, Daniel S.A., Alysson, Ricardo Torres e Rafael.

Aos professores da Universidade Federal de Uberlândia, pelo apoio, pelo incentivo e pela confiança que depositaram em mim. Um agradecimento especial àqueles que mais diretamente colaboraram para o início deste trabalho: Márcia Aparecida Fernandes, Sandra Aparecida de Amo, Pedro Frosi Rosa, Autran Macedo, Luis Carlos Rispoli, Luis Fernando

Faina e João Augusto Alves Pacheco.

Aos companheiros da XIV Turma de Bacharelado em Ciência da Computação da Universidade Federal de Uberlândia, pelo apoio e pelo incentivo de valor incalculável. Agradeço muito à Deus por tê-los conhecido.

Aos companheiros da Fundação André Tosello, pelo apoio e pela compreensão. Um agradecimento especial ao Ricardo pelos conselhos, pelos ensinamentos e pela troca de experiências.

Aos meus amigos e companheiros de moradia, Maikol e Edmar, por terem propiciado a criação de um ambiente amigável durante os vários meses de convivência.

Aos amigos e colegas que de diversas formas me apoiaram e proporcionaram momentos alegres. Em especial, aos Maringuetes, André Marcos, Jan Pfeifer, Delano Medeiros Beder, Guilherme Pimentel, Guilherme Albuquerque, Paulo Perez, Andréia Fagundes, Danival, José Augusto, Sérgio Yunes, Alessandro, Jó, Márcio, Adilson, Cesar, Valeska e Alexandre, pelo convívio, pelos conselhos e pela troca de experiências. Agradeço muito à Deus por tê-los conhecido.

A todas as pessoas que contribuíram de uma forma ou de outra para a produção deste trabalho, com dicas, apoio, incentivo e sugestões. Os nomes são tantos que se torna difícil citá-los todos aqui.

Aos funcionários desta universidade, que sempre receberam as solicitações de serviços com gentileza e competência, especialmente os integrantes da secretaria do IC.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pelo apoio financeiro que viabilizou o desenvolvimento deste trabalho.

Ao Projeto SAI (Sistemas Avançados de Informação) do PRONEX II - MCT pela infraestrutura que possibilitou o desenvolvimento do trabalho.

Conteúdo

Resumo	vii
Abstract	viii
Agradecimentos	ix
1 Introdução e Motivação	1
2 Conceitos básicos	5
2.1 Sistemas de Informação Geográfica	5
2.2 Sistemas que manipulam mapas	7
2.3 Operadores espaciais	8
2.4 Operações espaciais em um SIG	11
2.5 Sistemas de Informação Geográfica na Web	13
2.5.1 Interoperabilidade e SIG	13
2.5.2 Arquitetura SIG-Web	14
2.5.3 Exemplos de sistemas SIG-Web	15
2.5.4 Linguagens para desenvolvimento de sistemas Web	16
2.6 Metadados	20
2.7 Resumo '	22
3 Análise comparativa de alguns Sistemas de Informação Ambiental	23
3.1 Critérios de comparação de Sistemas de Informação Ambiental	23
3.2 Tipos de consultas e interfaces	24
3.3 Visão geral dos Sistemas de Informação	27
3.3.1 Gerenciamento e recuperação de dados	27
3.3.2 Gerenciamento e recuperação de metadados	36
3.4 Resumo	44

4	O Sistema de Informação de Biodiversidade do Estado de São Paulo	47
4.1	O Programa BIOTA/FAPESP	47
4.2	Visão Geral do Sistema de Informação	49
4.2.1	Interface com o usuário	49
4.2.2	Gerenciamento de dados e metadados	50
4.2.3	Repositórios de Dados	52
4.3	Estado atual do SINBIOTASP na Web	53
4.3.1	Arquitetura Web atual	54
4.3.2	Especificação funcional	56
4.3.3	Exemplos de operações	58
4.4	Comparação com sistemas existentes	60
4.5	Resumo	63
5	Caracterização de Consultas para o BIOTA	65
5.1	O Processamento de Consultas	66
5.1.1	Fontes de dados envolvidas no processamento da consulta	66
5.1.2	Tipologia dos objetos espaciais	68
5.1.3	Operadores de consultas possíveis e conjunto básico para o BIO- TA/FAPESP	70
5.1.4	Conjunto de transações primitivas para o BIOTA	71
5.1.5	Consultas básicas e sua classificação	72
5.1.6	Processamento via SGBD <i>versus</i> processamento via SIG	80
5.2	A Interface com o Usuário	81
5.3	Resumo	83
6	Aspectos de Implementação	85
6.1	O <i>Species Mapper</i>	85
6.1.1	Arquitetura do sistema	86
6.1.2	Exemplo de sessão	90
6.2	Arquitetura Web proposta para o SINBIOTASP	97
6.3	Resumo	99
7	Conclusões e extensões	101
7.1	Contribuições	101
7.2	Extensões	102
	Bibliografia	105

Lista de Tabelas

3.1	Classificação dos sistemas de informação ambiental quanto à abrangência dos objetivos dos módulos disponíveis e ao conteúdo do banco de dados . .	44
3.2	Tipos de consulta disponíveis nos sistemas de informação ambiental analisados	45
3.3	Formatos de exibição dos resultados disponíveis	45
5.1	Objetos espaciais e seus interrelacionamentos utilizados nas consultas do SINBIOTASP	69
5.2	Conjunto básico de operadores espaciais necessários para o processamento das consultas espaciais do SINBIOTASP	71
5.3	Transações primitivas necessárias para o processamento de consultas espaciais básicas no SINBIOTASP	72
5.4	Caracterização dos tipos de consultas	74
5.5	Caracterização dos tipos de consultas possíveis	79
5.6	Classificação de consultas em função dos formatos de entrada e saída . . .	81
6.1	Exemplo da documentação de uma função desenvolvida para o <i>Species Mapper</i>	87

Lista de Figuras

3.1	Exemplo de resultado gráfico de uma consulta ao sistema GCMD [GCM]	26
3.2	Gerenciamento de dados ECN	37
3.3	Exemplo de uma consulta no sistema WebCDS [CDS]	42
4.1	Arquitetura básica do SINBIOTASP.	50
4.2	Interface de seleção textual por campo específico no SINBIOTASP	57
4.3	Inserção de coleta (Ficha Padrão de coleta)	59
4.4	Interface de seleção textual por campo específico e por palavra-chave	60
4.5	Consulta gráfica no módulo Atlas Biota/SP versão 1.0 (descontinuado)	61
5.1	Modelagem do Banco de Coletas	67
5.2	Modelagem do Banco de Metadados	68
6.1	Diagrama das funcionalidades implementadas para o <i>Species Mapper</i>	88
6.2	Arquitetura do <i>Species Mapper</i>	90
6.3	Consulta textual por campo específico do <i>Species Mapper</i>	91
6.4	Resultado da consulta por campo específico - visualização textual	92
6.5	Visualização textual resumida das coletas de uma determinada espécie no <i>Species Mapper</i>	93
6.6	Visualização textual detalhada das informações referentes a uma coleta no <i>Species Mapper</i>	94
6.7	Visualização gráfica da distribuição de uma determinada espécie no <i>Species Mapper</i>	95

Capítulo 1

Introdução e Motivação

É crescente a preocupação com a conservação do meio ambiente. Para permitir iniciativas nesta área, é necessário disponibilizar um grande número de informações ambientais e softwares para gerenciá-las. O trabalho descrito nesta dissertação faz parte de uma iniciativa de desenvolvimento de um sistema de informação ambiental específico, para o programa BIOTA/FAPESP.

Mudanças ambientais têm emergido como uma importante questão na agenda global. Conseqüentemente, há uma demanda por acesso seguro, objetivo e atualizado das informações ambientais [RLW00]. Sistemas de informação ambiental são uma resposta a esta demanda. Eles enfocam o meio ambiente e os recursos naturais, visando o gerenciamento de dados sobre o solo, a água, o ar e sobre as diversas espécies de animais e plantas existentes [Fag99]. Estes sistemas visam suprir a necessidade do usuário ser capaz de localizar os conjuntos de dados relevantes e recuperá-los rapidamente. Em adição, as aplicações nas áreas ambientais e gerenciais têm como características marcantes o grande volume de dados envolvidos e o geo-referenciamento destes dados (aplicações geográficas).

O BIOTA/FAPESP [BIO] é um programa voltado às pesquisas de conservação da biodiversidade do Estado de São Paulo. Os projetos desenvolvidos dentro deste programa têm por objetivo estudar e conhecer a biodiversidade do Estado e prover referências para estudos de impacto ambiental, produzindo estimativas de perda de biodiversidade em diferentes escalas espaciais e temporais, e identificando áreas e componentes prioritários para conservação. O programa é único em termos do volume, variedade e abrangência de dados envolvidos, compreendendo fauna e flora, tanto terrestre quanto aquática. Um dos projetos do programa BIOTA/FAPESP é o SINBIOTASP, que é o sistema de informação que deverá integrar os dados produzidos pelos demais projetos.

A partir da disponibilização dos dados pela Internet, os pesquisadores serão capazes de buscar informações de diferentes temas, usando distintas fontes de informação, a partir de um *browser* Web. O uso efetivo da Internet demanda mecanismos eficientes e fáceis de

manusear, integrando e interrelacionando os dados, principalmente se estes forem heterogêneos, e fornecendo uma apresentação dos dados tanto textual quanto visual, visando atender as necessidades dos mais diversos usuários. Diversas iniciativas se propõem a fornecer estes mecanismos, possibilitando um ambiente multi-usuário que possa acessar dados heterogêneos, por exemplo [LP98, SGM98, BM98]. Apesar da grande quantidade de sistemas de biodiversidade existentes, não há estudos sobre perfis dos usuários e consultas. Isto dificulta a especificação e implementação de sistemas adequados.

O objetivo da dissertação é estabelecer uma base formal para permitir uma implementação adequada do processamento de consultas no SINBIOTASP. Estas consultas exigem combinar dados textuais e cartográficos e fornecer uma apresentação do resultado tanto textual quanto gráfica, visando atender todas as necessidades dos usuários. Devido à grande diversidade de usuários e de domínios de aplicação no que se refere a dados cartográficos, não há um consenso sobre um conjunto base de operações que devem ser tornadas disponíveis para a implementação deste tipo de aplicação [Cer96]. Para isto, os seguintes passos foram seguidos, resultando nas contribuições da dissertação, a saber:

- Levantamento de características de diversos sistemas de informação ambiental disponíveis. Este levantamento permitiu identificar características desejáveis para consultas no BIOTA (capítulo 3).
- Sistematização de tipos de consultas comuns em projetos de biodiversidade sob critério de processamento e interface. Esta caracterização foi usada como base para especificação do atual sistema de processamento de consultas do BIOTA (capítulo 5). Este estudo resultou na especificação de um conjunto básico de operadores espaciais, bem como interfaces para as consultas de uma maneira geral, envolvendo mapas e dados textuais, no contexto da informação ambiental para a biodiversidade.
- Projeto e implementação do atual sistema de consultas do BIOTA, que permite consultas integradas sobre alguns aspectos de biodiversidade, a partir do banco de dados já consolidado (capítulo 6).

A dissertação está organizada da seguinte forma. O Capítulo 2 apresenta alguns conceitos básicos necessários ao entendimento da dissertação. O Capítulo 3 analisa comparativamente diversos sistemas de informação ambiental disponibilizados via Internet. O Capítulo 4 descreve de forma simplificada o programa BIOTA/FAPESP e o projeto SINBIOTASP. O Capítulo 5 analisa e define os requisitos básicos e necessários para atender as consultas sobre o banco de dados de biodiversidade do SINBIOTASP. Além disso, é sugerida uma classificação das consultas visando estruturar o estudo do problema. O Capítulo 6 descreve os aspectos de implementação do protótipo desenvolvido a partir das análises realizadas nos capítulos 3 e 5, e propõe uma nova arquitetura Web para o

SINBIOTASP. Finalmente, o Capítulo 7 apresenta as conclusões finais da dissertação e as direções para trabalhos futuros.

Capítulo 2

Conceitos básicos

Este capítulo apresenta alguns conceitos básicos necessárias ao entendimento da dissertação. As seções estão organizadas da seguinte forma. A seção 2.1 apresenta conceitos relativos a Sistemas de Informação Geográfica (SIG). A seção 2.2 descreve alguns sistemas que permitem a manipulação de mapas. A seção 2.3 apresenta os principais operadores espaciais. A seção 2.4 discute diferentes grupos de operações espaciais em um SIG que fazem uso destes operadores. A seção 2.5 enfoca como SIGs são usados na Web. A seção 2.6 descreve o uso de metadados em SIG na Web para sistemas ambientais. A seção 2.7 conclui este capítulo.

2.1 Sistemas de Informação Geográfica

Um Sistema de Informação Geográfica (SIG) é um sistema automatizado usado para armazenar, analisar, manipular e visualizar dados geográficos, ou seja, dados que representam objetos e fenômenos em que a localização geográfica é uma característica inerente à informação e indispensável para analisá-la [CCH⁺96].

Os dados geográficos são caracterizados por sua posição geográfica, seus atributos, suas relações topológicas e seu aspecto temporal. Estes dados são obtidos de diversas fontes e são capturados por dispositivos diferentes, em formatos distintos. Eles ocupam um espaço de armazenamento considerável e requerem funções de análise e apresentação especializadas, não presentes em sistemas de banco de dados comerciais [Cer96].

A modelagem de dados geográficos difere da tradicional não apenas devido às características espaciais, mas também por envolver a questão da representação, que varia conforme a perspectiva do usuário ou aplicação, ou segundo fatores meramente técnicos. Isto frequentemente significa que diferentes representações coexistem para um mesmo objeto [CCH⁺96]. O conceito de *campo* se refere a dados contínuos (por exemplo, temperatura) enquanto *objeto* denota entidades discretas (por exemplo, artefatos humanos

como pontes).

A dicotomia de modelagem de campos ou objetos (fenômenos contínuos versus discretos) se reflete, no nível de representação, no chamado debate *raster* versus *vector* [CCH+96]. Campos são frequentemente representados no formato de *tesselação* ou *matricial*, ou seja, em uma matriz cujos elementos são unidades poligonais regulares do espaço (*células*). Os relacionamentos topológicos no espaço são implicitamente determinados a partir da vizinhança das células e as coordenadas geográficas são obtidas indiretamente a partir da posição da célula na matriz. A cada célula corresponde um valor do campo representado (ou seja, não pode haver dois valores distintos para uma mesma célula). O termo *raster* designa células retangulares, mas na maioria das vezes é usado como termo genérico para a representação matricial. Operações típicas neste tipo de representação são a sobreposição de matrizes, combinando os valores das células através de funções matemáticas, ou a abstração de conjuntos de várias células adjacentes em uma única célula, cujo valor é calculado a partir dos valores das células selecionadas (generalização) [CCH+96].

Já um objeto geográfico é tipicamente representado no formato *vetorial*, ou seja, sua geometria é descrita utilizando *pontos*, *linhas* e *polígonos*. Naturalmente, deve haver uma transformação bem definida entre o sistema de coordenadas utilizado na descrição geométrica e o sistema de coordenadas geográficas adotado. Linhas são formadas por sequências de pontos; e polígonos (abertos ou fechados) por sequências de linhas. Operações comuns neste tipo de representação são operações topológicas (por exemplo, adjacência, inclusão) e métricas (distância, área). Representações vetoriais podem ter diferentes níveis de sofisticação. Representações *não estruturadas* se restringem a armazenar listas de coordenadas sem qualquer preocupação com otimização do processamento. Representações *topológicas* permitem armazenar, associada à localização, informação sobre relacionamentos de contiguidade e vizinhança dos elementos armazenados [CCH+96].

As formas de interação de um usuário com um SIG podem ser divididas em dois paradigmas: textual (tradicional) e visual (utilizando cores, gráficos e ícones) [Oli97].

Linguagens Baseadas no Paradigma Textual Interfaces textuais têm a vantagem de prover suporte para exprimir ações complexas, que não podem ser facilmente expressas por meios gráficos. Dentre as linguagens baseadas neste paradigma, Oliveira [Oli97] destaca diferentes extensões espaciais do SQL e linguagens algébricas espaciais.

Linguagens Baseadas no Paradigma Visual Linguagens visuais apresentam a qualidade de serem mais naturais para o usuário. Existem, no entanto, poucos relatos de linguagens visuais, sendo a maioria relativa a propostas não implementadas. Estas linguagens sofrem, também, de problemas relativos à cognição para interpretação de um

certo comando. Um exemplo de linguagem visual que incorpora relacionamentos espaciais é a VISUAL [BOO96].

Visualização dos resultados Do ponto de vista de interface, várias pesquisas têm sido realizadas tanto no que se refere a visualização dos resultados (saída) [Oli97, CF98, RR98, VHN96], quanto à elaboração da consulta (entrada) [BLL⁺94, Jon98, LCHH97].

Existem duas abordagens de visualização dos resultados de consultas em sistemas geográficos, visualização textual e gráfica [Oli97]. Por exemplo, dada uma consulta que retorna um polígono, a visualização textual (tradicional) apresenta uma dada área através de uma lista de coordenadas de seus pontos limites, e a visualização gráfica apresenta o desenho do polígono que representa a área em questão.

Segundo Coors e Flick [CF98], a visualização do resultado de consultas geográficas no espaço 3D incrementa os detalhes e a facilidade de entendimento. Desta forma, Coors e Flick [CF98] acreditam em uma melhor apresentação dos dados geográficos, e propõem a utilização de um modelo de dados que suporta geometria 3D e topologia 3D, desenvolvendo um SIG 3D, incluindo uma interface com o usuário baseada na Web.

A visualização de um grande volume de informações textuais, por sua vez, é relegada a um segundo plano em SIG, mas pode ser necessária. Este tipo de manipulação é enfatizada por Roussinov e Ramsey [RR98], fora do contexto geográfico.

Definir boas interfaces com o usuário em SIG é crucial, uma vez que SIG requer frequentemente interação gráfica. Segundo Szmurlo *et al.* [SGM98], no passado, vários SIGs sugeriam boas ferramentas para visualização, e a parte da interface de consulta de um SIG estava limitada principalmente ou por um conjunto de menus, ou por uma janela textual onde o usuário especificava a consulta em uma linguagem formal. Contudo nenhuma destas soluções era aceitável: linguagens formais são muito difíceis para usuários não-especialistas, enquanto interfaces baseadas em menu não exploram completamente todas as fontes de informação. Hoje, as interfaces estão mais amigáveis. Vários sistemas utilizam *applets* de mapas interativos para permitir que o usuário possa interagir mais facilmente com o mapa, sobretudo na Web, como por exemplo o sistema GCMD [GCM].

Alguns trabalhos têm sido feito na área de interfaces para consultas a mapas visando prover ao usuário a possibilidade de executar consultas alfanuméricas [VHN96], consultas gráficas [BLL⁺94, Jon98] ou ambas [Voi95].

2.2 Sistemas que manipulam mapas

A disseminação do uso de SIG pode ser dimensionada, por exemplo, a partir da variada gama de estudos de caso apresentados em [Lan98]. O texto cobre aplicações que variam desde sistemas de informação para exploração de petróleo e gás até poluição do ar.

Além de SIG, há alguns softwares que permitem uma manipulação restrita de mapas, de forma a acessar informações (não necessariamente de uma natureza geográfica) por meio de uma busca espacial. O *Dynamic HomeFinder* [KO96] e o *Spatial Document Locator System* [KO96] são exemplos deste tipo de sistema.

Dynamic HomeFinder Este *software* auxilia seus usuários a encontrar casas disponíveis que satisfaçam a determinados requisitos especificados. Os usuários especificam os critérios, e um mapa com pontos é retornado. Os pontos apresentados representam as casas que satisfazem as especificações informadas. Clicando sobre estes pontos, os usuários conseguem obter informações adicionais sobre as casas [KO96].

Spatial Document Locator System O *Spatial Document Locator System* [KO96] é um *browser* de bibliotecas digitais baseado em hipermídia e um visualizador de documentos no qual os usuários navegam geograficamente (em mapas) para localizar e recuperar informações.

A idéia central é o conceito de zona de informação (IZ). Uma biblioteca digital de uma zona de informação (IZDL) é uma biblioteca digital que cobre uma área geográfica, onde as informações associadas de alguma forma à área estão catalogadas naquela biblioteca. Além do mais, Kacmar e Orendorf [KO96] afirmam que navegar entre regiões provê um método natural de aparar o espaço de informação. Para o *browser* IZDL determinar quais documentos estão acessíveis em uma determinada zona de informação existe um arquivo que contém a lista dos documentos existentes, por sua vez associados com o arquivo de metadados dos documentos.

2.3 Operadores espaciais

Uma consulta geográfica necessita ser expressa em uma linguagem de consulta que, além de incorporar operadores convencionais (operadores aritméticos, relacionais e lógicos), incorpore operadores com poder de expressão de propriedades topológicas (tais como interseção e inclusão), de propriedades geométricas (tais como área e comprimento) e de propriedades espaciais (tal como união, a qual gera como resultado o objeto geográfico resultante da união de vários outros objetos geográficos).

As aplicações SIG apresentam diferenças entre si devido à grande diversidade de usuários e domínios de aplicação. Em consequência, não há consenso sobre um conjunto base de operações que devem ser tornadas disponíveis para a implementação de aplicações [Cer96].

A enorme quantidade e a natureza dos dados armazenados por aplicações que utilizam SIGs implicam em alterações ou extensões nos métodos de acesso, otimizadores de consulta

e linguagens de consulta estabelecidos para sistemas gerenciadores de banco de dados convencionais. Com isto, diferentes propostas e soluções têm sido apresentadas, tornando-se imprescindível a definição de um conjunto básico de operadores para consultas espaciais [Cif95].

A formulação de consultas em banco de dados geográficos pode utilizar operadores espaciais, cujos parâmetros (operandos) são valores, objetos ou conjuntos destes. Os operadores espaciais podem ser de localização, de direção, métricos, e topológicos. Os operadores espaciais têm como operandos objetos geográficos (operador localização [Far98]) ou suas localizações (operadores de direção, métricos e topológicos [Gut94]). A sua aplicação depende da representação dos objetos [Far98].

O operador de localização definido por Faria [Far98] é responsável por retornar a localização do objeto passado como parâmetro. A localização de um objeto A corresponde a uma representação espacial do objeto, sendo composta por objetos geométricos (delimitados por um conjunto de coordenadas geográficas) que descrevem a geometria de A .

Segundo Guting [Gut94], os operadores espaciais ou relacionamentos espaciais podem ser separados nas classes de operadores de direção, métricos e topológicos.

Operadores de direção Os operadores de direção verificam se existe um determinado relacionamento de orientação entre dois conjuntos de objetos geométricos. Os relacionamentos de orientação (ou relacionamentos direcionais) descrevem como os objetos estão posicionados entre si [CCH⁺96]. Exemplos de relacionamentos de orientação encontrados são: *above*, *below*, *north*, *south*, *east*, *west*, *left*, *right*, *sameposition*, *northeast*.

Por exemplo, o relacionamento *east* é definido por $p \text{ east } q \rightarrow \forall p_i \forall q_i, X(p_i) \geq X(q_i)$, onde p e q são objetos, p_i e q_i são respectivamente pontos de p e q , e X é uma função que retorna a coordenada x de um ponto.

Operadores Métricos Os operadores métricos geram, a partir de um ou mais objetos, um escalar que representa uma propriedade intrínseca aos objetos analisados. Os operadores métricos podem ser classificados em:

- unários: calculam um valor escalar usando um único conjunto de objetos;
- binários: calculam um valor escalar usando dois conjuntos de objetos;

Exemplos de operadores métricos são: cálculo da área *Area*, comprimento *Length*, perímetro *Perimeter*, distância *Distance*.

Operadores Topológicos Os operadores topológicos retornam um valor lógico verdadeiro se há um determinado relacionamento topológico entre dois conjuntos de objetos geométricos. Os relacionamentos topológicos são invariantes face a transformações de escala, translação e rotação [CCH⁺96]. Estes relacionamentos são mutuamente exclusivos e são suficientes para representar todas as situações topológicas possíveis entre dois objetos bidimensionais.

Relacionamentos topológicos são os mais fundamentais e têm sido estudados em maior profundidade [Gut94]. Uma questão básica refere-se à possibilidade de enumerar todos os relacionamentos possíveis. Medeiros e Cilia [MC95] enumeram os relacionamentos topológicos baseando-se em Clementini *et al.* [CevO93]. O modelo utilizado por Medeiros e Cilia [MC95] considera todas as combinações binárias de interseção de regiões: a região A é um conjunto de pontos 2D com um *interior* A^0 e *limitante* δA . Se o *exterior* A^{-1} é considerado, obtém-se a matriz 9-interseção. A matriz a seguir mostra as matrizes 4-interseção e 9-interseção, representando todos os tipos de interseção entre duas regiões A e B .

$$\begin{pmatrix} A^0 \cap B^0 & A^0 \cap \delta B & A^0 \cap B^- \\ \delta A \cap B^0 & \delta A \cap \delta B & \delta A \cap B^- \\ A^- \cap B^0 & A^- \cap \delta B & A^- \cap B^- \end{pmatrix}$$

Clementini *et al.* [CevO93] estenderam esta matriz para considerar a dimensão *dim* da interseção entre quaisquer dois objetos e demonstraram formalmente que tal aproximação pode ser sintetizada em 5 relacionamentos topológicos mutuamente exclusivos e três operadores que permitem retornar os limitantes de um objeto (limites de uma região, ou pontos de início e fim de uma linha). Estes relacionamentos - *touch*, *in*, *cross*, *overlap*, *disjoint* - expressam quando a dimensão *dim* das interseções entre dois objetos pode ser vazia, um ponto (0D), uma linha (1D) ou uma região (2D). Os relacionamentos entre dois objetos A e B (de tipos linha, ponto e região) podem ser expressados da seguinte forma:

- $(A \text{ in } B) \Leftrightarrow (A \cap B \neq A) \wedge (A^0 \cap B^0 = \emptyset)$
- $(A \text{ touch } B) \Leftrightarrow (A^0 \cap B^0 = \emptyset) \wedge (A \cap B \neq \emptyset)$
- $(A \text{ cross } B) \Leftrightarrow (\dim(A^0 \cap B^0) = \max(\dim(A^0), \dim(B^0) - 1)) \wedge (A \cap B \neq A) \wedge (A \cap B \neq B)$
- $(A \text{ overlap } B) \Leftrightarrow (\dim(A^0) = \dim(B^0) = \dim(A^0 \cap B^0)) \wedge (A \cap B \neq A) \wedge (A \cap B \neq B)$
- $(A \text{ disjoint } B) \Leftrightarrow (A \cap B = \emptyset)$

Nem todos os relacionamentos podem ser aplicados a todos os tipos de objetos. O relacionamento *Overlap*, por exemplo, pode ser aplicado somente sobre região/região e

linha/linha [MC95]. Além disso, um relacionamento ou operador r é simétrico se e somente se $(A_1 r A_2) \Leftrightarrow (A_2 r A_1)$. Com a exceção de *in*, todos os relacionamentos definidos anteriormente são simétricos.

2.4 Operações espaciais em um SIG

Existem diferentes grupos de operações em SIG que fazem uso dos operadores descritos anteriormente. Vários trabalhos propõem um conjunto mínimo de operações primitivas (transações), as quais podem ser utilizadas para construir aplicações em SIGs genéricos [Cer96, Cif95].

Esta seção resume um conjunto de transações primitivas, especificadas em alto nível por Ciferri [Cif95], que pode ser utilizado para a formação de transações mais complexas. Estas transações primitivas são predominantemente orientadas aos dados espaciais, sendo, a priori, independentes do formato de dados utilizado (*raster* ou *vetorial*).

A utilização do termo objeto geográfico será feita independente do formato utilizado (*raster* ou *vetorial*). A especificação da dimensão espacial, quando necessária, será feita logo após o termo objeto geográfico (OG), sendo abreviada para OG-0D para objetos geográficos sem dimensão espacial (geometria na forma de um ponto), OG-1D para objetos geográficos unidimensionais (geometria na forma de uma linha) e OG-2D para objetos bidimensionais (geometria na forma de um polígono). Quando a dimensão espacial for irrelevante, será utilizado apenas o termo objeto geográfico (OG).

Reclassificação A reclassificação consiste em agrupar categorias de um tema, primitivo ou derivado, de modo a gerar um novo conjunto de categorias para o tema em questão. Em geral, reclassifica-se categorias de OGs-2D.

Superposição Existem várias variantes desta transação que possuem em comum o fato de superpor OGs-2D de temas distintos, diferindo somente no modo como as categorias ¹ resultantes são geradas. A superposição pode ser aplicada a vários temas, sendo efetuada geralmente aos pares, sucessivamente. A operação de superposição utiliza o operador topológico *overlay*.

Ponderação A ponderação é uma variante da superposição, diferindo principalmente desta em relação à geração das categorias dos OGs-2D do tema superposto. Esta geração é efetuada com base em pesos associados às categorias dos temas originais, ao invés de ser efetuada usando-se diretamente as próprias categorias destes temas. Para cada OG-2D do

¹SIGs utilizam o conceito de categoria para descrever o domínio de temas. Não existe nenhuma restrição que impeça que dois objetos geográficos distintos de um mesmo tema possuam a mesma categoria.

tema superposto, a categoria é obtida através de uma operação aritmética, frequentemente uma adição, com os pesos das categorias dos OGs-2D correspondentes nos temas originais. Esta operação utiliza o operador topológico *overlay*.

Análise de proximidade O tipo mais básico, chamado de análise de proximidade simples, consiste em gerar um OG-2D (designado como *buffer* simples, zona de *buffer* simples ou simplesmente *buffer*), na forma de um "corredor", cujos limites externos possuem uma distância fixa K em relação a um OG fonte, e cujos limites internos são formados pelos limites do próprio OG fonte a partir de onde a distância K começou a ser medida. Pode haver casos em que ocorra a interseção entre as diversas zonas de *buffer* simples geradas, sendo o resultado final a união de todas estas. Esta operação utiliza operadores métricos (distância) e topológicos (*overlay*).

Topológicas Ciferri [Cif95] propõe vários tipos de transações topológicas, todas baseadas na utilização dos operadores topológicos. Exemplos são as transações topológicas *booleanas* (que retornam um valor *booleano* a partir da aplicação de operadores topológicos), transações de busca topológica (que percorrem o banco de dados para recuperar objetos que satisfaçam determinado relacionamento topológico) e topológicas escalares (que determinam as coordenadas geográficas onde ocorre um determinado relacionamento topológico).

Decomposição Um OG-2D é composto por dois componentes básicos: seu interior e sua fronteira. Muitas vezes deseja-se saber se um dado OG-0D, por exemplo, está contido em um OG-2D. Entretanto este não pode estar localizado na fronteira. Para isto, primeiramente deve-se obter o interior do OG-2D, para depois efetuar-se a operação topológica de inclusão.

Transações baseadas em conjunto Estas transações retornam um conjunto de OGs baseadas em operações matemáticas de conjunto entre as fronteiras de vários OGs. As principais transações deste tipo são: união, interseção e diferença.

Transações diversas A seguir são descritas algumas transações comumente realizadas.

- conversão de formato de dados: *raster* para vetorial, vetorial para *raster*, além de conversões entre os formatos *raster* e vetorial e formatos específicos (*DXF*, *PostScript*, dentre outros);
- determinação das coordenadas geográficas de um conjunto de OGs que corresponde a aplicação do operador de localização;

- geração de um OG-1D a partir de dois OGs-0D; e de um OG-2D a partir de OGs-1D;
- seleção de OGs a partir dos valores de seus atributos convencionais, denominada de seleção convencional;
- operações de visualização (*pan*, *zoom*, etc);

2.5 Sistemas de Informação Geográfica na Web

A popularização da Internet, o advento da Web e o crescimento da importância dos Sistemas de Geoprocessamento em uma série de atividades apresentam novas perspectivas para as pessoas que necessitam utilizar dados geográficos. A disseminação da Internet vem levando ao desenvolvimento do processamento de aplicações SIG a partir da Web, muitas vezes conhecido como *Web-based GIS*.

2.5.1 Interoperabilidade e SIG

Quando o processamento de consultas em sistemas ambientais combina o uso de SGBD e SIG, torna-se importante o estudo de questões como heterogeneidade e interoperabilidade.

