Serviço Web para Anotação de Dados Geográficos Vetoriais e sua Aplicação em Sistemas de Informação de Biodiversidade

Este exemplar corresponde à redação final da Dissertação devidamente corrigida e defendida por Fabiana Bellette Gil e aprovada pela Banca Examinadora.

Campinas, 18 de dezembro de 2009.

Prof. Dr. Ricardo da Silva Torres
Instituto de Computação - UNICAMP
(Orientador)

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

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP
Bibliotecária: Crisllene Queiroz Custódio – CRB8 / 7966

Gil, Fabiana Bellette
G37s Serviço Web para anotação de dados geográficos vetoriais e sua aplicação em sistemas de informação de biodiversidade / Fabiana Bellette Gil -- Campinas, [S.P. : s.n.], 2009.

Orientador : Ricardo da Silva Torres
Dissertação (Mestrado) - Universidade Estadual de Campinas, Instituto de Computação.

1. Serviços Web. 2. Sistema de informação geografica. 3. Biodiversidade. I. Torres, Ricardo da Silva. II. Universidade Estadual de Campinas. Instituto de Computação. III. Título.

Título em inglês: Web service for annotating vector geographic data and its application in information systems for biodiversity.

Palavras-chave em inglês (Keywords): 1. Web service. 2. Geographical information system. 3. Biodiversity.

Área de concentração: Banco de Dados

Titulação: Mestre em Ciência da Computação

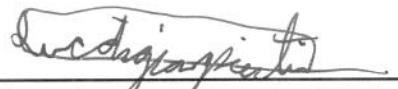
Banca examinadora: Prof. Dr. Ricardo da Silva Torres (IC-UNICAMP)
Prof. Dr. Geovane Cayres Magalhães (IC-UNICAMP)
Prof. Dr. Luciano Antonio Digiampietri (EACH-USP)

Data da defesa: 19/11/2009

Programa de Pós-Graduação: Mestrado em Ciência da Computação

TERMO DE APROVAÇÃO

Dissertação Defendida e Aprovada em 19 de novembro de 2009, pela Banca examinadora composta pelos Professores Doutores:



Prof. Dr. Luciano Antonio Digiampietri
EACH / USP.

Prof. Dr. Geovane Cayres Magalhães
IC / UNICAMP.

Prof. Dr. Ricardo da Silva Torres
IC / UNICAMP.

Instituto de Computação
Universidade Estadual de Campinas

Serviço Web para Anotação de Dados Geográficos Vetoriais e sua Aplicação em Sistemas de Informação de Biodiversidade

Fabiana Bellette Gil

Dezembro de 2009

Banca Examinadora:

- Prof. Dr. Ricardo da Silva Torres
  Instituto de Computação - UNICAMP (Orientador)
- Prof. Dr. Geovane Cayres Magalhães
  Instituto de Computação - UNICAMP
- Prof. Dr. Luciano Antonio Digiampietri
  Escola de Artes, Ciências e Humanidades - USP
- Profa. Dra. Islene Calciolari Garcia (Suplente)
  Instituto de Computação - UNICAMP
- Prof. Dr. André Santanchè (Suplente)
  Departamento de Ciência da Computação - UNIFACS

Resumo

Estudos na área de biodiversidade baseiam-se no uso de dados relacionados a coletas em campo. Normalmente estes dados estão associados a uma localização geográfica. Grande parte dos sistemas de biodiversidade oferece recursos básicos para armazenamento e consulta de dados geográficos.

Esta dissertação apresenta um modelo de dados e um serviço Web para correlacionar dados de biodiversidade e dados geográficos, baseados no conceito de Informação Sobreposta (*Superimposed Information*). Informação sobreposta é uma nova informação (ou nova interpretação) sobre uma informação existente. Pode estar na forma de um novo conteúdo ou novas estruturas. Esta dissertação descreve a utilização do conceito de *Superimposed Information* para a construção de um serviço Web para anotação de dados geográficos vetoriais.

A especificação do serviço Web inclui a definição de uma API genérica para manipulação de anotações e a definição de um modelo de dados para gerenciamento de informações sobrepostas. A solução foi validada a partir da implementação de um protótipo para a área de biodiversidade considerando um potencial cenário de uso.

Abstract

Biodiversity studies are often based on the use of data associated with field observations. These data are usually associated with a geographic location. Most of existing biodiversity information systems provide basic support for storing and quering geographic data.

This work presents a data model and a Web service to correlate biodiversity data and geographic information, based on the concept of Superimposed Information. Superimposed Information is a new information (or new interpretation) over existing information usually in the form of new content or new structures. This work uses the concept of Superimposed Information for constructing a Web service for annotating vector geographic data.

The Web service specification includes the definition of a generic API for handling annotations and the definition of a data model for managing Superimposed information. The solution was validated through the implementation of a prototype for the biodiversity area considering a potential usage scenario.

Agradecimentos

A Deus, pois sem Ele, nada disso teria sido possível. Obrigada, em primeiro lugar, pela vida e pela saúde. Agradeço também por me proporcionar oportunidade de ter uma família e um marido que estão sempre ao meu lado, por eu ter conhecido pessoas tão especiais durante esta caminhada e por eu trabalhar em uma empresa que apóia o estudo.

Ao Rodrigo, meu marido, pela compreensão e companheirismo. Obrigada por ter me incentivado e ter aberto mão de tantos compromissos por minha causa, sem nunca reclamar.

A minha família, por valorizar o estudo e pelo interesse em me acompanhar no andamento deste trabalho.

Aos meus familiares e amigos, pelo apoio.

Ao meu orientador, professor Dr. Ricardo da Silva Torres, por ter aceitado o desafio de me orientar, mesmo sabendo que eu não poderia me dedicar exclusivamente ao mestrado. Obrigada pela paciência, pela dedicação e por todos os ensinamentos, profissionais e pessoais.

À professora Claudia Bauzer Medeiros e aos colegas do LIS, em especial a Carla, Sidney, Joana, Edivaldo e Nadia, pelas contribuições que tornaram este trabalho melhor.

Aos professores do IC que me aceitaram como aluna especial em suas disciplinas. Sem essa ajuda eu não teria conseguido chegar até aqui.

Aos colegas do CPqD, pelas dicas e pela motivação. Agradeço especialmente ao Roger, João, Lin, Mirian e Sandro.

Aos funcionários do IC, pela gentileza com que sempre me atenderam.

Ao CPqD, através do FUNTTEL, pelo incentivo em melhorar minha formação.

Ao CNPq, CAPES, Projeto BioCORE do CNPq, FAPESP, Instituto Virtual FAPESP-Microsoft, pelo apoio financeiro na realização deste trabalho.

Sumário

Resumo **v**

Abstract **vi**

Agradecimentos **vii**

1 Introdução **1**

2 Conceitos Básicos e Trabalhos Correlatos **4**
2.1 Projeto BioCORE 4
2.2 Dados Geográficos 6
2.2.1 Conceitos Básicos 6
2.2.2 Padrões em Sistemas de Informações Geográficas 9
2.2.3 Bancos de Dados Espaciais 13
2.2.4 Ontologia 13
2.3 *Superimposed Information* (Informação Sobreposta) 15
2.3.1 Aplicações Sobrepostas e SPARCE 16
2.3.2 Exemplos de Aplicações Sobrepostas 17
2.4 Anotações 20
2.4.1 Anotação, *Superimposed Information* e Metadados 20
2.4.2 Anotação Semântica 22
2.4.3 Ferramentas para Anotação de Dados Geográficos 23

3 Anotações Sobrepostas Associadas a Objetos Geográficos Vetoriais **25**
3.1 Modelagem de Dados 25
3.2 Serviço Web para Gerenciamento de Anotações 32
3.2.1 Interface Genérica para Manipulação de Anotações 32
3.2.2 Inserção de Anotações 34
3.2.3 Consultas de Anotações 39
3.2.4 Remoção de Anotações 41

3.2.5 Atualização de Anotações 42
3.2.6 Uso de Anotações Semânticas 43

4 Aspectos de Implementação **45**
4.1 Arquitetura 45
4.1.1 Cliente 47
4.1.2 Servidor 49
4.1.3 Banco de Dados 51
4.2 Cenário de Uso 53
4.3 Exemplo de Utilização do Serviço Web 55
4.4 Operações Implementadas 65
4.4.1 Exemplos de Informações Armazenadas pela Operação de Inserção de Anotação 66
4.4.2 Exemplo de Consulta 71

5 Conclusões e Trabalhos Futuros **73**
5.1 Conclusões 73
5.2 Trabalhos Futuros 74
5.2.1 Extensões na Implementação 74
5.2.2 Extensões para Outros Domínios 74
5.2.3 Controle de Acesso 74
5.2.4 Testes de Desempenho: Tempo de Processamento e Custos de Armazenamento 74

Bibliografia **76**

Lista de Tabelas

2.1 Descrição de exemplos de aplicações sobrepostas. 19
2.2 Descrição dos elementos do Dublin Core. 21
2.3 Comparação entre ferramentas de anotação. Adaptada de [31]. 24

3.1 Atributos da entidade *GeographicObject*. 29
3.2 Atributos da entidade *Excerpt*. 29
3.3 Atributos da entidade *Mark*. 30
3.4 Atributos da entidade *Context*. 30
3.5 Atributos da entidade *Element*. 31
3.6 Atributos da entidade *Metadata*. 31
3.7 Atributos da entidade *SuperimposedInformation*. 32
3.8 Atributos da entidade *Text*. 32
3.9 Atributos da entidade *GeographicObject_has_Excerpt*. 33

Lista de Figuras

2.1 Arquitetura do sistema BioCORE. 5
2.2 Exemplos de relacionamentos topológicos *equals*, *disjoint*, *touches*, *crosses*, *overlaps* e *contains*. 8
2.3 Exemplo do uso das especificações do OGC. 10
2.4 Arquitetura conceitual para gerenciamento de informação sobreposta. Adaptada de [32]. 16
2.5 Modelo de referência do SPARCE. Adaptado de [38, 43]. 17
2.6 Arquitetura para Sistema de Informações de Biodiversidade (SIB) que utiliza informação sobreposta [41]. 18
2.7 Arquitetura para Sistema de Informações Biomédicas que utiliza informação sobreposta [40]. 20

3.1 Modelo do OGC para representação de geometria. 26
3.2 Exemplos de tipos de marcas. 27
3.3 Diagrama entidade-relacionamento proposto para gerenciar anotações associadas a objetos geográficos vetoriais. 28
3.4 Exemplo de *InsertAnnotation* (Text Anotacao, Geometry Regiao, Metadata Metadados, Context Contexto). 35
3.5 Exemplo de *InsertAnnotation* (Text Anotacao, Geometry Regiao, List {ObjectType Tipo}, Metadata Metadados, Context Contexto). 36
3.6 Exemplo de *InsertAnnotation* (Text Anotacao, Geometry Regiao, Query Consulta, Metadata Metadados, Context Contexto). 37
3.7 Exemplo de *InsertAnnotation* (Text Anotacao, GeoQuery GeoConsulta, Metadata Metadados, Context Contexto). 38
3.8 Exemplo de *InsertAnnotation* ((Text Anotacao, GeoQuery GeoConsulta, List {ObjectType Tipo}, Metadata Metadados, Context Contexto). 39
3.9 Exemplo de *InsertAnnotation* ((Text Anotacao, GeoQuery GeoConsulta, Query Consulta, Metadata Metadados, Context Contexto). 40

4.1 Arquitetura do serviço Web de criação e consulta de anotações. 46

4.2 Tecnologias utilizadas na arquitetura. 46
4.3 Tela de entrada da aplicação de inserção e consulta de anotação. 47
4.4 Tela com o seletor de tipo de objeto geográfico a ser apresentado. 48
4.5 Tela com o tipo de objeto geográfico Rios desabilitado. 48
4.6 Tela com o tipo de objeto geográfico Reservas desabilitado. 49
4.7 Tela com os detalhes alfanuméricos de um objeto do tipo Rio. 50
4.8 Tela com os detalhes alfanuméricos de um objeto do tipo Reserva. 51
4.9 Interface administrativa do Geoserver para configuração de *Feature Types*. . 52
4.10 Exemplo de dados de coletas sobre espécies de *Lepidoptera* capturadas em três habitats (B = Borda, T = Trilha e I = Interior da Mata) [20]. 54
4.11 Exemplo de criação de anotação associada a uma região de uma reserva ecológica. 55
4.12 Exemplo de criação de anotação próxima a fronteira de uma reserva ecológica. 56
4.13 Exemplo de consulta de informações de catálogo de espécies do BioCORE associadas a anotação. 58
4.14 Exemplo de consulta de anotação a partir de uma região selecionada no mapa. 59
4.15 Exemplo de consulta de anotação contida no interior da área de um objeto geográfico. 60
4.16 Exemplo de consulta de anotação associada a um objeto geográfico. 61
4.17 Exemplo de WSDL de consulta de anotação associada a um objeto geográfico. 62
4.18 Exemplo de consulta de geometria da marca de uma anotação. 63
4.19 Exemplo de exibição da marca de uma anotação no mapa. 64
4.20 Informações armazenadas na tabela Mark. 66
4.21 Informações armazenadas nas tabelas SuperimposedInformation e Text. . . 67
4.22 Informações armazenadas na tabela Metadata. 68
4.23 Informações armazenadas nas tabelas Excerpt e GeographicObject_Has_Excerpt. 69
4.24 Informações armazenadas na visão GeographicObjects e na tabela Reservas. 70
4.25 Informações armazenadas nas tabelas Context e Element. 71

Capítulo 1

Introdução

O termo biodiversidade - ou diversidade biológica - descreve a riqueza e a variedade do mundo natural. A fauna e a flora são partes do patrimônio de uma nação, produto de milhões de anos de evolução concentrada em um determinado local e momento e, portanto, são tão merecedoras de atenção quanto as particularidades da língua e da cultura [46]. Alguns problemas conhecidos na biodiversidade são a perda irreversível de espécies, a perda de serviços ambientais (produção de oxigênio pelas plantas, a capacidade de produção de água e o equilíbrio hidrológico, a fertilidade do solo e o equilíbrio climático) e a biopiratatia. É preciso que a sociedade entenda os valores da biodiversidade, de modo que haja uma conscientização da extensão da perda de espécies e suas conseqüências [46].