Vários termos são usados para referenciar ambientes heterogêneos de bancos de dados, ou seja, ambientes que integram bancos de dados diferentes. De um modo geral, o termo heterogêneo é usado para referenciar bancos de dados que podem diferir quanto ao esquema, protocolos de comunicação, hardware ou linguagens de consulta. Os termos *sistema integrado* ou *sistema de banco de dados federado* são usados para expressar a união de bancos de dados heterogêneos [Soa98].

Segundo Gardels [Gar97], interoperabilidade é a chave para o uso de dados heterogêneos e o processamento de recursos em todas as partes de um ambiente de rede de trabalho em uma simples sessão do usuário ou em um fluxo de trabalho. Interoperabilidade significa um mecanismo específico para troca de dados e um *software* de interação, junto com os protocolos de aceitação definidos [Gar97]. Os requisitos fundamentais de interoperabilidade em SIG são [Gar97]:

- espaço de dados compartilhado - um modelo de dados genérico suportando uma variedade de funções analíticas e cartográficas.
- aplicações compatíveis - um usuário de um pacote de ferramentas que é configurável para utilizar ferramentas específicas e dados necessários para resolver um problema.
- *browser* de recursos heterogêneos - um método para explorar e acessar informações espaciais e recursos analíticos disponíveis na rede.

Gonçalves [Gon97] apresenta abordagens de testes de interoperabilidade entre catálogos, utilizando componentes da linguagem do protocolo Z39.50 como linguagem intermediária e esse mesmo protocolo para fins de busca. O protocolo Z39.50 é um protocolo Internet de recuperação e pesquisa de informação. Este protocolo permite ao usuário acessar registros de banco de dados remotos pela especificação de critérios de seleção, e transmitir alguns ou todos os registros identificados.

Em adição, o Z39.50 é um protocolo de rede, orientado a sessão e "com estado" (isto é, *stateful*), em contraste ao conhecido protocolo Internet "sem estado" (isto é, *stateless*), o protocolo HTTP (usado pelo WWW) [Hak96].

O Z39.50 é baseado no modelo cliente/servidor. O protocolo define como o cliente e o servidor "conversarão" um com o outro. Geralmente, a aplicação cliente mantém a interface do usuário, converte as consultas do usuário final para a forma requerida pelo protocolo, envia a consulta convertida para a aplicação servidor, converte o retorno do servidor para a forma local e apresenta os resultados para o usuário final. A aplicação servidor espera pela consulta enviada para a aplicação servidor, processa a consulta quando ela chega e envia os resultados para a aplicação cliente [Hak96]. Como exemplos de sistemas de biodiversidade que utilizam o protocolo Z39.50 mencionamos: o *Species Analyst* [Uni] e o ERIN (*Environmental Australia On Line*) [ERI].

2.5.2 Arquitetura SIG-Web

As diferentes maneiras utilizadas para publicar dados geográficos na Internet, em geral, recaem no modelo clássico da Internet (cliente/servidor). Essa arquitetura é composta de um cliente com capacidade de exibição de dados geográficos e um programa de servidor de geoprocessamento sendo executado juntamente com o servidor Web para responder às solicitações sobre os dados geográficos.

Tipicamente o cliente é um *browser* que realiza solicitações a um servidor Web utilizando o protocolo HTTP. Quando o cliente realiza a solicitação de um mapa ou outra operação sobre os dados geográficos, o servidor Web a repassa para o servidor de geoprocessamento, que atende à solicitação e envia a resposta para o cliente através do servidor Web. A comunicação entre os dois servidores é feita, em geral, através de interfaces como CGI (*Common Gateway Interface*), ISAPI (*Internet Server Application Program Interface* - Microsoft) e NSAPI (API do servidor Netscape).

Sendo assim, as tecnologias comerciais disponíveis para facilitar a difusão de informação geográfica através da Internet e o acoplamento de um servidor de dados geográficos podem ser enquadradas genericamente em duas grandes classes [OPC00]:

- Servidores de mapas, que, respondendo a pedidos remotos, enviam uma imagem (matriz) de tamanho fixo nos formatos GIF ou JPEG. Esta solução permite con-

figurar o servidor para responder a diferentes tipos de consulta, sem requerer que todos os dados a ser transmitidos sejam pré-computados. Entretanto, o usuário consegue visualizar apenas as imagens enviadas; qualquer novo pedido é enviado de volta para o servidor, resultando em mais uma transferência pela Internet. Dependendo da velocidade de acesso, esta estratégia pode resultar em longos e sucessivos períodos de espera. Como exemplo temos o *Internet Map Server* da ESRI.

- Clientes de apresentação, que adotaram como solução a transmissão de todos os dados no formato vetorial para a máquina do cliente, com posterior visualização local. Estes servidores encapsulam a informação em formatos gráficos, que podem ser apresentados por meio de programas adicionais, *plug-ins* acoplados ao *browser* ou por meio de *applets* Java. Esta estratégia permite uma maior flexibilidade do lado do cliente, que pode realizar operações locais de visualização e consulta sob os dados transferidos. O tempo de acesso inicial para transferência é maior que no caso anterior, mas muitas das operações posteriores serão realizadas localmente, o que resulta usualmente em um tempo de resposta médio melhor. Exemplos são os produtos *Geomedia Web Map* da INTERGRAPH e *SpringWeb* do INPE.

No entanto, as duas alternativas apresentam problemas. No primeiro caso, todos os dados ficam armazenados no servidor, e no segundo, todos precisam ser transferidos para o cliente. Segundo Osses *et al.* [OPC00], seria mais conveniente dispor de configurações *cliente-servidor*, que pudessem balancear os pedidos de consulta, permitindo uma apresentação e navegação local em parte dos dados e realizando acessos remotos ao servidor apenas quando estritamente necessário.

O *Geographical Antserver* [SGM98] e o *GeoMedia Web* [Int00] são iniciativas recentes sobre arquiteturas para aplicações SIG na Web.

2.5.3 Exemplos de sistemas SIG-Web

Dentre as opções mais utilizadas para a publicação e manipulação de dados geográficos na Web, podemos citar o SpringWeb [INP00], o MapGuide [Aut00], o GeoMedia Web Map [Int00] e o MapObjects Internet Map [ESR00]. Exemplos de sistemas não comerciais e de uso limitado incluem o Baltimore County [FM], GeoMed [SRG98], Geo-meta [GL98], VQuery [Jon98] e o *software* desenvolvido por Fernández *et al.* [FNC⁺00]. Alguns destes serão descritos em maiores detalhes a seguir.

SpringWeb O SpringWeb é um exemplo de sistema que permite publicar dados geográficos na Web. Trata-se de um *applet* Java que roda no cliente e precisa da realização de uma cópia local da base de dados, apresentando um nível razoável de interação com

esses dados. Entre as funções disponibilizadas estão operações de visualização como *zoom*, *pan* e navegação sobre os mapas, além da possibilidade de consulta aos atributos de um objeto geográfico específico [VPG00].

MapGuide O MapGuide utiliza um formato proprietário para os mapas a serem visualizados chamado SDF. No entanto, existe uma ferramenta para importar dados de uma série de outras bases, como ArcView e MapInfo, que possibilita a navegação sobre os mapas, mas sem permitir gerá-los dinamicamente [VPG00].

GeoMedia Web Map O Geomedia Web Map apresenta como principal vantagem a capacidade de publicar dados na Web de diferentes formatos, como MGE e ArcView Shapefiles, entre outros. Neste programa, os mapas são exibidos no formato ActiveCGM para clientes Microsoft ou como JPEG para clientes em outros sistemas operacionais. Uma de suas principais desvantagens advém da considerável quantidade de programação em *scripts* CGI necessária para construir uma aplicação completa. Esse sistema permite a geração dinâmica de mapas e a navegação através deles [VPG00].

MapObjects Internet Map O MapObjects Internet Map apresenta como desvantagem principal o fato de publicar os mapas como imagens, em geral nos formatos JPEG ou GIF, e suportar somente as bases de dados nos formatos proprietários da ESRI. Embora o usuário possa realizar consultas sobre os dados, em razão dos mapas estarem em formato *raster* não é possível navegar através de entidades gráficas.

A disseminação do uso de SIG na Web pode ser avaliada também a partir de variada gama de estudos de caso feitos em [Har98]. O texto cobre aplicações que variam desde sistemas de informação viária e de transportes até sistemas para controle da poluição sonora provocada por aeronaves.

2.5.4 Linguagens para desenvolvimento de sistemas Web

Dentre as principais alternativas para a montagem de páginas dinâmicas destacamos: CGI, APIs proprietárias, ASP e Java.

CGI CGI é um padrão de interface utilizado entre servidores HTTP e programas para se comunicarem. O servidor executa um programa (chamado *scripts*) cujo resultado é então transmitido ao cliente. Normalmente, um *script* CGI gera como resultado uma página HTML para ser exibida no navegador. Os programas CGI são executados na máquina onde está localizado o servidor Web. Eles recebem os dados através de variáveis

de ambiente e *entrada padrão* e os transmitem de volta ao servidor através da *saída padrão*. Podem ser escritos em qualquer linguagem (C, C++, PERL, TCL, entre outras). A principal desvantagem de se utilizar *scripts* CGI é que para atender a cada solicitação do usuário o servidor Web precisa criar um novo processo, o que resulta numa baixa eficiência. Cada processo produz uma nova conexão com o banco de dados utilizado e o servidor Web tem que esperar até que os resultados lhe sejam enviados. Outro ponto negativo é com relação à segurança, uma vez que os arquivos (*scripts*) não ficam inteiramente protegidos, pois devem ser armazenados nos subdiretórios CGI-bin do servidor. Ainda outra desvantagem está relacionada à baixa taxa de reutilização de código, em função das linguagens usadas para a programação [VPG00].

APIs proprietárias Muitos servidores Web incluem APIs proprietárias que expandem a sua funcionalidade. Os exemplos mais difundidos são a NSAPI e a ISAPI. A NSAPI é uma API do Servidor da Netscape na qual os programadores podem criar módulos binários que irão acrescentar e/ou substituir elementos para autenticação, autorização ou geração dinâmica de conteúdos. A ISAPI tem como princípio básico criar uma DLL que é carregada no servidor quando HTTP é inicializado e permanece instalada enquanto for necessária. Esta DLL irá gerenciar a conexão com a base de dados sem a necessidade de se criarem novas conexões, como ocorre com o CGI. O grande problema em adotar tais soluções é que elas são proprietárias, perdendo-se a portabilidade [VPG00].

Java A linguagem de programação Java tem ganhado rápida e larga aceitação nos últimos anos. Segundo Fernández *et al.* [FNC⁺00], a capacidade de programação, implementação multiplataforma, facilidades para o desenvolvimento de aplicações distribuídas, dentre outras características têm feito de Java uma das ferramentas mais desejáveis para o desenvolvimento de *software*. Para Sorokine *et al.* [SM98], Newton *et al.* [NGS97], JAVA é uma linguagem ideal quando se necessita incrementar a portabilidade e permitir o acesso a partir de qualquer *browser* Web.

Um programa escrito em Java pode ser desenvolvido para executar como um *applet*, um *servlet* ou como uma aplicação (um programa que reside no lado do cliente).

O *applet* é um programa baixado da Internet e executado no cliente. No caso de SIG-Web, o controle de apresentação do modelo e dos dados geográficos e seus atributos aos usuários são funções básicas deste programa. Além disso, o *applet* se preocupa em manter um cache dos dados do lado do cliente de forma a minimizar a transferência de dados de forma desnecessária entre o cliente e o servidor [OPC00].

Um *servlet*, por sua vez, é um módulo (programa escrito em Java) que é carregado dinamicamente para atender às solicitações de um servidor Web, ou seja, é uma extensão acrescentada ao servidor que aumenta a sua funcionalidade. Os servidores Web respondem

às solicitações dos usuários, geralmente, usando o protocolo HTTP através do envio de documentos escritos em HTML. Os *servlets* podem ser carregados em diversos servidores, pois a API utilizada para escrevê-los usa apenas o ambiente da Máquina Virtual do servidor [VPG00].

No caso de SIG-Web, o *servlet* fica localizado no servidor, sendo responsável por manter uma conexão com o banco de dados (relacional) e com o *applet*. Na conexão com o *applet* é necessário estabelecer um protocolo para que o *servlet* seja capaz de interpretar as solicitações, efetuar as atividades corretas e enviar as respostas adequadas. O protocolo utilizado pelo *servlet* para comunicação com o banco de dados relacional é o padrão SQL. A conexão é efetuada através da chamada de rotinas pertencentes à API JDBC. O servidor carrega o *servlet* que pode então aceitar diversas requisições dos clientes e retornar dados a eles [OPC00].

Segundo Vianna *et al.* [VPG00], as principais razões que indicam o uso de Java *servlets* são as seguintes:

- Eficiência: a inicialização de um *servlet* é feita apenas uma vez, ou seja, após ele ser carregado pelo servidor Web, as novas solicitações são apenas chamadas do método de serviço.
- Persistência: os *servlets* podem manter o estado entre solicitações.
- Portabilidade: os *servlets* são portatéis graças à escrita destes em Java.
- Robustez: Java possui um método bem definido para o tratamento de erros que minimiza a perda de memória por alocações indevidas (*garbage collector*).
- Segurança: os *servlets* herdam toda a segurança que é peculiar a um servidor Web, porém os *servlets* Java também podem contar com o Java Security Manager.
- Reutilização: como os *servlets* são escritos em Java, eles possuem as vantagens da utilização de uma linguagem orientada a objetos, como a possibilidade de reutilização.

Em suma, o *servlet* é um aplicativo que permanece em execução no servidor aguardando por solicitações dos clientes e tem a capacidade de atender diversas solicitações simultâneas. O *applet* pode estabelecer uma conexão com qualquer servidor onde exista um *servlet* preparado para receber suas solicitações.

Em adição, Java é visto como uma excelente linguagem de programação de banco de dados por causa da independência da plataforma, possibilidade do trabalho em rede, e

devido ao JDBC ². A API JDBC da linguagem Java é um conjunto de especificações que define como um programa escrito em Java pode se comunicar e interagir com um banco de dados, enquanto o JDBC é um módulo flexível de conectividade de banco de dados e acesso a SQL dinâmico para a linguagem de programação Java [BM98, CSH⁺98, WCF98].

Em SIG na Web, a linguagem Java permite distribuir os componentes SIG desenvolvidos através da rede, do servidor para o cliente, e permite a construção de três diferentes arquiteturas adaptadas a diferentes usuários: uma local com todos os dados integrados, uma Intranet e uma arquitetura Internet com um repositório remoto de dados. Em função disto, a comunidade de SIG está vivendo um aumento do interesse pela tecnologia Java-Web [FNC⁺00]. Vários sistemas utilizam *applets* de mapas interativos para permitir que o usuário possa interagir mais facilmente com o mapa na Web, como por exemplo o sistema GCMD [GCM].

ASP A solução da Microsoft para a criação de páginas dinâmicas é o ASP (Active Server Pages - Páginas de Servidor Ativas). O ASP é um ambiente para programação por *scripts* no servidor, que o desenvolvedor pode usar para criar páginas dinâmicas e interativas. Ele fica embutido nas páginas HTML, não sendo pré-compilado. As páginas ASP executam no servidor e não no cliente. É o próprio servidor que transforma os *scripts* em HTML padrão, fazendo com que qualquer *browser* do mercado seja capaz de acessar um site que usa ASP.

Entre os recursos que podem ser implementados via ASP, podemos citar:

- Programação em VBScript ou JavaScript.
- Acesso a banco de dados.
- Sessões (persistência de informações no servidor).

A grande vantagem da ASP é que pré-requisitos são necessários apenas do lado do servidor. No lado do cliente, o usuário pode utilizar qualquer *browser*, mesmo os que não suportam VBScript ou JavaScript. Por outro lado, o grande problema desta solução é que ela está ligada a um tipo particular de servidor - o Internet Information Server (IIS) - ou a alguma ferramenta que consiga fazer a tradução do ASP antes de enviar a resposta para o cliente.

Alguns sistemas optam por utilizar uma combinação destas alternativas. Por exemplo, o *National Comprehensive Cancer Network* (NCCN) [CNOP98] é um sistema de banco de

²JDBC é um pacote responsável por oferecer um *framework* para banco de dados SQL genérico, definindo uma API uniforme para uma variedade de fontes de dados.

dados baseado na Web que foi escrito usando ASP em combinação com HTML, JavaScript 1.1 e *applets* Java.

Como se verá no decorrer deste texto a versão atual do SINBIOTASP utiliza PERL, mas uma das propostas da dissertação é que as próximas versões utilizem Java.

2.6 Metadados

Metadados são dados sobre dados [ERI, FGD, TRK96]. Os metadados ajudam a decidir se os dados de uma fonte específica são úteis para outros propósitos que não aqueles para os quais eles foram originalmente coletados [TRK96]. Sendo assim, metadados e SIG podem ser relacionados, por exemplo, para facilitar consultas, encorajar o uso comum de informações geográficas [TRK96], ou mesmo definir controles de acesso aos dados. O sistema de informação do BIOTA/FAPESP, como se verá no capítulo 4, é fortemente calçado na noção de metadados. Grande parte dos sistemas de informação de biodiversidade disponíveis também se baseia no uso de metadados para agilizar consultas.

A noção de metadados é usada em muitas áreas diferentes com objetivos similares: melhor integração, intercâmbio, acesso e interpretação dos dados [TS97]. Metadados podem ser vistos como uma documentação sobre o dado, a partir da qual são representadas informações que descrevem o conteúdo, contexto, qualidade ou estrutura de um determinado conjunto de dados [Fag99]. Além disso, a utilização de metadados também auxilia o tratamento de dados heterogêneos, facilitando o acesso e atualizações destes dados, a manutenção de consistência, e a uniformidade e qualidade dos dados [Gon97].

De forma geral, o uso de metadados dentro das aplicações ambientais satisfaz vários propósitos [Fag99]:

- Gerenciamento da heterogeneidade dos dados ambientais. Metadados ajudam a contornar a heterogeneidade, descrevendo como os dados estão armazenados, permitindo que sejam desenvolvidas rotinas de conversão quando necessárias. Além do uso de metadados, uma nova vertente de pesquisa para integração de bancos de dados heterogêneos na Web é o uso de padrões do tipo XML [ABS00]. Esta vertente, no entanto, é aplicável tipicamente a sistemas textuais.
- Imprecisão dos dados ambientais. Metadados são utilizados para especificar a precisão dos dados.
- Diferença entre nomenclaturas. Metadados são úteis para unificar os nomes dos atributos, e para registrar os relacionamentos entre diferentes conjuntos de dados.
- Gerenciamento de um grande volume de dados. Estruturas de metadados servem como índice intermediário de acesso aos dados, agilizando sua recuperação.

Os metadados podem ser classificados de acordo com o tipo da informação armazenada, caracterizados da seguinte forma [URB97]:

- **Metadado Estrutural:** refere-se à estrutura (esquema) das fontes de dados sendo integradas. Pode ser usado para facilitar a tarefa de integração de *software*, e construção de analisadores léxicos e sintáticos para o tratamento de tipos de dados retornados por processos externos.
- **Metadado Semântico:** refere-se às informações contidas dentro da estrutura de dados acima mencionada, e especifica as informações sobre seus tipos semânticos e o inter-relacionamento entre estes tipos.
- **Metadado de Custo:** refere-se às informações sobre o desempenho (tempo ou custo de processamento) associado a diferentes fontes de dados. Este tipo é usado em otimização de consultas.
- **Metadado de Confiabilidade:** refere-se à informação de confiabilidade das fontes de dados, relacionando frequentemente o tempo em que foi processada uma alteração ou a manutenção dos dados no tempo. Tal como o metadado de custo, o metadado de confiabilidade é frequentemente usado na avaliação e reescrita de um esquema de consulta.
- **Metadado Ancestral:** refere-se às informações genéricas sobre a fonte dos dados (quem criou, quando, onde e por quê). Pode ser usado para avaliar a confiabilidade da fonte, ou avaliar a capacidade de uma fonte de dados de prover certos tipos de informação.

Moura *et al.* [MCB98], por sua vez, apresentam uma taxonomia de metadados voltada ao contexto de recursos em ambientes distribuídos, visando principalmente busca na Web. Esta taxonomia foi desenvolvida a partir do estudo de diferentes categorias de padrões de metadados e se baseia em suas características funcionais:

- **Metadados para descoberta de recursos:** é o conjunto de atributos necessários à descoberta e identificação de recursos na Web. Inclui, por exemplo, campos de descrição de assunto, formato e referências bibliográficas.
- **Metadados para disponibilidade de recursos:** são os campos que descrevem as condições para acessar e recuperar recursos. Incluem, por exemplo, atributos que determinam como usar, acessar e assegurar a autenticidade dos recursos.
- **Metadados para uso de recursos:** são os atributos que permitem determinar o uso adequado dos recursos. Exemplos são descritores de qualidade, propósito e contexto.

- Metadados para administração e controle de recursos: são os itens que permitem controlar e fazer auditoria sobre as informações sobre os recursos e seus metadados. Incluem, por exemplo, campos de histórico de uso, controle de atualização e administração.

2.7 Resumo

Este capítulo apresentou alguns conceitos básicos que serão utilizados e necessários para o entendimento do restante do texto.

Alguns aspectos de Sistemas de Informação Geográfica foram abordados, por serem relevantes ao contexto da dissertação, tais como: modelagem de campos ou objetos, paradigmas de interação de um usuário com um SIG, exemplos de sistemas que manipulam mapas, dentre outros.

Foram apresentados os operadores espaciais disponíveis, segundo Faria [Far98] e Gutting [Gut94], para a implementação de aplicações, bem como as operações primitivas de acordo com Cereja [Cer96] e Ciferri [Cif95].

Em adição, este capítulo discutiu questões referentes a Sistemas de Informação Geográfica na Web, tais como: arquitetura, interoperabilidade e SIG, exemplos de sistemas e as principais alternativas de linguagens para desenvolvimento na Web. Finalmente, foram abordadas algumas noções sobre metadados em SIG para sistemas ambientais.

Capítulo 3

Análise comparativa de alguns Sistemas de Informação Ambiental

Este capítulo analisa sistemas de informação ambiental [SLF⁺98, Uni, LP98, UDK, CDS, GCM, ERI, EOS, FGD, GIL, Spe, Mal], visando principalmente identificar as consultas comumente utilizadas. Esta análise servirá de base para o restante do trabalho da dissertação.

O estudo dos sistemas foi realizado a partir do uso (via Internet) dos diferentes sistemas mencionados e análise das documentações correlatas. Os serviços de busca disponibilizados pelos sistemas foram analisados sob os seguintes aspectos: objetivo, conteúdo do banco de dados, características específicas, arquitetura, tipos de consulta e interfaces de exibição disponíveis. A avaliação dos tipos de consulta e interfaces de exibição disponíveis considerou as seguintes questões: interfaces apresentadas, natureza dos dados solicitados e formato de saída dos documentos recuperados.

O capítulo está organizada da seguinte forma. As seções 3.1 e 3.2 estabelecem as principais categorias para classificação de sistemas de informação ambiental, enquanto a seção 3.3 subsequente avalia alguns sistemas existentes em função destas categorias. A seção 3.4 apresenta um resumo dos assuntos discutidos neste capítulo.

3.1 Critérios de comparação de Sistemas de Informação Ambiental

Os sistemas de informação podem atender a diferentes finalidades e possuir diversos tópicos de interesse, não se focando em um único objetivo. Do ponto de vista de escopo, sistemas de informação ambiental são aqui classificados como sistemas com enfoque único e sistemas com enfoque múltiplo.

- **Sistemas com enfoque único.** Os sistemas assim classificados possuem um único objetivo. No entanto, o fato de possuir um foco único não diminui a importância dos sistemas enquadrados neste grupo. O *Herbarium Specimen Browser* [SLF+98], *Species Analyst* [Uni], *Species 2000* [Spe] e o *Malaysia Biodiversity Online* [Mal] são exemplos de sistemas voltados para o objetivo específico de atender as consultas de especialistas em biologia relativas a biodiversidade. Em outras palavras, todos os módulos disponíveis visam um único objetivo.
- **Sistemas com enfoque múltiplo.** Sistemas com enfoque múltiplo são aqueles que congregam mais de um tópico de interesse. O ECN [LP98], UDK [UDK], CDS [CDS], GCMD [GCM], *Environmental Australia On Line* [ERI], EOSDIS [EOS], GILS [GIL] e o FGDC [FGD] são exemplos de sistemas que visam atender a mais de uma finalidade, com funções totalmente independentes. Por exemplo, o ERIN possui um conjunto de módulos de biodiversidade, outro conjunto relativo a legislação ambiental, outro conjunto relativo a acervos bibliográficos e assim por diante.

Um outro critério de comparação entre os sistemas avaliados refere-se ao gerenciamento e recuperação de metadados e/ou dados. A partir deste critério de comparação, os sistemas apresentados neste capítulo podem ser divididos em duas categorias: sistemas limitados ao gerenciamento e recuperação de metadados e sistemas responsáveis pelo gerenciamento e recuperação de dados.

- **Gerenciamento e recuperação de metadados.** Este tipo de sistema manipula metadados relacionados a informação como: *quem organiza, faz o que e onde está* determinado arquivo de dados. Estes metadados apontam para os arquivos, que podem a seguir ser recuperados de forma monolítica. O ECN [LP98], UDK [UDK], CDS [CDS], GCMD [GCM] e o GILS [GIL] são exemplos de sistemas que gerenciam e recuperam metadados.
- **Gerenciamento e recuperação de dados.** Os sistemas desta categoria permitem acessar dados a partir do uso de mapas, de metadados ou de consultas diretas às fontes de dados. O *Herbarium Specimen Browser* [SLF+98], *Species Analyst* [Uni], *Environmental Australia On Line* [ERI], *Species 2000* [Spe], *Malaysia Biodiversity Online* [Mal], EOSDIS [EOS] e o FGDC [FGD] são exemplos de sistemas de informação que manipulam dados.

3.2 Tipos de consultas e interfaces

Os vários sistemas ambientais analisados nesta dissertação foram examinados quanto aos tipos de consulta permitidos e aspectos da visualização do resultado da consulta. As

principais categorias de consulta disponíveis são:

- **Consultas por palavras-chave.** UDK [UDK], CDS [CDS], GILS [GIL], GCMD [GCM], *Environmental Australia On Line* [ERI], EOSDIS [EOS], FGDC [FGD] e *Malaysia Biodiversity Online* [Mal] permitem a realização de consultas a partir de uma dada palavra-chave, retornando os dados que a contenham.
- **Consultas por hierarquia de palavras.** *Herbarium Specimen Browser* [SLF+98], CDS [CDS], FGDC [FGD], GCMD [GCM] e UDK [UDK] permitem que o usuário realize consultas a partir de uma hierarquia de palavras pré-definidas. Com isto, é possível realizar consultas mais específicas, descendo até os níveis mais baixos da hierarquia. Por exemplo, o sistema *Herbarium Specimen Browser* [SLF+98] permite ao usuário navegar através de uma hierarquia de classificação taxonômica (família, gênero e espécie), onde a partir de qualquer nível da hierarquia é possível obter informações referentes à distribuição das espécies e às coletas realizadas.
- **Consultas por campos específicos.** *Herbarium Specimen Browser* [SLF+98], GILS [GIL], UDK [UDK], CDS [CDS], GCMD [GCM], EOSDIS [EOS], *Environmental Australia On Line* [ERI], *Species Analyst* [Uni], ECN [LP98], FGDC [FGD], *Species 2000* [Spe] e *Malaysia Biodiversity Online* [Mal] possibilitam que o usuário realize consultas especificando termos para os campos de seu interesse. Por exemplo, o sistema *Species 2000* [Spe] permite ao usuário encontrar organismos cujo nome científico (gênero e/ou espécie) seja igual ao especificado pelo usuário.
- **Consultas por índice alfabético.** *Environmental Australia On Line* [ERI] apresenta todos os conjuntos de dados organizados através de um índice alfabético. Desta forma, o usuário pode selecionar uma letra e recuperar os conjuntos de dados. Por exemplo, o sistema ERIN (*Environmental Australia On Line*) [ERI] fornece uma lista de letras de A a Z, quando o usuário seleciona uma determinada letra, todos os termos iniciados pela letra selecionada são apresentados para que o usuário possa selecionar o termo de interesse. Como exemplo de termos inicializados com a letra E temos: eco-eficiência, economia, educação, dentre outros.
- **Consultas gráficas.** *Herbarium Specimen Browser* [SLF+98], FGDC [FGD], GCMD [GCM], *Environmental Australia On Line* [ERI], EOSDIS [EOS], ECN [LP98] e *Malaysia Biodiversity Online* [Mal] disponibilizam interfaces gráficas (mapas) que facilitam a especificação dos atributos espaciais (coordenadas geográficas) relacionados ao conjunto de dados descrito.

Os formatos de exibição do resultado podem ser classificados em:

- **Formato textual.** As informações recuperadas podem ser: (a) metadados que apontam um conjunto de metadados detalhado, (b) metadados que apontam para documentos, (c) um conjunto de metadados detalhado, (d) um documento inteiro direto, ou (e) um conjunto resumido e/ou detalhado de dados. As formas de recuperação descritas em (a) e (b) são descritas no texto como lista de *links*. *Herbarium Specimen Browser* [SLF+98], CDS [CDS], GILS [GIL], UDK [UDK], GCMD [GCM], *Environmental Australia On Line* [ERI], EOSDIS [EOS], *Species Analyst* [Uni], ECN [LP98], FGDC [FGD], *Species 2000* [Spe] e *Malaysia Biodiversity Online* [Mal] são exemplos de sistemas que permitem uma exibição textual do resultado das consultas.
- **Formato gráfico.** O resultado de uma consulta é apresentado sob a forma de pontos ou regiões em um mapa, indicando a localização do conjunto de dados. *Herbarium Specimen Browser* [SLF+98], GCMD [GCM], *Environmental Australia On Line* [ERI] e *Species Analyst* [Uni] são exemplos de sistemas que permitem uma exibição gráfica do resultado das consultas. Nos casos específicos do ECN [LP98] e do GCMD [GCM], os resultados de uma consulta são apresentados através de gráficos de barra. A figura 3.1 permite visualizar um exemplo da interface de gráfico de barras.

3.3 Visão geral dos Sistemas de Informação

Esta seção apresenta características gerais dos sistemas de informação, tais como: o que é, objetivo, arquitetura, característica específica e um maior detalhamento dos tipos de consulta e formato de visualização dos resultados. Os sistemas detalhados a seguir são apresentados de forma agrupada, onde o primeiro grupo relaciona os sistemas que gerenciam e recuperam dados e metadados, enquanto o segundo grupo relaciona os sistemas que gerenciam e recuperam metadados.

3.3.1 Gerenciamento e recuperação de dados

O *Herbarium Specimen Browser* [SLF+98], *Species Analyst* [Uni], *Environmental Australia On Line* [ERI], *Species 2000* [Spe], *Malaysia Biodiversity Online* [Mal], EOSDIS [EOS] e o FGDC [FGD] são exemplos de sistemas de informação que manipulam dados e metadados.