A computação pode ser uma grande aliada de biólogos, fornecendo a eles ferramentas para analisar e registrar descobertas sobre espécimes e seus comportamentos. Sistemas de informação podem também ser utilizados para detectar e alertar sobre os crimes praticados contra a natureza [26].

O conhecimento sobre a diversidade biológica é incipiente e desagregado. As informações existentes estão dispersas e, via de regra, inacessíveis. Existe uma enorme quantidade de dados em bibliotecas tradicionais (papel) ou em bases de dados digitais não integradas. Um volume enorme de informações associadas a amostras (espécimes) de material biológico está depositado em herbários e coleções zoológicas. Estima-se que as coleções de história natural contenham mais de 2,5 bilhões de espécimes por todo o mundo. Cada espécime dessas coleções representa a evidência física de ocorrência do organismo no passado, e fornece registro de informações que compreendem desde a presença histórica e geográfica até detalhes morfológicos e ecológicos. Quando bem documentados, os registros de observações, assim como os gerados por levantamentos de fauna e flora e estudos de impacto ambiental, podem também representar uma valiosa fonte de dados [7].

Alguns desafios para os Sistemas de Informação de Biodiversidade são [11]: (i) manipular grandes volumes de informações; (ii) integrar informações provenientes de diferentes

fontes e formatos (heterogeneidade); (iii) manipular dados e imagens; (iv) manipular informações geo-refenciadas.

Grande parte dos sistemas de biodiversidade gerenciam informações sobre espécies ou dados coletados em campo. Normalmente essas informações estão relacionadas a uma localização geográfica. Muitas vezes a localização geográfica é representada de forma simplificada, como por exemplo, contendo o nome e coordenadas da localidade onde uma coleta em campo foi realizada. Além disso, o armazenamento é feito em atributos alfanuméricos tradicionais.

Este trabalho de mestrado faz parte do projeto Biocore [4]. O Biocore tem como objetivo especificar e desenvolver modelos e ferramentas computacionais para permitir que biólogos possam gerenciar e compartilhar os seus dados, promovendo a investigação colaborativa da biodiversidade.

O projeto de pesquisa visa ao desenvolvimento de uma ferramenta para permitir a associação entre informações sobrepostas e dados geográficos (objetos vetoriais). Informação sobreposta é uma nova informação ou nova interpretação sobre uma informação base, como por exemplo anotações ou comentários. As informações sobrepostas podem ser representadas em diversos formatos como texto, áudio, video ou imagem, porém esta dissertação trata somente o formato texto e o referencia como anotação sobreposta ou simplesmente como anotação.

Existem aplicações que manipulam informações sobrepostas, permitindo reunir trechos de informação base em diferentes formatos. Em algumas destas aplicações é possível reunir partes de imagens, trechos de arquivos no formato PDF, planilhas do Excel, etc. Para lidar com informação base em diferentes formatos, estas aplicações requerem gerenciadores de informações sobrepostas próprios para cada tipo de dado. Neste trabalho, foi desenvolvido um serviço de anotações sobrepostas associadas a dados geográficos vetoriais.

A implementação do sistema de informações sobrepostas associadas a dados geográficos utiliza padrões abertos. A solução contempla modelagem do Banco de Dados para o gerenciamento de informações sobrepostas, desenvolvimento de uma interface para fornecer operações de inserção, remoção, consulta e atualização de informações sobrepostas associadas a dados geográficos e uma interface visual Web para apresentação dos dados geográficos e chamada do serviço.

O sistema permite a especificação de consultas do tipo "Retorne anotações associadas à espécimes da espécie *Anaea ryphea* que tenham sido observados na região da Mata Atlântica pelo biólogo Arlindo Gomes Filho no período de janeiro/2009 até dezembro/2009". Com o uso do sistema, biólogos podem inserir textos associados a pontos, linhas ou regiões do mapa, selecionar uma região e adicionar o mesmo texto a todos os elementos que fazem parte dessa região ou visualizar todas as anotações que estão associadas a uma região.

Na área de biodiversidade a ferramenta pode ser utilizada para associar textos a trechos de rios, reservas ecológicas, ou áreas no mapa, para que biólogos registrem fenômenos naturais, dados de coletas em campo ou ainda identifiquem o habitat de algumas espécies.

O sistema foi validado a partir do desenvolvimento de um protótipo para a área de biodiversidade considerando um potencial cenário de uso. Neste cenário, um biólogo insere e consulta dados de coletas sobre borboletas realizadas em diferentes regiões de uma reserva ecológica.

Em resumo, as principais contribuições deste projeto são:

- Especificação e implementação de um modelo de dados para armazenamento de informações sobrepostas;
- Especificação e implementação parcial de serviço Web para anotação de dados geográficos vetoriais que utiliza operadores e relacionamentos espaciais;
- Construção de um protótipo voltado para a área de biodiversidade que utilize dados reais.

O restante do texto está organizado da seguinte forma: o capítulo 2 apresenta conceitos básicos, assim como alguns trabalhos relacionados. A modelagem de dados e interfaces são descritos no capítulo 3. No capítulo 4 são apresentados aspectos da implementação e um cenário de uso da aplicação desenvolvida. O capítulo 5 apresenta as considerações finais e possíveis extensões deste trabalho de pesquisa.

Capítulo 2

Conceitos Básicos e Trabalhos Correlatos

A seção 2.1 descreve uma visão geral do projeto BioCORE, ao qual este trabalho está relacionado. O desenvolvimento deste projeto abrangeu pesquisa em várias áreas. Dentre os assuntos envolvidos, destacam-se modelo de dados geográficos (seção 2.2) e *Superimposed Information* (seção 2.3). A seguir, são descritos os conceitos básicos, arquiteturas e exemplos de aplicações relacionados a estes temas. A seção 2.4 apresenta alguns conceitos e ferramentas envolvendo Anotações.

2.1 Projeto BioCORE

Sistemas de Informação de Biodiversidade normalmente integram informações sobre espécies, registros de coletas em campo e dados geográficos. Grande parte dos sistemas associa a posição geográfica a registros de coletas ou observação de espécies. Há ainda os que permitem visualização de mapas de distribuição de espécies. A posição geográfica é bastante relevante nesse tipo de sistema, porém normalmente se limita aos dados da localidade e coordenadas do local pesquisado.

O BioCORE [4] é um sistema de biodiversidade desenvolvido em parceria entre pesquisadores do Instituto de Biologia e pesquisadores do Instituto de Computação da UNICAMP. Teve sua origem no projeto WeBios. Seu objetivo é prover um sistema que permita a execução de consultas exploratórias multimodais sobre fontes de dados de biodiversidade heterogêneas e distribuídas na Web. Estas fontes incluem dados textuais, imagens (fotos de seres vivos e seus habitats), dados geográficos, ontologias e metadados específicos do domínio. O BioCORE possui arquitetura orientada a serviços.

A Figura 2.1 apresenta a arquitetura do sistema BioCORE. Ela é composta de quatro camadas principais:

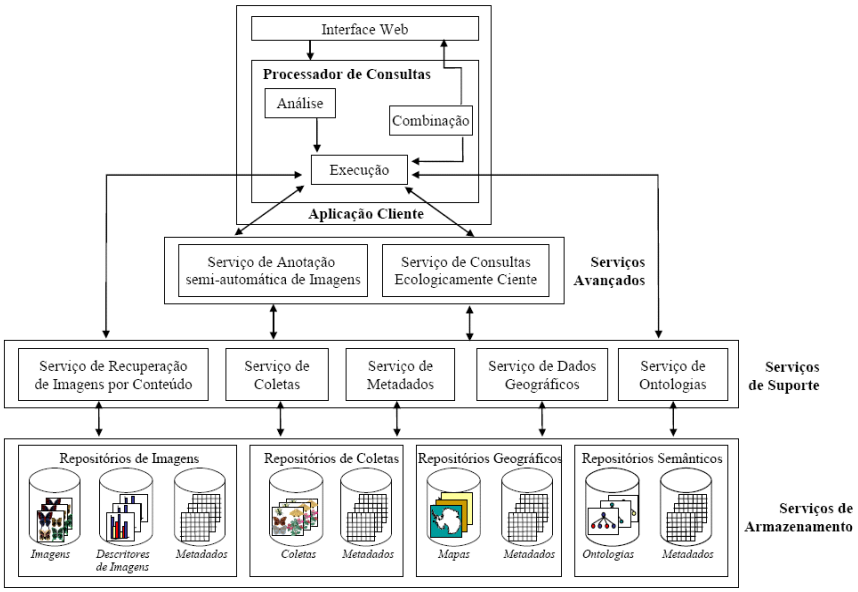


Figura 2.1: Arquitetura do sistema BioCORE.

Camada de Aplicação: camada responsável pela interação com o usuário. Nesta camada, uma requisição realizada pela interface Web é traduzida em uma série de novas requisições pelo Processador de Consultas. As requisições são encaminhadas para as camadas de Serviços Avançados e serviços de Suporte. Após a execução dos serviços, o Processador integra os resultados obtidos e retorna a resposta final à interface.

Camada de Serviços Avançados: camada que contém serviços Web que invocam os serviços da Camada de Serviços de Suporte para atender requisições que demandam acesso a distintos tipos de fontes de dados, provendo funcionalidades mais complexas.

Camada de Serviços de Suporte: camada que contém serviços Web básicos de acesso à camada de armazenamento. Compreende cinco serviços Web, cada um dedicado a uma modalidade específica de recuperação de dados:

- Serviço de Recuperação de Imagens por Conteúdo: processa requisições baseadas em conteúdo de imagens.
- Serviço de Metadados: recupera informações com base em parâmetros de metadados.

- Serviço de Coletas: fornece acesso a registros de observações de seres vivos e dados coletados em campo [33, 34, 55].
- Serviço de Dados Geográficos: recupera informação espacial, usada para criar mapas e distribuição das espécies em uma determinada área.
- Serviço de Ontologia: fornece acesso a ontologias de domínio, integrando dados heterogêneos em consultas distribuídas [12, 13].

Camada de Persistência: responsável pelo armazenamento e gerenciamento de dados. Possui serviços web que encapsulam repositórios que contêm imagens (de habitats, espécies), metadados, ontologias de domínio, dados espaciais e os registos de coleções de espécies.

A pesquisa desenvolvida teve como objetivo especificar e implementar um serviço Web na camada de Serviços Avançados que possibilite a criação, atualização, remoção e consulta de anotações associadas a dados geográficos. Na camada de Serviços de Suporte foi especificado e implementado um serviço de anotações e na camada de Persistência foi definido um repositório para armazenamento das anotações associada a dados geográficos.

2.2 Dados Geográficos

Esta seção tem por objetivo descrever alguns conceitos importantes relacionados a dados geográficos, entre eles, alguns padrões utilizados para modelagem e troca de informações.

2.2.1 Conceitos Básicos

Dados Geográficos

O termo dado espacial denota qualquer tipo de dado que descreve fenômenos aos quais esteja associada alguma dimensão espacial. Dados geo-referenciados ou dados geográficos pertencem a uma classe particular de dados espaciais, que descrevem fatos, objetos e fenômenos no globo terrestre associados à sua localização sobre a superfície terrestre, em um certo instante ou período de tempo [8].

Modelagem de Dados Geográficos

No contexto das aplicações que tratam informações geográficas, o mundo real é frequentemente modelado segundo duas visões complementares: o modelo de campos e o modelo de objetos. O modelo de campos (em geral implementado segundo a chamada estrutura *raster*) enxerga o mundo como uma superfície contínua, sobre a qual os fenômenos

geográficos a serem observados variam segundo diferentes distribuições (e.g., pressão atmosférica). Cada camada corresponde a um tema diferente (e.g., vegetação, solo) [29].

Já o modelo de objetos (em geral implementado usando representação vetorial) representa o mundo como uma superfície ocupada por objetos identificáveis, com geometria e características próprias e que existem independentemente de qualquer definição (e.g., um edifício ou um rio). No modelo de objetos, um objeto geográfico é tipicamente representado no formato vetorial, ou seja, sua geometria é descrita utilizando pontos, linhas e polígonos. Linhas são formadas por seqüências de pontos; e polígonos (abertos ou fechados), por seqüência de linhas [29]. Relacionamentos métricos, topológicos e direcionais são comuns nesse tipo de representação e são obtidos a partir de funções aplicadas a um ou mais objetos geográficos [25].

Relacionamentos topológicos:

- *equals*(geometria1, geometria2): Retorna verdadeiro quando a geometria1 é espacialmente igual a geometria2. As duas geometrias têm a mesma borda e mesmo interior.
- *disjoint*(geometria1, geometria2): Retorna verdadeiro quando a geometria1 não possui nenhum ponto em comum com a geometria2. As bordas e interiores das geometrias não se interceptam.
- *intersects*(geometria1, geometria2): Retorna verdadeiro quando existe intersecção entre a geometria1 e a geometria2.
- *touches*(geometria1, geometria2): Retorna verdadeiro quando a geometria1 toca espacialmente a geometria2. As bordas das geometrias se interceptam, mas os interiores não.
- *crosses*(geometria1, geometria2): Retorna verdadeiro quando a geometria1 cruza espacialmente a geometria2.
- *overlaps*(geometria1, geometria2): Retorna verdadeiro quando a geometria1 sobrepõe espacialmente a geometria2.
- *contains*(geometria1, geometria2): Retorna verdadeiro quando a geometria1 contém espacialmente a geometria2.

A Figura 2.2 mostra o uso dos principais relacionamentos topológicos citados. Na maioria dos exemplos foram utilizados dois polígonos, mas outros tipos de geometrias também são permitidos, dependendo do tipo do relacionamento topológico.

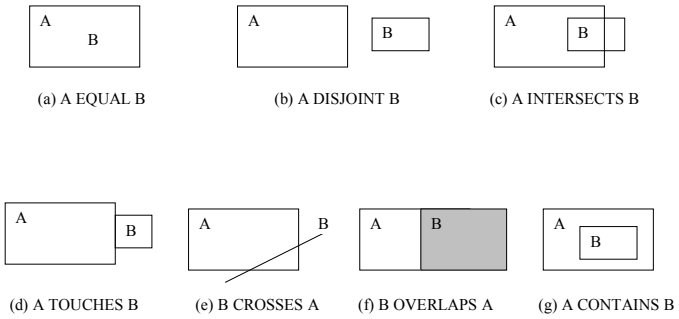


Figura 2.2: Exemplos de relacionamentos topológicos *equals*, *disjoint*, *touches*, *crosses*, *overlaps* e *contains*.

Relacionamentos direcionais:

- *above*(geometria1, geometria2): Retorna verdadeiro se a geometria1 está acima da geometria2.
- *north_of*(geometria1, geometria2): Retorna verdadeiro se a geometria1 está ao norte da geometria2.

Relacionamentos métricos:

- *distance*(geometria1, geometria2): Retorna a distância entre a geometria1 e a geometria2.
- *area*(geometria1): Retorna a área da geometria1.

Sistema de Informações Geográficas

SIGs são sistemas automatizados usados para armazenar, analisar e manipular 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 [8].

A arquitetura utilizada por Sistemas de Informações Geográficas que disponibilizam informações através da Internet segue o modelo cliente/servidor. O servidor de mapas é executado juntamente com o servidor Web e responde às solicitações do cliente (browser).

2.2.2 Padrões em Sistemas de Informações Geográficas

Em Sistemas de Informações Geográficas, a comunicação entre servidor e cliente pode ser feita de várias maneiras, assim como, a modelagem dos dados geográficos pode variar, dificultando a interoperabilidade entre diferentes aplicações. O Open Geospatial Consortium (OGC) [48] é uma organização voluntária internacional que trata desses problemas a partir da definição de padrões para simplificar o compartilhamento, comercialização e reuso de dados e sistemas geográficos na Web.

O modelo de referência especificado pelo OGC utiliza o conceito de *feature* geográfica como o ponto de partida para a modelagem de informação geoespacial. *Feature* é a unidade fundamental de informação espacial e pode ser definida como uma abstração de um fenômeno do mundo real (ISO 19101), correspondendo à noção de objeto espacial.

Uma *feature* geográfica é uma *feature* associada a uma localização geográfica na Terra. Uma instância de uma feature corresponde a um fenômeno discreto. Instâncias individuais são agrupadas em classes com características comuns: os chamados *feature types*.

Em geral, os padrões existentes podem ser divididos em dois grupos: padrões de dados e padrões de serviços. Os padrões de dados são utilizados para modelar, trocar e visualizar informações geográficas. Os padrões de serviços servem para dar suporte à realização de consultas, análises espaciais, entre outras. A Figura 2.3 ilustra alguns exemplos de uso de padrões de dados e de serviços. Na camada de armazenamento, pode-se utilizar o padrão *Simple Features Specification* (SFS) (ver seção 2.2.2). Na camada de serviços, os padrões *Web Map Service* (WMS), *Web Feature Service* (WFS) e *Web Coverage Service* (WCS) (ver seção 2.2.2) são usados para disponibilizar os serviços. Entre a camada de aplicação e de serviços, informações são trocadas utilizando-se os padrões *Styled Layer Descriptor* (SLD) ou *Geography Markup Language* (GML) (ver seção 2.2.2).

Padrões para Serviços

As principais especificações de serviços conhecidas foram propostas pelo OGC. São elas:

- Web Feature Service (WFS) [48]: esta especificação apresenta uma forma de acesso (inserção, atualização, exclusão e análise) a features usando o ambiente Web, com chamadas HTTP. As operações entre clientes e servidores são baseadas no formato GML (ver seção 2.2.2).

- Web Map Service (WMS) [48]: esta especificação define quatro protocolos (*GetCapabilities, GetMap, GetFeatureInfo e DescribeLayer*) que permitem a leitura de múltiplas camadas de informações (*layers*) georreferenciadas, contendo vetores e/ou imagens. Essa conexão permite somente consulta de dados, sendo todo o processo de renderização do mapa feito no servidor. Com isso, o cliente recebe uma imagem

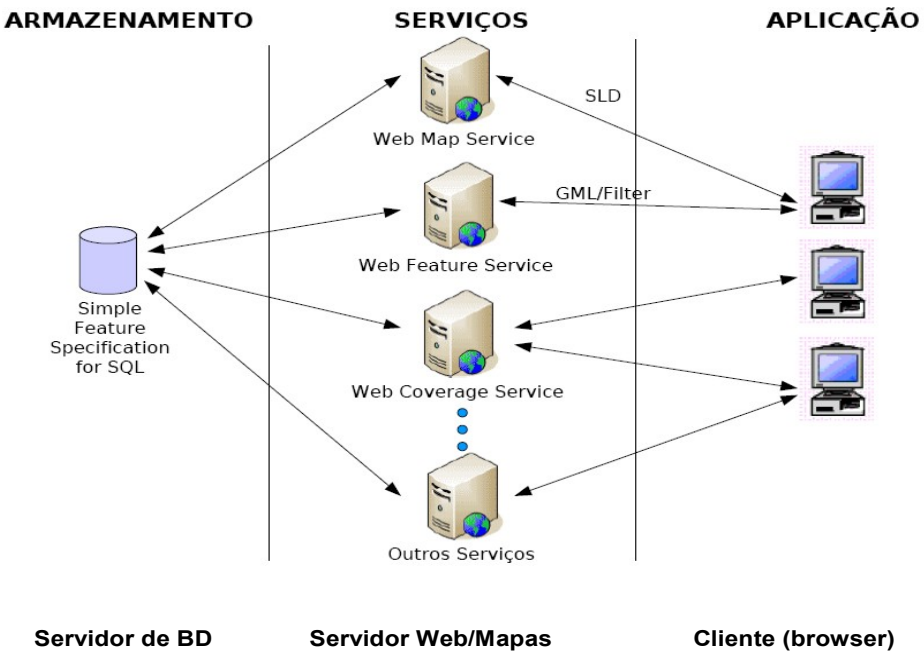


Figura 2.3: Exemplo do uso das especificações do OGC.

que corresponde a uma visualização do mapa, de acordo com as camadas (vetoriais ou matriciais) solicitadas.

- Web Coverage Service (WCS) [48]: esta especificação define três operações (*GetCapabilities, DescribeCoverage e GetCoverage*) que permitem a disponibilização de coverages em um ambiente Web (HTTP). A renderização dos dados ocorre no cliente.

Implementações de serviços do OGC

Geoserver: A partir de 2003, o OGC adotou o projeto GeoServer [21] como sua implementação de referência para WFS. O objetivo do projeto é prover uma implementação dos padrões WFS e WMS com código aberto e disponibilizado livremente. O GeoServer visa possibilitar a criação de uma rede em que usuários possam pesquisar e navegar pelos dados espaciais.

Geotools: GeoTools [22] é uma biblioteca de código livre que permite desenvolver soluções adaptadas aos padrões internacionais em vigor no desenvolvimento de software para informação geográfica. Proporciona uma implementação das especificações do OGC conforme estas vão sendo estabelecidas.

Deegree: Deegree [15] é um framework Java totalmente baseado em padrões de domínio público que fornece suporte a um subconjunto de especificações do OGC. O projeto Deegree é coordenado pela empresa Lat/lon e pelo Departamento de Geografia da Universidade de Bonn.

Mapserver: O MapServer [35] é um ambiente de desenvolvimento de código aberto para construção de aplicativos espaciais na Internet. Foi originalmente desenvolvido pelo projeto ForNet da Universidade de Minnesota (UMN) em cooperação com a NASA e o Departamento de Recursos Naturais de Minnesota (MNDNR). Atualmente o projeto MapServer é mantido pelo projeto TerraSIP, um projeto patrocidado conjuntamente pela NASA e a UMN.

Padrões para Descrição de Dados Geográficos

As principais especificações de dados geográficos propostas pelo OGC são:

Simple Features Specification (SFS): Esta especificação define um formato, de acordo com o SQL padrão, para armazenamento e leitura de "features simples" (dados geográficos) utilizando-se de uma API (ODBC[1]). Estas features são baseadas em geometrias 2D com interpolação linear entre os vértices. O documento 99-049 do OGC define os detalhes dessa *interface* que deve conter, entre outras coisas, análises espaciais/geográficas e topológicas. Este padrão já está sendo substituído pelo SFA (Simple Feature Access), que entre outras melhorias, prevê o tratamento de geometrias 3D.

Geography Markup Language: A GML [5, 21] foi definida pelo OGC e trata-se de uma linguagem baseada em XML para modelagem, transporte e armazenamento de informações geográficas, compreendendo propriedades espaciais e não espaciais das features geográficas. Sua especificação define a sintaxe XML Schema, mecanismos e convenções que provêem um arcabouço aberto e independente de fabricante para a definição de esquemas e objetos de aplicações geoespaciais. O objetivo da GML é oferecer um conjunto de regras com as quais um usuário passa a definir sua própria linguagem para descrever

[1]Open Data Base Connectivity é um padrão para acesso a sistemas gerenciadores de bancos de dados.

seus dados. GML implementa geometrias, com as seguintes classes: Point, LineString, LinearRing, Polygon, MultiPoint, MultiLineString, MultiPolygon e MultiGeometry. Possui ainda os elementos $< coordinates >$ e $< coord >$ para representação de coordenadas.

KML: KML [48] é uma linguagem XML centrada na visualização geográfica, incluindo anotação de mapas e imagens. Foi desenvolvida originalmente para gerenciar a visualização de dados geoespaciais no Google Earth. Em abril de 2008, foi aprovada pelo OGC e passou a ser chamada de OGC KML, marcando sua transição para um padrão aberto a ser mantido pelo OGC. KML é complementar a maior parte dos principais padrões da OGC existentes, incluindo *Geography Markup Language* (GML), *Web Feature Service* (WFS) e *Web Map Service* (WMS).

Styled Layer Descriptor (SLD): Esta especificação permite a representação gráfica de objetos espaciais, permitindo a definição da simbologia e personalização dos dados geográficos.

Existem vários padrões ou formatos para descrever dados geográficos. Algumas iniciativas são comerciais e outras públicas. Há padrões usados exclusivamente para modelar dados geográficos, enquanto outros são focados nos aspectos de visualização. Alguns dos padrões e formatos mais conhecidos são detalhados a seguir.