Herbarium Specimen Browser O *Herbarium Specimen Browser* [SLF+98] é um sistema baseado na Web que gerencia uma das maiores coleções de espécies de plantas protegidas do herbário do Texas, valorizando principalmente a distribuição geográfica de

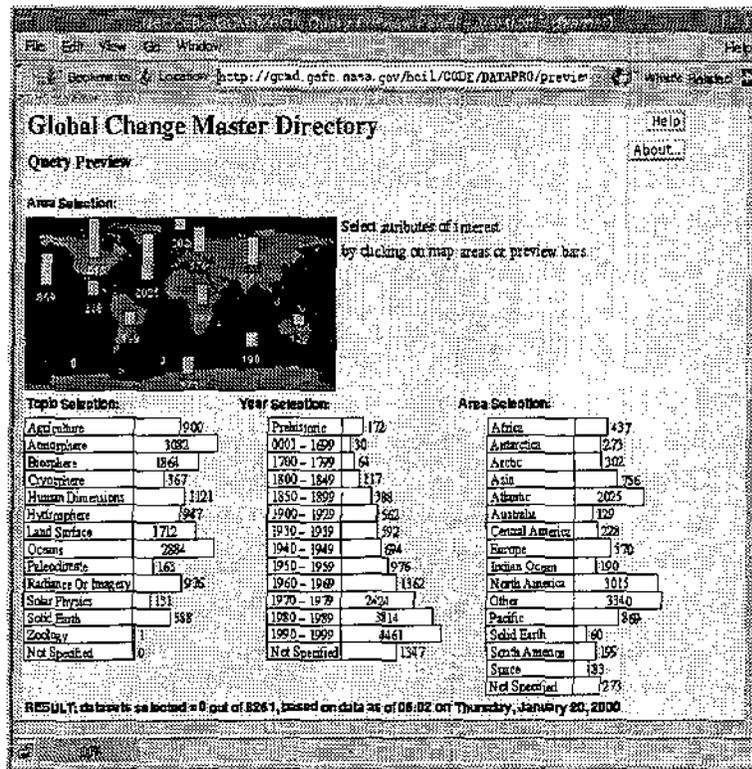


Figura 3.1: Exemplo de resultado gráfico de uma consulta ao sistema GCMD [GCM]

vários grupos de plantas e permitindo a realização de consultas sobre mapas. Ele visa permitir que qualquer tipo de pessoa acesse as informações via Internet. O banco de dados do sistema armazena todas as informações relacionadas a coletas de espécies. Estas informações estão georeferenciadas, possibilitando, assim, o interrelacionamento destas informações com mapas.

Este sistema é classificado como um sistema com enfoque único, devido ao fato de ter como único foco de interesse o gerenciamento das espécies de plantas protegidas do herbário do Texas. Em adição, ele permite o gerenciamento e recuperação de dados a partir do uso de mapas, de metadados ou de consultas diretas às fontes de dados.

Características específicas A interface é composta por três *frames* gerados dinamicamente:

- **Frame superior** Este *frame* contém a indicação do modo de operação do *browser*, além de permitir a seleção de outros modos de operação. O *browser* pode operar em modo texto ou gráfico (mapas).
- **Frame de controle** Este *frame* contém os controles usados para alterar a exibição

da informação no *frame* de exibição (lista de coletas ou dados completos). Todos os controles são textuais. Na parte superior do *frame* de exibição são exibidos os critérios de seleção: município, herbário, critério de filtragem do cultivo e a identidade dos itens listados pela categoria taxonômica.

- **Frame de exibição** Exibe as informações contidas no banco de dados de coletas, que podem ser visualizadas em forma de lista ou dados completos. No formato de lista, o usuário conta com funções de expansão, contração e filtragem dos dados que compõem a lista. Cada item da lista provê um *link* para acessar os mapas relacionados à distribuição do item em questão sobre o território do Texas. No formato de dados completos, o usuário pode visualizar todos os metadados armazenados.

O sistema pode apresentar resultados de vários modos, dentro do *frame* de exibição: visão geral taxonômica de alto nível, listas de coletas detalhadas, e mapas mostrando a distribuição de coletas e espécies por município.

Apesar de todas as funcionalidades oferecidas, o *Herbarium Specimen Browser* não disponibiliza uma maneira de visualizar e selecionar o conjunto de coletas a partir de nomes vulgares.

Arquitetura O sistema se baseia em uma coleção de programas CGI (*Common Gateway Interface*), baseado em *frames* e coordenado por Javascript. Sendo assim, para utilizar o *Herbarium Specimen Browser* é necessário um *browser* WWW que possa tratar Javascript 1.1.

Em adição, o sistema usa *cookies* para se manter informado do estado corrente da interação do usuário; estes *cookies* não são expiráveis, não havendo desta forma escrita no disco local do usuário. Este mecanismo pode ser usado para desviar alguns avisos sobre *cookies* enquanto o sistema está em uso, visando evitar uma inundação de caixas de diálogo informando o usuário sobre a chegada de *cookies*.

Tipos de consulta e interface de exibição O sistema provê consultas por campo específico, consultas por hierarquia de palavras e consulta gráfica, descritas a seguir.

- **Consulta por campo específico** A consulta por campo específico permite que o usuário defina a pesquisa de acordo com o campo de interesse. A atualização do *frame* de exibição é automática a partir da seleção de um valor para o campo de interesse no *frame* de controle.

Como exemplo de consultas por campo específico, existe a opção de seleção pelo campo *número de acesso da coleta*, disponível a partir da informação do *nome do herbário* e do *número de acesso da coleta*. As informações referentes àquela coleta

armazenadas no banco de dados são recuperadas. Se nenhum *número de acesso da coleta* for informado, todas as coletas do herbário selecionado são exibidas.

- **Consulta por hierarquia de palavras** A consulta hierárquica possibilita ao usuário realizar consulta a partir de uma hierarquia de palavras pré-definida ou definida por uma seleção anterior (consulta gráfica ou por campo específico).

Um exemplo de geração de hierarquia é a seleção de um município no mapa, de tal forma que todas as coletas relacionadas ao município selecionado sejam automaticamente exibidas de forma hierárquica no *frame* de exibição.

- **Consulta gráfica** A consulta gráfica disponibiliza um mapa, visando ajudar a especificação dos atributos espaciais (região). A partir da seleção de uma região ou município no mapa é possível visualizar, em modo de lista, todas as coletas individuais da região ou município especificado.

O modo de seleção gráfica permite interagir com o mapa de diferentes formas, tais como: (a) selecionando um ou mais municípios em uma caixa de seleção, que são automaticamente projetados no mapa, (b) selecionando uma região pré-definida ou (c) clicando diretamente sobre o mapa. Este mapa encontra-se dividido por municípios. As regiões são conjuntos pré-definidos de municípios.

A seleção de uma região (conjunto de municípios) resulta na seleção dos nomes dos municípios constituintes na caixa de seleção e a exibição em um mapa menor (prévia da exibição no mapa principal), sendo possível transferir o estado do mapa menor para o mapa central.

O sistema permite uma visualização textual e gráfica dos resultados. Ambas as formas de visualização são exibidas no *frame* de exibição. O formato textual está disponível quando o sistema define a forma de visualização de lista ou dados completos. O formato gráfico permite a apresentação do mapa com os municípios ou regiões selecionados, de acordo com o filtro definido. Selecionando na interface o *link* mapa é possível ver a localização e a distribuição das espécies ou amostras coletadas, e ter uma idéia de quantas (intervalo numérico) espécies ou coletas foram registradas no município em questão.

Species Analyst O *Species Analyst* [Uni] é um sistema de análise de biodiversidade desenvolvido pela Universidade do Kansas. Basicamente, esta ferramenta provê acesso simultâneo, de forma transparente, a múltiplos bancos de dados distribuídos de coleções biológicas a partir de um *browser* Web, onde o usuário pode selecionar as fontes de dados e definir o conjunto de dados de interesse.

Este sistema é classificado como um sistema com enfoque único, uma vez que tem como único foco de interesse a análise de coleções biológicas. Em adição, ele permite o

gerenciamento e recuperação de dados a partir de metadados ou de consultas diretas às fontes de dados.

Características específicas Este sistema e o *Herbarium Specimen Browser* são bastante distintos dos demais descritos nesta dissertação (como por exemplo GILS [GIL] e CDS [CDS]), pois enquanto os demais visam recuperar dados globais (em geral arquivos), estes permitem o acesso a tuplas. As tuplas do *Species Analyst* referem-se a diferentes tipos de espécies (animais ou vegetais) observadas e catalogadas em diversos bancos de dados de coleções. O sistema dispõe de dois tipos básicos de função:

- **Recuperação.** Permite recuperar dados específicos relativos a uma única espécie (animal ou vegetal) retornados em tuplas padronizadas. Este processamento é realizado em duas etapas: em primeiro lugar é feita uma busca aos vários *sites*, recuperando os registros de interesse a partir de suas descrições de metadados. Uma vez recuperados, estes registros são processados pelo *Species Analyst* visando uma análise espacial.
- **Análise.** Permite executar funções de análise espacial e diferentes tipos de simulação sobre o conjunto de tuplas recuperado.

O sistema tem um módulo que permite gerar o modelo de prognóstico de distribuição de espécies (ver <http://biodi.sdsc.edu> para maiores detalhes). O resultado é visualizado através do SIG ArcView da ESRI. O modelo de prognóstico gera um mapa de probabilidade indicando a distribuição de uma espécie particular baseado em um conjunto de dados da coleta. Para utilizar a extensão do *Species Analyst*, o cliente deve ter instalado ArcView 3.0a e a extensão do *Spatial Analyst*. A extensão foi testada somente na plataforma Microsoft Windows NT 4.0.

Arquitetura O *Species Analyst* permite montar um documento HTML para a apresentação de todas as informações armazenadas no banco de dados referente a cada nome científico, inclusive com os dados de localização espacial (formato texto), tempo, dentre outros. Em adição, este sistema permite baixar os dados no formato XML e Microsoft Excel ou qualquer outra aplicação que suporta interfaces *script* apropriadas [Uni].

Tipos de consulta e interface de exibição O sistema provê consultas apenas por **campo específico (nome da espécie)**. A seleção por campo específico exige a definição de pelo menos uma fonte de dados, para só depois restringir o campo de pesquisa e o valor do critério de busca. Durante a pesquisa nos servidores especificados pelo usuário, um

feedback constante do *status* da conexão é exibido. Quando os resultados são retornados, o usuário pode visualizar os resultados de cada servidor individualmente.

O formato de apresentação dos resultados pode ser textual ou gráfico. No formato textual, a informação é apresentada primeiramente de forma resumida, facilitando a visualização geral do resultado da consulta. O resultado é apresentado sob a forma de uma lista. Cada item da lista é um *link* direto para informações mais detalhadas, possibilitando acesso a todos os metadados relacionados ao item selecionado.

Já o formato gráfico permite visualizar um mapa da distribuição. Este mapa provê uma indicação da distribuição global do local das coletas para os registros identificados pela consulta, selecionando uma região no mapa.

Environmental Australia On Line O ERIN (*Environmental Australia On Line*) [ERI] é um sistema de informação desenvolvido pelo departamento do meio ambiente australiano. Este sistema apresenta como principal objetivo o desenvolvimento de mecanismos que auxiliem as tomadas de decisão ambiental, os planos de recuperação de espécies em extinção, e a manutenção de um repositório nacional de informação sobre *sites* ambientais. Ele é composto por vários subsistemas ambientais específicos (por exemplo, preservação da água, eucaliptos, espécies aquáticas).

Este sistema é classificado como um sistema com enfoque múltiplo, uma vez que é composto por vários subsistemas ambientais específicos. Em adição, ele permite o gerenciamento e recuperação de dados a partir do uso de mapas, metadados ou de consultas diretas às fontes de dados.

Características específicas O *EnviroMaps Online* do ERIN é uma interface para caracterizar e exibir informações ambientais espaciais. Esta interface é uma aplicação SIG para WWW, provendo uma interface HTML simples para gerenciar operações espaciais [ERI].

A interface do *EnviroMaps Online* está dividida em três *frames*: cabeçalho, mapa e legenda. O *frame* cabeçalho exibe o título *EnviroMaps Online* e uma descrição da tecnologia utilizada para o seu desenvolvimento, no caso, SDE v3.0.

O *frame* mapa exibe um mapa central e um outro da região em foco no mapa central, e gerencia o alcance do mapa. O mapa central é gerado a partir de camadas especificadas no *frame* legenda, clicando sobre qualquer região deste mapa é possível obter informação quanto ao alcance do mapa. O mapa central permite realizar operações usuais como: *pan*, *zoom in*, *zoom out* e *zoom box*. Em adição, a interface conta ainda com uma barra de escala que fornece a indicação de distâncias no mapa.

O *frame* legenda exibe a legenda do mapa. Este *frame* é responsável por controlar as camadas exibidas no mapa (adição e remoção), além de como estas camadas estão

representadas no mapa. As camadas possíveis são pré-definidas, contendo informações topográficas (por exemplo, rios e aeroportos), de localização de espécies (por exemplo, eucaliptos), dentre outras. O mapa central exibido por esta interface pode ser gravado como um arquivo GIF ou PostScript.

O ERIN permite especificar as fontes de dados nas quais a consulta será realizada. Os nós são acessados simultaneamente. Quando todos os resultados das consultas processadas nos nós são retornados, o usuário pode visualizar os resultados de servidores individuais, clicando no hipertexto do nome do banco de dados.

Arquitetura O ERIN é baseado no ASDD (*Australian Spatial Data Directory*). O ASDD provê as interfaces de pesquisa para descrição de um conjunto de dados geoespaciais (metadados) de todas as jurisdições da Austrália. O protocolo Z39.50 é usado para simultaneamente pesquisar cada uma das listas de dados que forma o ASDD.

O ERIN tem vários subsistemas, cada qual com a sua arquitetura. O monitoramento do progresso da consulta por um *gateway* WWW responsável por combinar os resultados das consultas processadas nos nós quando todas as pesquisas são completadas é um exemplo de arquitetura implementada.

Tipos de consulta e interface de exibição Os serviços de busca disponíveis variam para cada módulo do ERIN, são eles: por palavras-chave, por campo específico, por índice alfabético e gráficas. As consultas por palavras-chave, campo específico e por índice alfabético seguem a característica padrão definida para estes tipos de consulta. A consulta gráfica permite selecionar uma região no mapa, através da definição do canto superior esquerdo e do canto inferior direito de um retângulo que irá fazer parte dos critérios de seleção, além de disponibilizar interfaces gráficas que permitem a composição de diversas camadas de informação, tais como: topografia, localização de espécies e área. Em adição, o sistema possui um módulo de consulta capaz de retornar uma informação textual quando o usuário clica sobre o mapa.

O sistema também permite a combinação da pesquisa textual (consulta por palavras-chave, por campo específico ou por índice alfabético) com a pesquisa gráfica.

A apresentação dos resultados das consultas pode ser realizada em dois formatos: textual e gráfico. O formato textual possibilita ver todos os registros retornados, com uma breve descrição e *links* para o registro completo. Já no formato gráfico os resultados são apresentados em um mapa sob a forma de pontos que indicam a localização dos conjuntos de dados que satisfazem a consulta especificada.

Species 2000 O *Species 2000* [Spe] é uma "federação" de organizações trabalhando com usuários, taxonomistas e agências patrocinadoras e tem o objetivo de enumerar todas as

espécies de seres vivos conhecidos na Terra como um conjunto de dados base para os estudos da biodiversidade global.

O programa *Species 2000* foi estabelecido pela IUBS (*International Union of Biological Sciences*) em co-operação com o CODATA (*Committee on Data for Science and Technology*) e o IUMS (*International Union of Microbiological Societies*) em setembro de 1994.

O *Species 2000* é classificado como um sistema com enfoque único, uma vez que tem como único foco de interesse o estudo da biodiversidade global. Em adição, ele permite o gerenciamento e recuperação de dados a partir de metadados ou de consultas diretas às fontes de dados.

Características específicas A implementação do programa *Species 2000* envolve:

- a formação de uma "federação" dos bancos de dados taxonômicos existentes: *Species 2000*.
- o estímulo ao estabelecimento de bancos de dados de espécies globais para todos os grupos de organismos, para acelerar a conclusão dos bancos de dados existentes e o desenvolvimento de novos.
- o desenvolvimento de procedimentos para manter os bancos de dados e atualizar a taxonomia.
- o trabalho de prover, como objetivo final, um índice de todas as espécies conhecidas baseado em computação.

Tipos de consulta e interface de exibição O *Species 2000* [Spe] permite somente consultas por campo específico. O sistema fornece duas formas para realizar estas consultas: na primeira forma, o usuário pode especificar o banco de dados que será pesquisado (vertebrados, plantas e/ou micro-organismos), e selecionar o campo de interesse (gênero, ou gênero e espécie).

Na segunda forma, o usuário não pode especificar o banco de dados, e a consulta pode ser processada de acordo com os seguintes campos: nome científico (gênero e/ou espécie), nome comum e referência (autor, ano e/ou título).

A visualização textual dos resultados das consultas exhibe uma lista dos resultados com uma breve descrição dos registros que satisfazem os critérios de seleção com *links* para metadados encarregados de prover uma informação mais detalhada.

Malaysia Biodiversity Online O *Malaysia Biodiversity Online* [Mal] é um sistema de informação ambiental dedicado ao avanço e a disseminação do conhecimento sobre a biodiversidade da Malásia. Em adição, este projeto foi proposto tendo em mente facilitar o compartilhamento de informações no campo de plantas medicinais e aromáticas da Malásia entre cientistas.

Este sistema foi estabelecido no início de 1997, fundado através de contribuições financeiras do NEDO do Japão e do MOSTE da Malásia para o *Malaysian-Japanese Tropical Bioresources Program*.

O *Malaysia Biodiversity Online* é classificado como um sistema com enfoque único, uma vez que tem como único foco de interesse a disseminação da biodiversidade da Malásia. Em adição, ele permite o gerenciamento e recuperação de dados a partir do uso de mapas, metadados ou de consultas diretas às fontes de dados.

Tipos de consulta e interface de exibição O *Malaysia Biodiversity Online* permite realizar consultas por palavra-chave, campo específico e gráfica. As consultas por palavra-chave e por campo específico seguem a característica padrão definida para estes tipos de consulta. A consulta gráfica permite ao usuário obter informações textuais clicando sobre o mapa.

Este sistema disponibiliza apenas uma visualização textual dos resultados das consultas. O formato do resultado (conjunto de dados ou somente metadados sem *links* para informações mais detalhadas) depende da opção de pesquisa escolhida.

EOSDIS O EOSDIS [EOS] é um sistema de informação ambiental desenvolvido pela NASA. Seu principal objetivo é gerenciar os dados vindos de satélites e programas responsáveis pelas pesquisas sobre a Terra.

Este sistema é classificado como um sistema com enfoque múltiplo, uma vez que visa gerenciar vários programas responsáveis pelas pesquisas sobre a Terra. Em adição, ele permite o gerenciamento e recuperação de dados a partir do uso de mapas, metadados ou de consultas diretas às fontes de dados.

Características específicas Este sistema possui um *browser* de acrônimos escrito em Java. O sistema permite ao usuário pesquisar por acrônimos, ou listar acrônimos definindo a categoria prefixo, sufixo ou parte de uma palavra, no texto, cuja opção retorna uma lista de acrônimos. Em adição, os acrônimos sugeridos pelos especialistas da NASA como também abreviações específicas para a Internet, WWW e computação foram incorporados ao sistema. Por exemplo, o acrônimo EOSDIS é proveniente da palavra *Earth Observing System/Data and Information System*.

Tipos de consulta e interface de exibição Após a utilização do EOSDIS, concluímos que o sistema provê os seguintes serviços de busca: consultas por campos específicos, por palavras-chave e gráfica. As consultas por campo específico e por palavras-chave seguem a característica padrão definida para estes tipos de consulta. As consultas gráficas permitem definir uma região geográfica como critério de seleção, através da especificação de um retângulo sobre o mapa. O usuário deve informar o canto superior esquerdo e o canto inferior direito do retângulo sobre o mapa, ou definir o intervalo de longitude e latitude manualmente.

Os resultados das consultas realizadas no EOSDIS podem ser visualizados somente no formato textual. Neste caso, os resultados são apresentados na forma de uma lista de *links*, onde um ou vários itens podem ser selecionados para obter documentos detalhados.

FGDC O FGDC [FGD] é uma comissão responsável por especificar padrões e definir um sistema que possibilite acesso via Internet a um conjunto de repositórios de dados geoespaciais provenientes de aproximadamente 80 servidores de dados espaciais dos Estados Unidos. Em adição, o FGDC é um exemplo de um padrão de metadados específico para a catalogação de dados ambientais ou geográficos.

O FGDC é classificado como um sistema com enfoque múltiplo, uma vez que tem como objetivos especificar padrões e definir um sistema que permita acesso a dados geoespaciais. Além disso, permite o gerenciamento e recuperação de dados a partir do uso de mapas, metadados ou de consultas diretas às fontes de dados.

Características específicas As coleções de dados podem ser pesquisadas através de interfaces baseadas em metadados. Estes metadados incluem descrições do conjunto de dados e informações de como adquirir a informação através da Internet por meio de *links* hipertexto.

Arquitetura O sistema usa uma interface baseada em Java [FGD]. Durante a pesquisa nos servidores especificados pelo usuário, um *feedback* constante do *status* da conexão é exibido. Quando os resultados são retornados, o usuário pode visualizar os resultados de cada servidor individualmente.

Tipos de consulta e interface de exibição As seguintes formas de consulta são disponibilizadas [FGD]: consultas por hierarquia de palavras, por campo específico, por palavra-chave e consultas gráficas. O usuário também pode submeter consultas com restrições temporais (consulta por campo específico) e espacial (consultas gráficas), através de *applets*.

As consultas por hierarquia de palavras, campo específico e por palavras-chave seguem a característica padrão definida para estes tipos de consulta. Na consulta gráfica, a restrição espacial pode ser feita de quatro formas distintas: clicando no mapa, definindo uma região no mapa, selecionando um item (região) em uma caixa de seleção ou informando as coordenadas numéricas desejadas. A região definida no mapa pode ser usada para realizar um *zoom in* ou *zoom out*. Em adição, esta interface permite acrescentar no mapa camadas de detalhes, tais como: limites políticos, rios, ou ambos.

A visualização textual dos resultados das consultas exibe uma lista dos resultados com uma breve descrição dos registros que satisfazem os critérios de seleção com *links* para informações mais detalhadas.

3.3.2 Gerenciamento e recuperação de metadados

O ECN [LP98], UDK [UDK], CDS [CDS], GCMD [GCM] e o GILS [GIL] são exemplos de sistemas que gerenciam e recuperam metadados.

Environmental Change Network O sistema UK (*umweltdatenkatalog*) *Environmental Change Network* (ECN) [LP98] é um catálogo de dados ambientais, baseado em acesso a metadados. Este sistema faz uso da Internet para disseminar a maioria dos dados e informação corrente para sua ampla comunidade de usuários. O ECN é patrocinado por um consórcio de agências e departamentos ambientais e co-ordenado pelo *UK Natural Environmental Research Council*.

Este sistema é classificado como um sistema com enfoque múltiplo, uma vez que funciona como um catálogo de dados ambientais. Em adição, ele permite o gerenciamento e recuperação de metadados ambientais.

Características específicas Os dados são obtidos através dos gerenciadores das estações de monitoramento do sistema ECN, que conta com 54 estações de monitoramento, sendo 12 estações terrestres e 42 estações aquáticas.

A captura dos dados ECN requer principalmente métodos de entrada manual de dados. Métodos automáticos são correntemente usados para registrar variáveis meteorológicas através de estações meteorológicas automatizadas, dentre outras estações. Fotos aéreas e imagens de sensoriamento remoto de estações ECN estão sendo incorporadas dentro do sistema de informação geográfica do ECN [LP98].

As estações ECN são responsáveis por enviar os dados coletados, usando formatos pré-definidos que incluem regras para tratamento de valores ausentes e informação de qualidade dos dados. Alguns dos formulários de entrada são validados automaticamente. Quase todos os dados são enviados por e-mail. O carregamento dos dados, a transfor-

mação, validação e o *software* de entrada estão sendo desenvolvidos para cada tipo de conjunto de dados ECN, com regras para tratamento das transações de atualização do banco de dados. Os dados estão armazenados em um banco de dados, e os gráficos e planilhas de dados disponíveis são automaticamente gerados para exibição na Web.

Em adição, o sistema provê um controle de acesso e permite análises espaço-temporal de uma série de variáveis em uma variedade de escalas.

Arquitetura A figura 3.2 mostra a arquitetura do sistema. Ele usa um banco de dados Oracle, com ligações para os SIGs Arc/Info e ArcView, para o tratamento de dados espaciais. Entretanto, segundo Lane *et al.* [LP98], uma questão técnica que causou problemas para o ECN foi a falta de integração ou interoperabilidade entre banco de dados, sistema de informação geográfica, sistema de análise estatística e ferramentas de programação. A transição entre os sistemas é, idealmente, escondida do usuário, provendo uma interface transparente para facilitar a consulta a dados, manipulação e análise [LP98]. Uma tabela central de metadados forma a ligação principal entre dados e metadados através do identificador do conjunto de dados. Esta tabela provê a metainformação essencial para o conjunto de dados. Metainformações de todos os níveis estão sendo armazenadas em banco de dados Oracle.

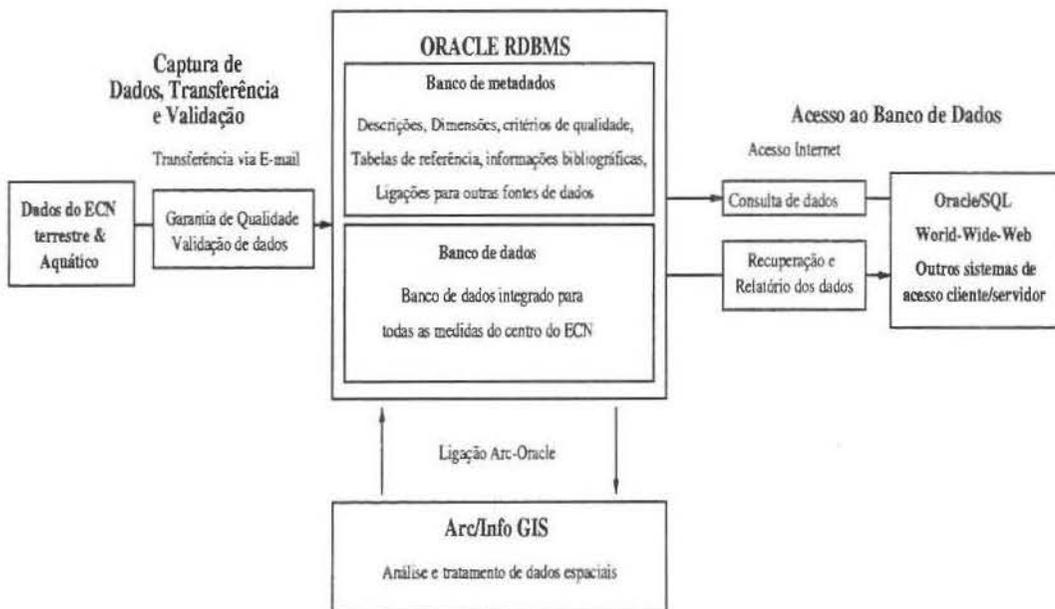


Figura 3.2: Gerenciamento de dados ECN

O banco de dados, SIG e software executam em servidores baseados em UNIX com *links* para a Internet para acesso remoto dos dados. O sistema permite recuperação e

armazenamento integrado para todos os dados ECN, através da incorporação de metainformações. O modelo escolhido para integração de dados com o padrão de qualidade de dados requerido foi o de um banco de dados ECN central com *links* de acesso remoto.

O sistema trabalha usando ASP com uma conexão ADO/ODBC para o banco de dados Oracle. A conexão ADO/ODBC é necessária uma vez que os desenvolvedores ainda não codificaram uma API ODBC capaz de permitir o acesso direto da interface aos dados. O ODBC é uma API (*Application Programming Interface*) para dados SQL. O ADO (*Active Data Objects*), por sua vez, é usado como a interface para ODBC. O ADO é uma interface de programação de aplicação [RLW00].

Os critérios de seleção informados nos formulários WWW são convertidos em SQL e enviados para o banco de dados. Desta forma, o banco de dados Oracle do ECN é consultado usando SQL. O dado retornado ao usuário é formatado em páginas HTML através do uso de VBScript em ASP (*Active Server Pages*) e exibido automaticamente como gráficos e planilhas via o *browser* do cliente. Os dados recuperados da consulta são armazenados como um conjunto de registros que é temporariamente armazenado em uma "variável de sessão" tal que o banco de dados não precisa ser re-consultado se o usuário escolher visualizar os mesmos dados em um formato diferente [RLW00].

Tipos de consulta e interface de exibição O ECN permite realizar consultas por campo específico e gráficas.

- **Consulta por campo específico**

A consulta por campo específico permite ao usuário construir suas próprias consultas ao banco de dados através da escolha dos parâmetros de filtragem desejados em caixas de seleção e da definição de intervalos de datas, para em tempo de execução gerar planilhas e gráficos dos resultados. Os usuários podem especificar também a exibição de metadados.

- **Consulta gráfica**

A consulta gráfica permite ao usuário clicar em *links* simbólico ou textual em um mapa para obter informações adicionais sobre vários sites diretamente.

Neste sistema, o usuário pode escolher visualizar os dados de forma gráfica ou textual. A forma gráfica disponibiliza um gráfico de barras dos dados selecionados. Já a forma textual é apresentada como uma lista onde cada item tem um *link* para informações mais detalhadas.

UDK O UDK [UDK] é um catálogo de dados ambientais que fornece dados coletados e armazenados por instituições austríacas. Esta ferramenta visa qualquer cidadão interessado em informações sobre dados ambientais e, especialmente, pessoas trabalhando nas unidades administrativas relacionadas ao ambiente e no campo de pesquisa ambiental. Em adição, o UDK é um exemplo de um padrão específico para a catalogação de dados ambientais ou geográficos.

Este sistema é classificado como um sistema com enfoque múltiplo, uma vez que desempenha o papel de um catálogo de dados ambientais. Em adição, ele permite o gerenciamento e recuperação de metadados ambientais.

Características específicas Ele foi desenvolvido para coleção e recuperação de metadados ambientais, com informação sobre *quem tem onde, qual ordem* de dados, em *qual formato*. Os registros de metadados são armazenados como objetos UDK. Os atributos dos objetos UDK estão relacionados com a identificação do objeto, e com o alcance técnico, espacial e temporal do dado a ser descrito. Todo objeto UDK está associado a um WWW-UDK.

A organização dos dados está baseada em um modelo de objetos da seguinte forma: objetos ambientais (isto é, rios, fábricas), conjunto de dados que descrevem os objetos ambientais, e a descrição dos conjuntos de dados (isto é, metadados).

Hiperlinks provêm a navegação entre os objetos. Além do mais, um *thesaurus* polihierárquico com termos ambientais presta ajuda na determinação de palavras-chave de pesquisa. Os termos de pesquisa do usuário são mapeados para (uma sequência de) consultas SQL, cujos resultados ou são usados como entrada para a geração dinâmica de páginas HTML ou são diretamente apresentados por *applets* Java.

Arquitetura O UDK utiliza a tecnologia WWW clássica (clientes WWW, HTTP, *scripts* CGI) e programas C com expressões SQL embutidas para acesso e integração de fontes de informação. Os *scripts* CGI são utilizados para acesso a banco de dados.

Tipos de consulta e interface de exibição O UDK permite os seguintes tipos de consulta: consulta por palavras-chave, consultas por campos específicos e consulta por hierarquia de palavras.

- **Consulta por palavra-chave**

A consulta por palavra-chave permite ao usuário entrar com um termo de busca que é por sua vez mapeado para vários atributos de várias relações. Nesta opção, o usuário pode definir se o termo informado é interpretado como parte de uma palavra ou como uma palavra completa.

- **Consulta por campo específico**

A consulta por campo específico permite que o usuário defina a busca de acordo com o campo de interesse. Adicionalmente, o UDK permite definir restrição espacial (via caixa de seleção) e/ou restrição temporal. A restrição espacial permite especificar um país, uma região ou uma área de interesse.

- **Consulta por hierarquia de palavras**

A consulta por hierarquia de palavras permite uma pesquisa por tema ambiental. Sendo assim, uma lista de temas organizados em ordem alfabética é apresentada ao usuário.

Os resultados das consultas submetidas ao UDK só podem ser visualizados no formato textual. Para todos os objetos UDK, a interface apresenta uma arvore de pesquisa que possibilita a visualização tanto do predecessor quanto de uma lista de sucessores do tema corrente. Desta forma, a interface permite ao usuário ter uma visão geral de sua posição corrente na arvore de pesquisa.

CDS O CDS [CDS] é um sistema de catálogo ambiental estabelecido com o objetivo de fornecer um sistema de informação ambiental para a Europa. O sistema de catálogo está baseado em duas ferramentas: o WinCDS (sistema para catalogar dados) e o WebCDS (*browser* de banco de dados baseado na Web). Ambas as aplicações estão conectadas a um *thesaurus* ambiental multilíngue que provê indexação, funções de busca e tradução entre idiomas de consulta. O WebCDS é uma ferramenta que visa permitir o acesso global aos dados via Internet, provendo recuperação/distribuição *on-line*.