Scalable Vector Graphics (SVG): SVG [5, 28, 51] é uma linguagem para descrição de gráficos vetoriais bi-dimensionais e aplicações gráficas em XML. Foi criado pela World Wide Web Consortium (W3C), responsável pela definição de outros padrões, como o HTML e o XHTML. Uma das principais características dos gráficos vetoriais é que não perdem qualidade em diferentes escalas. SVG contém algumas formas pré-definidas que podem ser usadas e manipuladas pelos desenvolvedores. São elas: retângulos (rect), círculos (circle), elipses (ellipse), linhas (line), formas que contêm somente linhas retas (polyline), polígonos - formas que contêm pelo menos três lados (polygon) e caminhos (path). Essa linguagem tem uma vasta aplicação em desenvolvimento para a Web e em Sistemas de Informações Geográficas.

Shapefile: O shapefile [19] é um formato idealizado pela empresa ESRI para suportar tanto dados geográficos quanto atributos não geográficos. É composto de um conjunto de arquivos, dos quais três são obrigatórios: Shp - shape format (arquivo principal) - contém a feature geográfica propriamente dita; Shx - shape index format - contém o endereço e o tamanho de cada registro do arquivo principal; Dbf - attribute format - arquivo no formato de tabela dBase, onde estão localizados os atributos de cada feature. O formato

shapefile admite cinco tipos básicos de dados: ponto, polilinha, polígono, multiponto e multipath.

2.2.3 Bancos de Dados Espaciais

Bancos de Dados Espaciais são bancos de dados preparados para armazenar a geometria e os atributos alfanuméricos de objetos. Podem manipular grande volume de informações e fornecer suporte a consultas espaciais, como por exemplo, buscar determinado objeto localizado nas proximidades de um ponto. Existem dois Sistemas de Gerenciamento de Banco de Dados (SGBD) que são muito utilizados no desenvolvimento de soluções baseadas em software livre: MySQL[2] e PostgreSQL[3].

O MySQL não possui uma licença totalmente livre, pois quando é utilizado em soluções comerciais, a licença exige o pagamento pelo uso. Inicialmente foi projetado para trabalhar com aplicações de pequeno e médio porte. Atualmente tem sido utilizado em aplicações Web e lojas virtuais, pelo fato de demandarem acesso rápido.

O PostgreSQL é um sistema gerenciador de banco de dados objeto-relacional, gratuito e de código fonte aberto. Possui uma extensão espacial que possibilita o tratamento, com eficiência, de grandes volumes de dados geográficos: PostGIS[4]. Um ponto forte deste SGBD é a sua capacidade de escalabilidade, ou seja, a sua arquitetura pode ser continuamente ampliada de acordo com a demanda dos usuários [50, 52].

O Oracle[5] é uma extensão espacial desenvolvida sobre o modelo objeto-relacional do SGDB Oracle. Este modelo permite definir novos tipos de dados, utilizando-se da linguagem de definição de dados SQL DDL, e implementar operações sobre esses novos tipos utilizando-se a linguagem PL/SQL, que é uma extensão da SQL. Esta extensão contém um conjunto de funcionalidades e procedimentos que permite armazenar, acessar, modificar e consultar dados espaciais [50].

2.2.4 Ontologia

O termo Ontologia vem sendo utilizado para representar uma "especificação explícita formal de uma conceitualização compartilhada" [24]. A utilização de ontologias como estratégia de representação do conhecimento sobre um determinado domínio de interesse, fornece um esquema semântico que vem se mostrando eficaz, já que se pode especificar de maneira explícita e formal os termos do domínio bem como o relacionamento entre eles [3].

[2]http://www.mysql.com
[3]http://www.postgresql.org
[4]http://www.postgis.org
[5]http://www.oracle.com/technology/products/spatial

A *Web Ontology Language* (OWL) [49] é a especificação mais recente do *World Wide Web Consortium* (W3C) para representação de ontologias e é compatível com a arquitetura da Web em geral e da Web Semântica em particular [49]. Foi desenvolvida a partir de RDF e RDF *Schema*. RDF [6] é um arcabouço para representação e correlação de recursos na Web. O modelo RDF define triplas que contêm um sujeito, um predicado e um objeto. Sujeito refere-se a um recurso que compreende qualquer elemento que possa ser representado por uma *Universal Resource Identification* (URI). Predicado compreende uma propriedade do recurso e objeto representa o valor dessa propriedade. RDF Schema é uma linguagem para a definição de esquemas para os vocabulários (termos) utilizados nas declarações (triplas). Fornece os mecanismos necessários para a definição de classes, subclasses, propriedades, subpropriedades e instâncias.

A Web Semântica é uma extensão da Web atual, que acrescenta semântica a dados e serviços, de modo a viabilizar métodos mais eficazes para a busca, automação, integração e reutilização dos recursos disponíveis via rede. A ambição da Web Semântica é possibilitar a agentes de software selecionar e compor recursos em ambientes virtuais para a solução de problemas, visando descobrir novos fatos e realizar tarefas mais sofisticadas [54].

Ontologias estão sendo desenvolvidas sobre este aparato tecnológico para estabelecer consenso sobre o significado de conceitos e termos específicos de diversos domínios do conhecimento [54].

O conhecimento do domínio é formalizado em uma ontologia utilizando quatro tipos de componentes [12]:

- Classes: representam conjuntos ou tipos de objetos (conceitos ou categorias de conceitos do domínio), comumente organizados em taxonomias;
- Instâncias: materializam os objetos do domínio e são representados por instâncias das classes;
- Propriedades: modelam as características das classes e instâncias. Propriedades podem expressar atributos (características ou parâmetros que uma classe deve possuir) ou expressar como as classes e instâncias se relacionam entre si;
- Restrições: são definições abstratas que utilizam propriedades para descrever os conceitos do domínio usando condições.

A construção de uma ontologia é um processo iterativo. Para realizar essa atividade é necessário entrevistar profissionais da área de domínio para coletar termos importantes e relacionamentos entre eles, e assim, contruir e modelar as ontologias [53].

Ontologias Geográficas (também conhecidas como ontologias espaciais, ontologias espaço-temporais, geo-ontologias ou ainda ontologias geo-espaciais) estão se tornando um importante campo de aplicação e pesquisa relacionado à comunidade de Sistemas de Informações

Geográficas (SIG) [27]. Ontologia geográfica é a conceitualização de um fenômeno ou objeto geográfico no mundo real [53].

Uma ontologia relacionando os conceitos e objetos relativos ao território permite localizar o objeto em foco numa base de conhecimento. Com isso, um usuário pode se referir a um objeto (e.g., Estado de São Paulo) de acordo com uma ontologia definida consensualmente [54].

2.3 *Superimposed Information* (Informação Sobreposta)

Informação sobreposta (SI) [32] é uma nova informação (ou nova interpretação) associada a uma informação existente ou base (como, por exemplo, uma anotação, um termo ou palavra-chave, um comentário sobre uma imagem). Este tipo de informação pode conter dados que não estão presentes explicitamente no conteúdo original. Além disso, pode estar na forma de um novo conteúdo (como anotações, comentários) ou novas estruturas (tabela de conteúdo). Em geral, esse tipo de informação não é estruturada, ou seja, não possui um esquema pré-definido. É possível haver diversos tipos diferentes de informação sobreposta sobre a mesma informação base [16].

As informações sobrepostas podem ser utilizadas para várias finalidades:

- Ajudar na localização de uma informação, na forma de índices ou tabelas de conteúdo;
- Explicar, acrescentar, corrigir ou refutar uma informação;
- Relacionar e comparar informações diferentes;
- Classificar e organizar informações.

A Figura 2.4 ilustra a arquitetura conceitual para gerenciamento de informação sobreposta. Cada informação fonte (information source) pode ser vista como uma coleção pré-existente de informações. A informação fonte pode ter uma estrutura simples, como uma coleção de páginas HTML, ou estruturas mais complexas, especificadas por um DTD[6] XML ou esquema de um banco de dados relacional.

A camada de informação sobreposta, assim como a de informação base, consiste de um esquema e uma instância. O esquema descreve a possível estrutura da informação. A camada de informação sobreposta é o mecanismo que permite destacar, interconectar e elaborar informações na camada base. Informação sobreposta referencia a informação base

[6]Document Type Definition: trata-se de um esquema pré-definido a partir do qual as instâncias XML são construídas [56].

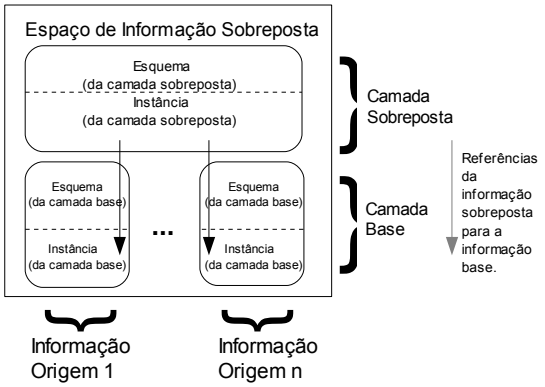


Figura 2.4: Arquitetura conceitual para gerenciamento de informação sobreposta. Adaptada de [32].

usando uma abstração chamada "marca". Uma marca pode ser ativada para se chegar à informação referenciada (*excerpt*) em seu contexto original (*context*). Informação de contexto é um conjunto de informações obtidas usando uma marca. Por exemplo, número de página e características da fonte fazem parte das informações de contexto de uma seleção em um arquivo PDF.

As informações sobrepostas podem ser representadas em diversos formatos (texto, áudio, video ou imagem), porém esta dissertação trata somente o formato texto e o referencia como anotação sobreposta ou simplesmente como anotação.

2.3.1 Aplicações Sobrepostas e SPARCE

Superimposed applications (SAs) são aquelas que permitem que se manipule e se crie informação sobreposta, tipicamente para destacar, anotar, elaborar, selecionar, organizar, conectar ou reusar elementos de informação [44].

Superimposed applications podem fazer uso do *Superimposed Pluggable Architecture for Contexts and Excerpts* (SPARCE) [38, 43], que é um middleware que fornece gerenciamento de marcas e outros serviços. Seu modelo de referência, que pode ser visto na Figura 2.5, é composto dos seguintes módulos:

- Gerenciamento de marcas: implementa operações como criação de marcas. Também mantém um repositório de marcas.
- Gerenciamento de contexto: responsável por recuperar o contexto da informação

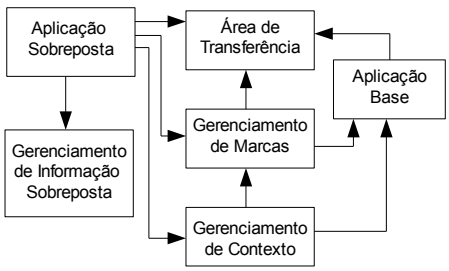


Figura 2.5: Modelo de referência do SPARCE. Adaptado de [38, 43].

base. Depende do módulo de gerenciamento de marcas para localizar informações dentro da camada de base.

- Módulo de área de transferência: responsável por recuperar o conteúdo da região marcada.
- Gerenciamento de informação sobreposta: fornece serviços de armazenamento para aplicações sobrepostas.

2.3.2 Exemplos de Aplicações Sobrepostas

Na área de Biodiversidade, foi proposta uma aplicação para descrição e recuperação de imagens de seres vivos chamada SIERRA [41]. Com essa aplicação, usuários podem associar imagens a outros conteúdos multimídia em diferentes granularidades de um documento (sub-documento ou documento todo). Além disso, também podem realizar consultas utilizando parte da imagem ou da descrição e recuperar o conteúdo original. A arquitetura da aplicação SIERRA, exibida na Figura 2.6, é composta de dois módulos principais: módulo de anotação e módulo de consulta. SIERRA faz uso do *Content-Based Image Search Component* (CBISC) [11], um componente que suporta consultas por conteúdo em coleções de imagens. Este componente recupera imagens similares a um dado padrão fornecido pelo usuário baseado em propriedades do conteúdo que são codificadas por descritores de imagem. Além disso, prevê integração com outros tipos de componentes, como a ontologia WordNet e o sistema SPARCE.

O *Superimposed Image Description and Retrieval Tool* (SuperIDR) [42] é uma extensão do sistema SIERRA, desenvolvida para trabalho com Tablet PC [47]. Tablet PC é um notebook que possui sistema operacional Windows XP Tablet PC Edition e tela sensível ao toque criada para interagir com uma caneta complementar. Com o SuperIDR, o usuário

pode navegar através de informação das espécies ou através da organização taxonômica das espécies com base na família e nos gêneros.

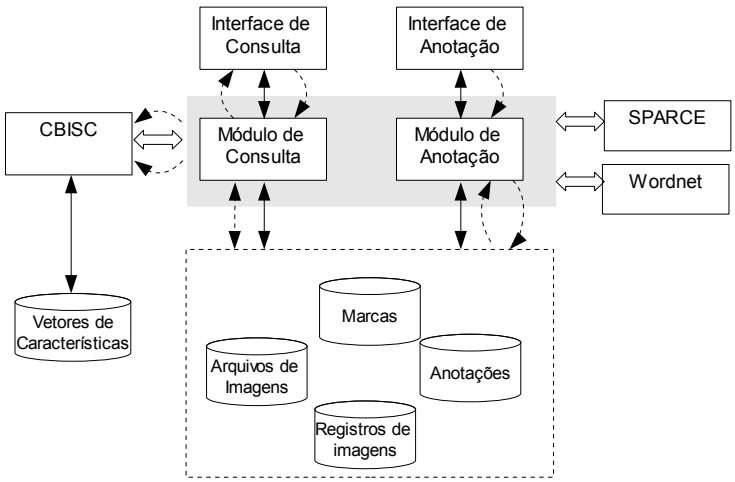


Figura 2.6: Arquitetura para Sistema de Informações de Biodiversidade (SIB) que utiliza informação sobreposta [41].

Informações biomédicas são heterogêneas e geralmente estão em diferentes formatos, como por exemplo, textos (registros de pacientes, procedimentos médicos, etc), imagens ou vídeos (imagens de raio-x, ultrassom, vídeos de cirurgias). Visando ajudar a fornecer suporte para aplicações sobrepostas em sistemas de informações biomédicas, foi proposta uma arquitetura [40], como exibido na Figura 2.7. A arquitetura é composta por quatro camadas: interface com usuário, serviços, gerenciamento da informação e persistência. A camada de interface com o usuário é composta de componentes usados para acessar os serviços das camadas mais baixas. A camada de serviços contém componentes que fornecem funcionalidades ao usuário final (e outro software) para criar, acessar, e gerenciar informação base, marcas e informação sobreposta. O módulo de gerenciamento de informações é responsável por gerenciar dados da camada de armazenamento, existindo componentes distintos para tratar informação base, marcas e informação sobreposta. Três tipos de dados são manipulados na camada de armazenamento: informação base ou existente, marcas descrevendo seleções sobre a informação base e informação sobreposta.

A Tabela 2.1 descreve outras aplicações que possuem serviços de Informação Sobreposta.

Aplicação	Uso	Descrição
RIDPad [38]	Departamento de Agricultura (Serviço de Florestas - USFS) no processo de apelação de decisões	Aplicação utilizada para colecionar e organizar informações necessárias para preparar uma carta RID (Records, Information, Documentation), que é um dos documentos necessários para se recorrer de uma decisão do USFS.
Schematics Browser [43]	Departamento de Agricultura (Serviço de Florestas - USFS) no processo de apelação de decisões	Os processos de apelação tendem a possuir informações comuns, como nome e endereço da pessoa que entra com um recurso. Isso sugeriu o uso de esquemáticos superimpostos, que é um esquema Entidade-Relacionamento sobre a informação base. Schematics Browser foi desenvolvido para demonstrar o uso de esquemáticos superimpostos.
SIMPEL [39]	Variadas áreas - por exemplo, ensino a distância	Ferramenta para edição e execução de apresentações multimídia. Permite ao usuário, por exemplo selecionar um *clip* com áudio, algumas imagens e texto e criar uma apresentação que pode ser executada em uma determinada ordem. Com a ferramenta é possível referenciar informações de muitos tipos, de granularidade variada, sem replicar a informação.
SLIMPad [17]	Área médica	Permite que usuários da área médica possam criar pacotes (bundles) digitais estruturados. Pacote é o agrupamento de informações selecionadas, coletadas, elaboradas e estruturadas por um clínico durante a resolução de um problema.
Mash-o-matic [37]	Aplicações Web	Aplicações Web chamadas mash-ups combinam informações de variada granularidade provenientes de diversas origens. *Mash-o-matic* é um utilitário que extrai, separa e combina fragmentos de informações e gera dados para *mash-ups*.

Tabela 2.1: Descrição de exemplos de aplicações sobrepostas.

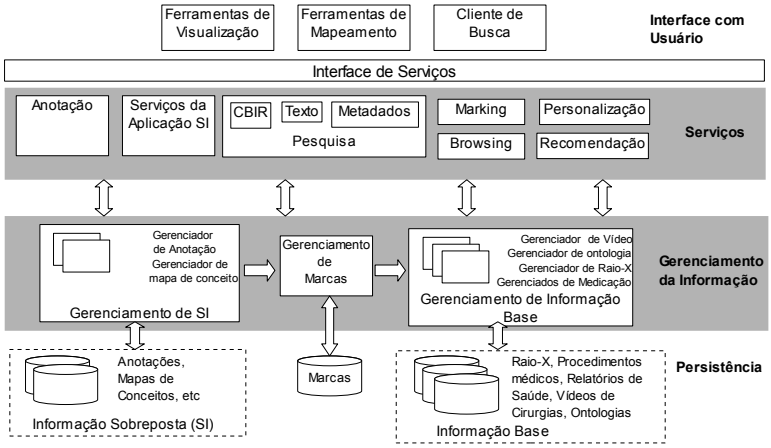


Figura 2.7: Arquitetura para Sistema de Informações Biomédicas que utiliza informação sobreposta [40].

2.4 Anotações

A pesquisa desenvolvida neste trabalho está ligada ao conceito de anotação. A seção 2.4.1 descreve os conceitos de anotação, *superimposed information* e metadados e mostra as similaridades e diferenças entre eles. A seção 2.4.2 mostra conceitos de anotação semântica. Na seção 2.4.3 são apresentadas características de algumas ferramentas para anotação de dados geográficos.

2.4.1 Anotação, *Superimposed Information* e Metadados

Anotações não são somente uma maneira de explicar e enriquecer uma informação com observações pessoais, mas também um meio de transmitir e compartilhar idéias para melhorar as práticas de trabalho colaborativo. Além disso, anotações permitem a usuários mesclar e ligar conteúdos pessoais com recursos de informação disponíveis em bibliotecas digitais de modo que um contexto comum, que unifique todos esses conteúdos, possa ser criado [1].

Anotações podem ser privadas, compartilhadas ou públicas, dependendo do tipo de trabalho que está sendo realizado. Anotações privadas exigem envolvimento dos autores, enquanto anotações compartilhadas ou públicas envolvem a participação de uma comunidade toda [1].

Metadados são considerados dados sobre os dados. Com o uso de metadados, é possível

Elemento	Descrição
Title	Título. Nome dado ao recurso. Tipicamente, o nome pelo qual o recurso é formalmente conhecido.
Creator	Criador. A entidade responsável em primeira instância pela existência do recurso.
Subject	Assunto. Tipicamente, um Assunto deverá ser expresso por palavras chave, frases, ou códigos de classificação que descrevem o conteúdo do recurso.
Description	Descrição. Descrição do conteúdo do recurso.
Publisher	Editor. Entidade responsável por tornar o recurso acessível.
Contributor	Contribuinte. Uma entidade responsável por qualquer contribuição para o conteúdo do recurso.
Date	Data. Data da criação ou disponibilidade do recurso.
Type	Tipo. A natureza ou gênero do conteúdo do recurso.
Format	Formato. A manifestação física ou digital do recurso.
Identifier	Identificador. Referência não ambígua ao recurso, definida num determinado contexto.
Source	Fonte. Referência a um recurso de onde o presente recurso possa ter derivado.
Language	Língua. A língua do conteúdo do recurso.
Relation	Relação. Referência a um recurso relacionado.
Coverage	Cobertura. Extensão ou alcance do recurso.
Rights	Direitos. Informação de direitos sobre o recurso ou relativos ao mesmo.

Tabela 2.2: Descrição dos elementos do Dublin Core.

adicionar informações importantes a um recurso de modo que as ambiguidades possam ser evitadas e a recuperação do recurso seja feita de forma mais fácil. Ausência de metadados pode conduzir à falta de confiança e retrabalho quando ocorre interoperabilidade entre sistemas, dificultando a troca de dados e integração [45].

Na área de biodiversidade, foram desenvolvidos trabalhos envolvendo metadados. Em [10], é apresentada a modelagem e implementação do banco de metadados do sistema de informação para um programa que visa permitir a cooperação entre os diferentes pesquisadores em biodiversidade e a disseminação dos seus trabalhos, buscando dar subsídios à criação de programas de preservação ambiental para o estado de São Paulo. Em [9] foi especificado um padrão de metadados que, associados a workflows científicos, pudessem documentar atividades de planejamento ambiental.

A *Dublin Core Metadata Initiative* (DCMI) é uma organização dedicada a promover a adoção de padrões de interoperabilidade de metadados e desenvolver vocabulários especializados para descrever fontes que tornem mais inteligentes sistemas de descoberta de informações.

O padrão *Dublin Core* [18] inclui dois níveis: Simples e Qualificado. O *Dublin Core* Simples inclui quinze elementos, o Qualificado inclui três elementos adicionais (Audiência, Proveniência e Detentor de Direitos), assim como um grupo de refinamentos de elementos (também chamados qualificadores), que refinam a semântica dos elementos de maneiras que sejam úteis na descobertas de recursos. O conjunto de metadados simples (*Core Metadata Element Set*, DCMES) consiste de quinze elementos, descritos na Tabela 2.2.

Cada elemento *Dublin Core* é opcional e pode ser repetido. O DCMI estabeleceu maneiras padronizadas para refinar os elementos e encorajar o uso de esquemas de codificação e vocabulário. Não há ordem no *Dublin Core* para apresentar ou usar os elementos. Em geral, sistemas que armazenam metadados representam apenas dois níveis de informação, o esquema dos metadados e os metadados em si. O esquema deve ser definido antes da instância dos metadados.

Metadados e informações sobrepostas são similares pois ambos tratam de informações que dependem de uma informação base. Não são portanto entidades totalmente autônomas. Enquanto só existe um conjunto de metadados associado a informação base, podem existir inúmeras associações entre informações sobrepostas e informação base.

Anotações podem ser consideradas metadados, isto é, dado adicional que diz respeito a um conteúdo existente e esclarece as propriedades e a semântica de um conteúdo anotado. Partindo deste princípio, anotações têm que respeitar alguma especificação que defina a estrutura, a semântica e até mesmo valores que podem assumir. Os destinatários desse tipo de anotação podem ser tanto pessoas como máquinas [1].

Anotações podem ser tratadas apenas como um conteúdo adicional a respeito de um conteúdo existente. Nesse caso, não são entidades autônomas e, na verdade, dependem de uma informação pré-existente para justificar sua existência. Anotações também podem ser autônomas e possuírem algum tipo de conexão com o conteúdo existente [1].

Anotações são um tipo de *superimposed information*. *Superimposed information* não segue necessariamente uma especificação e pode estar associada a um fragmento da informação base e seu contexto original. Desta forma, *superimposed information* pode ser considerado um conceito mais amplo do que os conceitos de anotação ou metadados.

2.4.2 Anotação Semântica

Anotações semânticas fazem uso de ontologias e são descritas em linguagens próprias, como OWL ou RDF. As linguagens são tipicamente representadas no formato XML. Anotações semânticas utilizam ontologias para eliminar ambiguidades e promover um entendimento comum dos conceitos. Além disso, promovem a interoperabilidade semântica entre quem produz e quem consome dados [30].

Anotações semânticas geralmente são associadas a conteúdos textuais, como páginas da Web, por exemplo. Porém, as anotações podem ser empregadas em diferentes tipos de dados, como vídeos e imagens.

O *World Wide Web Consortium* (W3C) propôs o uso de *Resource Description Framework* (RDF) para descrever recursos disponíveis na Web como uma iniciativa para fornecer interoperabilidade semântica. RDF descreve recursos através de declarações do tipo $< sujeito, predicado, objeto >$. Sujeito é o recurso que está sendo anotado, predicado

é o atributo do recurso anotado e objeto é o valor do atributo do recurso anotado [1].

Alguns trabalhos envolvendo anotação semântica foram desenvolvidos na área de agricultura. Um deles [31] define um framework para suportar gerenciamento de anotações semânticas para conteúdo digital na Web para planejamento e monitoramento agrícola. Em [14] foi proposta uma abordagem baseada em RDF, metadados geográficos e ontologias para descrever recursos geoespaciais, unindo tecnologias de Web Semântica e padrões geográficos.

2.4.3 Ferramentas para Anotação de Dados Geográficos

Dados geográficos são a base para tomada de decisão em uma série de domínios. Diversas ferramentas oferecem funcionalidades para anotação de dados geográficos, algumas usando ontologia. Em [31] algumas ferramentas de anotação foram testadas. A Tabela 2.3 mostra as características analisadas nos testes. Para cada ferramenta, foi verificado o formato usado nas anotações, se a ferramenta utilizava ontologia, os tipos de dados anotados e se a ferramenta possuía um componente espacial.

Dentre as ferramentas analisadas, nem todas utilizam ontologia. A maioria das ferramentas que possuem componente espacial, anotam principalmente dados textuais. Somente uma delas fornece opção de anotação de dados geográficos.

As ferramentas da Google, Google Maps e Google Earth, permitem associar texto, fotos, vídeos a pontos do mapa. O armazenamento das anotações é feito pela própria API do Google. A API do Google não é open source. As consultas de anotações não oferecem opções para utilização de relacionamentos espaciais. Além disso não possibilitam a especificação de consultas não geográficas mais avançadas.