Este sistema é classificado como um sistema com enfoque múltiplo, uma vez que funciona como um catálogo ambiental. Em adição, ele permite o gerenciamento e recuperação de metadados ambientais.

Arquitetura O CDS está baseado no Sistema Localizador de Informação Ambiental Global (GELOS). GELOS é um esquema de metainformação internacionalmente usado para descrição e localização de recursos de informação, sendo compatível com o protocolo de recuperação e pesquisa de informação Z39.50.

Os dados CDS estão armazenados em um banco de dados Oracle. Este sistema *on-line* está disponível para Sun, HP e servidor Linux, e as características de interatividade com o usuário são programadas em JAVA ou HTML simples. Segundo a *European Environmental Agency* [CDS], uma nova versão integrada de pesquisa em Java estará disponível em breve. Esta versão incluirá uma ferramenta para navegar através do banco de dados CDS e uma interface de pesquisa geográfica, dentre outras características.

Tipos de consulta e interface de exibição O sistema WebCDS provê os seguintes serviços de busca: consultas por palavra-chave, campo específico e por hierarquia de palavras.

- **Consulta por palavra-chave**

O usuário pode realizar uma consulta por palavra-chave informando um termo de pesquisa livre e submetendo-o à execução.

- **Consulta por campo específico**

Esta opção de consulta permite definir critérios de busca nos seguintes campos do CDS: título, primeiro e último nome de uma pessoa, e nome de uma organização, restrição espacial (país) e temporal, dentre outros.

- **Consulta por hierarquia de palavras**

A figura 3.3 mostra uma cópia de tela de consulta CDS por hierarquia de palavras. A consulta por hierarquia de palavras é processada a partir de uma lista de tópicos ambientais exibidos, sendo possível selecionar um elemento através de um clique. Cada um dos 40 tópicos mostra o número de fontes de dados relacionadas.

Os resultados das consultas do sistema CDS são apresentados no formato textual, exibindo uma lista de *links* com uma breve descrição do conjunto de dados recuperado no *frame* mais a direita da figura 3.3. A partir destes *links*, é possível obter informações mais detalhadas sobre o conjunto de dados, sendo que estas informações são disponibilizadas no *frame* Informação detalhada da figura 3.3.

GCMD O GCMD [GCM] é um programa da NASA que visa a descrição de conjuntos de dados relevantes às pesquisas em mudanças no clima, hidrosfera e oceanos, biosfera, geologia, geografia, e dimensões humanas das mudanças globais. No GCMD, os registros de metadado provêm informações sobre a natureza dos dados e onde estão armazenados. Para utilizá-la, o usuário deve ter um *browser* Java.

Este sistema é classificado como um sistema com enfoque múltiplo, uma vez que visa a descrição de todo um conjunto de dados relevantes às pesquisas sobre as mudanças ambientais. Em adição, ele permite o gerenciamento e recuperação de informações sobre a natureza dos dados e onde estão armazenados a partir do acesso a metadados.

Tipos de consulta e interface de exibição As formas de consulta atualmente disponíveis são: consultas textuais por palavras-chave, por campo específico e por hierarquia de palavras, e consultas gráficas.

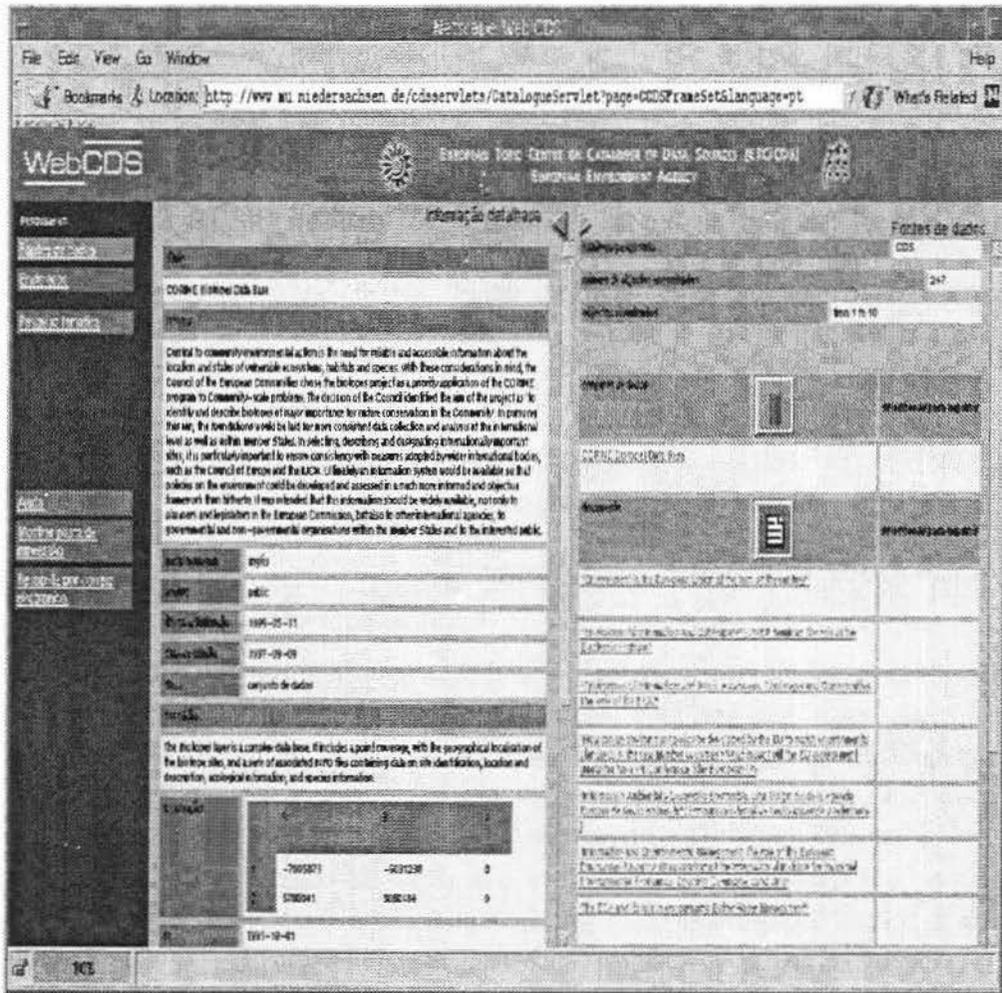


Figura 3.3: Exemplo de uma consulta no sistema WebCDS [CDS]

- **Consulta por palavra-chave**

A consulta por palavra-chave percorre o texto pesquisando o documento inteiro.

- **Consulta por campo específico**

As consultas por campo específico seguem a característica padrão definida para este tipo de consulta.

- **Consulta por hierarquia de palavras**

A consulta por hierarquia de palavras permite ao usuário navegar através de uma hierarquia de palavras-chave até chegar ao documento desejado.

Nesta modalidade de consulta, é possível realizar uma pesquisa textual por campo específico para restringir o universo de pesquisa. Um exemplo disto é quando

o usuário ao percorrer a hierarquia de palavras resolve refinar as informações do hierarquia corrente antes de prosseguir a pesquisa por hierarquia.

- **Consulta gráfica**

Um mapa *applet* permite selecionar um ponto ou uma região. Em adição, o sistema disponibiliza ainda as funções usuais de *zoom in* e *zoom out*.

A seleção dos atributos (regiões) de interesse do usuário é realizada através de cliques nas regiões do mapa ou nos nós gráficos das regiões. Exemplos de regiões são África, Ásia, Antártica, Atlântico, Pacífico, Europa, América do Sul, América Central. A especificação de uma área particular no mapa pode ser realizada também através da caixa de seleção de áreas.

O GCMD possui dois formatos de exibição dos resultados de consultas: textual e gráfica. No formato textual, a página de resultado exibe uma lista dos resultados, onde cada item da lista é um *link* para documentos, permitindo obter uma informação mais detalhada.

Na interface de consulta gráfica, os resultados são apresentados sob a forma de pontos em um mapa, indicando a localização dos conjuntos de dados recuperados. Os gráficos de barra são atualizados dinamicamente, quando alguma seleção é definida. A figura 3.1 mostra uma cópia de tela de consulta gráfica GCMD, bem como a visualização gráfica do resultado da consulta.

GILS O GILS [GIL], Serviço de Localização de Informação Governamental, é um ferramenta eletrônica pública que faz parte da Infra-estrutura de Informação Nacional dos Estados Unidos. O GILS visa documentar dados públicos vindos das agências governamentais; descrever as informações disponíveis; e ajudar na consulta e recuperação destas informações.

Este sistema é classificado como um sistema com enfoque múltiplo, uma vez que visa documentar diversos dados públicos vindos das agências governamentais. Em adição, ele permite o gerenciamento e recuperação de documentos e arquivos oficiais considerados relevantes a partir do acesso a metadados.

Características específicas A descrição dos dados provida pelo GILS informa aos usuários que dados estão disponíveis, onde eles estão localizados, e como eles podem ser acessados. Este sistema permite seus usuários buscarem dados em bancos de dados públicos e acessar documentos e arquivos oficiais, além de conduzir a informação governamental disponível.

	Tópico de interesse		Gerenciamento e Recuperação	
	Sistema com enfoque único	Sistema com enfoque múltiplo	Metadados	Dados e Metadados
<i>Herbarium Specimen Browser</i>	✓			✓
<i>Species Analyst</i>	✓			✓
<i>Environmental Australia On Line</i>		✓		✓
<i>Species 2000</i>	✓			✓
<i>Malaysia Biodiversity Online</i>	✓			✓
EOSDIS		✓		✓
FGDC		✓		✓
ECN		✓	✓	
UDK		✓	✓	
CDS		✓	✓	
GCMD		✓	✓	
GILS		✓	✓	

Tabela 3.1: Classificação dos sistemas de informação ambiental quanto à abrangência dos objetivos dos módulos disponíveis e ao conteúdo do banco de dados

Tipos de consulta e interface de exibição O GILS provê os seguintes serviços de pesquisa: consulta por palavra-chave e por campos específicos. Ambos os serviços de busca, por campo específico ou por palavra-chave, seguem a característica padrão definida para estes tipos de consulta e permitem, em adição, que a pesquisa seja baseada na disjunção ou conjunção de palavras ou frases.

O sistema GILS permite visualizar os resultados das consultas apenas no formato textual. Sendo assim, as consultas retornam *links* para documentos considerados relevantes.

A interface permite ao usuário refinar a busca através da informação de um texto de pesquisa adicional. O refinamento é uma pesquisa por palavra-chave sobre o conjunto de documentos recuperado previamente.

3.4 Resumo

Este capítulo analisou diversos sistemas de informação ambiental a partir do uso (via Internet) destes e da análise das documentações correlatas, visando principalmente identificar as consultas comumente utilizadas e os formatos de visualização disponíveis.

A tabela 3.1 apresenta a classificação dos sistemas analisados de acordo com o objetivo ou abrangência dos objetivos dos módulos disponíveis (sistema com enfoque único ou com enfoque múltiplo) e o conteúdo do banco de dados que cada sistema gerencia e

	Tipos de consulta				
	Palavras-chave	Hierarquia de palavras	Campos específicos	Índice alfabético	Gráfica
<i>Herbarium Specimen Browser</i>		✓	✓		✓
<i>Species Analyst</i>			✓		
<i>Environmental Australia On Line</i>	✓		✓	✓	✓
<i>Species 2000</i>			✓		
<i>Malaysia Biodiversity Online</i>	✓		✓		✓
EOSDIS	✓		✓		✓
FGDC	✓	✓	✓		✓
ECN			✓		✓
UDK	✓	✓	✓		
CDS	✓	✓	✓		
GCMD	✓	✓	✓		✓
GILS	✓		✓		

Tabela 3.2: Tipos de consulta disponíveis nos sistemas de informação ambiental analisados

	Formato de exibição	
	Textual	Gráfico
<i>Herbarium Specimen Browser</i>	✓	✓
<i>Species Analyst</i>	✓	✓
<i>Environmental Australia On Line</i>	✓	✓
<i>Species 2000</i>	✓	
<i>Malaysia Biodiversity Online</i>	✓	
EOSDIS	✓	
FGDC	✓	
ECN	✓	✓
UDK	✓	
CDS	✓	
GCMD	✓	✓
GILS	✓	

Tabela 3.3: Formatos de exibição dos resultados disponíveis

UNICAMP
BIBLIOTECA CENTRAL
SEÇÃO CIRCULANTE

recupera. A tabela 3.2 resume os tipos de consulta disponível para cada sistema. A tabela 3.3 apresenta, por sua vez, um resumo dos formatos de visualização dos resultados das consultas.

Em suma, a análise dos sistemas permitiu concluir que não existe um modelo padrão para todo sistema de informação ambiental, mas sim que existem características em comum entre estes sistemas e que as soluções adotadas por alguns sistemas podem ser mecanismos recomendáveis para outros sistemas. A identificação de necessidades similares e de funcionalidades já implementadas pode contribuir para melhoramento dos sistemas já existentes e para a agilização do processo de desenvolvimento de novos sistemas. O capítulo 4 se baseia nestas observações para descrever o sistema de informação ambiental do BIOTA/FAPESP enquanto que o capítulo 6 mostra como aplicá-las a um protótipo desenvolvido para o BIOTA.

Capítulo 4

O Sistema de Informação de Biodiversidade do Estado de São Paulo

Este capítulo descreve de forma simplificada o sistema de informação ambiental para o BIOTA/FAPESP, denominado Sistema de Informação de Biodiversidade do Estado de São Paulo.

O capítulo está organizado da seguinte forma. A seção 4.1 apresenta uma breve descrição do programa BIOTA/FAPESP. A seção 4.2 descreve o sistema de informação para o programa. A seção 4.3 fornece uma visão geral do sistema de informação atualmente disponível na Web. A seção 4.4 realiza uma comparação entre os sistemas apresentados no capítulo 3 e o Sistema de Informação para o BIOTA/FAPESP. A seção 4.5 resume o conteúdo deste capítulo.

4.1 O Programa BIOTA/FAPESP

O BIOTA/FAPESP [BIO] é um programa voltado a pesquisas em conservação sustentável de biodiversidade do Estado de São Paulo lançado oficialmente em 1999 pela FAPESP, com o objetivo de ampliar o conhecimento sobre a biodiversidade do Estado, incluindo seus aspectos estruturais e funcionais. A abrangência deste programa é única em termos do volume, variedade e cobertura de dados envolvidos, compreendendo fauna e flora, tanto terrestre quanto aquática. Não há no mundo um sistema que considere uma quantidade tão ampla e tão diversificada de seres vivos (atualmente, 166 grupos taxonômicos, desde bactérias e fungos até vegetais superiores e mamíferos).

Os projetos desenvolvidos dentro deste programa têm por objetivo mapear e analisar os inúmeros aspectos da biodiversidade do Estado de São Paulo e divulgar este conhecimento

desenvolvendo bases metodológicas e padrões de referência para estudos de impacto ambiental, produzindo estimativas de perda de biodiversidade em diferentes escalas espaciais e temporais, e identificando áreas e componentes prioritários para conservação.

A maioria dos projetos do BIOTA/FAPESP é voltado a pesquisas em biologia englobando aproximadamente 200 pesquisadores organizados em cerca de 30 grupos de pesquisa em biodiversidade do Estado de São Paulo, responsáveis por realizar, através de coletas, o levantamento da biodiversidade em diversas regiões do Estado. Um dos projetos do programa - o SINBIOTASP - diz respeito ao desenvolvimento do sistema de informação ambiental específico para o BIOTA/FAPESP. Diante do fato que o processamento de consultas do SINBIOTASP irá combinar o uso de SGBD e SIG, processando múltiplos tipos de dados (espaciais e não espaciais), as considerações quanto heterogeneidade e interoperabilidade realizadas na seção 2.6 são aplicadas também ao SINBIOTASP.

Esta dissertação está ligada ao processamento de consultas no projeto SINBIOTASP. Este projeto é coordenado pelo Prof. C. A. Joly do Instituto de Biologia (Unicamp), e envolve três grupos: (1) Instituto de Computação; (2) Instituto de Geociências (Unicamp), Faculdade de Engenharia Agrícola (Unicamp) e Instituto Florestal (USP) responsáveis pelo projeto, especificação e desenvolvimento da base cartográfica; e (3) Fundação André Tosello (Base de Dados Tropical) responsável pela integração dos esforços dos demais grupos, manutenção e disponibilização do sistema aos usuários.

As diretrizes definidas para o sistema de informação são [BIO]:

- Informatização das coleções científicas do estado de São Paulo com possibilidade de acesso via Internet;
- Adoção de padrões para a integração dos dados;
- Disponibilização de informações de forma fácil e eficiente para apoiar os pesquisadores;
- Acesso a mapas para o planejamento e realização de coletas;
- Criação de um banco de metadados contendo informações básicas sobre os dados armazenados.

Os fornecedores de dados são principalmente biólogos e ecologistas, que necessitam de informações sobre a biodiversidade do estado de São Paulo. A partir da disponibilização dos dados pela Internet, os pesquisadores serão capazes de buscar informações de diferentes temas (tais como distribuição espacial de espécies, topografia, clima e vegetação), usando distintas fontes de informação (banco de dados textuais, mapas digitalizados e imagens de satélites), necessitando para isto apenas um *browser* Web.

As principais contribuições do projeto SINBIOTASP serão:

- Disponibilização de uma base cartográfica para a integração e consolidação dos dados disponíveis, permitindo correlações espaciais entre as diversas pesquisas em biodiversidade do estado de São Paulo. Em [Fag99], é ressaltado a importância em prover ao BIOTA/FAPESP a integração com um sistema de informação geográfica.
- Desenvolvimento de um conjunto de ferramentas que permita aos pesquisadores de biodiversidade a verificação dos dados e projetos existentes, provendo acesso às informações pela Internet.

4.2 Visão Geral do Sistema de Informação

O sistema de informação está centrado na noção de *coleta*. Uma coleta é, neste contexto, um registro de algum tipo de observação sobre seres vivos. A esta observação são associados o local em que estes organismos foram observados, assim como o método de observação aplicado. Registros de coleta podem ser encontrados sob diversas formas: em catálogo de museus; em textos científicos; em relatos de expedições científicas e em registros de banco de dados. Em grande parte destas observações, a especificação do local de observação é imprecisa (por exemplo, "Cataratas do Iguaçu", "Floresta da Tijuca"). Ao contrário, no sistema do BIOTA/FAPESP, as coletas serão sempre associadas a uma localização precisa, determinada a partir do uso de GPS (*Global Positioning System*) [Fag99].

No sistema, toda coleta é caracterizada por dois tipos de conjuntos de dados: conjunto básico (fixo e obrigatório) e conjunto complementar (variável). O conjunto básico caracteriza a coleta segundo os quesitos: *quem* realizou (autor), *onde* (localização) e *quando* (tempo). O conjunto complementar descreve *o que* foi observado (lista de seres vivos enumerados segundo uma taxonomia básica). Além disso, o sistema deve gerenciar dados referentes aos mapas armazenados e aos metadados (associados tanto às coletas, quanto aos mapas). Para isto, o sistema é baseado em uma arquitetura composta por três blocos principais: *interface com o usuário*, *gerenciadores de dados e metadados*, e *repositórios de dados* [Fag99]. A figura 4.1 mostra de forma esquematizada a arquitetura do sistema. Esta dissertação refere-se ao bloco pontilhado da figura 4.1, denominado Sistema de Processamento de Consultas, que é capaz de estabelecer comunicação com os gerenciadores de dados e metadados e prover uma interface com o usuário para a elaboração de consultas e visualização dos resultados.

4.2.1 Interface com o usuário

Um dos objetivos da dissertação é desenvolver a interface de consultas com o usuário. Basicamente o módulo de interface fornece duas funcionalidades:

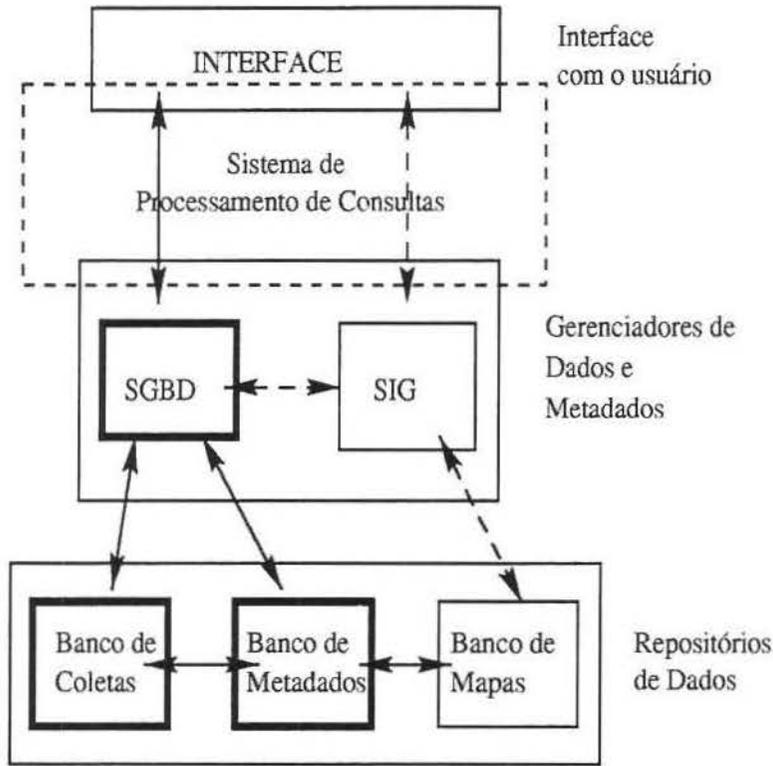


Figura 4.1: Arquitetura básica do SINBIOTASP.

- Entrada de dados. Disponibiliza funções que permitem a inserção e atualização de conjuntos de dados para serem armazenados no sistema.
- Consulta e visualização dos dados. Permite a realização de consultas aos dados armazenados no sistema, que serão processadas pelos gerenciadores de dados (SGBD e/ou SIG). Estão previstas consultas tanto textuais (realizadas a partir de formulários pré-definidos), como também consultas interativas sobre mapas.

4.2.2 Gerenciamento de dados e metadados

Através do SGBD e SIG, o sistema oferece funções para o gerenciamento dos dados (coletas, metadados e mapas) armazenados nos repositórios. São exemplos de funcionalidades disponibilizadas: funções de inserção/modificação de dados, validação de formato e conversão [Fag99].

O sistema de busca e navegação analisado por esta dissertação (ver capítulo 5) permite realizar consultas tanto textuais, a partir de formulários pré-definidos, como também consultas interativas sobre mapas [SM98, KO96]. Desta forma, o sistema disponibiliza dois tipos básicos de formas de acesso: as que irão acessar os repositórios usando funções já

especificadas e pré-computadas (semelhantes a visões) e aqueles que permitem a navegação nos repositórios. Esta segunda forma provê acesso direto aos bancos de dados de coletas ou metadados (via SGBD) e mapas (via SIG). Supõe-se, no entanto, que em geral o acesso será processado inicialmente a partir do banco de metadados, que permitirá realizar uma primeira filtragem dos dados. O resultado desta filtragem será então utilizado para fazer consultas combinadas usando o Sistema Gerenciador de Banco de Dados (SGBD) e o Sistema de Informação Geográfica (SIG).

As consultas previstas dentro do SINBIOTASP podem assim ser classificadas em dois tipos:

Navegação Este tipo de consulta corresponde a fazer um folheio (*browsing*) dos dados, exigindo estrutura de *hiperlinks*. Como exemplo, poderíamos mencionar: navegação entre mapas, entre coletas, etc. Para isto, é necessário desenvolver em paralelo aos bancos de dados toda uma estrutura de ligação, possivelmente via HTML ou XML. O problema de ligações deste tipo é que são estáticas e só permitem navegação entre nós pré-estabelecidos. Este tipo de trabalho representa, por si só, tópico de pesquisa [GM98] e foge ao escopo da dissertação. A dissertação se concentra em desenvolver ferramentas para facilitar consultas através de predicados específicos, descritos a seguir.

Busca através de predicados específicos As consultas deste tipo têm um predicado específico capaz de selecionar dados textuais e informações contidas em mapas (pontos, polígonos ou regiões). No caso do formato convencional (textual), o predicado varia apenas com relação aos parâmetros de seleção em função de uma lista de valores possíveis. No caso de consulta sobre mapas, o sistema deve consultar o banco de dados em função de regiões pré-definidas (Bacia do Mogi-Mirim, por exemplo), polígonos definidos pelo usuário, ou pontos definidos em função dos dados georeferenciados.

- "Quais as coletas realizadas na data X?" (consulta sobre predicado textual - data, retornando texto - coleta)
- "Qual o mapa de vegetação da bacia do Mogi-Mirim?" (consulta envolvendo região e retornando mapa, mas não envolve análise espacial)
- "Exibir a distribuição do grupo taxonômico *artrópoda* dentro da Bacia do Mogi-Mirim." (consulta envolvendo predicado textual - grupo taxonômico e análise espacial - distribuição de espécies, retornando mapa com a distribuição do grupo taxonômico)

4.2.3 Repositórios de Dados

Os repositórios armazenam dados relativos aos levantamentos e trabalhos de biodiversidade de diferentes tipos (levantamentos em campo, referências bibliográficas, artigos científicos), mapas e metadados. A utilização destes repositórios permitirá que as coletas realizadas pelos pesquisadores sejam associadas à região (coordenadas geográficas) em que o levantamento foi realizado. Desta forma, será possível realizar diferentes tipos de cruzamentos e análises de dados que envolvam tanto os dados textuais sobre as espécies observadas quanto os fenômenos geográficos na região em que os dados foram levantados. Para isto, os repositórios de dados incluem um Sistema Gerenciador de Banco de Dados (SGBD) que armazena e gerencia os conjuntos de dados de biodiversidade resultantes das coletas (banco de coletas) e os metadados (banco de metadados) que descrevem estes dados de acordo com seu conteúdo, contexto, adequação ao uso, entre outros, e um Sistema de Informação Geográfica (SIG) responsável pelo armazenamento e manipulação de dados cartográficos (banco de mapas) relacionados aos conjuntos de dados coletados.

Banco de Coletas O projeto lógico do banco de coletas [Fag99] foi resultado da análise dos perfis dos usuários envolvidos no projeto, dos mecanismos de trabalho empregados pelos grupos de pesquisa em biodiversidade, e dos resultados obtidos através de coletas já realizadas. Um dos primeiros problemas enfrentado na modelagem deste banco foi a diversidade de usuários e fontes de dados, além da grande gama de metodologias de levantamento de dados sobre as espécies (flora e fauna).

De um modo geral, o trabalho de coleta envolve a ida dos grupos de pesquisa a campo, e o emprego de métodos específicos que permitam identificar elementos da flora e fauna encontrados em uma determinada área geográfica. Estes grupos de pesquisa devem utilizar um GPS para georeferenciar seus dados. Isto visa padronizar a forma de associação geográfica de espécies aos mapas do banco de mapas. Com isto, espera-se permitir que os usuários "cruzem" informações auxiliando, principalmente, a formulação de programas de impacto ambiental nas áreas pesquisadas. A modelagem do banco de coletas é apresentada em maiores detalhes por Fagundes [Fag99].

O banco de coletas armazena todas as informações relacionadas a taxonomia e as coletas das espécies, tais como: domínio, reino, divisão, sub-divisão, filo, sub-filo, superclasse, classe, família, gênero, espécie, autor, bacia, ecossistema, a coleta em si, a relação de quais espécies foram encontradas em um determinada coleta, dentre outras.

Banco de Mapas As coletas são georeferenciadas e, no sistema, associadas ao chamado Banco de Mapas. Este banco é responsável pelo armazenamento e manipulação dos dados cartográficos do Estado.

Deverão ser produzidos cerca de 410 mapas, cobrindo todo o Estado de São Paulo, em uma escala de 1:50.000. Os mapas deverão apresentar os seguintes temas: mapas base (limite de municípios, curvas de nível e hidrografia), mapas pedológicos, mapas de declividade, mapas de risco de erosão e mapas geomorfológicos [BIO].

Parte do conjunto de metadados definido para o sistema é responsável pela descrição dos mapas armazenados, apresentando informações sobre o processo utilizado para sua criação, a categoria coberta pelo mapa, entre outros.

As consultas que retornam visualização cartográfica podem ter como fonte apenas o banco de mapas, como também interrelacionamentos entre os bancos de metadados, de coletas e de mapas.

Banco de Metadados O banco de metadados contém registros que descrevem as coletas e mapas mantidos pelo sistema. O sistema deve ser integrado a outros sistemas de biodiversidade existentes. O principal meio de integração previsto é através do uso de metadados.

O padrão de metadados proposto para o BIOTA possui as seguintes classes [Fag99]: metadados descritivos, metadados de referência espacial (localização), metadados de referência temporal (data da coleta), metadados de qualidade, metadados de contexto científico, metadados de método de pesquisa e metadados de informações adicionais. A dissertação de Fagundes [Fag99] contém maiores detalhes sobre cada uma das classes que compõem o padrão de metadados para o BIOTA/FAPESP, bem como a modelagem conceitual das mesmas.

Usando a classificação de metadados definida por Ullman *et al.* [URB97] e apresentada na seção 2.6, no caso específico do SINBIOTASP, esta dissertação trabalha com 3 tipos de metadados: semântico, de confiabilidade e ancestral. Os metadados semânticos (descritivos, informações adicionais, contexto científico) têm como objetivo completar e acrescentar um maior significado aos dados armazenados no banco de coletas e no banco de mapas. Os metadados de confiabilidade (qualidade), por sua vez, visam prover informações sobre a qualidade dos dados aos usuários do sistema. Já os metadados ancestrais (referência espacial, referência temporal, método de pesquisa, descritivos) fornecem referências sobre a fonte dos dados, tais como: pesquisador, quando, onde, como, dentre outras.

4.3 Estado atual do SINBIOTASP na Web

O sistema de informação de biodiversidade do Estado de São Paulo atualmente disponível na Web (novembro 2000) visa permitir a extração de informações textuais e cartográficas referentes às coletas. Hoje, o sistema permite realizar somente seleção textual a partir

de um campo específico ou palavra-chave, enquanto os formatos disponíveis de visualização dos dados podem ser tanto textual quanto gráfico. As consultas gráficas não estão disponíveis.

Este sistema de informação provê funcionalidades referentes à entrada e consulta de dados. As ferramentas de consulta são capazes de combinar os dados contidos nos bancos de coletas, de metadados e de mapas, e prover uma exibição textual e gráfica dos resultados. A exibição gráfica de resultados sobre mapas é disponibilizada pelo módulo denominado Atlas Biota/SP versão 2.0. Este módulo utiliza algumas funções desenvolvidas como parte desta dissertação, dentro do módulo *Species Mapper* descrito no capítulo 6.

Para o acesso aos dados, foi definida a seguinte hierarquia de usuários: administrador/curador (gerencia o conjunto básico de dados, gerencia os mapas armazenados, realiza a curadoria dos mapas e da classificação taxonômica dos conjuntos de espécies observados, entre outros), usuário pesquisador (insere e atualiza os dados de coletas) e usuário geral (consulta dados e metadados armazenados nos bancos de coletas, de metadados e de mapas) [Fag99]. O acesso ao sistema está disponível na Web a partir do site do BIOTA (<http://www.biotasp.org.br>).

Esta seção está organizada da seguinte forma. A subseção 4.3.1 descreve a arquitetura Web adotada atualmente pelo SINBIOTASP. A subseção 4.3.2 apresenta uma especificação funcional do sistema. A subseção 4.3.3 apresenta um conjunto de operações que representam as funcionalidades disponibilizadas.

4.3.1 Arquitetura Web atual

A arquitetura atual do sistema de informação possui dois tipos de processamento: convencional e espacial.

Processamento convencional O processamento convencional se aplica quando uma consulta possui critérios textuais e o resultado textual da consulta é gerado sem a necessidade da utilização de operações e operadores espaciais. Este processamento interrelaciona os bancos de coletas e de metadados.