Ferramenta	Formato	Ontologia	Dado Anotado	Componente Espacial
Agência de Informação Embrapa	XML, utilizando metadados Dublin Core	Não	Páginas textuais da Web, vídeos, imagens e documentos	Não
Amaya	XML, RDF	Não	Páginas textuais da Web	Não
Kim	RDF, OWL	Sim	Páginas textuais da Web	Não
AKTive	RDF	Sim	Páginas textuais da Web e imagens	Não
CREAM	RDF, OWL	Sim	Páginas textuais da Web, vídeos e imagens	Sim, mas muito limitado
E-Culture	RDF, OWL utilizando metadados VRA	Sim	Imagens	Sim
OnLocus	XML	Sim	Páginas textuais da Web	Sim
SPIRIT		Sim	Páginas textuais da Web	Sim
Geodata Annotation	XML, utilizando metadados ISO 19115	Sim	Dados geográficos	Sim

Tabela 2.3: Comparação entre ferramentas de anotação. Adaptada de [31].

Capítulo 3

Anotações Sobrepostas Associadas a Objetos Geográficos Vetoriais

Este capítulo apresenta a proposta para inserção, consulta, atualização e remoção de anotações sobrepostas associadas a objetos geográficos vetoriais. A seção 3.1 descreve o modelo e o dicionário de dados. A seção 3.2 define as interfaces para gerenciamento de anotações.

3.1 Modelagem de Dados

Esta seção apresenta detalhes da modelagem de dados para anotações sobrepostas de objetos geográficos vetoriais. Nela são apresentados o diagrama entidade-relacionamento e o dicionário de dados.

A modelagem de dados para a aplicação foi baseada nos conceitos de *Superimposed Information*. As entidades *Excerpt*, *Mark* , *SuperimposedInformation*, *Context* e *Element* foram inspiradas no modelo do SPARCE [38, 43]. A modelagem dos metadados seguiu atributos do modelo *Dublin Core* [18]. Dos quinze elementos do *Dublin Core* utilizados para descrição de recursos eletrônicos, quatro foram aplicados no modelo aqui proposto.

A modelagem dos dados espaciais seguiu o modelo da OGC [48]. O OGC define a entidade *Geometry* como a mais genérica, de acordo com a Figura 3.1. A geometria está associada a um Sistema de Referenciamento Espacial, que determina o sistema de coordenadas que está sendo utilizado. A geometria pode ser um ponto, uma curva, uma superfície ou uma coleção. A entidade *GeometryCollection* representa uma coleção de geometrias heterogêneas. As entidades *MultiPoint*, *MultiLineString*, *MultiPolygon*, *MultiCurve* e *MultiSurface* são coleções de geometrias homogêneas dos tipos *Point*, *LineString*, *Polygon*, *Curve* e *Surface*, respectivamente. A entidade *Point* representa uma localização simples no espaço de coordenadas. Possui coordenada representada por x e

y que referem-se respectivamente à sua abscissa e ordenada respectivamente. A entidade *LineString* representa uma seqüência de pontos conectados com uma interpolação linear entre dois pontos consecutivos. Cada par de pontos consecutivos define um segmento de linha. A entidade *Curve* é uma generalização de *LineString*. A entidade *Line* é uma especialização de *LineString* que possui apenas dois pontos. A entidade *LinearRing* representa uma *LineString* que possui as características de ser fechada e simples. A entidade *Polygon* representa um objeto geométrico bidimensional definido por um caminho e um contorno externo e uma ou mais contornos internos. A entidade *Surface* é uma generalização de *Polygon*. Os objetos espaciais do modelo de dados da aplicação contêm um atributo do tipo *Geometry* para representar sua geometria.

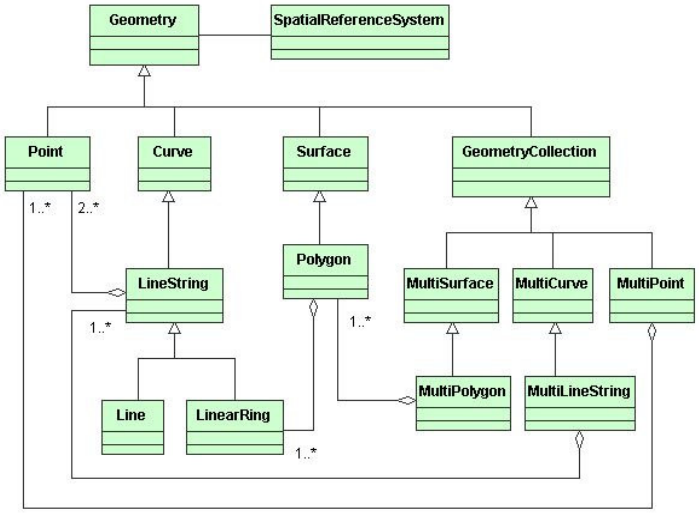


Figura 3.1: Modelo do OGC para representação de geometria.

Quando uma anotação é inserida, uma marca é criada. A marca é um meio de cruzar do espaço de informação sobreposta para elementos existentes no espaço de informação base [2].

Em se tratando de mapas, diferentes tipos de marcas podem ser criadas. A Figura 3.2 mostra quatro exemplos de tipos de marcas possíveis, para:

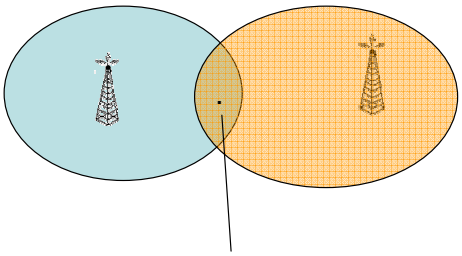
- Delimitar uma região contendo vários pontos ou linhas;
- Englobar uma região composta de partes de vários polígonos;

- Delimitar uma região pertencente à intersecção de dois polígonos;
- Sobrepor uma outra marca.

A Figura 3.2(a) ilustra a criação de uma marca em uma região de um mapa que contém pontos associados a cidades e linhas que representam estradas e rios. Na Figura 3.2(b), a marca criada no mapa do Brasil, engloba polígonos que representam os estados brasileiros. Na Figura 3.2(c), a marca está associada a região resultante da intersecção entre duas regiões que representam os raios de alcance de antenas de rede sem fio. A Figura 3.2(d), a marca sobrepõe uma mesma região do mapa marcada por outro usuário.

(a) Marca em uma região com vários pontos e linhas

(b) Marca em uma região com vários polígonos



(c) Marca em um ponto onde há intersecção de dois polígonos

(d) Marca de um usuário X — Marca de um usuário Y

Figura 3.2: Exemplos de tipos de marcas.

O diagrama entidade-relacionamento do modelo de dados proposto é apresentado na

Figura 3.3. As entidades são representadas por retângulos. Os relacionamentos são representados por losangos. As linhas representam as conexões entre as entidades e relacionamentos. Ao lado dos losangos é apresentada a cardinalidade do relacionamento.

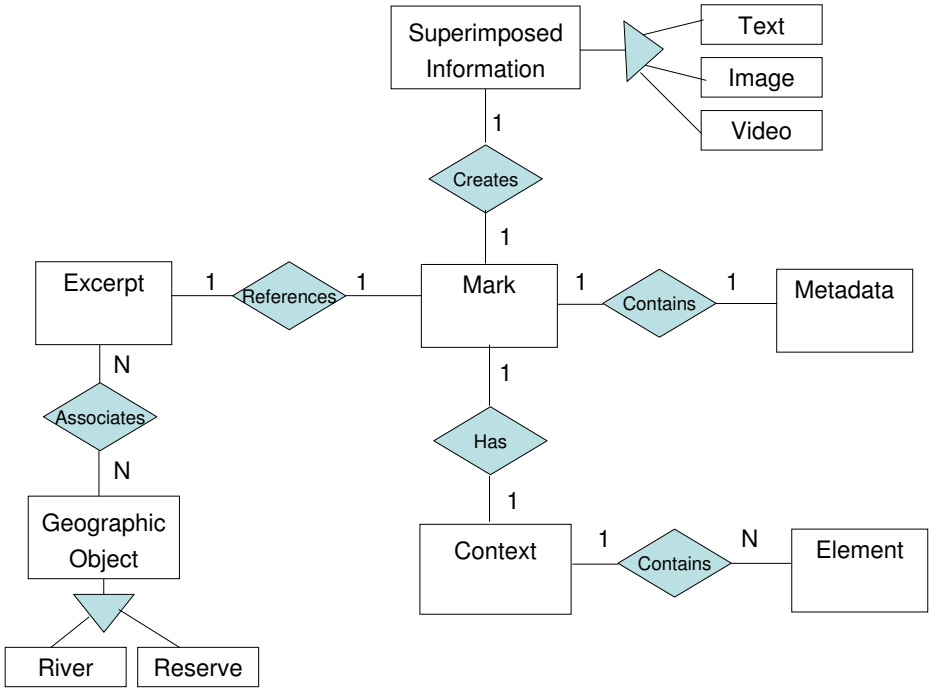


Figura 3.3: Diagrama entidade-relacionamento proposto para gerenciar anotações associadas a objetos geográficos vetoriais.

No dicionário de dados, as entidades e relacionamentos do diagrama entidade-relacionamento foram detalhados utilizando-se uma descrição e uma tabela de atributos.

GeographicObject: representa uma entidade do mundo real que possui coordenadas geográficas, como Cidade, Rio, Reserva Ecológica, Floresta. É uma entidade genérica que possui os atributos comuns a todo objeto geográfico. Os atributos específicos de um determinado tipo de objeto geográfico serão armazenados em tabelas próprias. A coluna Type indica qual é o tipo de objeto geográfico que está armazenado na linha da tabela. No diagrama entidade-relacionamento foi ilustrado como as entidades River e Forest, que representam entidades Rios e Florestas e que possuem dados específicos que não existem

em GeographicObject, poderiam ser adicionadas ao modelo de dados. Os atributos da entidade estão descrito na Tabela 3.1.

Atributo	Descrição
IdGeographicObject	Chave primária da entidade.
Name	Nome do Objeto Geográfico.
Type	Tipo do Objeto Geográfico.
The_Geometry	Geometria do objeto geográfico. Atributo usado para armazenar as coordenadas geográficas e tipo de objeto geográfico (ponto, linha, etc).

Tabela 3.1: Atributos da entidade *GeographicObject*.

Excerpt: conteúdo da região marcada. No caso de um mapa, o conteúdo da região marcada pode ser um ponto, uma linha, parte de uma linha, um polígono, parte de um polígono, vários polígonos, parte de vários polígonos, uma coleção de objetos diferentes como pontos, linhas, polígonos ou parte deles. Os atributos da entidade estão descrito na Tabela 3.2.

Atributo	Descrição
IdExcerpt	Chave primária da entidade.
Mark_IdMark	Chave estrangeira para a entidade Mark.

Tabela 3.2: Atributos da entidade *Excerpt*.

A região marcada pode englobar vários objetos geográficos. Um objeto geográfico pode estar contido em mais de uma região marcada. Existe então a necessidade da criação de um relacionamento $n : n$ entre as entidades GeographicObject e Excerpt.
A entidade Excerpt também se relaciona com Mark. Uma marca sempre referencia um conteúdo.

Mark: representa a referência para o elemento da camada base. Neste caso, o elemento da camada base é o objeto geográfico. Os atributos da entidade estão descrito na Tabela 3.3.

Atributo	Descrição
IdMark	Chave primária da entidade.
The_Geometry	Geometria da marca.

Tabela 3.3: Atributos da entidade *Mark*.

A marca possui relacionamento com contexto, metadados e informação sobreposta. Toda marca possui um contexto. O contexto contém um ou vários elementos de contexto. Toda marca contém metadados, que são as informações complementares, como autor, data, etc. A marca é criada quando uma informação sobreposta é inserida.

Context: informação relacionada à região marcada. Existem vários tipos de contexto, dependendo do conteúdo da informação base. Se a informação base for um documento, o contexto pode ser o número da página, parágrafo, fonte, tamanho da fonte, etc. Diferentes componentes para manipulação das informações de contexto foram criados, como componentes para Word, Acrobat, etc. Se a informação base for um mapa, ou objetos geográficos, o contexto será composto pelas características que o mapa deve apresentar ao ser exibido. Os atributos da entidade estão descrito na Tabela 3.4.

Atributo	Descrição
IdContext	Chave primária da entidade.
Mark_IdMark	Chave estrangeira para a entidade Mark.

Tabela 3.4: Atributos da entidade *Context*.

Element: elementos do contexto, como por exemplo no caso de mapas, nível de zoom utilizado no momento da anotação. Nesta tabela, cada elemento de contexto é armazenado em uma linha que contém seu tipo e valor. Os atributos da entidade estão descrito na Tabela 3.5.

Metadata: Metadados sobre a marca baseado no padrão Dublin Core. Os atributos da entidade estão descrito na Tabela 3.6.

SuperimposedInformation: informação sobreposta ou nova informação associada a informação base. Pode estar no formato texto, imagem, vídeo. É uma entidade genérica que possui os atributos comuns a todo objeto *Superimposed Information*. Os atributos específicos de um determinado tipo de *Superimposed Information* serão armazenados em tabelas próprias. A coluna Type indica qual é o tipo de *Superimposed Information* que

Atributo	Descrição
IdElement	Chave primária da entidade.
Context_IdContext	Chave estrangeira para a entidade Context.
Context_Type	Tipo de elemento de contexto.
Context_Value	Valor do elemento de contexto.

Tabela 3.5: Atributos da entidade *Element*.

Atributo	Descrição
IdMetadata	Chave primária da entidade.
Mark_IdMark	Chave estrangeira para a entidade Mark.
Author	Responsável pela criação da *Superimposed Information*.
AnnotationDate	Data da criação da *Superimposed Information*.
Language	Idioma utilizado na criação da *Superimposed Information*.
Rights	Direitos sobre a *Superimposed Information*.

Tabela 3.6: Atributos da entidade *Metadata*.

está armazenado na linha da tabela. Os atributos da entidade estão descrito na Tabela 3.7.

Esta dissertação trata somente informações sobrepostas no formato texto. Por isso, somente a tabela Text foi modelada. Quando uma informação sobreposta é inserida, uma anotação é criada. Nesta tabela são armazenadas as anotações no formato texto. Os atributos da entidade estão descrito na Tabela 3.8.

No modelo físico gerado a partir deste diagrama entidade-relacionamento, foram criadas duas tabelas: Rios e Reservas. Elas foram utilizadas para armazenar os dados dos objetos geográficos pertencentes ao protótipo desenvolvido para a área de biodiversidade. Foi criada também uma tabela chamada GeographicObject_has_Excerpt para resolver o relacionamento $n : n$ entre as entidades GeographicObject e Excerpt. Os atributos da entidade estão descrito na Tabela 3.9.

Atributo	Descrição
IdSuperimposedInformation	Chave primária da entidade.
Mark_IdMark	Chave estrangeira para a entidade Mark.
Type	Tipo da informação sobreposta.

Tabela 3.7: Atributos da entidade *SuperimposedInformation*.

Atributo	Descrição
IdText	Chave primária da entidade.
SuperimposedInformation _IdSuperimposedInformation	Chave estrangeira para a entidade SuperimposedInformation.
Annotation	Anotação no formato texto.

Tabela 3.8: Atributos da entidade *Text*.

3.2 Serviço Web para Gerenciamento de Anotações

As interfaces das operações de inserção, atualização, remoção e consulta de anotações estão detalhadas nesta seção. Foi definida uma interface genérica para manipulação de anotações e derivações dela para simplificar o seu uso.

3.2.1 Interface Genérica para Manipulação de Anotações

A interface genérica para manipulação (inserção, atualização, remoção e consulta) de anotações é a seguinte:

< *retorno* > *operation* (**Text** Anotacao, **Query** Consulta, **Metadata** Metadados, **Context** Contexto)

onde *operation* é a operação a ser realizada e pode ser do tipo *InsertAnnotation, QueryAnnotation, DeleteAnnotation* ou *UpdateAnnotation*, que referem-se respectivamente às operações de inserção, consulta, remoção ou atualização de anotações. O parâmetros de entrada Anotacao, do tipo *Text*, é o conteúdo textual da anotação. O parâmetro Consulta, do tipo *Query*, define a consulta que será executada e a partir da qual são recuperados os objetos geográficos que serão associados às anotações. Matias [36] define que, de maneira geral, uma consulta pode ser expressa da seguinte forma:

Atributo	Descrição
IdGeographicObject_has_Excerpt	Chave primária da entidade.
Excerpt_IdExcerpt	Chave estrangeira para a entidade Excerpt.
GeographicObject_IdGeographicObject	Chave estrangeira para a entidade GeographicObject.
The_Geometry	Geometria da região marcada do objeto geográfico.

Tabela 3.9: Atributos da entidade *GeographicObject_has_Excerpt*.

SELECT <resultado>
FROM <fonte de dado>
WHERE <predicado>

Essa sintaxe indica que serão recuperados resultados (ou dados) a partir de uma fonte de dados (por exemplo, tabelas ou visões) que satisfaçam a um predicado (conjunto de condições). Os componentes *fonte de dados, predicado* e *resultado* dependem diretamente dos dados a que estão ligados e podem ser dos tipos Convencional, Espacial ou Misto. A fonte de dados convencional é aquela que contém dados alfanuméricos tradicionais. A fonte de dados espacial contém dados que descrevem as características espaciais de uma entidade geográfica. A fonte de dados mista contém dados do tipo convencional e espacial.

O tipo de um predicado é definido em função dos operadores utilizados. 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. O resultado é definido em função do tipo de dados retornado. O resultado é convencional quando somente dados alfanuméricos tradicionais são retornados. O resultado é espacial quando são retornados somente dados espaciais. O resultado é misto quando são retornados dados convencionais e espaciais.

Utilizando essa classificação, é possível definir um objeto do tipo consulta, com o qual diversas combinações podem ser feitas entre os tipos e componentes. A interface para manipulação (inserção, atualização, remoção e consulta) de anotação de objetos geográficos se torna bastante flexível com o uso de um parâmetro do tipo consulta.

O parâmetro Metadados, do tipo *Metadata*, é um objeto que contém atributos como nome do autor, data, idioma, conforme entidade descrita no dicionário de dados (ver seção 3.1). O parâmetro Contexto, do tipo *Context*, é um objeto que contém as características do mapa que foi marcado, conforme entidade descrita no dicionário de dados (ver seção 3.1).

Esta interface genérica define uma operação entre uma anotação em um conjunto de objetos que façam parte do resultado da Consulta. Por exemplo, se for atribuído o comando SQL "SELECT IdGeographicObject FROM GeographicObject WHERE Name = 'Campinas'" ao parâmetro Consulta, e for atribuído valor "InsertAnnotation" a operation, será inserida uma anotação no objeto que representa a cidade de Campinas.

3.2.2 Inserção de Anotações

Esta seção define as interfaces para inserção de anotações. As interfaces são derivadas da interface genérica. Os diferentes tipos de parâmetros definidos servem para auxiliar o uso da interface.

- **boolean** *InsertAnnotation* (**Text** Anotacao, **Geometry** Regiao, **Metadata** Metadados, **Context** Contexto)

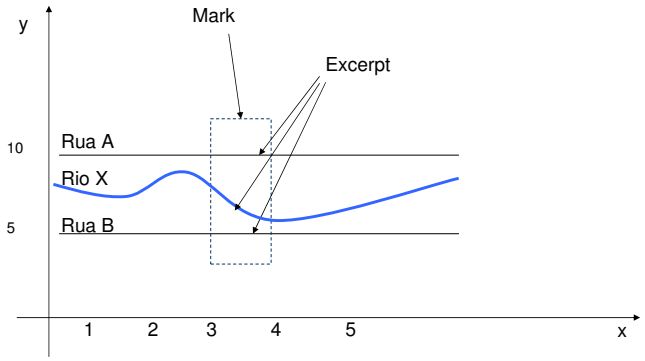
  Esta operação insere uma anotação em um conjunto de objetos geográficos que estejam contidos dentro de uma região. Devem ser considerados objetos totalmente ou parcialmente contidos dentro da região. Objeto totalmente contido é aquele cuja área encontra-se dentro da região marcada. Objeto parcialmente contido é aquele cuja parte da área está dentro da região marcada. No caso de objetos parcialmente contidos, somente a parte que estiver dentro da região marcada será associada à anotação.

  Esta operação é derivada da interface genérica. Ao ser executada, gera uma consulta conforme padrão a seguir, em que são selecionados os identificadores dos objetos (resultado convencional). Duas fontes de dados são utilizadas, uma convencional, para a busca dos identificadores e uma espacial, para recuperar os objetos que estão contidos na Região. O predicado é espacial, pois a condição é buscar os identificadores dos objetos que estão contidos em uma região.

  SELECT <resultado> (resultado convencional - identificação dos objetos geográficos, ou seja, IdGeographicObject)
  FROM <fonte de dado> (mista)
  WHERE <predicado> (espacial)

  A Figura 3.4 ilustra um exemplo em que é feita uma marca (Mark) em uma região de um mapa contendo duas ruas (A e B) e um rio (X), para inserção de uma anotação. As tabelas indicam os valores armazenados no banco de dados. A primeira tabela contém os dados e coordenadas das ruas A e B e do rio X. A segunda tabela contém

os identificadores e coordenadas dos objetos que estão dentro da marca, ou seja, o conteúdo da região marcada (Excerpt). A anotação associa-se com o identificador do conteúdo da região marcada (IdExcerpt). Para simplificação, as coordenadas foram representadas no plano cartesiano.



IdGeographic Object	Name	Type	The_Geometry
1	A	Rua	0,10; 10,10
2	B	Rua	0,5; 10,5
3	X	Rio	0,7; 10,7

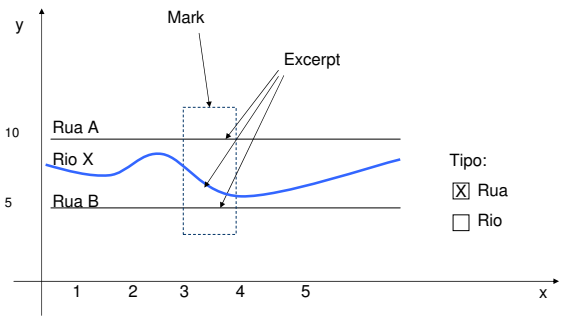
IdGeographic Object_has_ Excerpt	IdExcerpt	Id Geographic Object	The_ Geometry
101	100	1	3,10; 4,10
102	100	2	3,5; 4,5
103	100	3	3,7; 4,6

Figura 3.4: Exemplo de *InsertAnnotation* (Text Anotacao, Geometry Regiao, Metadata Metadados, Context Contexto).

- **boolean** *InsertAnnotation* (**Text** Anotacao, **Geometry** Regiao, **List** {ObjectType Tipo}, **Metadata** Metadados, **Context** Contexto)

  Esta operação insere uma anotação em um conjunto de objetos geográficos que estejam contidos dentro de uma região e que sejam de um (ou mais) determinado tipo. Por exemplo, se Type for igual a "Rua", inserir uma anotação em todas as ruas de uma região.

  A consulta que será executada é similar à consulta anterior. A diferença é que, nesse caso, o predicado é misto, uma vez que além de recuperar os identificadores dos objetos geográficos, será necessário recuperar seus atributos alfanuméricos, a fim de retornar somente objetos de determinados tipos.



IdGeographic Object	Name	Type	The_Geometry
1	A	Rua	0,10; 10,10
2	B	Rua	0,5; 10,5

IdGeographic Object_has_ Excerpt	IdExcerpt	Id Geographic Object	The_ Geometry
101	100	1	3,10; 4,10
102	100	2	3,5; 4,5

Figura 3.5: Exemplo de *InsertAnnotation* (Text Anotacao, Geometry Regiao, List {ObjectType Tipo}, Metadata Metadados, Context Contexto).

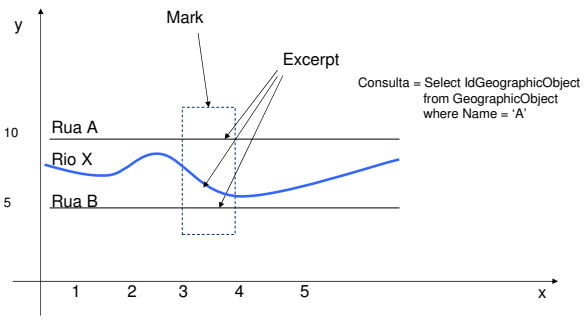
SELECT <resultado> (resultado convencional - identificação dos objetos geográficos, ou seja, IdGeographicObject)
FROM <fonte de dado> (mista)
WHERE <predicado> (misto)

No exemplo da Figura 3.5, somente os objetos do tipo Rua serão associados à anotação.

- **boolean** *InsertAnnotation* (**Text** Anotacao, **Geometry** Regiao, **Query** Consulta, **Metadata** Metadados, **Context** Contexto)

  Esta operação insere uma anotação em um conjunto de objetos geográficos que estejam contidos dentro de uma região e façam parte do resultado da Consulta.

  Na Figura 3.6, a consulta recupera o identificador da Rua A. Assim, somente a Rua A é associada com a anotação.



IdGeographic Object	Name	Type	The_Geometry
1	A	Rua	0,10; 10,10

IdGeographic Object_has_ Excerpt	IdExcerpt	Id Geographic Object	The_ Geometry
101	100	1	3,10; 4,10

Figura 3.6: Exemplo de *InsertAnnotation* (Text Anotacao, Geometry Regiao, Query Consulta, Metadata Metadados, Context Contexto).