Este sistema foi desenvolvido utilizando CGI escrito em PERL. Os papéis do cliente e do servidor estão definidos da seguinte forma:

- Ao cliente cabe capturar os critérios de seleção através de uma interface desenvolvida em HTML e JavaScript, solicitar a execução do *script* CGI ao servidor e exibir as páginas geradas dinamicamente através da consulta (via Oracle) aos bancos de coletas e de metadados.

- Ao servidor cabe a execução de um programa escrito em PERL contendo a consulta sobre os bancos de dados, cujo resultado é transmitido ao cliente. O *script* CGI gera como resultado uma página HTML para ser exibida no navegador.

A execução do *script* CGI exige que a cada solicitação do usuário o servidor Web crie um novo processo. Cada processo produz uma nova conexão com o banco de dados utilizado e o servidor Web tem que esperar até que os resultados lhe sejam enviados.

Processamento espacial O processamento espacial se aplica quando uma consulta possui critérios textuais e o resultado gráfico (mapa) é gerado a partir da utilização de operações e operadores espaciais. As operações são executadas sobre os bancos de coletas, de metadados e de mapas.

O módulo Atlas Biota/SP versão 2.0 foi desenvolvido utilizando CGI escrito em PERL, Delphi 5.0 e Map Objects da ESRI. Os bancos de dados de coletas e de metadados são pesquisados através da execução de cláusulas SQL. O acesso ao banco de mapas, por sua vez, é realizado pelo Map Objects da ESRI. Os componentes que compõem esta arquitetura, bem como suas funcionalidades são descritos a seguir. A interface do usuário é com um módulo denominado *Cliente HTML*, que passa as solicitações ao módulo *Cliente PERL*. Este, por sua vez, aciona o módulo *Map Client* que realiza uma consulta espacial no SIG e devolve seu resultado ao *Cliente PERL*. A conexão entre o *Cliente PERL* e o *Map Client* é gerida por dois outros módulos, denominados *Map Server* e *Map Statistic*.

- *Cliente HTML* é responsável por capturar o predicado de uma consulta através de uma interface desenvolvida em HTML e JavaScript, solicitar a sua execução ao *Cliente PERL* e exibir a resposta como uma página gerada dinamicamente.
- *Cliente PERL* é um *script* CGI escrito em PERL responsável por solicitar ao *Map Server* a geração de um mapa contendo o resultado da consulta, receber uma imagem JPEG como resposta e gerar como resultado uma página HTML para ser exibida no navegador.
- *Map Server* é um *script* CGI escrito em PERL responsável por estabelecer uma comunicação entre o *Cliente PERL* e o *Map Client* (mediante solicitação do *Cliente PERL*), solicitar o identificador de um *Map Client* disponível para uso e sinalizar no arquivo de *Map Client* existentes quando um determinado *Map Client* foi instanciado ou liberado. Além disso, este componente é responsável por enviar a imagem JPEG gerada pelo *Map Client* para o *Cliente PERL*.
- *Map Statistic* é um *script* CGI escrito em PERL responsável por acessar um arquivo que contém a identificação de todos os *Map Client* existentes, selecionar uma

instância do *Map Client* para a execução da consulta solicitada pelo *Map Server*. Hoje, a escolha do *Map Client* é feita percorrendo sequencialmente o arquivo que contém os *Map Client* existentes com o intuito de encontrar alguma instância disponível.

- *Map Client* é responsável por processar a consulta espacial solicitada e retornar uma imagem JPEG como resultado da consulta. O *Map Client* foi desenvolvido em Delphi 5.0 utilizando o componente Map Objects da ESRI para pesquisar o banco de mapas e o padrão SQL para acessar as informações necessárias armazenadas nos banco de coletas e de metadados. Hoje, existem diversas instâncias do *Map Client* em execução. A disponibilidade de diversas instâncias permite o processamento de diversas consultas espaciais simultaneamente.

Considere a consulta "Exibir a distribuição das coletas onde foram encontradas *Acacia glomerosa*, *Dalbergia miscolobium* e *Salvia nervosa*". Inicialmente, o cliente (*Cliente HTML*) solicita a execução do processamento espacial ao *Cliente PERL*. Depois, o servidor (através do *Cliente PERL*) solicita a instanciação de um *Map Client* para o processamento da consulta espacial. O *Map Statistic* seleciona um *Map Client* disponível, que é instanciado pelo *Map Server* e executado, retornando uma imagem JPEG para o *Map Server*. O *Map Server* envia a imagem JPEG obtida como resultado para o *Cliente PERL* para a geração de uma página HTML como resultado. Por último, o cliente (*Cliente HTML*) exibe a página gerada dinamicamente através da consulta aos bancos de dados.

Parte do trabalho de implementação desta dissertação envolveu o desenvolvimento do componente *Map Client*, responsável pelo processamento da consulta espacial para até 16 espécies diferentes. A visualização de no máximo 16 espécies diferentes simultaneamente é resultado da combinação de 4 cores distintas das utilizadas para os temas existentes (branco, azul, verde e vermelho) e 4 formatos diferentes de apresentação de um ponto (círculo, quadrado, triângulo e cruz). Na consulta-exemplo "Exibir a distribuição das coletas onde foram encontradas *Acacia glomerosa*, *Dalbergia miscolobium* e *Salvia nervosa*", o *Map Client* é o responsável por permitir a visualização da distribuição espacial de cada espécie em um mesmo mapa, estabelecendo uma correlação espacial entre elas.

4.3.2 Especificação funcional

A especificação funcional do SINBIOTASP pode ser dividida em dois módulos, a saber: módulo de atualização e módulo de consulta. Estes módulos são responsáveis por fornecer funcionalidades e ferramentas que permitem aos diferentes tipos de usuário gerenciar todos os dados mantidos pelo sistema. A seguir são apresentadas as descrições de cada um destes módulos.

Módulo de Atualização O módulo de atualização permite a inserção, eliminação e alteração de todos os dados e metadados mantidos pelo sistema. A inserção de dados e metadados é tratada conjuntamente, para garantir sua associação natural, já que grande parte dos metadados precisa ser fornecida explicitamente pelo usuário pesquisador, não podendo ser derivada a partir dos dados de coleta [Fag99]. Atualmente, a alteração de dados é considerada como uma eliminação seguida de uma inserção por questões de integridade dos dados.

Módulo de Consulta O módulo de consulta do SINBIOTASP é responsável por fornecer funcionalidades e ferramentas que permitam aos usuários gerenciar e analisar todos os dados mantidos pelo sistema. Os seguintes mecanismos de consulta são disponibilizados, cada um com uma interface de consulta específica:

- Consultas exclusivas ao banco de coletas. Este tipo de processamento não necessita acessar o conjunto de dados contidos no banco de metadados nem no banco de mapas. Desta forma, torna-se necessário apenas acessar diretamente o banco de coletas. A figura 4.2 ilustra a interface que permite especificar uma consulta a ser processada exclusivamente sobre o banco de coletas. Por exemplo, "Quais os gêneros cadastrados no sistema cujos nomes começam com a letra A?".

Entre com o nome da espécie:

Gênero:

Espécie:

Para obter uma lista dos gêneros cadastrados no sistema, escolha a letra inicial e clique no botão enviar, abaixo.

Primeira letra do gênero:

Figura 4.2: Interface de seleção textual por campo específico no SINBIOTASP

- Consultas feitas aos bancos de coletas e de metadados. Para possibilitar a visualização do conjunto de informações desejado, a consulta é executada interrelacionando os bancos de coletas e de metadados. As figuras 6.5 e 6.6 seguem o mesmo formato de visualização do SINBIOTASP e exemplificam resultados que têm informações dos bancos de coletas e de metadados. Por exemplo, "Quais as coletas onde foram encontradas *Acacia polyphylla*?". Esta consulta-exemplo envolve metadados, pois alguns dados exibidos na figura 6.5 são provenientes do banco de metadados,

são eles: a data da coleta, o método de coleta, grupo taxonômico e localização (município).

- Consultas feitas aos bancos de coletas, metadados e mapas. Por exemplo, a consulta "Exibir a distribuição das coletas onde foram encontradas *Acacia glomerosa*, *polyphylla* e *Acacia paniculata*" cujo processamento é dividido nas seguintes etapas:
 1. Primeiramente, é realizada uma consulta interrelacionando as informações armazenadas no banco de coletas e metadados com o intuito de obter a localização (banco de metadados) da espécie desejada (banco de coletas);
 2. A segunda etapa recupera o mapa sobre o qual será exibida a distribuição da espécie desejada. O banco de metadados permite determinar quais os mapas do banco de mapas que devem ser recuperados, bem como as características de cada mapa. Os seguintes temas podem ser escolhidos para visualização: divisa estadual de São Paulo, bacia do Mogi-Mirim, uso do solo e drenagem;
 3. A última etapa diz respeito à visualização dos pontos que representam a localização da espécie desejada através do interrelacionamento dos resultados das etapas anteriores. A distribuição de cada espécie é representada por um par (formato, cor). Conforme mencionado anteriormente, há possibilidade de representar até 16 espécies simultaneamente sem repetir a combinação entre formato e cor.

4.3.3 Exemplos de operações

Esta subseção apresenta exemplos de operações, a partir das quais são realizadas inserção de coletas e consultas aos bancos de coletas, de metadados e de mapas. Mais detalhes são mostrados no capítulo 6.

Suponha um usuário pesquisador que pretende realizar uma série de operações. Inicialmente, ele deseja inserir uma nova coleta, com dados de organismos pertencentes ao grupo taxonômico "*Angiospermae*". Para isto, ele deve acessar a página para inserção de coletas do SINBIOTASP a partir de qualquer navegador Web, validar sua senha e entrar com os dados. A figura 4.3 mostra a página de inserção de coleta, onde os campos em negrito (por exemplo, Autor, Data) representam os atributos obrigatórios. Uma vez preenchidos todos os campos necessários, o usuário solicita a inserção (botão Enviar) e o sistema valida os dados, enviando mensagem de erro ou de sucesso, conforme o caso.

Posteriormente, o usuário pode querer consultar coletas em geral através das interfaces de consulta disponíveis. Inicialmente, ele deseja visualizar todas as coletas realizadas pelo método de "pesca". A figura 4.4 mostra a página onde o usuário pode selecionar um campo específico para o processamento da consulta, no caso o campo "METODO". Vale

BIOTA/FAPESP - Ficha Padrão Para Coleta/Registro

Usuário: Ricardo Scachetti Pereira (ricardo)
 Projeto: Sistema de Informação Ambiental
 Autor da Coleta: Ricardo Scachetti Pereira
 Referência Bibliográfica da Coleta:

Data da Coleta: Início: 10/01/2000 formato: dd/mm/aaaa (todos numéricos)
 Fim: 12/02/2000 formato: dd/mm/aaaa (todos numéricos)

Município: Campinas
 Localidade:
 Unidade de conservação:
 Ambiente: Terrestre

Bacia Hidrográfica: Piracicaba/Capivari/Jundiaí
 Coordenadas: Latitude: 23S 10' 56" UTM X:
 Longitude: 43W 50' 11" UTM Y:
 Zona:
 Precisão do GPS: m

Sistema
 Floresta Ombrófila Densa
 Floresta Ombrófila Mista
 Floresta Estacional Semidecidual
 Floresta Estacional Decidual
 Cerrado "alto serrão" (savana)
 Área de influência floresta marinha, Marinha (Atolado)
 Área de influência floresta marinha, Marinha (Estuário)
 Litoral consolidado
 Litoral não consolidado

Habitat/Localização Geral
 Planície aluvial costeira
 Formação montana
 Formação sub-montana
 Formação de altitude

Microhabitat/Localização Particular
 Ambiente de borda Área grossa
 Escarpilhada Cascalho
 Sombria Árida
 Área de influência floresta marinha, Marinha (Atolado)
 Área de influência floresta marinha, Marinha (Estuário)
 Litoral consolidado
 Litoral não consolidado

Método de Coleta: Contínuo
 Descrição do método:
 Escala da Coleta: 1m
 Observações Finais:
 Palavras Chave:

Lista de Espécies
 Lista de espécies 1:
 Lista de espécies 2:
 Lista de espécies 3:
 Lista de espécies 4:
 Lista de espécies 5:

Figura 4.3: Inserção de coleta (Ficha Padrão de coleta)

observar, também, que esta interface permite a execução de uma consulta por palavra-chave, o que implica em pesquisar um conjunto de campos (LOGIN, METODO, AUTOR, PROJETO, PALAVRAS_CHAVE, LOCALIDADE, MUNICIPIO, USUARIO, BIBLIOGRAFIA, NOME_COMPLETO, UNIDADE, BACIA e IDENTIFICADOR) em busca do termo especificado.

Suponha, agora, que o usuário queira obter um mapa com a distribuição das coletas. O resultado desta consulta é um conjunto de mapas, parcialmente mostrado na figura 4.5. Esta figura corresponde a uma versão atualmente fora de funcionamento e que foi substituída pelos módulos implementados nesta dissertação.

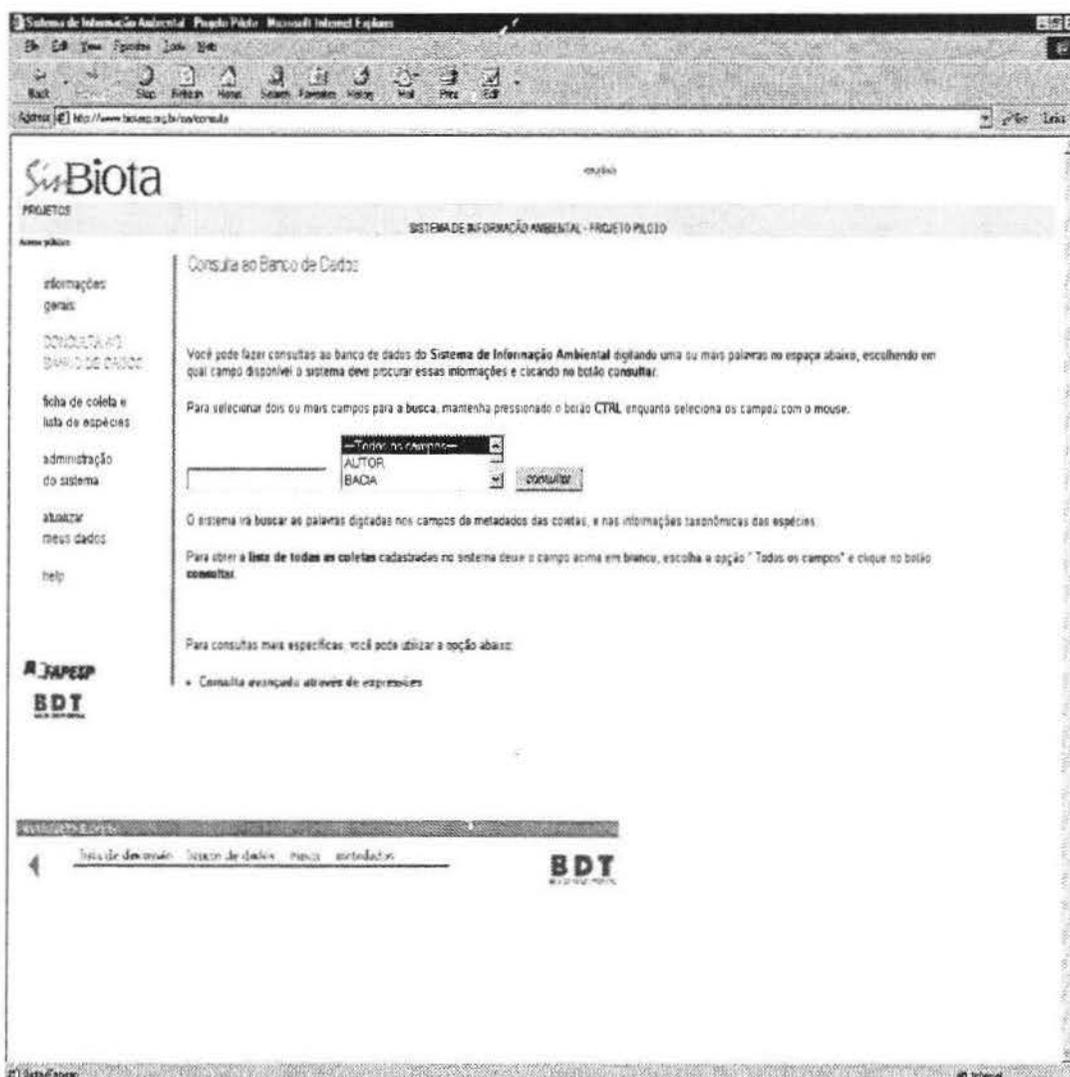


Figura 4.4: Interface de seleção textual por campo específico e por palavra-chave

4.4 Comparação com sistemas existentes

Esta seção compara o SINBIOTASP aos sistemas de informação ambiental apresentados no capítulo 3. Tipos de consulta, formatos de visualização, fontes de dados, formas de interação, dentre outros são alguns dos tópicos avaliados.

Enfoque Conforme descrito no capítulo 3, os sistemas de informação ambiental podem ser divididos em sistemas com enfoque único e sistemas com enfoque múltiplo. O ECN [LP98], UDK [UDK], CDS [CDS], GCMD [GCM], *Environmental Australia On Line* [ERI], EOSDIS [EOS], GILS [GIL] e o FGDC [FGD] se diferenciam do SINBIOTASP,

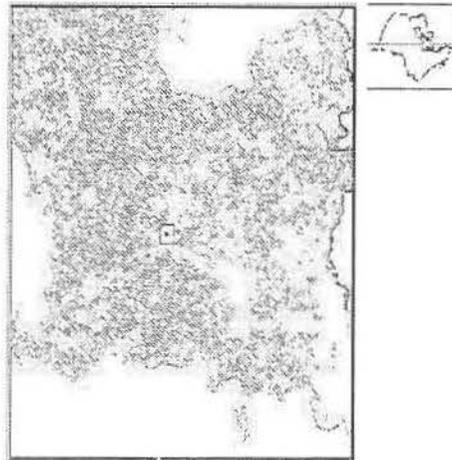


Figura 4.5: Consulta gráfica no módulo Atlas Biota/SP versão 1.0 (descontinuado)

uma vez que os módulos disponíveis para cada um destes sistemas visam mais de um objetivo (sistemas com enfoque múltiplo) e o SINBIOTASP é um sistema com enfoque único, ou seja, todos os módulos visam um único objetivo, gerenciamento dos dados sobre a biodiversidade do Estado de São Paulo, sendo semelhante neste quesito a sistemas como: *Herbarium Specimen Browser* [SLF⁺98], *Species Analyst* [Uni], *Species 2000* [Spe] e o *Malaysia Biodiversity Online* [Mal].

Dados gerenciados O SINBIOTASP gerencia e recupera tanto dados quanto metadados, sendo neste ponto semelhante ao *Herbarium Specimen Browser*, *Species Analyst*, *Environmental Australia On Line*, *Species 2000*, *Malaysia Biodiversity Online*, EOSDIS e o FGDC.

O *Species Analyst* e o *Herbarium Specimen Browser* são bastante distintos dos outros sistemas descritos nesta dissertação, pois enquanto os demais visam recuperar dados globais (em geral arquivos), estes permitem o acesso a tuplas que armazenam informações sobre espécies. Sendo assim, neste sentido, estes estão também mais próximos do SINBIOTASP do que os outros sistemas.

Fontes de dados O *Species Analyst*, o ERIN e o *Species 2000* permitem selecionar as fontes de dados que serão pesquisadas, uma vez que são capazes de acessar múltiplos bancos de dados. Já o SINBIOTASP não disponibiliza esta funcionalidade visto que todos os dados estão armazenados em um único banco de dados.

Interação com mapas As formas de interação com mapas variam de sistema para sistema. O FGDC permite interagir com o mapa de diferentes formas, tais como: clicando

no mapa, definindo uma região no mapa, selecionando um item (região) em uma caixa de seleção. O ERIN e o EOSDIS, por sua vez, permitem definir uma região geográfica como critério de seleção, através da especificação de um retângulo sobre o mapa. O SINBIOTASP, por sua vez, permitia definir critérios de seleção através de um clique sobre o mapa (ponto). No entanto, atualmente, o SINBIOTASP não disponibiliza nenhuma forma de interação com mapas por questão de cronograma de desenvolvimento.

Visualização dos resultados O SINBIOTASP necessita prover tanto uma visualização textual quanto gráfica (mapas) dos dados. Esta é uma preocupação também de sistemas como o *Herbarium Specimen Browser*, o *Species Analyst* e o *Environmental Australia On Line*. Apesar dos sistemas ECN e do GCMD disponibilizarem uma exibição tanto textual quanto gráfica dos dados, estes diferem dos anseios do SINBIOTASP uma vez que a visualização gráfica não é disponibilizada baseando-se em mapas, mas em gráficos de barras. Já o UDK, CDS, EOSDIS, GILS e o FGDC, por sua vez, se preocupam apenas com a visualização textual dos dados.

O *Malaysia Biodiversity Online* e o ERIN possuem a capacidade de retornar uma informação textual quando o usuário clica sobre o mapa, o que ainda não está disponível no SINBIOTASP. Por outro lado, o sistema *Malaysia Biodiversity Online* não permite visualizar o mapa da distribuição de uma determinada espécie, enquanto que o SINBIOTASP disponibiliza a distribuição simultânea de até 16 espécies. A visualização da distribuição de espécies disponíveis no SINBIOTASP é uma das contribuições desta dissertação, conforme descrito no capítulo 6.

O SINBIOTASP contém uma ferramenta mais elaborada do que o *Species 2000* para a consulta e visualização dos dados, permitindo consultas textuais, e disponibilizando os resultados em formato textual e gráfico. Já o *Species 2000* permite realizar apenas consulta textual (campo específico) e visualizar os resultados das consultas textualmente.

Dentre todos os sistemas de informação ambiental analisados no capítulo 3, o sistema *Herbarium Specimen Browser* é o sistema que mais se assemelha ao SINBIOTASP. Contudo, as consultas sobre mapas e a visualização do resultado de consultas sobre mapas no *Herbarium Specimen Browser* se prendem a um conjunto de polígonos fixos de limites pré-definidos (os condados do Texas). Já as regiões pré-definidas no SINBIOTASP são bacias. O sistema *Species Analyst* provê também uma forma de visualizar a distribuição global do local das coletas para os registros identificados por uma consulta através da seleção de uma região no mapa.

Objetivos As funcionalidades do *Species Analyst* parecem se aproximar dos objetivos do BIOTA/FAPESP, com algumas diferenças principais: as fontes de dados do *Species Analyst* são distribuídas e especializadas segundo o tipo de ser vivo coletado; os dados

analisados se referem a número limitado de seres vivos; as análises espaciais são realizadas pelos usuários a partir das tuplas retornadas; as consultas são todas baseadas em especificação textual. A integração com o *Species Analyst* permitiria a ativação dos módulos de simulação e modelagem da distribuição de espécies.

4.5 **Resumo**

Este capítulo descreveu de forma simplificada o sistema de informação ambiental para o BIOTA/FAPESP, bem como estabeleceu uma comparação entre o SINBIOTASP e os sistemas analisados no capítulo 3, frisando as semelhanças e diferenças.

O sistema foi descrito enfatizando os objetivos, as contribuições, os principais componentes que constituem o sistema (*interface com o usuário, gerenciadores de dados e metadados, e repositórios de dados*) e como este projeto se relaciona com esta dissertação.

Esta visão geral do sistema de informação para o BIOTA/FAPESP irá nortear o entendimento e os estudos realizados nos capítulos 5 e 6. Estes capítulos estão centrados no processamento de consultas e gerenciamento dos dados nos diferentes repositórios.

Capítulo 5

Caracterização de Consultas para o BIOTA

Este capítulo apresenta uma sistematização de tipos de consultas que podem ser realizadas no BIOTA, classificando-as sob diferentes critérios. Este estudo foi utilizado na especificação do atual módulo de processamento de consultas do BIOTA - o *Species Mapper* - descrito no capítulo 6.

As consultas foram analisadas em função de dois pontos de vista: processamento e interface com o usuário. Do ponto de vista de processamento da consulta, foram estudados a tipologia dos objetos espaciais e as fontes de dados envolvidas neste processamento, bem como o conjunto básico de operadores e transações espaciais primitivas necessárias. Do ponto de vista de interface, tais consultas diferem entre si com relação ao formato de entrada dos seus predicados, formato de exibição do resultado e à forma de interação do usuário com o sistema.

Nos exemplos dados no decorrer do capítulo serão consideradas as seguintes entidades: municípios, bacias e reservas florestais. Os municípios, bacias e as reservas florestais são regiões pré-definidas, cujo polígono limitante está armazenado no banco de mapas. Supõe-se que a localização de uma coleta realizada e, conseqüentemente, a localização de uma ou mais espécies, seja armazenada como entidades do tipo ponto. Há 18 bacias no estado e a bacia do Mogi-Mirim vem servindo de base para todos os testes.

Este capítulo está organizado da seguinte forma. A seção 5.1 analisa e sistematiza o processamento de consultas. A seção 5.2 examina a interface com o usuário. A seção 5.3 apresenta as conclusões deste capítulo.

5.1 O Processamento de Consultas

O processamento de consultas no SINBIOTASP exige uso do SGBD Oracle e também do SIG ArcInfo. O processamento de consultas em SIG é um assunto bastante pesquisado. Vários trabalhos discutem problemas que são relevantes para o SINBIOTASP enfocando questões como o processamento propriamente dito [SLF⁺98] e a integração entre dados espaciais e não espaciais [SA95].

As consultas no BIOTA podem, de uma forma geral, ser divididas em: consultas *temáticas*, consultas *espaciais* e consultas *mistas*, uma vez que os componentes fundamentais de informações geográficas são atributos temáticos (ou convencionais) e componente espacial. Consultas mistas são aquelas que envolvem componentes espaciais e temáticos, ou seja, objetos inteiros. Em qualquer caso, as consultas podem recuperar atributos temáticos, espaciais, valores calculados a partir destes atributos e objetos inteiros. Podem, igualmente, criar novos objetos (por exemplo, através de visões [Cer96]).

Para facilitar o entendimento dos exemplos, foram copiadas da dissertação de Fagundes [Fag99] duas figuras, que mostram os diagramas de modelagem dos bancos de coletas (figura 5.1) e de metadados (figura 5.2).

5.1.1 Fontes de dados envolvidas no processamento da consulta

Os dados no BIOTA podem ser extraídos diretamente do banco de coletas, ou diretamente do banco de mapas, ou passando por filtros que exploram o interrelacionamento do banco de metadados com estes. Para consultar o banco de mapas é necessário acessar o banco de metadados, uma vez que a descrição do conteúdo do banco de mapas se encontra armazenada no banco de metadados. O banco de metadados armazena a referência espacial dos dados das coletas. Assim, o interrelacionamento coletas e mapas só pode ser realizado a partir do banco de metadados.

Desta forma, as consultas aos bancos de dados do SINBIOTASP podem ser realizadas diretamente sobre os bancos de coletas e de metadados ou através das seguintes combinações: banco de coletas/metadados, banco de metadados/mapas e banco de coletas/metadados/mapas. O banco de metadados contém características relevantes de cada mapa armazenado no banco de mapas, tais como: descrição, escala, datum, sistema de coordenadas, classe, sub-formato, formato, autor, dentre outras.

A seguir, algumas consultas são classificadas em função das fontes de dados requeridas.

- i. **Banco de coletas** Consultas que acessam somente informações contidas no banco de coletas, como por exemplo, "Quais as coletas realizadas pelos membros da instituição X?". Os dados referentes ao autor da coleta (por exemplo, sua instituição) e à coleta estão armazenados no banco de coletas.

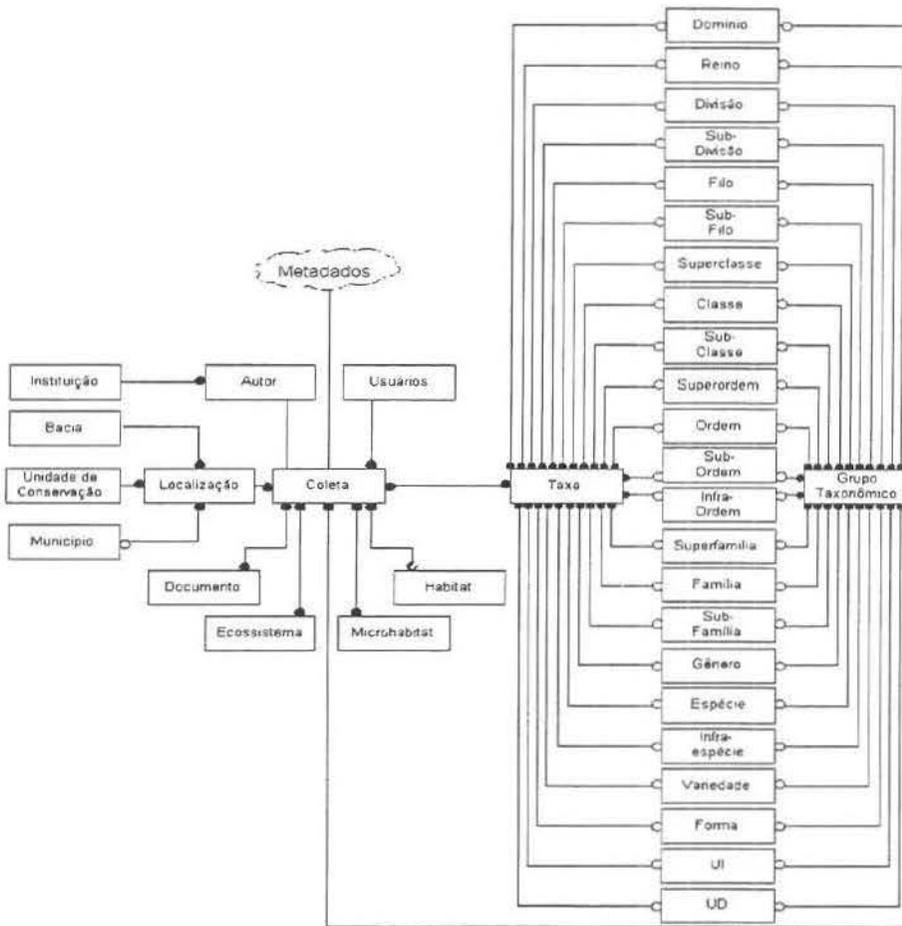


Figura 5.1: Modelagem do Banco de Coletas

- ii. **Banco de metadados** Consultas que acessam somente informações contidas no banco de metadados, como por exemplo, "Quais os métodos de coleta utilizados?" (método de coleta é metadado).
- iii. **Banco de coletas e metadados** Consultas que requerem informações contidas nos bancos de coletas e metadados, como por exemplo, "Quais as espécies coletadas entre 11/01/2000 e 21/01/2000?". A consulta-exemplo requer acesso aos seguintes repositórios de dados: banco de metadados (classe metadado de referência temporal) e, a seguir, banco de coletas (nome da espécie e identificador da coleta).
- iv. **Banco de mapas e metadados** Consultas que operam sobre as informações contidas nos bancos de mapas e de metadados, como por exemplo, "A Reserva do Camapoã está dentro do município de Mogi Guaçu?". Através de uma consulta ao banco de metadados se determina o mapa que contém a Reserva do Camapoã e o município de Mogi-Mirim (metadados de mapas).

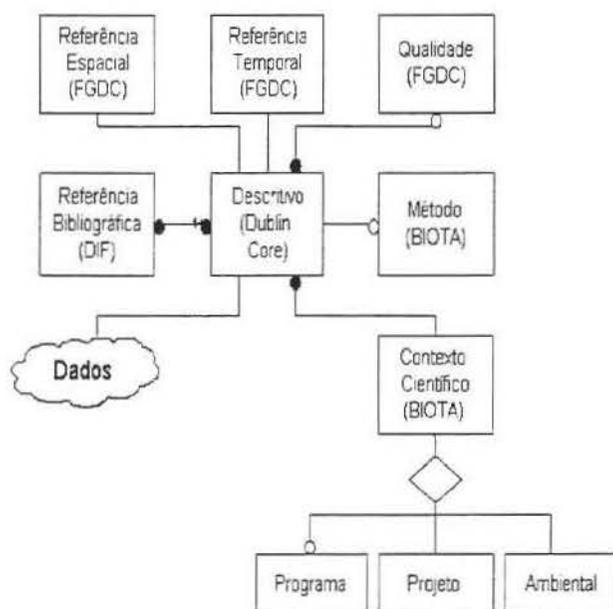


Figura 5.2: Modelagem do Banco de Metadados

- v. **Banco de coletas, mapas e metadados** Consultas que interrelacionam informações contidas nos bancos de coletas, de metadados e de mapas, como por exemplo, "Quais as espécies coletadas pelo método de Dragagem e que estão localizadas dentro da bacia do Mogi Mirim?". A consulta-exemplo requer acesso aos seguintes repositórios de dados: banco de metadados (método de coleta, localização, identificador da coleta e identificador do mapa), banco de coletas (nome da espécie e identificador da coleta) e banco de mapas (limites da bacia do Mogi Mirim e identificador do mapa).

5.1.2 Tipologia dos objetos espaciais

A determinação dos objetos espaciais e seus interrelacionamentos envolvidos no processamento das consultas contribui para sistematizar os operadores espaciais necessários à execução de consultas espaciais. Os três tipos de objetos espaciais utilizados no SINBIOTASP são:

- **Ponto.** O ponto é um objeto espacial muito utilizado e de extrema importância na tomada de decisões para a conservação ambiental. O ponto indica a localização de uma coleta (e das espécies correspondentes). Sendo assim, um conjunto de pontos referentes a uma determinada espécie caracteriza a distribuição das observações feitas sobre a espécie em uma certa região.

	Processamento da consulta	SINBIOTASP
Ponto	O1	✓
Linha	O2	✓
Região	O3	✓
Relação Ponto/Ponto	O4	✓
Relação Ponto/Linha	O5	✓
Relação Ponto/Região	O6	✓
Relação Linha/Linha	O7	✓
Relação Linha/Região	O8	✓
Relação Região/Região	O9	✓

Tabela 5.1: Objetos espaciais e seus interrelacionamentos utilizados nas consultas do SINBIOTASP

- **Linha.** A linha é um objeto espacial utilizado para representação de rios e, futuramente, de rodovias no SINBIOTASP.
- **Região.** Há dois tipos de regiões no sistema: pré-definidas e definidas pelo usuário. As regiões pré-definidas são polígonos especificados a priori no banco de mapas, com identidade própria, como por exemplo: as 18 bacias do Estado, os 465 municípios do Estado de São Paulo. Já as regiões definidas pelo usuário são retângulos especificados textual ou graficamente ("*window query*").

A tabela 5.1 analisa quais objetos espaciais estão envolvidos no processamento das consultas. Cada célula corresponde a uma possível combinação de objetos. Por exemplo, a consulta "Listar as espécies encontradas no ponto selecionado" usa um ponto durante o processamento da consulta (célula **O1**). A consulta "Mostre no mapa a região correspondente à bacia do Mogi-Mirim" utiliza uma região pré-definida durante o processamento da consulta (célula **O3**). As consultas "Quais as localizações onde coexistem as espécies *Acacia polyphylla* e *Acacia glomerosa*", "Selecione as coletas que estão contidas na bacia do Mogi-Mirim." e "Qual a região resultante da interseção das regiões onde existem as espécies *Acacia polyphylla* e *Acacia glomerosa*?" usam, respectivamente, as seguintes relações entre objetos espaciais: ponto/ponto (célula **O4**), ponto/região (célula **O6**) e região/região (célula **O9**).

As consultas "Mostre a distribuição hidrografia atual", "Quais as coletas realizadas a menos de 5 Km do Rio Grande?", "Quais os afluentes do Rio Grande?" e "Quais as regiões atravessadas pelo Rio Mogi?", por sua vez, são exemplos da utilização de objetos espaciais do tipo linha e correspondem, respectivamente, às células **O2**, **O5**, **O7** e **O8**.

5.1.3 Operadores de consultas possíveis e conjunto básico para o BIOTA/FAPESP

Em geral, uma consulta tem mais de uma solução de processamento. A escolha da forma a ser adotada depende do custo de cada solução. O texto subsequente apresenta exemplos que mostram esta multiplicidade de soluções para o SINBIOTASP. A tabela 5.2 mostra os operadores espaciais que podem ser utilizados no BIOTA (coluna 1) e aqueles que efetivamente devem ser considerados (coluna 2). As consultas analisadas envolvem coletas. A consulta "Quais as regiões atravessadas pelo Rio Mogi onde há *Acacia glomerosa*?" exemplifica o uso do operador topológico *cross* (célula Op8), pois este operador requer uma linha (Rio Mogi) como um dos operandos.

A consulta "Quais as regiões de plantação de cana localizadas ao leste da bacia do Mogi-Mirim?" utiliza o operador *fuzzy* de direção (célula Op2). A consulta "Quais as coordenadas correspondentes à distribuição da espécie *Acacia glomerosa*?" exemplifica o uso do operador de localização (célula Op1). O operador métrico (célula Op3) é utilizado para determinar relação de distância entre objetos espaciais, como por exemplo, "Quais as espécies que se encontram a uma distância menor que 10 km da espécie *Acacia polyphylla*?".

O operador topológico *touch* (célula Op5) é utilizado para determinar uma relação de adjacência entre uma coleta e uma região ou entre duas regiões, como por exemplo, "A bacia do Mogi-Mirim é adjacente à Reserva do Camapoã?". O operador *overlap* (célula Op6), por sua vez, determina uma região resultante da interseção de duas outras, como por exemplo, "Quais as regiões resultantes da interseção das regiões onde são encontradas as espécies *Acacia polyphylla* e *Acacia glomerosa*?".

Os operadores *disjoint* (célula Op4) e *in* (célula Op7) são de uma certa forma complementares e podem ser utilizados para uma mesma consulta quando está sendo procurado estabelecer uma relação de coincidência entre objetos espaciais do tipo ponto. Caso se deseje determinar a relação de "está dentro de" entre duas regiões somente o operador *in* pode ser utilizado. Por último, caso seja estabelecido uma relação de pertinência entre ponto e região, ambos operadores podem ser utilizados. Por exemplo, na consulta "Exibir as localizações onde existe a espécie *Acacia polyphylla* dentro da bacia do Mogi-Mirim", a utilização do operador *in* verifica se os pontos estão contidos ou não dentro da bacia do Mogi-Mirim. Já o uso de *disjoint* (ponto, região) também pode ser adotado no mesmo caso. Este é um exemplo "Exibir as localizações onde existe a espécie *Acacia polyphylla* dentro da bacia do Mogi-Mirim".

A ausência de *benchmarks* adequados e a dificuldade de definir parâmetros de custo para processamento de consultas espaciais torna muito complicado o problema de seu processamento. A solução adotada foi a definição de um conjunto básico de operadores a

	Consultas	SINBIOTASP
Operador de localização	Op1	✓
Operador de direção	Op2	-
Operador métrico	Op3	✓
Operador topológico - <i>disjoint</i>	Op4	✓
Operador topológico - <i>touch</i>	Op5	✓
Operador topológico - <i>overlap</i>	Op6	✓
Operador topológico - <i>in</i>	Op7	✓
Operador topológico - <i>cross</i>	Op8	✓

Tabela 5.2: Conjunto básico de operadores espaciais necessários para o processamento das consultas espaciais do SINBIOTASP

ser usado em qualquer situação.

A segunda coluna da tabela 5.2 descreve o conjunto básico de operadores espaciais definido a partir das considerações anteriores. A tabela identifica como básicos os operadores de localização e métrico, além dos cinco operadores topológicos *disjoint*, *touch*, *overlap*, *in* e *cross*. Operadores *fuzzy* devem ser descartados por dificuldades de interpretação e implementação. No entanto, uma avaliação mais precisa para cada caso pode ser realizada a partir da execução de um *benchmark* para medir o desempenho de cada operador espacial para cada consulta em específico. A construção deste *benchmark* é proposta como uma das extensões a esta dissertação.

5.1.4 Conjunto de transações primitivas para o BIOTA

A tabela 5.3 apresenta as transações primitivas necessárias para o processamento das consultas do BIOTA, dentre as apresentadas no capítulo 2.

A operação de superposição (célula **T2**) é usada por consultas que precisam superpor polígonos (OGs-2D) de diferentes temas, como por exemplo, "A Reserva do Camapoã está dentro da bacia do Mogi-Mirim?". A operação de análise de proximidade (célula **T4**), por sua vez, é necessária em função de algumas consultas que requerem a especificação de relações de distância entre dois objetos espaciais, como por exemplo, "Quais as espécies que se encontram a uma distância menor que 10 km da espécie *Acacia polyphylla*?".

A transação topológica (célula **T5**) é aplicada quando se deseja determinar a existência de uma relação topológica entre dois ou mais objetos espaciais, como por exemplo, "Quais as espécies que foram encontradas dentro da bacia do Mogi-Mirim?". A utilização da operação de decomposição (célula **T6**) se justifica em função de algumas consultas requererem somente os objetos interiores a um determinado polígono (região), excluindo do conjunto

	Consultas	SINBIOTASP
Reclassificação	T1	-
Superposição	T2	✓
Ponderação	T3	-
Análise de proximidade	T4	✓
Topológicas	T5	✓
Decomposição	T6	✓
Transações baseadas em conjunto	T7	✓
Transações diversas	T8	✓

Tabela 5.3: Transações primitivas necessárias para o processamento de consultas espaciais básicas no SINBIOTASP

de soluções os objetos que estão sobre os limites do polígono. Por exemplo, "Quais as regiões de plantação de cana situadas dentro da bacia do Mogi-Mirim?".

As transações baseadas em conjunto (célula T7) retornam, por sua vez, um conjunto de polígonos baseados em operações matemáticas de conjunto entre as fronteiras de vários polígonos, como por exemplo, "Quais os tipos de uso do solo encontrados na bacia do Mogi-Mirim?". As transações diversas (célula T8) também são utilizadas para atender, por exemplo, operações de visualização, conversão de formatos de dados, dentre outras.

Finalmente, as transações primitivas de reclassificação (célula T1) e ponderação (célula T3) são típicas de simulações e modelagens de distribuição de espécies, ainda não disponíveis no BIOTA. Vale a pena, no entanto, ressaltar que para definir alguns dados referentes a um mapa (por exemplo, uso do solo), o mapa passou por reclassificação (célula T1) e ponderação (célula T3).

5.1.5 Consultas básicas e sua classificação

De maneira geral, uma consulta pode ser expressa em função de:

```

select {resultado}
from {fonte de dado}
where {predicado}

```

Os componentes fonte de dado, predicado e resultado de uma consulta dependem diretamente dos dados a que estão ligados. A vantagem em distinguir estes componentes é facilitar a compreensão da semântica da consulta.

As funções de processamento são naturalmente dependentes dos tipos de dados envolvidos. A ligação entre os dados convencionais e espaciais e as funções de processamento

de consultas espaciais é feita por mecanismos que definem restrições sobre o conjunto de dados, que podem ser espaciais ou não. Câmara *et al.* [CCH⁺96] mostram como algumas políticas específicas para otimização de consultas requerem o tratamento em separado de dados convencionais e espaciais. Desta forma, a classificação de consultas aqui realizada visa especificar os tipos de consultas possíveis em função dos tipos de dados manipulados.

Baseando-se nas definições apresentadas por Câmara *et al.* [CCH⁺96], Medeiros e Pires [MP94], Medeiros e Botelho [MB96] e Faria [Far98], esta dissertação propõe a definição dos seguintes tipos de dados: convencionais, espaciais e geográficos. Dados convencionais são atributos alfanuméricos tradicionais. Estes dados descrevem as propriedades temáticas de uma entidade geográfica. Já os dados espaciais descrevem as características espaciais de uma entidade geográfica. Este tipo de dado é responsável por informar a localização espacial de um fenômeno, associada a propriedades geométricas e topológicas.

Os dados geográficos são comumente caracterizados a partir de três componentes fundamentais: atributo (dado convencional), localização (dado espacial) e tempo. O componente tempo descreve os períodos em que os valores daqueles dados geográficos são válidos. Esta dissertação trata o tempo como um dado convencional. Logo, o dado geográfico é formado, aqui, a partir da combinação de dados convencionais e dados espaciais. Este tipo de dado geográfico é denominado neste texto de *dado misto*.

A partir da definição dos tipos de dados é possível caracterizar os tipos de fonte de dado, predicado e resultado.

- **Fonte de dado** A fonte de dados é definida em função do tipo do dado. As fontes de dados podem ser: convencionais, espaciais ou mistas. A fonte de dados convencional é aquela que contém dados do tipo convencional (banco de coletas). A fonte de dados espacial, por sua vez, contém dados do tipo espacial (componentes espaciais retirados do banco de metadados ou de mapas). Já a fonte de dados mista contém dados dos tipos convencional e espacial (combinação dos diferentes bancos de dados).
- **Predicado** O tipo de um predicado é definido em função dos operadores utilizados. O predicado de uma consulta pode ser: convencional, espacial ou misto. Predicados convencionais não utilizam operadores espaciais. Predicados espaciais utilizam somente operadores espaciais (de direção, métrico, topológico e de localização). O predicado misto combina predicados convencional e espacial.
- **Resultado** O resultado é definido em função do tipo de dado retornado. Os resultados podem também ser classificados em: convencionais, espaciais ou mistos. O resultado é convencional quando somente dados convencionais são retornados. O resultado é espacial quando são retornados somente dados espaciais. O resultado é misto quando são retornados dados convencionais e dados espaciais.

Resultado	Predicado	Fonte de Dado	Fonte de Dado	Fonte de Dado
		Convencional	Espacial	Mista
		Tipo de Dado	Tipo de Dado	Tipo de Dado
		Convencional	Espacial	Misto
		Banco de Coletas	Componentes Espaciais	Combinação entre os BDs
Convencional	Convencional	C1	C2	C3
	Espacial	C4	C5	C6
	Misto	C7	C8	C9
Espacial	Convencional	C10	C11	C12
	Espacial	C13	C14	C15
	Misto	C16	C17	C18
Misto	Convencional	C19	C20	C21
	Espacial	C22	C23	C24
	Misto	C25	C26	C27

Tabela 5.4: Caracterização dos tipos de consultas

O SINBIOTASP exige resultados que envolvam mapas. Um mapa com legenda é um exemplo de resultado do tipo misto.

Os bancos de dados do SINBIOTASP manipulam tanto dados convencionais quanto espaciais. O banco de coletas armazena somente dados convencionais. O banco de metadados, por sua vez, armazena dados convencionais e espaciais (localização das coletas). Já o banco de mapas armazena tanto dados convencionais (por exemplo, legenda) quanto dados espaciais (objetos espaciais, tais como ponto, linha e polígono). Sendo assim, o banco de coletas é uma fonte de dado convencional e os bancos de metadados e de mapas são fontes de dado mistas.

A tabela 5.4 apresenta os diversos tipos de consultas possíveis de serem processadas. Cada célula da tabela corresponde a um tipo de consulta particular. Os tipos de consulta, descritos a seguir, estão agrupados de acordo com a fonte de dado exigida.

Esta dissertação considera que o usuário tem a possibilidade de definir dois tipos de regiões como predicado de seleção: (a) seleção de polígonos pré-definidos (por exemplo, bacia, município) e (b) definição de um retângulo. Neste texto, considera-se um retângulo sempre que for descrito como predicado: "região definida pelo usuário". Quando o predicado contém um retângulo como parâmetro, o processamento não requer obrigatoriamente a utilização de operadores espaciais, como será visto a seguir.

Fonte de dado convencional As consultas deste tipo manipulam somente dados convencionais, no caso apenas banco de coletas.

- i. **Célula C1** Esta célula corresponde a consultas que têm fonte de dado convencional, predicado convencional e resultado convencional.

"Quais as espécies da família X já coletadas?" é um exemplo de consulta do tipo **C1**. O processamento da consulta seleciona, no banco de coletas, as espécies pertencentes à família X . Este processamento é executado pelo SGBD Oracle. A fonte de dado é convencional (banco de coletas), o predicado é convencional (família X) e o resultado é convencional (nomes de espécies).

As demais consultas da primeira coluna da tabela 5.4 (fonte de dado convencional) não são possíveis porque seus predicados ou resultados envolvem algum tipo de dado espacial, mas a fonte de dado é convencional.

Fonte de dado espacial As consultas deste tipo manipulam somente dados espaciais. Nenhum dos bancos de dados do SINBIOTASP é estritamente espacial. Para analisar as consultas do tipo **C2**, **C5**, **C8**, **C11**, **C14**, **C17**, **C20**, **C23** e **C26**, suponha que previamente tenha sido realizada uma consulta que retorna como resultado um conjunto de pontos \mathcal{P} representando a distribuição de uma espécie qualquer. Este conjunto de pontos \mathcal{P} é uma fonte de dado estritamente espacial. As consultas que possuem fonte de dado estritamente espacial são via de regra processadas pelo SIG ArcInfo.

- i. **Célula C5** Esta célula corresponde a consultas que têm fonte de dado espacial, predicado espacial e resultado convencional.

Um exemplo seria "Qual a menor distância entre dois pontos quaisquer em \mathcal{P} ?". O processamento da consulta executa uma operação métrica de distância entre todos os possíveis pares de pontos. A fonte de dado é espacial (conjunto de pontos \mathcal{P}), o predicado é espacial (distância) e o resultado é convencional (valor numérico - distância).

- ii. **Célula C14** Esta célula corresponde a consultas que têm fonte de dado espacial, predicado espacial e resultado espacial.

Um exemplo é "Quais são os dois pontos de \mathcal{P} mais próximos?". O processamento da consulta executa uma operação métrica de distância entre todos os possíveis pares de pontos. A fonte de dado é espacial (conjunto de pontos), o predicado é espacial (mais próximos - distância) e o resultado é espacial (dois pontos).

Note-se que a diferença entre os exemplos de **C14** e **C5** é o resultado - em um caso localização e no outro um escalar.

- iii. **Célula C23** Esta célula corresponde a consultas que têm fonte de dado espacial, predicado espacial e resultado misto. Um exemplo seria combinar as consultas **C5** e **C14**: "Qual a menor distância entre dois pontos de \mathcal{P} e quais são estes pontos?".

As demais consultas da coluna 2 - **C2**, **C8**, **C11**, **C17**, **C20** e **C26** - não são possíveis, visto que não é possível aplicar predicado convencional sobre dados do tipo espacial e portanto também não se pode aplicar predicado misto. Note-se que esta afirmação depende das definições adotadas nesta dissertação. Seja, por exemplo, a consulta "Quais os pontos em \mathcal{P} com latitude igual a Y ?". Esta consulta pode ser interpretada de duas formas:

- (a) Fonte de dado espacial (\mathcal{P}), predicado convencional ("latitude = Y ") e resultado espacial (conjunto de pontos);
- (b) Fonte de dado e resultado espacial, porém predicado espacial ("latitude = Y "). Nesta segunda interpretação, adotada na dissertação, o que é considerado é a consulta do usuário do ponto de vista semântico e não sua especificação em alguma linguagem. Neste caso, a consulta corresponde a "Quais os pontos que estão em alguma faixa de latitude do globo terrestre?", ou seja, $p \mid p \in \mathcal{P} \wedge p \text{ in região}(\text{latitude} = Y)$.

Raciocínio semelhante se aplica aos demais elementos descartados da segunda coluna da tabela 5.4 (fonte de dado espacial).

Fonte de dado mista As consultas deste tipo manipulam dados convencionais e dados espaciais. Isto significa que as fontes de dado podem ser os bancos de mapas ou de metadados, ou ainda qualquer combinação destes com o banco de coletas (por exemplo, bancos de coletas e de metadados ou bancos de coletas, de metadados e de mapas).

- i. **Célula C3** Esta célula corresponde a consultas que têm fonte de dado mista, predicado convencional e resultado convencional.

A consulta "Quais os métodos de coleta utilizados?" é um exemplo de consulta do tipo **C3**. O processamento da consulta seleciona, no banco de metadados, os métodos de coleta utilizados. Este processamento é executado pelo SGBD Oracle. A fonte de dado é mista (banco de metadados), o predicado é convencional (seleção dos métodos de coleta) e o resultado é convencional (métodos de coleta).

Em adição, as seguintes consultas do tipo **C3** exemplificam que alguns tipos de consulta são satisfeitos por mais de uma alternativa de combinação de banco de dados, por exemplo: "Quais os métodos de coleta utilizados?" requer somente o banco de metadados, "Quais os tipos de uso do solo existentes?" requer somente o banco de mapas e "Quais os métodos de coletas utilizados para encontrar organismos do gênero *Acacia*?" requer os bancos de coletas/metadados.

- ii. **Célula C6** Esta célula corresponde a consultas que têm fonte de dado mista, predicado espacial e resultado convencional.

A consulta "Qual a menor distância entre duas coletas realizadas?". O processamento da consulta executa uma operação métrica de distância entre as localizações das coletas armazenadas no banco de metadados. Este processamento é executado pelo SGBD Oracle e pelo SIG ArcInfo. A fonte de dado é mista (banco de metadados), o predicado é espacial (distância) e o resultado é convencional (valor numérico - distância).

- iii. **Célula C9** Esta célula corresponde a consultas que têm fonte de dado mista, predicado misto e resultado convencional.

A consulta "Quais as espécies da família *X* localizadas na bacia selecionada?" é um exemplo do tipo **C9**. O processamento pode ser realizado da seguinte forma: (a) seleciona, no banco de metadados, a localização das coletas (classe metadado de referência espacial); (b) seleciona, no banco de mapas, os limites da bacia selecionada; (c) executa uma operação topológica de inclusão a fim de determinar quais as coletas estão contidas dentro dos limites da bacia selecionada; (d) seleciona, no banco de coletas, as espécies da família *X* destas coletas. Este processamento é executado pelo SGBD Oracle e pelo SIG ArcInfo. A fonte de dados é mista (banco de coletas, de metadados e de mapas), o predicado é misto (família *X* e localizadas na bacia selecionada) e o resultado é convencional (espécies).

- iv. **Célula C12** Esta célula corresponde a consultas que têm fonte de dado mista, predicado convencional e resultado espacial.

Um exemplo de consulta deste tipo é "Quais as localizações das coletas onde foram encontradas *Acacia polyphylla*?". O processamento pode ser dividido em dois passos: (a) seleciona, no banco de coletas, as coletas que encontraram *Acacia polyphylla*; (b) seleciona, no banco de metadados, as localizações das coletas selecionadas no passo (a). Este processamento é executado pelo SGBD Oracle. Esta consulta possui as seguintes características: fonte de dado mista (banco de coletas e de metadados), predicado convencional (*Acacia polyphylla*) e resultado espacial (localização das coletas - coordenadas).

- v. **Célula C15** Esta célula corresponde a consultas que têm fonte de dado mista, predicado espacial e resultado espacial.

Como exemplo de consulta do tipo **C15**, podemos mencionar: "Quais as localizações das coletas realizadas dentro da bacia selecionada?". O processamento desta consulta pode ser realizado da seguinte forma: (a) seleciona, no banco de mapas, os

limites da bacia selecionada; (b) seleciona, no banco de metadados, a localização das coletas; (c) executa uma operação topológica de inclusão a fim de determinar todas as localizações situadas dentro dos limites da bacia selecionada. Este processamento é executado pelo SGBD Oracle e pelo SIG ArcInfo. A fonte de dado é mista (banco de metadados e de mapas), o predicado é espacial (dentro da bacia selecionada) e o resultado é espacial (localização das coletas - coordenadas).

- vi. **Célula C18** Esta célula corresponde a consultas que têm fonte de dado mista, predicado misto e resultado espacial.

Um exemplo de consulta do tipo **C18** é "Quais as localizações das coletas do gênero *Acacia* realizadas a menos de 5 km do ponto selecionado?". O processamento pode ser dividido nas seguintes etapas: (a) cria uma região de *buffer* de raio igual a 5 km em torno do ponto; (b) seleciona, no banco de coletas, as coletas do gênero *Acacia*; (c) seleciona, no banco de metadados, as localizações destas coletas; (d) executa uma operação topológica de inclusão a fim de determinar quais destas coletas estão situadas dentro da região do *buffer*. Este processamento é executado pelo SGBD Oracle e pelo SIG ArcInfo. Esta consulta possui as seguintes características: fonte de dado mista (banco de coletas e de metadados), predicado misto (gênero *Acacia* e menos de 5 km do ponto selecionado) e resultado espacial (localização das coletas - coordenadas).

- vii. **Célula C21** Esta célula corresponde a consultas que têm fonte de dado mista, predicado convencional e resultado misto.

Um exemplo é "Quais os métodos de coleta utilizados e as localizações onde foram encontrados organismos do gênero *Acacia*?". O processamento pode ser executado da seguinte forma: (a) seleciona, no banco de coletas, as coletas que encontram organismos do gênero *Acacia*; (b) seleciona, no banco de metadados, as localizações e os métodos de coleta referentes às coletas selecionadas no item (a). Este processamento é executado pelo SGBD Oracle. Esta consulta possui as seguintes características: fonte de dado mista (banco de coletas e metadados), predicado convencional (gênero *Acacia*) e resultado misto (localização - coordenada e método de coleta).

- viii. **Célula C24** Esta célula corresponde a consultas que têm fonte de dado mista, predicado espacial e resultado misto.

Como exemplo de consulta do tipo **C24**, podemos mencionar: "Qual a menor distância entre duas coletas e qual a localização destas coletas?". O processamento desta consulta executa uma operação métrica de distância entre as localizações das coletas armazenadas no banco de metadados. Este processamento é executado pelo SGBD Oracle e pelo SIG ArcInfo. A fonte de dados é mista (banco de metadados),

Resultado	Predicado	Fonte de Dado	Fonte de Dado	Fonte de Dado
		Convencional	Espacial	Mista
		Tipo de Dado	Tipo de Dado	Tipo de Dado
		Convencional	Espacial	Misto
		Banco de Coletas	Componentes Espaciais	Combinação entre os BDs
Convencional	Convencional	✓	-	✓
	Espacial	-	✓	✓
	Misto	-	-	✓
Espacial	Convencional	-	-	✓
	Espacial	-	✓	✓
	Misto	-	-	✓
Misto	Convencional	-	-	✓
	Espacial	-	✓	✓
	Misto	-	-	✓

Tabela 5.5: Caracterização dos tipos de consultas possíveis

o predicado é espacial (distância) e o resultado é misto (valor numérico - distância e localização das coletas - coordenadas).

- ix. **Célula C27** Esta célula corresponde a consultas que têm fonte de dado mista, predicado misto e resultado misto.

Um exemplo é "Quais as localizações e os autores das coletas do gênero *Acacia* realizadas a menos de 5 Km do ponto selecionado?". O processamento pode ser dividido nas seguintes etapas: (a) seleciona, no banco de coletas, as coletas do gênero *Acacia* com seus respectivos autores; (b) cria uma região de *buffer* de raio igual a 5 Km em torno do ponto selecionado; (c) seleciona, no banco de metadados, as localizações das coletas selecionadas na etapa (a); (d) executa uma operação topológica de inclusão a fim de determinar quais destas coletas estão situadas dentro da região do *buffer*. Este processamento é executado pelo SGBD Oracle e pelo SIG ArcInfo. Esta consulta apresenta as seguintes características: fonte de dado mista (banco de coletas e de metadados), predicado misto (gênero *Acacia* e menos de 5 Km do ponto selecionado) e resultado misto (localização das coletas - coordenadas e autor).

A análise anterior de cada tipo de consulta permite concluir que somente os tipos de consulta sinalizados por ✓ na tabela 5.5 são possíveis. Esta conclusão parte das definições adotadas, neste capítulo, para cada tipo de fonte de dado, predicado e resultado.

5.1.6 Processamento via SGBD *versus* processamento via SIG

No SINBIOTASP, algumas consultas podem ser resolvidas tanto através da utilização de operadores espaciais (processamento espacial no SIG) quanto a partir de expressões SQL (processamento convencional no SGBD relacional). Este fato é exemplificado através das consultas apresentadas a seguir.

- i. **"Quais as localizações onde coexistem os grupos taxonômicos *mammalia* e *mollusca*?"** Esta consulta pode ser resolvida através da aplicação do operador topológico *in*, aplicado quando se deseja determinar a coincidência entre dois pontos, no caso os pontos de ocorrência dos grupos taxonômicos *mammalia* e *mollusca*. Por outro lado, a execução de consultas não espaciais SQL pode também satisfazer a consulta. Uma das consultas SQL deve recuperar as localizações onde existem organismos pertencentes ao grupo taxonômico *mammalia* a partir dos bancos de coletas e de metadados e a outra consulta deve recuperar as localizações onde existem organismos pertencentes ao grupo taxonômico *mollusca*. O resultado é a interseção dos resultados das duas consultas SQL.

```
select sp.localizacao
from espacial sp, ambiental amb, coleta
where (sp.identificador = coleta.identificador)
and (amb.identificador = coleta.identificador)
and (amb.grupo_taxonomico = "mammalia")
and
(sp.localizacao in
      (select sp.localizacao
      from espacial sp, ambiental amb, coleta
      where (sp.identificador = coleta.identificador)
      and (amb.identificador = coleta.identificador)
      and (amb.grupo_taxonomico = "mollusca"))
);
```

- ii. **"Selecione as coletas que estão contidas dentro de um retângulo definido pelo usuário."** Esta consulta pode ser resolvida por exemplo através da aplicação do operador topológico *in*, pois este operador pode ser aplicado quando se deseja determinar a relação existente entre um ponto e um polígono, ou seja, os pontos de ocorrência das coletas e o retângulo definido pelo usuário. No entanto, a execução de uma expressão SQL também pode responder a consulta. A expressão SQL requer uma transformação das coordenadas do canto superior e inferior do retângulo para a mesma escala das coordenadas das coletas armazenadas no banco de dados (banco de metadados), para a seguir verificar se as coordenadas das coletas estão entre os limites do retângulo definido pelo usuário.

	Saída Textual	Saída Gráfica	Saída Textual e Gráfica
Formulação Textual	F1	F2	F3
Formulação Gráfica	F4	F5	F6
Formulação Textual e Gráfica	F7	F8	F9

Tabela 5.6: Classificação de consultas em função dos formatos de entrada e saída

Considerando uma especificação eficiente das consultas, a utilização de expressões SQL é a opção ideal para responder as consultas que podem ser resolvidas tanto via expressões SQL quanto através do uso de operadores espaciais.

5.2 A Interface com o Usuário

O projeto SINBIOTASP visa fornecer um suporte de interface textual e gráfico ao processamento de consultas sobre o banco de dados de biodiversidade. Esta seção analisa a especificação da interface necessária para as consultas do SINBIOTASP. Esta análise classifica as consultas em função do formato de entrada para predicados e fontes de dados (textual, gráfico ou ambos), e das formas de visualização dos resultados. Os formatos de consulta necessários estão resumidos na tabela 5.6.

Consultas a partir de formulação textual (via formulários) Trata-se de consultas formuladas fornecendo parâmetros dentro de um formulário apresentado ao usuário. Lidam apenas com predicados do tipo textual, podendo ser aplicadas aos bancos de dados de coletas, de metadados e/ou de mapas, no caso dos parâmetros informados serem coordenadas geográficas.

Consultas a partir de formulação gráfica em mapas Trata-se de consultas em que os parâmetros fornecidos são objetos e/ou entidades geográficas selecionadas, e/ou regiões definidas pelo usuário sobre um mapa. Além disso, a formulação gráfica só faz sentido se partir do pressuposto que a tela exibe o resultado de alguma consulta prévia. Esta consulta tanto pode ser explícita (de uma interação anterior do usuário) quanto implícita (no caso em que a aplicação é iniciada pela exibição de um mapa inicial ao usuário).

Consultas combinando formulação textual em formulários e gráfica em mapas Trata-se de consultas especificadas a partir da informação de parâmetros dentro de um formulário apresentado ao usuário, combinadas com interação em mapas.

Cada célula da tabela 5.6 será avaliada a seguir em função do SINBIOTASP.

- i. **Célula F1** Esta célula corresponde a uma consulta onde o usuário especifica uma restrição textual, e os resultados são retornados textualmente. Por exemplo, "Quais os identificadores das coletas realizadas entre 11/01/2000 e 21/01/2000?"
- ii. **Célula F2** As consultas "Exibir a distribuição das coletas que estão contidas na bacia do Mogi-Mirim" e "Qual o mapa da vegetação da bacia do Mogi-Mirim?" são exemplos deste tipo de consulta cuja formulação é semelhante à realizada nas consultas do tipo **F1**, mas o resultado retornado é um mapa.
- iii. **Célula F3** A consulta "Exibir a distribuição das espécies que estão contidas na bacia do Mogi-Mirim, juntamente com uma legenda do mapa" é um exemplo de consulta do tipo **F3**. Esta célula corresponde a consultas onde o usuário especifica uma restrição textual (bacia do Mogi-Mirim) e os resultados retornados são textual (legenda) e gráfico (mapa com a distribuição das espécies).
- iv. **Célula F4** Esta célula corresponde às consultas que permitem a especificação gráfica dos predicados e uma visualização textual do resultado da consulta. A formulação gráfica da consulta pode ser especificada de três diferentes formas: através de um ponto, ou de uma região definida pelo usuário, ou da seleção de uma região pré-definida.
 - (a) **Pontual.** O usuário deve clicar sobre um ponto em um mapa para obter as informações convencionais referentes àquele ponto. No caso de existir um conjunto de pontos sobrepostos e/ou visivelmente próximos, o sistema retorna as informações referentes a todos estes pontos.

Suponha que o mapa disponível seja resultante de interações anteriores do usuário, onde a localização de uma coleta é representada por um ponto. Este ponto é resultado de uma consulta sobre os bancos de coletas e de metadados, sendo possível através de um clique sobre ele recuperar as informações da coleta localizada no ponto especificado, como por exemplo: método de coleta, espécies encontradas, autor da coleta, dentre outras.
 - (b) **Região definida pelo usuário.** O usuário deve definir os cantos superior e inferior de um retângulo que delimitará a região de interesse.
 - (c) **Região pré-definida.** O usuário seleciona uma região pré-definida através de um clique sobre o limitante ou no interior da região já existente no mapa (bacia do Mogi-Mirim, por exemplo).

- v. **Célula F5** Este tipo de consulta permite uma formulação gráfica semelhante à consulta **F4** e uma exibição gráfica dos resultados. Por exemplo, "Exibir a distribuição das coletas realizadas a menos de 5 km do ponto selecionado", "Exibir a distribuição das coletas realizadas dentro da região definida pelo usuário" e "Exibir a distribuição das espécies encontradas dentro da bacia selecionada".
- vi. **Célula F6** A consulta "Exibir a distribuição das espécies que estão contidas na bacia selecionada, juntamente com uma legenda do mapa" é um exemplo de consulta do tipo **F6**. Esta célula corresponde a consultas onde o usuário especifica uma restrição gráfica (bacia selecionada) e os resultados retornados são textual (legenda) e gráfico (mapa com a distribuição das espécies).
- vii. **Célula F7** Esta célula corresponde às consultas que permitem uma combinação entre a formulação textual e gráfica descritas anteriormente e disponibilizam uma visualização textual do resultado. A formulação combina formulários e mapas. Por exemplo, "Quais as coletas realizadas por Ana Eugênia Farinha que estão a menos de 5 km do ponto selecionado?" e "Quais as coletas realizadas entre 11/01/2000 e 21/01/2000 e que estão dentro da região selecionada no mapa?".
- viii. **Célula F8** As consultas deste tipo são formuladas textual e graficamente, semelhante às descritas anteriormente, e os resultados são exibidos graficamente. Por exemplo, "Exibir a distribuição das coletas realizadas por Ana Maria Felgar a menos de 5 km do ponto selecionado", "Exibir a distribuição das coletas realizadas pelos membros da instituição *X* dentro da região definida pelo usuário" e "Exibir a distribuição das espécies encontradas dentro da bacia selecionada no período de 11/01/2000 a 21/01/2000".
- ix. **Célula F9** A consulta "Exibir a distribuição das espécies coletadas entre 11/01/2000 e 21/01/2000 e que estejam contidas na bacia selecionada, juntamente com uma legenda do mapa" é um exemplo de consulta do tipo **F9**. Esta célula corresponde a consultas onde o usuário especifica restrições textual (entre 11/01/2000 e 21/01/2000) e gráfica (bacia selecionada) e os resultados retornados são textual (legenda) e gráfico (mapa com a distribuição das espécies).

5.3 **Resumo**

Este capítulo discutiu o conjunto de operadores espaciais necessários para a processamento das consultas espaciais do projeto SINBIOTASP. Além disso, foram sistematizados: as fontes de dados requeridas pelas consultas, os tipos de objetos espaciais, as operações

primitivas necessárias, bem como os operadores espaciais disponíveis e passíveis de uso. Em adição, a interface com o usuário para a entrada de critérios de seleção, as formas de interação e os formatos de visualização dos resultados também foram discutidos.

O próximo capítulo descreve o módulo *Species Mapper* de processamento de consultas de distribuição de espécies, cuja implementação foi parcialmente baseada nesta sistematização.

Capítulo 6

Aspectos de Implementação

Este capítulo descreve os aspectos de implementação do módulo desenvolvido a partir das análises realizadas nos capítulos 3 e 5. Como contribuição adicional, propõe uma alternativa para a atual arquitetura Web do SINBIOTASP. É importante lembrar que esta dissertação está voltada ao processamento de consultas. Desta forma, os aspectos referentes ao projeto do banco de dados e modelagem ambiental fogem ao escopo deste trabalho, havendo sido discutidos na dissertação de Fagundes [Fag99].

A implementação desenvolvida como parte desta dissertação teve em vista os aspectos descritos no capítulo 5 referentes ao processamento das consultas. As fontes de dados envolvidas no processamento da consulta são: banco de coletas, bancos de coletas e de metadados e bancos de coletas, de metadados e de mapas. Os objetos espaciais utilizados são do tipo ponto e região. Além disso, não foram utilizados operadores espaciais e foi possível implementar a execução de transações diversas (conversão de formato e operação de visualização de pontos e regiões). As consultas passíveis de processamento são classificadas como dos tipos C1, C3, C12 e C21. Quanto ao ponto de vista de interface de consulta, o módulo disponibiliza consultas por campo específico e permite uma visualização tanto textual quanto gráfica dos resultados.

Este capítulo está organizado da seguinte forma. A seção 6.1 descreve o sistema *Species Mapper* desenvolvido como parte do trabalho da dissertação. A seção 6.2 apresenta a arquitetura Web proposta para o SINBIOTASP. A seção 6.3 apresenta um resumo deste capítulo.

6.1 O *Species Mapper*

O *Species Mapper* é o módulo do SINBIOTASP desenvolvido como parte do trabalho desta dissertação. Seu desenvolvimento foi norteado pelo estudo de outros sistemas, descrito no capítulo 3 e pelas necessidades dos usuários do SINBIOTASP. O módulo é responsável

pelo processamento de consultas referentes a espécies, tais como informações textuais referentes às coletas e informações gráficas que permitem visualizar a distribuição da espécie sobre um mapa. O sistema foi desenvolvido usando CGI, PERL, Borland Delphi 5.0 e Map Objects da ESRI.

Sob o prisma do estudo realizado no capítulo 5, a interface provê entrada de dados somente textual, enquanto que os formatos disponíveis de visualização da saída podem ser tanto textual quanto gráfico. A figura 6.1 mostra o diagrama das funcionalidades implementadas para a realização de consultas textuais, bem como as funcionalidades necessárias para o desenvolvimento futuro de consultas gráficas e mistas. A figura mostra que uma consulta é inicialmente formulada a partir de um conjunto de critérios de seleção (predicado). Estes critérios são formatados pelo sistema em um conjunto de comandos em SQL que são enviados ao SGBD Oracle e em Map Objects, enviados ao SIG ArcInfo. O acesso ao *Species Mapper* está disponível na Web a partir do endereço http://www.biotasp.org.br/sia/espacial/sp_mapper.

Após processamento, o resultado é exibido conforme especificação do usuário, sob forma textual, gráfica ou mista. A figura também mostra que o resultado de uma consulta ao *Species Mapper* pode conter dados de coletas (fonte de dado mista), taxa (fonte de dado convencional) e mapas de distribuição de coletas/espécies (fonte de dado mista).

A tabela 6.1 exemplifica o formato utilizado para documentar as funções e procedimentos do *Species Mapper*. Basicamente, a documentação de cada função aponta para o módulo correspondente, descreve o objetivo da função, suas entradas e saídas, as variáveis globais utilizadas e adiciona observações gerais.

A seguir, o texto descreve a arquitetura do *Species Mapper* e apresenta um exemplo de sessão mostrando as funcionalidades disponibilizadas pelo sistema.

6.1.1 Arquitetura do sistema

A arquitetura do sistema *Species Mapper*, mostrada na figura 6.2, contempla dois tipos de processamento: convencional e espacial. A função dos módulos mostrados na arquitetura já foi descrito no capítulo 4 (processamento da arquitetura Web).

Processamento convencional O processamento convencional se aplica no *Species Mapper* quando uma consulta possui critérios convencionais e o resultado textual da consulta é gerado sem a necessidade da utilização de operações e operadores espaciais. As consultas, em SQL, são executadas somente sobre o banco de coletas ou sobre o banco de coletas juntamente com o banco de metadados.

Esta parte do sistema utiliza módulos escritos em PERL, já que esta é a forma de processamento convencional atualmente disponível na BDT. A solicitação do usuário é

Função	Sp_mapper_formata_where_gen_esp
Módulo	/system/webiota/docs/sia/espacial/sp_mapper
Descrição	Define a cláusula <i>Where</i> diante dos critérios textuais informados.
Entrada	busca <i>string</i> ou <i>substring</i> do nome da espécie pesquisada pelo usuário; reino nome do reino a que busca pertence.
Saída	Cláusula <i>Where</i> de uma consulta SQL com os critérios de seleção devidamente formatados.
Globais	Nula
Observação	Esta função retorna uma expressão em SQL que irá buscar as espécies cujo nome seja igual a busca e que pertençam ao reino reino . Além disso, ela avalia se a pesquisa será realizada somente a partir do campo gênero ou sobre gênero e espécie.

Tabela 6.1: Exemplo da documentação de uma função desenvolvida para o *Species Mapper*

formatada pelo *Cliente HTML*, e enviada ao *Cliente PERL*, que formata uma consulta SQL enviada ao Oracle. O resultado da consulta segue o caminho inverso, sendo formatado pelo *Cliente PERL* e exibido pelo *Cliente HTML*.

Processamento espacial O processamento espacial se aplica no *Species Mapper* quando o resultado é gerado a partir da utilização de operadores espaciais. Este processamento é executado sobre os bancos de coletas, de metadados e de mapas.

Os bancos de dados de coletas e metadados são pesquisados através da execução de cláusulas SQL. Por exemplo, a consulta "Quais os nomes das espécies cujo gênero começa com *Ac* e que pertençam ao reino *Plantae*?" é executada através da seguinte cláusula SQL.

```
select distinct (initcap(g.genero) || ' ' || lower(sp.especie)) as Especie, r.reino as Reino
from taxa t, especie sp, genero g, reino r
where (t.cod_especie = sp.cod_especie)
and (t.cod_genero = g.cod_genero)
and (t.cod_reino = r.cod_reino)
and (r.cod_reino = "Plantae")
and (lower(g.genero) like 'Ac%')
order by Especie;
```

O acesso ao banco de mapas, por sua vez, é realizado pelo Map Objects da ESRI através do Delphi 5.0. O procedimento *AddLayer*, descrito a seguir, adiciona a camada (tema) especificado pelo usuário (*strFilename*) ao mapa (*Map*) que será posteriormente visualizado pelo usuário. O tema adicionado é formatado pelos seguintes parâmetros de entrada do procedimento: *CoordType* e *CoordSysType* (sistema de coordenadas especificado segundo

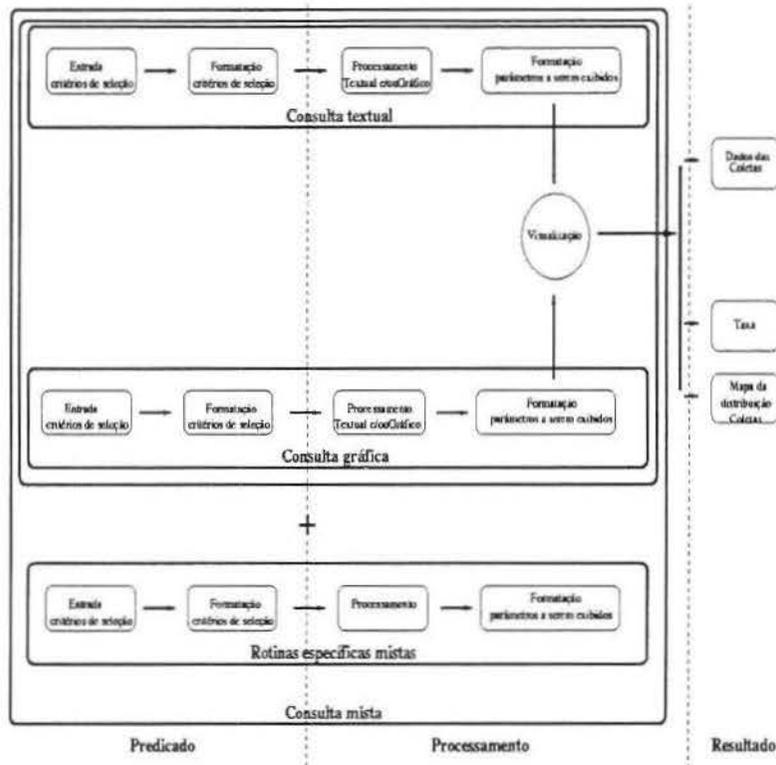


Figura 6.1: Diagrama das funcionalidades implementadas para o *Species Mapper*

o padrão Map Objects), *color* (cor), *style* (estilo de apresentação) e *size* (largura do objeto espacial).

```

procedure AddLayer(Map: TMap; strPath, strFilename: string; CoordType: string; CoordSysType: integer; color, style,
size: integer);
var
dc: IMoDataConnection; layer: IMoMapLayer; ProjCoordSys: IMoProjCoordSys; GeoCoordSys: IMoGeoCoordSys;
begin
{Estabelece conexão com o Map Objects.}
dc := IMoDataConnection(CreateOleObject('MapObjects2.DataConnection'));
dc.Database := strPath;
{Cria um objeto do tipo camada para armazenar as informações recuperadas da camada especificada pelo usuário.}
layer := IMoMapLayer(CreateOleObject('MapObjects2.MapLayer'));
{Busca a camada com o nome fornecido pelo usuário.}
layer.GeoDataset := dc.FindGeoDataSet(strFilename);
layer.Symbol.Color := color;
layer.Symbol.Style := style;
layer.Symbol.Size := size;
if CoordType = 'proj' then
begin
ProjCoordSys := IMoProjCoordSys(CreateOleObject('MapObjects2.ProjCoordSys'));
ProjCoordSys.Type := CoordSysType;
layer.CoordinateSystem := ProjCoordSys;
end
else if CoordType = 'geo' then
begin

```

```

GeoCoordSys := IMoGeoCoordSys(CreateOleObject('MapObjects2.GeoCoordSys'));
GeoCoordSys.Type_ := CoordSysType;
layer.CoordinateSystem := GeoCoordSys;
end;
{Formata o mapa com a nova camada.}
Map.Layers.Add(layer);
end;

```

Voltando à figura 6.2, o fluxo de controle e dados é o seguinte. A interface do usuário é provida por um módulo denominado *Cliente HTML*, que passa as solicitações ao módulo *Cliente PERL*. Este, por sua vez, aciona o módulo *Map Client* que realiza uma consulta espacial no SIG e devolve seu resultado ao *Cliente PERL*. A conexão entre o *Cliente PERL* e o *Map Client* é gerida por dois outros módulos, denominados *Map Server* e *Map Statistic*.

Considere a consulta "Exibir a distribuição das coletas onde foram encontradas *Acacia glomerosa*". Inicialmente, o cliente (*Cliente HTML*) solicita a execução do processamento espacial ao *Cliente PERL*. Depois, o servidor (através do *Cliente PERL*) solicita a instanciação de um *Map Client* para o processamento da consulta espacial. O módulo *Map Statistic* seleciona um *Map Client* disponível, que é instanciado pelo *Map Server*. O *Map Client* gera a consulta ao ArcInfo (Map Objects). O resultado desta consulta é uma imagem JPEG que é retornada para o *Map Server*. O *Map Server*, por sua vez, envia esta imagem para o *Cliente PERL* para a geração de uma página HTML contendo esta imagem. Por último, o *Cliente HTML* exibe a página. O procedimento *MapAfterLayerDraw*, descrito a seguir, desenha os pontos correspondentes à distribuição da *Acacia glomerosa* sobre o mapa composto pelos temas definidos.

```

procedure TfrmAtlasBiotaSP.MapAfterLayerDraw(Sender: TObject; index: Smallint; canceled: WordBool; hDC: Cardinal);
var
point: IMoPoint; sym: IMoSymbol; fids: IMoFields; fid: IMoField; s: IMoPoint; i: integer;
begin
if session.N_SpeciesLayer = 0 then exit;
if index = 0 then
begin
i := 1;
{Repetir enquanto existir uma espécie cuja distribuição não tenha sido exibida sobre o mapa.}
while (i := session.N_SpeciesLayer) do
begin
sym := IMoSymbol(CreateOleObject('MapObjects2.Symbol'));
sym.SymbolType := moPointSymbol;
//define simbolo.
if session.SpeciesLayer[i].symbol = '-1' then
sym.Style := StrToInt(session.SpeciesSymbol[i])
else
sym.Style := StrToInt(session.SpeciesLayer[i].symbol);
//define cor.
if session.SpeciesLayer[i].color = -1 then
sym.color := session.SpeciesColor[i]
else
sym.color := session.SpeciesLayer[i].color;

```

```

//define tamanho do simbolo.
sym.size := 3 + session.Definesize(i);
if Not RecSet[i].Eof then RecSet[i].MoveFirst;
  point := IMoPoint(CreateOleObject('MapObjects2.Point'));
  {Repetir enquanto existir coletas (pontos) correspondentes a espécie RecSet[i] que não tenham sido
  desenhadas sobre o mapa.}
  while not RecSet[i].eof do
    begin
      fids := IMoFields(RecSet[i].fields);
      fld := IMoField(fids.item('Shape'));
      s := IMoPoint(IDispatch(fld.value));
      Map.DrawShape(s, sym);
      RecSet[i].MoveNext;
    end;
  Inc(i);
end;
end;
end;

```

Todos os componentes, cliente e servidor, desta arquitetura, exceto o *Map Server* e o *Map Statistic*, foram desenvolvidos como parte da dissertação. O componente *Map Client* é também utilizado por outro sistema desenvolvido pela BDT - o Atlas Biota/SP versão 2.0 - atualmente em manutenção (novembro 2000), descrito no capítulo 4.

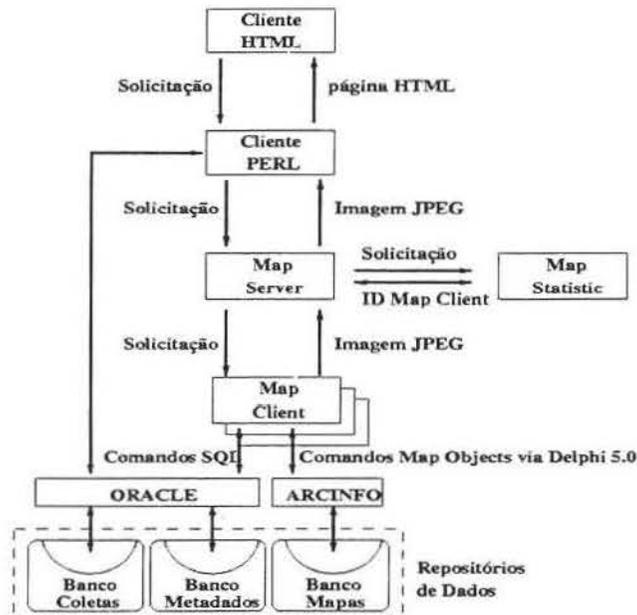


Figura 6.2: Arquitetura do *Species Mapper*

6.1.2 Exemplo de sessão

Esta subseção apresenta exemplo de uma sessão no sistema *Species Mapper*.

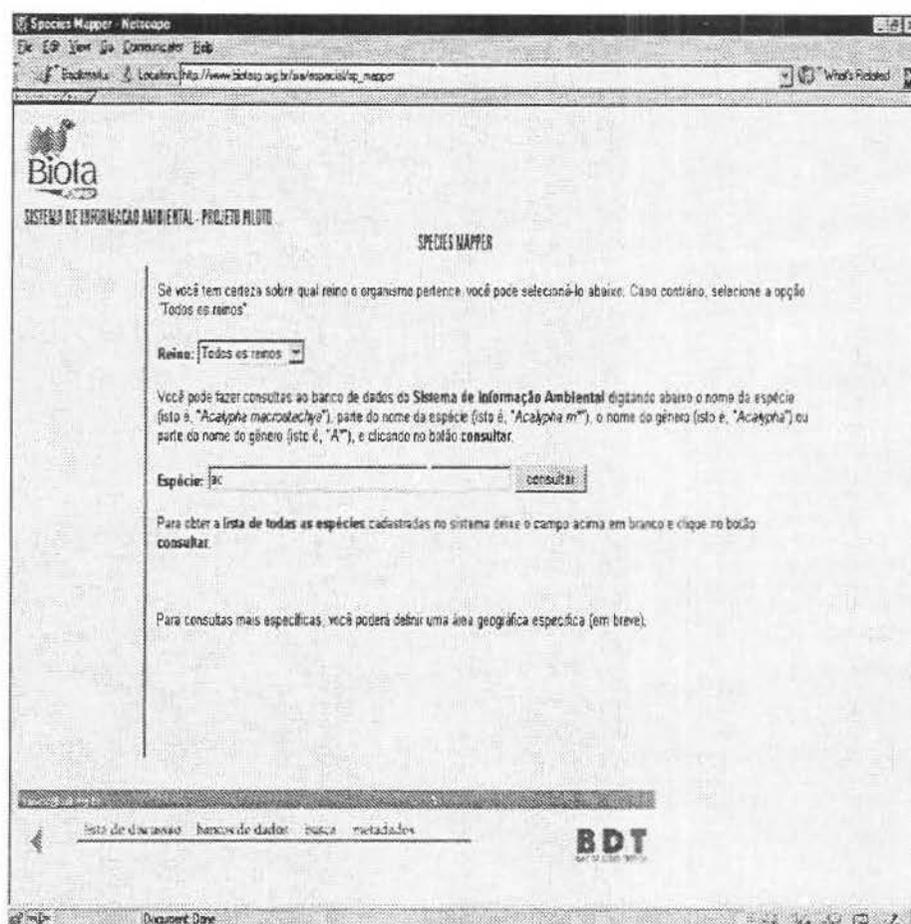


Figura 6.3: Consulta textual por campo específico do *Species Mapper*

Suponha um usuário que pretende realizar uma série de consultas referentes às coletas. Inicialmente, o usuário solicita o nome de espécies iniciado com *Ac* para espécies de todos os reinos. A figura 6.3 mostra como esta consulta é formulada. O resultado da consulta é mostrado na figura 6.4. Os nomes iniciados com *Ac* são exibidos em forma de lista. Os *links* [coleta] e [mapa] permitem, respectivamente, visualizar as informações das coletas de cada espécie e a distribuição destas sobre o mapa. O código implementado procura pelo termo informado no banco de coletas; caso o sistema não obtenha nenhum resultado, ele tentará encontrar o termo mais próximo do informado. O conteúdo destes *links* é gerado dinamicamente em função da espécie e do objetivo do *link* (acesso aos bancos de coletas e de metadados ou aos bancos de coletas, de metadados e de mapas). Desta forma, o acionamento de um *link* corresponde, a grosso modo, à execução de uma consulta estruturada dinamicamente sobre fonte de dado convencional (banco de coletas) e mista (banco de metadados e de mapas).

Assim, se o usuário quiser visualizar algumas informações textuais referentes às coletas,

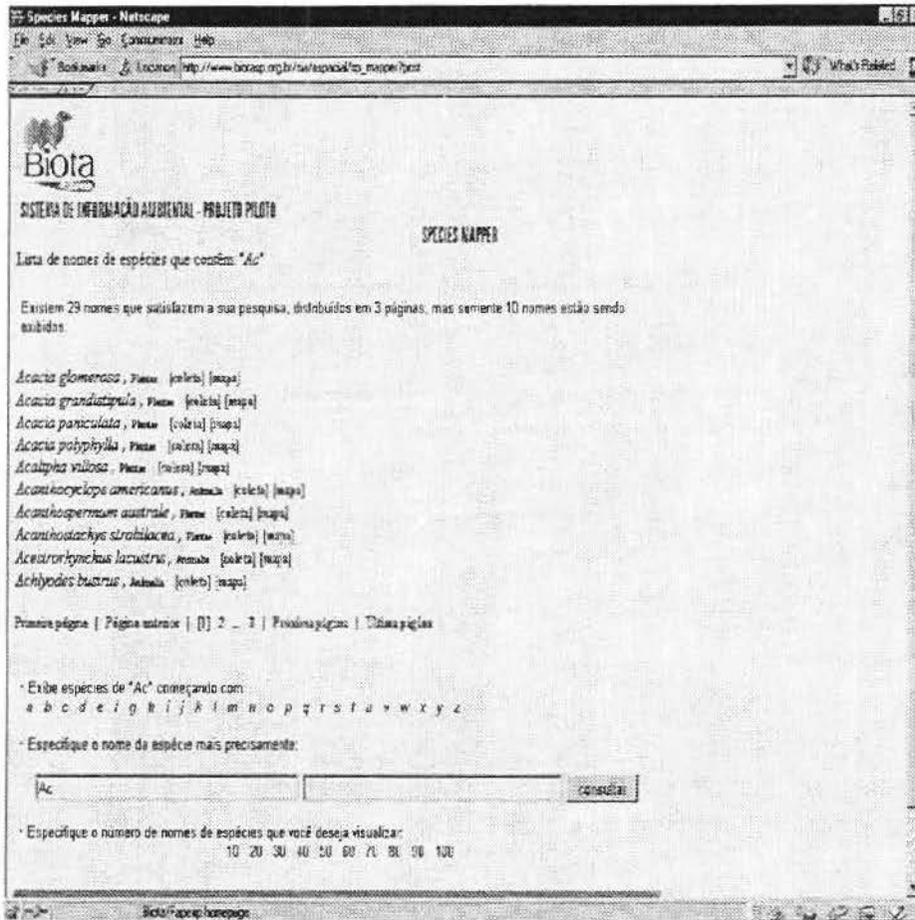


Figura 6.4: Resultado da consulta por campo específico - visualização textual

ele pode selecionar o *link* [coleta]. Exemplos são os campos: código de identificação da coleta, usuário, data, autor, método de coleta utilizado, grupo taxonômico e localização. Este *link* permite obter uma visão resumida e simplificada dos dados de uma coleta. A figura 6.5 mostra o que acontece se o usuário clica o *link* [coleta] relativo a *Acacia paniculata* (terceira na lista da figura 6.4). A figura 6.5 mostra, por exemplo, que a coleta de código 750 foi feita em 8/10/99 e contém espécies do grupo taxonômico *Angiospermae* (ao qual pertence a *Acacia paniculata*). Os campos data da coleta, método da coleta, grupo taxonômico e localização correspondem a dados do banco de metadados e os demais são provenientes do banco de coletas.

Continuando sua sessão, suponha que o usuário seleciona a coleta 750 para exibição, pressionando o botão *exibir* (coluna da direita, figura 6.5). Como resultado, aparece a figura 6.6. Nesta última, os campos coletor, unidade de conservação, ambiente, bacia hidrográfica, outras observações, conteúdo e os dados sobre o autor da coleta correspondem a dados do banco de coletas e os demais são metadados.

Species Mapper - Netscape
 Location: http://www.biota.org.br/sia/especial/sp_mapper?coleta=Acacia+paniculata

Biota
 SISTEMA DE INFORMAÇÃO AMBIENTAL - PROJETO PILOTO
 SPECIES MAPPER

Espécie: *Acacia paniculata* [mapa]

Código/ Usuário	Data	Autor/ Método	Grupo	Localização	Oper.
<input type="checkbox"/> 750 marinez	08/10/1999 a 08/10/1999	Giselda Dungen Levantamento Botânico Expedite	Angiospermae	São Pedro do Turvo, SP	abr
<input type="checkbox"/> 751 marinez	07/10/1999 a 07/10/1999	Giselda Dungen Levantamento Botânico Expedite	Angiospermae	São Pedro do Turvo, SP	abr
<input type="checkbox"/> 749 marinez	08/10/1999 a 08/10/1999	Giselda Dungen Levantamento Botânico Expedite	Angiospermae	São Pedro do Turvo, SP	abr
<input type="checkbox"/> 754 marinez	09/11/1999 a 09/11/1999	Giselda Dungen Levantamento Botânico Expedite	Angiospermae	Ubirajara, SP	abr

Document Done

Figura 6.5: Visualização textual resumida das coletas de uma determinada espécie no *Species Mapper*

A seguir, suponha que o usuário queira visualizar a distribuição da espécie "*Acacia paniculata*" sobre o território do Estado de São Paulo. Retornando à tela da figura 6.4, e acionando o *link* [mapa] associado ao nome da espécie desejada, obtém-se o resultado mostrado na figura 6.7 que ilustra a visualização gráfica da distribuição da espécie "*Acacia paniculata*".

O processamento desta consulta é dividido nas seguintes etapas:

- i. Primeiramente, é realizada uma consulta interrelacionando as informações armazenadas no banco de coletas e metadados com o intuito de obter a localização (banco de metadados) da espécie desejada (banco de coletas). Ao mesmo tempo, recupera o identificador das coletas, que estão relacionados a pontos já cadastrados no banco de mapas;
- ii. A segunda etapa recupera o mapa sobre o qual será exibida a distribuição da espécie

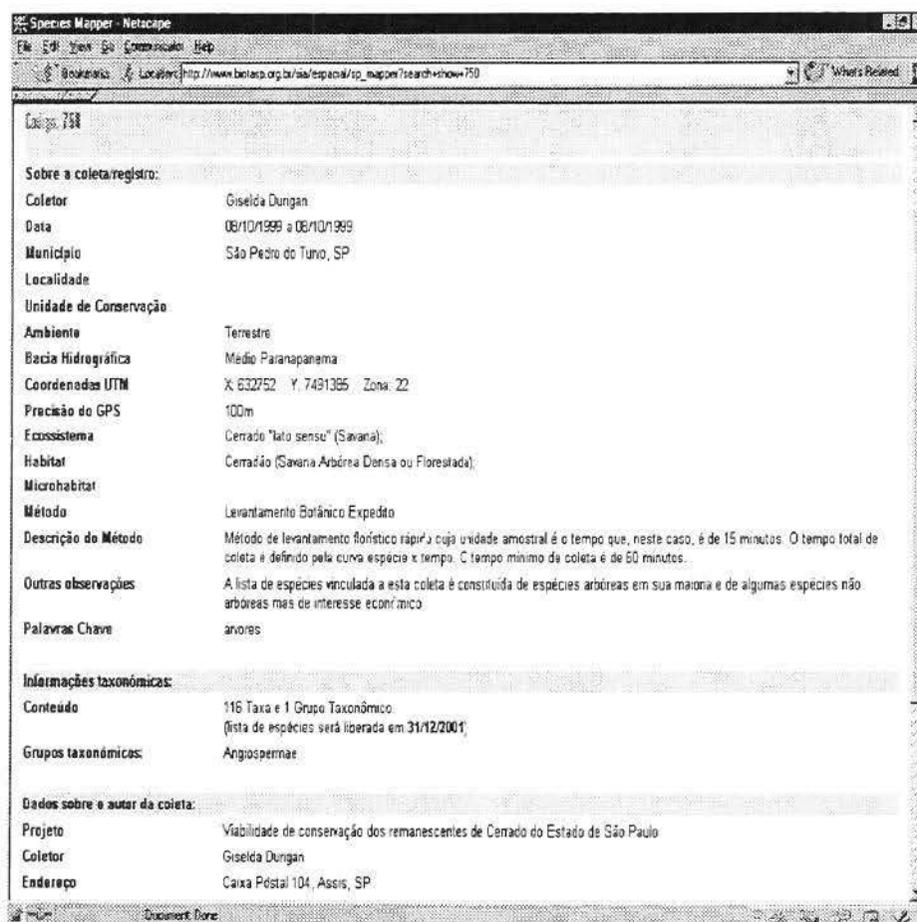


Figura 6.6: Visualização textual detalhada das informações referentes a uma coleta no *Species Mapper*

desejada, aproveitando também a ligação com as coletas via identificador. O banco de metadados indica quais mapas do banco de mapas devem ser recuperados, bem como as características de cada mapa recuperado;

- iii. A última etapa diz respeito à especificação sobre o mapa dos pontos que representam a localização da espécie desejada através do interrelacionamento dos resultados das etapas anteriores.

É importante analisar também a sessão do usuário sob o prisma de questões referentes ao processamento da consulta, tratadas no capítulo 5, tais como: fontes de dados envolvidas no processamento, tipos de objetos espaciais manipulados, operadores espaciais e operações primitivas utilizadas e classificação da consulta.

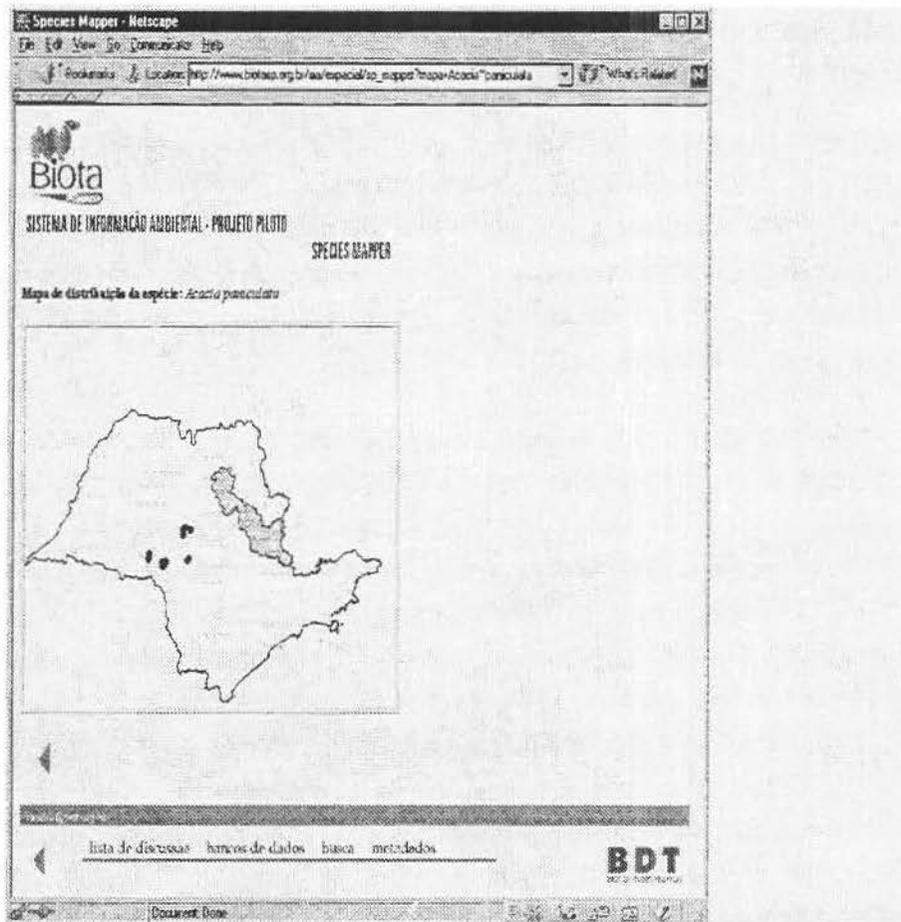


Figura 6.7: Visualização gráfica da distribuição de uma determinada espécie no *Species Mapper*

Fontes de dados As fontes de dados envolvidas no processamento das consultas no *Species Mapper* são (vide para detalhes as figuras 5.1 e 5.2 referentes aos esquemas dos bancos de coletas e de metadados, respectivamente):

- Consultas diretamente ao banco de coletas. Este tipo de consulta não necessita acessar o conjunto de dados contidos no banco de metadados nem no banco de mapas. Envolvem dados de nome de espécie, gênero, espécie e reino. A figura 6.3 mostra um exemplo deste tipo de consulta correspondendo à solicitação de nomes de espécies coletadas iniciando por *Ac*.
- Consultas feitas aos bancos de coletas e metadados. Neste tipo de consulta, o processamento pode passar por um primeiro filtro (de metadados). A figura 6.5 corresponde ao resultado da consulta "Quais as coletas onde foram encontradas *Acacia polyphylla*?". Este tipo de consulta envolve metadados, pois alguns dados

exibidos na figura 6.5 são provenientes do banco de metadados, são eles: a data da coleta, o método de coleta, grupo taxonômico e localização (município).

- Consultas feitas aos bancos de coletas, metadados e mapas. A figura 6.7 corresponde ao resultado da consulta "Exibir a distribuição das coletas onde foram encontradas *Acacia paniculata*", executada quando o usuário clica o *link* [mapa] relativo a *Acacia paniculata* (terceira na lista da figura 6.4). Este tipo de consulta envolve metadados (localização e identificador da coleta), coletas (nome da espécie e identificador da coleta) e mapas (temas selecionados).

O *Species Mapper* não possui um mecanismo de consulta que consulta diretamente o banco de metadados, embora este tipo de consulta seja possível de ser implementado.

Objetos espaciais A consulta "Exibir a distribuição das coletas onde foram encontradas *Acacia paniculata*" executada quando o usuário clica o *link* [mapa] relativo a *Acacia paniculata* (terceira na lista da figura 6.4) envolve objetos espaciais. O resultado desta consulta é exibido na figura 6.7 e envolve objetos espaciais do tipo ponto (localização da coleta/espécie) e região (divisa do Estado de São Paulo e bacia do Mogi-Mirim).

Operadores espaciais e operações primitivas O *Species Mapper* não utiliza operadores espaciais. Por outro lado, o processamento da figura 6.7 requer a execução de *transações diversas* referentes à conversão de formato e operação de visualização de pontos (localização da coleta/espécie) e regiões (divisa do Estado de São Paulo e bacia do Mogi-Mirim).

Classificação da consulta A classificação das consultas processadas no *Species Mapper* é realizada em função do tipo de fonte de dado, predicado e resultado.

A figura 6.3 mostra um exemplo de consulta do tipo **C1**, visto que a fonte de dado envolvida é convencional (banco de coletas), o predicado é convencional (seleção por campo específico) e o resultado é convencional (nome de espécies). A figura 6.4, por sua vez, permite especificar dois tipos diferentes de consulta. O primeiro tipo executado através do *link* [coleta] é do tipo **C3**, pois a fonte de dado envolvida é mista (banco de coletas e de metadados), o predicado é convencional (seleção por campo específico) e o resultado é convencional (data de coleta, autor, método de coleta, grupo taxonômico, dentre outros). Já o segundo tipo, executado através do *link* [mapa] é do tipo **C12**, visto que a fonte de dado envolvida é mista (banco de coletas, de metadados e de mapas), o predicado é convencional (seleção por campo específico) e o resultado é espacial (mapa com a distribuição de coleta/espécie).

Por último, a figura 6.5 permite executar, pressionando o botão exibir (coluna da direita da figura 6.5), uma consulta do tipo **C21** visto que a fonte de dado é mista (banco de coletas e metadados), o predicado é convencional (seleção por campo específico) e o resultado é misto (autor, localização da coleta - coordenadas, método de coleta, dentre outros). O resultado desta consulta é exibido pela figura 6.6.

6.2 Arquitetura Web proposta para o SINBIOTASP

Em um ambiente Web para manipulação de bancos de dados e geoprocessamento é desejável que as páginas HTML sejam geradas dinamicamente em uma aplicação integrada ao servidor Web. Existem vários mecanismos para realizar a integração entre essa aplicação, o cliente e o servidor, como já foi visto no capítulo 2.

Dentre os mecanismos avaliados, *applets* e *servlets* são as alternativas recomendadas para o desenvolvimento futuro do sistema de informação ambiental do programa BIOTA/FAPESP, em função das vantagens apresentadas na subseção 2.5.4 e da utilização destes mecanismos por outros sistemas de informação ambiental, como por exemplo, UDK [UDK], GCMD [GCM] e FGDC [FGD]. Desta forma, a dissertação propõe uma nova arquitetura Web para o processamento de consultas no SINBIOTASP, contemplando dois tipos de processamento: textual (convencional) e espacial.

Processamento espacial O processamento espacial se aplica quando uma consulta utiliza operações espaciais. A implementação proposta utiliza *applet*, *servlet*, Borland Delphi 5.0 e Map Objects da ESRI. No caso de SIG-Web, o *servlet* fica localizado no servidor, sendo responsável por manter uma conexão com o banco de dados (banco de coletas e de metadados) e com o *applet*. O protocolo utilizado pelo *servlet* para comunicação com os bancos de coletas e de metadados é o padrão SQL. O acesso ao banco de mapas, por sua vez, é realizado pelo Map Objects da ESRI através do Borland Delphi 5.0.

Este processamento está baseado em um modelo de três camadas. A primeira camada é um navegador para Internet que serve como um Cliente. Acoplado ao navegador há um *applet* Java. O *applet* é responsável por funções ligadas ao cliente, tais como apresentação de mapas e objetos geográficos, controle de diálogos e interfaces e o controle de cache de dados.

A segunda camada é constituída por um servidor para protocolos HTTP com capacidade de executar *servlets* em Java. O *servlet* é encarregado de interpretar as solicitações dos clientes, acessar os bancos de dados, executar os comandos necessários, gerar dinamicamente as páginas HTML para a publicação dos dados e enviar a resposta adequada de volta ao cliente. Desta forma, o *servlet* exerce um papel de mediador entre o cliente e os gerenciadores de mapas (SIG) e dados (SGBD).

A terceira camada é composta por um SGBD Relacional (Banco de Coletas e de Metadados) que armazena todos os dados utilizados pelo sistema e um SIG (Banco de Mapas). O *servlet* acessa as informações do banco de dados relacional utilizando a API JDBC da linguagem Java e do banco de mapas utilizando alguns dos componentes desenvolvidos para o processamento espacial no sistema *Species Mapper* (subseção 6.1.1). Os componentes utilizados são: *Map Server*, *Map Statistic* e o *Map Client*. No entanto, o comportamento do *Map Server* apresenta as seguintes diferenças: o *Map Server* estabelece uma comunicação entre o *servlet* e o *Map Client* (mediante solicitação do *servlet*) e o *Map Server* é responsável por enviar a imagem JPEG gerada pelo *Map Client* para o *servlet*.

Considere a seguinte consulta do tipo C12: "Exibir a distribuição das coletas onde foram encontradas *Acacia paniculata*". Inicialmente, o cliente (*applet*) solicita a execução do processamento espacial ao *servlet*. Depois, o servidor (através do *servlet*) solicita a instanciação de um *Map Client* para o processamento da consulta espacial. O *Map Statistic* seleciona um *Map Client*, que é instanciado pelo *Map Server* e executado, retornando uma imagem JPEG para o *Map Server*. O *Map Server* envia a imagem JPEG obtida como resultado para o *servlet* para a geração de uma página HTML como resultado. Por último, o cliente (*applet*) exibe a página gerada dinamicamente através da consulta aos bancos de dados.

Processamento convencional O processamento convencional ocorre quando uma consulta processada é gerada sem a necessidade da utilização de operações espaciais. Esta arquitetura é executada somente sobre os bancos de coletas e/ou de metadados.

Neste processamento, os papéis do cliente e do servidor são simplificados e executados da seguinte forma:

- Ao cliente cabe capturar os critérios de seleção através de uma interface desenvolvida em HTML e JavaScript, e solicitar a execução do *servlet* ao servidor.
- Ao servidor cabe a execução do *servlet*, cujo resultado é transmitido ao cliente. O *servlet* interpreta as solicitações, gera uma página HTML dinamicamente através da consulta aos bancos de coletas e de metadados e envia a página-resposta para ser exibida no navegador.

O protocolo utilizado pelo *servlet* para comunicação com o banco de dados relacional é o padrão SQL. A conexão é efetuada através da chamada de rotinas pertencentes à API JDBC. O servidor carrega o *servlet* que pode então aceitar diversas requisições dos clientes e retornar dados a eles. A inicialização de um *servlet* é feita apenas uma vez, ou seja, após ele ser carregado pelo servidor Web, as novas solicitações são apenas chamadas do método de serviço.

6.3 Resumo

Este capítulo apresentou duas das contribuições desta dissertação. Primeiramente, foram descritos alguns aspectos da implementação do sistema *Species Mapper* desenvolvido na BDT para consulta sobre as espécies coletadas. Esta implementação foi desenvolvida utilizando HTML, JavaScript, *script CGI* escrito em PERL, Delphi 5.0 e Map Objects. Algumas funções desenvolvidas para o *Species Mapper* foram utilizadas por um sistema desenvolvido na BDT denominado Atlas Biota/SP versão 2.0, descrito no capítulo 4.

Em seguida, o capítulo descreveu a arquitetura Web proposta como recomendável para uma futura arquitetura do SINBIOTASP, definida a partir da análise realizada no capítulo 2. *Applets* e *servlets* são as alternativas recomendadas para o desenvolvimento do sistema de informação, em função das vantagens apresentadas na subseção 2.5.4.

O próximo capítulo apresenta as conclusões e contribuições desta dissertação e propõe alguns trabalhos futuros.

Capítulo 7

Conclusões e extensões

7.1 Contribuições

Esta dissertação abordou o processamento de consultas no Sistema de Informação de biodiversidade do Estado de São Paulo. Este tipo de processamento envolve combinar dados textuais e cartográficos e fornecer uma apresentação dos dados tanto textual quanto visual. Isto acarreta uma série de problemas de programação, agravados pela heterogeneidade dos dados.

Alguns aspectos importantes de SIG e conceitos básicos de diferentes grupos de operações espaciais primitivas em um SIG de acordo com Cereja [Cer96] e Ciferri [Cif95] foram apresentados, bem como os operadores espaciais básicos, segundo Faria [Far98] e Guting [Gut94]. Em adição, foram discutidas questões referentes a Sistemas de Informação Geográfica na Web, tais como: arquitetura, interoperabilidade e SIG, o uso de metadados, exemplos de sistemas e as principais alternativas de linguagens para desenvolvimento na Web.

Seguiu-se uma análise de aspectos relevantes de sistemas de informação ambiental [SLF+98, Uni, LP98, UDK, CDS, GCM, ERI, EOS, FGD, GIL, Spe, Mal]. Esta análise permitiu concluir que não existe um modelo padrão o qual todo sistema de informação ambiental tenha que seguir, mas sim que existem funções em comum entre estes sistemas, que podem ser adotadas.

Este estudo de consultas e interfaces de sistemas de biodiversidade permitiu identificar os principais componentes de consultas em sistemas de biodiversidade, e por conseguinte determinar a base para formulação de consultas no SINBIOTASP. A partir desta análise, foi desenvolvido um sistema de consultas, apresentado no capítulo 6. O mesmo capítulo também apresenta sugestões para uma futura arquitetura do SINBIOTASP no que tange o uso da Web.

As principais contribuições da dissertação foram:

- a. levantamento de características de diversos sistemas de informação ambiental disponíveis, tais como: objetivo, dados gerenciados e recuperados, arquitetura quando disponível, tipos de consulta e interfaces de exibição disponíveis;
- b. análise detalhada de tipos de consultas comuns em projetos de biodiversidade e sua formulação em termos de bancos de dados espaciais (e não em termos do usuário);
- c. sistematização das características básicas das consultas, envolvendo mapas e dados textuais, no contexto da informação ambiental para a biodiversidade;
- d. desenvolvimento do sistema *Species Mapper* para o BIOTA que permite consultas integradas sobre alguns aspectos de biodiversidade, a partir do banco de dados já consolidado.

7.2 Extensões

Diversas extensões podem ser propostas a esta dissertação. Dentre as futuras direções possíveis desta pesquisa, destacamos:

Desenvolvimento de técnicas que levem à otimização das consultas realizadas As consultas que mesclam características de predicado convencional e predicado espacial podem ser respondidas de duas maneiras: optando-se primeiramente pela pesquisa do predicado convencional, ou pela pesquisa do predicado espacial. Esta escolha é uma extensão desta dissertação e deve ser efetuada a partir de módulos adicionais que façam as vezes de otimizador de consultas. Uma escolha errada pode afetar significativamente o desempenho, uma vez que estes predicados atuam sobre diferentes tamanhos e tipos de conjuntos de dados. Ainda neste contexto, deve ser feita uma análise do uso de índices espaciais, para os predicados espaciais, de novo exigindo um conjunto de módulos dedicados.

Incorporação de técnicas de processamento de imagens Uma outra classe de consultas não explorada é aquela que envolve acesso a imagens (por exemplo, fotos de animais coletados, armazenadas como documentos). Isto envolve aspectos de armazenamento e acesso, pois devido ao grande tamanho da imagem e à banda limitada, há o problema de deterioração da taxa de resposta. As soluções existentes em geral esquecem que os usuários precisam somente de um pequeno pedaço da imagem total em um momento do tempo. Uma solução é cortar a imagem fonte em outras menores, criando um repositório de imagens que pode ser usado para uma aplicação específica (por exemplo, via segmentação).

Incorporação de novos tipos de interação do usuário e de operadores espaciais

Continuação do desenvolvimento do sistema de consultas para incorporar todas as facilidades discutidas no capítulo 5. Atualmente, o módulo de consulta está limitado por questões de cronograma, de qualidade de dados e forma de trabalho do usuário.

Formato do arquivo de resposta do usuário Atualmente, o sistema utiliza arquivos no formato JPEG para exibir o resultado de consultas que demandam mapas. No entanto, existem diversos outros formatos de arquivo possíveis que devem ser avaliados.

Para Sorokine *et al.* [SM98], do ponto de vista do programador, existem três tipos de solução para o mapeamento WWW destas imagens: imagem GIF com formulários HTML ou JavaScript, *applets* Java e controles ActiveX. Contudo, segundo Sorokine *et al.* [SM98], cada uma destas soluções apresenta alguns problemas. O uso de imagem GIF, a mais popular e simples, é inapropriado devido à frequente necessidade de recarregamento mesmo para operações simples, tais como *pan* e *zoom*. *Applets* JAVA e controles ActiveX têm desvantagens em função do grande tamanho do *download* necessário, código e dados, que sobrecarrega o sistema do cliente devido à demanda de recursos de memória e CPU. O controle ActiveX está limitado à plataforma Windows95/NT. Sorokine *et al.* [SM98] estudam apenas imagens GIF desconsiderando outras imagens, como por exemplo JPEG.

Utilização de um arquivo vetorial para resposta a consultas espaciais

A utilização de arquivo vetorial permite seleção de objetos, análise espacial do resultado, operações de visualização no cliente sem acessos ao servidor e geração de arquivos menores do que a maioria dos outros formatos disponíveis. Este tipo de solução exigiria no entanto outras opções de processamento e envio de dados.

Uma hipótese seria usar o TWF (*Tecgraf Web Format*) um formato de armazenamento de dados criado para representar um desenho que contenha linhas, regiões, textos e imagens para ser transmitido via Web [GFVG99], embora este formato ainda esteja em estudo.

Desenvolvimento de um *benchmark*

Elaboração de um *benchmark* para a avaliação do desempenho dos diferentes operadores espaciais possíveis para uma mesma consulta espacial. Esta avaliação pode identificar o conjunto básico de operadores espaciais a ser priorizado para o sistema obter um melhor desempenho.

Diminuição do tráfego de rede

Elaboração de uma estratégia que permite adicionar e remover camadas de mapas exibidas como resultado ao cliente, com o intuito de diminuir a quantidade de informação transmitida entre o cliente e o servidor.

Bibliografia

- [ABS00] S. Abiteboul, P. Buneman, and D. Suciu. *Data on the Web - From Relations to Semistructured Data and XML*. Morgan Kaufmann, 2000.
- [Aut00] Autodesk. MapGuide, 2000. <http://www.mapguide.com> (consulta em 21/06/2000).
- [BIO] BIOTA/FAPESP. BIOTA (Sistema de Informação de Biodiversidade do Estado de São Paulo). <http://www.biotasp.org.br> (consulta em 14/09/1999).
- [BLL⁺94] A. Brossier-Wansek, B. Langou, C. Lecocq, D. Calcinelli, J-M. Augier, and M. Mainguenaud. CIGALES : A Visual Language for Geographical Information Systems. *Visual Languages and Computing*, 1994.
- [BM98] B.F. Burton and V.W. Marek. Applications of JAVA programming language to database management. In *ACM SIGMOD Record*, volume 27(1), pages 27-34, Março 1998.
- [BOO96] N.H. Balkir, G. Ozsoyoglu, and Z.M. Ozsoyoglu. A Graphical Query Language: VISUAL. In *Proceedings of ICDE*, 1996.
- [CCH⁺96] G. Câmara, M. Casanova, A. Hemerly, G. Magalhães, and C.M.B. Medeiros. *Anatomia de Sistemas de Informação Geográfica*. 10^a Escola de Computação, 1996.
- [CDS] CDS - Catalogue of Data Source. <http://www.mu.niedersachsen.de/system/cds> (consulta em 14/09/1999).
- [Cer96] N. Cereja. Visões em Sistemas de Informações Geográficas-modelo e mecanismos. Master's thesis, Universidade Estadual de Campinas, Dezembro 1996.
- [CevO93] E. Clementini, P. Di Felice, and P. van Oosterom. A Small Set of Formal Topological Relationships Suitable for End-user Interaction. In *Proceedings of the 3rd International Symposium Spatial Databases - SSD*, pages 277-295, 1993.

- [CF98] V. Coors and S. Flick. Integrating Levels of Detail in a Web-based 3D-GIS. In *ACM-GIS'98*, pages 40–45, 1998.
- [Cif95] R.R. Ciferri. Um benchmark voltado à análise de desempenho de sistemas de informações geográficas. Master's thesis, Universidade Estadual de Campinas, Maio 1995.
- [CNOP98] J. Chen, J. Niland, R. Ottesen, and W. Paul. A Web-Based System for Conducting Outcomes Research via the Internet: the National Comprehensive Cancer Network System. In *10th International Conference on Scientific and Statistical Database Management*, pages 233–237, Julho 1998.
- [CSH⁺98] G. Clossman, P. Shaw, M. Hapner, J. Klein, R. Pledereder, and B. Becker. Java and Relational Databases: SQLJ. In *Proceedings of ACM SIGMOD International Conference on Management of Data*, page 500, Junho 1998.
- [EOS] EOSDIS - Earth Observing System Data and Information. <http://red-hook.gsfc.nasa.gov/~imswww/pub/imswelcome/> (consulta em 19/09/2000).
- [ERI] ERIN - Environmental Resources Information Network. <http://www.environment.gov.au/net/asdd/advanced/> (consulta em 11/04/2000).
- [ESR00] ESRI. MapObjects Internet Map Server, 2000. <http://www.esri.com/software/mapobjects/index.html> (consulta em 21/06/2000).
- [Fag99] A.S. Fagundes. Bancos de Metadados para Biodiversidade do Estado de São Paulo. Master's thesis, Universidade Estadual de Campinas, 1999. Orientador C.M.B. Medeiros.
- [Far98] G. Faria. Um Banco de Dados Espaço-Temporal para Desenvolvimento de Aplicações em Sistemas de Informação Geográfica. Master's thesis, Universidade Estadual de Campinas, Fevereiro 1998.
- [FGD] Geospatial Data Clearinghouse. <http://130.11.52.178/FGDCgateway.html> (consulta em 14/09/1999).
- [FM] M. Field and E. Meckel. From Start to Finish: Steps in a Comprehensive Technical GIS. <http://www.esri.com/library/userconf/proc97/proc97/-to150/pap132/p132.htm> (consulta em 29/03/2000).
- [FNC⁺00] P. Fernández, J. Nogueras, O. Cantán, J. Zarazaga, and P.R. Muro-Medrano. Java Application Architectures to Facilitate Public Access to Large Remote Sensed and Vector Geographic Data. In *TELEGEO 2000*, Maio 2000. França.

- [Gar97] K. Gardels. Open GIS and On-Line Environmental Libraries. In *ACM SIGMOD Record*, volume 26(1), pages 32–38, Março 1997.
- [GCM] GCMD - Global Change Master Directory. <http://gcmd.gsfc.nasa.gov> (consulta em 14/09/1999).
- [GFVG99] M. Gattass, C.C.F. Ferreira, A.S. Vilar, and M. Glasberg. Efficient Map Visualization on the Web. In *Geoinfo'99*, Campinas, São Paulo, 1999.
- [GIL] GILS - Government Information Locator Service. <http://www.epa.gov/gils> (consulta em 14/09/1999).
- [GL98] S. Gobel and K. Lutze. Development of meta databases for geospatial data in the WWW. In *ACM-GIS'98*, pages 94–99, 1998.
- [GM98] M.A. Gonçalves and C.M.B. Medeiros. Construção de Bibliotecas Digitais Geográficas a partir de um Modelo Hipermídia. In *Brazilian Symposium on Multimedia and Hypermedia Systems*, Maio 1998.
- [Gon97] M.A. Gonçalves. Uso de modelos hipermídia em bibliotecas digitais para dados geográficos. Master's thesis, Universidade Estadual de Campinas, Outubro 1997.
- [Gut94] R. H. Guting. An Introduction to Spatial Database Systems. *The International Journal on Very Large Data Bases*, 3(4):357–399, Outubro 1994.
- [Hak96] J. Hakala. Z39.50-1995: Information retrieval protocol, Janeiro 1996. <http://www.linnea.helsinki.fi/z3950/z3950pr.html> (consulta em 23/06/2000).
- [Har98] C. Harder. *Serving Maps on the Internet*. ESRI Press, 1998.
- [INP00] INPE - Divisão de Processamento de Imagem. SpringWeb, 2000. <http://www.dpi.inpe.br/spring/springweb/springweb.html> (consulta em 21/06/2000).
- [Int00] Intergraph Corporation. GeoMedia Web Applications - GIS For The Web, Setembro 2000. <http://www.intergraph.com> (consulta em 19/09/2000).
- [Jon98] S. Jones. Dynamic Query Result Previews for a Digital Library. In *Digital Libraries 98*, pages 291–292, 1998.
- [KO96] C. Kacmar and J. Orendorf. A Spatial Approach to Organizing and Locating Digital Libraries and Their Content. In *Digital Libraries 96*, pages 83–89, 1996.

- [Lan98] L. Lang. *Managing Natural Resources with GIS*. ESRI Press, 1998.
- [LCHH97] W. Li, K.S. Candan, K. Hirata, and Y. Hara. Facilitating Multimedia Database Exploration through Visual Interfaces and Perpetual Query Reformulations. In *Proceedings of the 23rd VLDB*, pages 538–547, 1997.
- [LP98] A.M.J. Lane and T.W. Parr. Providing Information on Environmental Change: Data Management Strategies and Internet Access Approaches within the UK Environmental Change Network. In *10th International Conference on Scientific and Statistical Database Management*, pages 244–253, Julho 1998.
- [Mal] Malaysia Biodiversity Online. <http://biodiversity.ukm.my/> (consulta em 24/04/2000).
- [MB96] C.M.B. Medeiros and M. Botelho. Tratamento do tempo em SIG. In *Proceedings of GIS Brazil '96, Curitiba, Brazil*, Maio 1996.
- [MC95] C.M.B. Medeiros and M.A. Cilia. Maintenance of Binary Topological Constraints through Active Databases. In *Proceedings of the 3rd ACM Workshop on Advances in GIS*, Dezembro 1995. Baltimore, USA.
- [MCB98] A.M.C. Moura, M.L.M. Campos, and C.M. Barreto. A survey on metadata for describing and retrieving Internet resources. In *World Wide Web*, pages 221–240, 1998.
- [MP94] C.M.B. Medeiros and F. Pires. Databases for GIS. In *ACM SIGMOD Record*, volume 23, No.1, pages 107–115, Março 1994.
- [NGS97] A. Newton, B. Gittings, and N. Stuart. *Designing a scientific database query server using the World Wide Web: The example of Tephabase*, chapter 18, pages 251–266. In *Innovations in GIS 4*, Taylor and Francis. London, 1997.
- [Oli97] J.L. Oliveira. Projeto e Implementação de Interfaces para Sistemas de Aplicações Geográficas. Master's thesis, Universidade Estadual de Campinas, Dezembro 1997.
- [OPC00] J.R. Osses, J.A.C Paiva, and G. Câmara. Arquiteturas Cliente-Servidor para Bibliotecas Geográficas Digitais. In *Geoinfo'2000*, pages 60–67, Centro Anhembi, São Paulo, SP, Brazil, Junho 2000.
- [RLW00] S.C. Rennie, A.M.J. Lane, and M. Wilson. Web Access to Environmental Databases: a Database Query and Presentation System for the UK Environmental Change Network. In *Proceedings of the 2000 ACM Symposium on Applied Computing*, 2000.

- [RR98] D. Roussinov and M. Ramsey. Information Forage through Adaptive Visualization. In *Digital Libraries 98*, pages 303–304, 1998.
- [SA95] H. Samet and W.G. Aref. Spatial Data Models and Query Processing. In *Modern Database Systems: The Object Model, Interoperability, and Beyond*, pages 338–360, 1995.
- [SGM98] M. Szmurlo, M. Gaio, and J. Madelaine. The Geographical Antserver: a Client/Server Architecture for GIS. In *Proceedings of EOGEO'98 Workshop*, 1998. Salzburg, Austria.
- [SLF+98] E.R. Schneider, J.J. Leggett, R.K. Furuta, H.D. Wilson, and S.L. Hatch. Herbarium Specimen Browser: A Tool for Accessing Botanical Specimen Collections. In *Digital Libraries 98*, pages 227–234, 1998.
- [SM98] A. Sorokine and I. Merzliakova. Interactive Map Applet for Illustrative Purposes. In *ACM-GIS'98*, pages 46–51, 1998.
- [Soa98] H.R. Soares. Uma Metodologia para Integração de Sistemas Legados e Banco de Dados Heterogêneos. Master's thesis, Universidade Estadual de Campinas, Dezembro 1998.
- [Spe] Species 2000. <http://www.sp2000.org> (consulta em 24/04/2000).
- [SRG98] B. Schmidt-Belz, C. Rinner, and T.F. Gordon. GeoMed for Urban Planning - First User Experiences. In *ACM-GIS'98*, pages 82–87, 1998.
- [TRK96] S. Timpf, M. Raubal, and W. Kuhn. Experiences with Metadata. In *7th Int. Symposium on Spatial Data Handling*, volume 2, pages 31–43, 1996.
- [TS97] A. Tomasic and E. Simon. Improving Access to Environmental Data using Context Information. In *ACM SIGMOD Record*, volume 26(1), pages 11–15, Março 1997.
- [UDK] UDK - Environmental Data Catalogue. <http://udk.bmu.gv.at> (consulta em 14/09/1999).
- [Uni] University of Kansas. Species Analyst. <http://habanero.nhm.ukans.edu/SpeciesAnalyst/> (consulta em 13/04/2000).
- [URB97] J. Ullman, L. Raschid, and P. Buneman. Mediator Languages - a Proposal for a Standard. In *SIGMOD Record*, volume 26(1), pages 39–44, Março 1997.

- [VHN96] A. Veerasamy, S. Hudson, and S. Navathe. Querying, Navigating and Visualizing an Online Library Catalog. Technical report, GVU, 1996. <ftp://ftp.cc.gatech.edu/pub/people/veerasam/sir.ps>.
- [Voi95] A. Voisard. Mapgets: A Tool for Visualizing and Querying Geographic Information. *Visual Languages and Computing*, pages 1–32, 1995.
- [VPG00] A.C. Vianna, A.C. Paiva, and M. Gattass. Servlets e COM para a Visualização de Dados Geográficos na Web. In *Geoinfo '2000*, pages 76–81, Centro Anhembi, São Paulo, SP, Brazil, Junho 2000.
- [WCF98] S. White, R. Cattell, and S. Finkelstein. Enterprise Java Platform Data Access. In *Proceedings of ACM SIGMOD International Conference on Management of Data*, pages 504–505, Junho 1998.