- **boolean** *InsertAnnotation* (**Text** Anotacao, **GeoQuery** GeoConsulta, **Metadata** Metadados, **Context** Contexto)

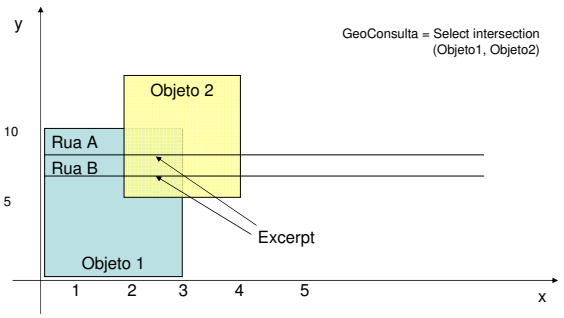
  Esta operação insere uma anotação associada aos objetos geográficos que estejam contidos em uma região que faça parte do resultado de uma consulta geográfica.

  SELECT <resultado> (resultado convencional - identificação dos objetos geográficos, ou seja, IdGeographicObject)
  FROM <fonte de dado> (mista)
  WHERE <predicado> (misto)

  Na Figura 3.7, a anotação é associada aos objetos que estão contidos na intersecção do Objeto 1 e do Objeto 2, que são dois polígonos.

- **boolean** *InsertAnnotation* (**Text** Anotacao, **GeoQuery** GeoConsulta, **List** {ObjectType Tipo}, **Metadata** Metadados, **Context** Contexto)

  Esta operação insere uma anotação relacionada aos objetos geográficos que estejam contidos em uma região que faça parte do resultado de uma consulta geográfica e



IdGeographic Object	Name	Type	The_Geometry
1	A	Rua	0,9; 10,9
2	B	Rua	0,8; 10,8

IdGeographic Object_has_ Excerpt	IdExcerpt	Id Geographic Object	The_ Geometry
101	100	1	2,9; 3,9
102	100	2	2,8; 3,8

Figura 3.7: Exemplo de *InsertAnnotation* (Text Anotacao, GeoQuery GeoConsulta, Metadata Metadados, Context Contexto).

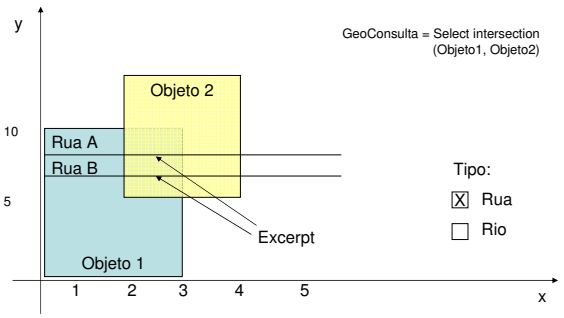
que sejam de determinados tipos.

Na Figura 3.8, a anotação é associada aos objetos do tipo igual a "Rua" que estão contidos na intersecção do Objeto 1 e do Objeto 2.

- **boolean** *InsertAnnotation* (**Text** Anotacao, **GeoQuery** GeoConsulta, **Query** Consulta, **Metadata** Metadados, **Context** Contexto)

  Esta operação insere uma anotação associada aos objetos geográficos que estejam contidos em uma região que faça parte do resultado de uma consulta geográfica e que façam parte do resultado de uma consulta não geográfica.

  Na Figura 3.9, a anotação é associada ao objeto Rua A que está contido na intersecção do Objeto 1 e do Objeto 2.



IdGeographic Object	Name	Type	The_Geometry
1	A	Rua	0,9; 10,9
2	B	Rua	0,8; 10,8

IdGeographic Object_has_ Excerpt	IdExcerpt	Id Geographic Object	The_ Geometry
101	100	1	2,9; 3,9
102	100	2	2,8; 3,8

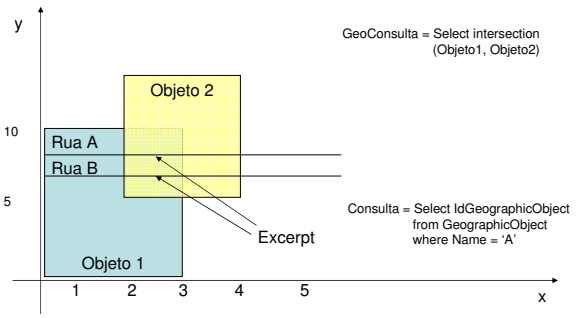
Figura 3.8: Exemplo de *InsertAnnotation* ((Text Anotacao, GeoQuery GeoConsulta, List {ObjectType Tipo}, Metadata Metadados, Context Contexto).

3.2.3 Consultas de Anotações

Esta seção descreve as interfaces para consulta de anotações. As interfaces são derivadas da interface genérica. Os diferentes tipos de parâmetros definidos servem para auxiliar o uso da interface.

- **List** {**Text, Metadata** } *QueryAnnotation* (**List** {**MetadataField** Campo, **Operator** Operador, **MetadataValue** Valor, **Connector** Conector})
  Esta operação retorna os textos e metadados das anotações cujos metadados coincidam com os parâmetros informados. O parâmetro de entrada desta operação é uma lista com os nomes dos metadados a serem utilizados na consulta, os operadores a serem usados nas condições (por exemplo, =, <, >) e seus respectivos valores. Os conectores (por exemplo, AND, OR) fazem as ligações entre as condições.

- **List** {**Text, Metadata** } *QueryAnnotation* (**List** {**String** PalavraChave})
  Esta operação retorna os textos das anotações que contenham as palavras-chaves informadas.



IdGeographic Object	Name	Type	The_Geometry
1	A	Rua	0,9; 10,9
2	B	Rua	0,8; 10,8

IdGeographic Object_has_ Excerpt	IdExcerpt	Id Geographic Object	The_ Geometry
101	100	1	2,9; 3,9
102	100	2	2,8; 3,8

Figura 3.9: Exemplo de *InsertAnnotation* ((Text Anotacao, GeoQuery GeoConsulta, Query Consulta, Metadata Metadados, Context Contexto).

- **List** {**Text, Metadata** } *QueryAnnotation* (**String** NomeObj)
  Esta operação retorna os textos das anotações que estejam associadas ao objeto geográfico cujo nome foi informado no parâmetro NomeObj.

- **List** {**Text, Metadata** } *QueryAnnotation* (**Geometry** Regiao)
  Esta operação retorna os textos das anotações de uma região. Podem ser retornadas anotações de diversos autores e datas. Os metadados auxiliam a diferenciar cada anotação.

- **List** {**Text, Metadata** } *QueryAnnotation* (**Query** Consulta)
  Esta operação retorna os textos das anotações de objetos que façam parte do resultado da Consulta (não geográfica).

- **List** {**Text, Metadata** } *QueryAnnotation* (**GeoQuery** GeoConsulta)
  Esta operação retorna os textos das anotações de objetos geográficos que estejam contidos em uma região que faça parte do resultado de uma Consulta Geográfica.

- **MarkGeometry** *QueryMark* (**Text** Anotacao)
  Esta operação retorna a geometria da marca referente a uma anotação.

As consultas acima definidas podem ser combinadas, dando origem a novos tipos de consultas complexas. Uma consulta complexa pode ser realizada com a combinação dos resultados de duas ou mais consultas simples. Por exemplo, suponha que um biólogo deseja realizar a seguinte consulta: "Retorne todas as anotações do biólogo Arlindo que contenham as palavras armadilhas e *Anaea ryphea*". Neste caso, pode-se executar a operação de consulta que recebe como parâmetro uma lista de metadados e seus valores. Nos parâmetros, passar como MetaDataField o valor "Autor", o operador "=" e como MetaDataValue o valor "Arlindo" e guardar o resultado. Em seguida executar a operação de consulta que recebe uma lista de palavras-chave. Nos parâmetros passar "armadilhas","*Anaea ryphea*" e guardar o resultado. Como ambas operações retornam listas de anotações e metadados, percorrer as duas listas, separando em uma nova lista os elementos que aparecem nas duas.

3.2.4 Remoção de Anotações

Esta seção define as interfaces para remoção de anotações. As interfaces são derivadas da interface genérica. Os diferentes tipos de parâmetros definidos servem para auxiliar o uso da interface.

- **boolean** *DeleteAnnotation* (**List** {**MetadataField** Campo, **Operator** Operador, **MetadataValue** Valor, **Connector** Conector})
  Esta operação remove as anotações cujos metadados coincidam com os parâmetros informados. O parâmetro de entrada desta operação é uma lista com os nomes dos metadados a serem utilizados na consulta, os operadores a serem usados nas condições (por exemplo, =, <, >) e seus respectivos valores. Os conectores (por exemplo, AND, OR) fazem as ligações entre as condições.

- **boolean** *DeleteAnnotation* (**List** {**String** PalavraChave})
  Esta operação remove as anotações que contenham as palavras-chaves informadas.

- **boolean** *DeleteAnnotation* (**Geometry** Regiao)
  Esta operação remove as anotações de uma região.

- **boolean** *DeleteAnnotation* (**Query** Consulta)
  Esta operação remove as anotações de objetos geográficos que estejam contidos em uma região que faça parte do resultado de uma consulta não geográfica definida no parâmetro Consulta.

- **boolean** *DeleteAnnotation* (**GeoQuery** GeoConsulta)
  Esta operação remove as anotações de objetos geográficos que estejam contidos em uma região que faça parte do resultado de uma consulta geográfica definida no parâmetro GeoConsulta.

Todos os parâmetros das operações podem ser combinados, dando origem a novos tipos de operações.

3.2.5 Atualização de Anotações

Nesta seção foram definidas as interfaces para atualização de anotações. As interfaces são derivadas da interface genérica. Os diferentes tipos de parâmetros definidos servem para auxiliar o uso da interface.

- **boolean** *UpdateAnnotation* (**Text** novaAnotacao, **List** {**MetadataField** Campo, **Operator** Operador, **MetadataValue** Valor, **Connector** Conector})
  Esta operação atualiza as anotações cujos metadados coincidam com os parâmetros informados. O parâmetro de entrada desta operação é uma lista com os nomes dos metadados a serem utilizados na consulta, os operadores a serem usados nas condições (por exemplo, =, <, >) e seus respectivos valores. Os conectores (por exemplo, AND, OR) fazem as ligações entre as condições.

- **boolean** *UpdateAnnotation* (**Text** novaAnotacao, **List** {**String** PalavraChave}))
  Esta operação atualiza as anotações que contenham as palavras-chaves informadas.

- **boolean** *UpdateAnnotation* (**Text** novaAnotacao, **Geometry** Regiao)
  Esta operação atualiza as anotações de uma região.

- **boolean** *UpdateAnnotation* (**Text** novaAnotacao, **Query** Consulta)
  Esta operação atualiza as anotações de objetos geográficos que estejam contidos em

uma região que faça parte do resultado de uma consulta não geográfica.

- **boolean** *UpdateAnnotation* (**Text** novaAnotacao, **GeoQuery** GeoConsulta)
  Esta operação atualiza as anotações de objetos geográficos que estejam contidos em uma região que faça parte do resultado de uma consulta geográfica.

- **boolean** *UpdateAnnotation* (**Text** antigaAnotacao, **Text** novaAnotacao)
  Esta operação atualiza as anotações trocando o(s) termo(s) contido(s) em antigaAnotacao por novaAnotacao.

- **boolean** *UpdateMark* (**MarkGeometry** novaMarca, **MarkGeometry** antigaMarca)
  Esta operação atualiza as coordenadas da marca referente a uma anotação. Nesse caso, dependendo da forma de armazenamento do conteúdo da região marcada, é necessário reassociar os objetos associados a marca.

Todos os parâmetros das operações podem ser combinados, dando origem a novos tipos de operações.

3.2.6 Uso de Anotações Semânticas

Esta dissertação trata informações sobrepostas no formato texto. Por isso, existe uma tabela chamada *Text* no modelo de dados e as operações para gerenciamento das anotações possuem parâmetros de entrada e saída do tipo *Text*. O modelo de dados e as operações podem ser adaptados para tratarem outros formatos, como imagens ou vídeos.

Uma possível adaptação seria integrar o serviço de anotação a um serviço e uma fonte de ontologias e estender o modelo e operações para permitir o uso de anotações semânticas.

Uma opção para tratar anotações semânticas no modelo de dados seria armazenar a anotação no formato RDF/XML. Muitos gerenciadores de bancos de dados suportam formato XML, como é o caso do PostgreSQL. Os dados podem ser recuperados utilizando-se linguagens como XPath e XQuery. Neste caso, no modelo de dados deve ser criada uma tabela nova, que estende da tabela *SuperimposedInformation* e que possua uma coluna do tipo text. Considerando a interface genérica para manipulação de anotação, o parâmetro do tipo *Text* deve ser substituído pelo tipo XML.

< *retorno* > *operation* (**XML** AnotacaoSemantica, **Query** Consulta, **Metadata** Metadados, **Context** Contexto)

Outra opção seria armazenar diretamente as triplas que contêm informações sobre o sujeito, o predicado e o objeto. O sujeito é o objeto geográfico, o predicado e objeto são as informações da anotação. Para armazenar as triplas, seria necessário criar uma nova tabela no modelo de dados associada a tabela GeographicObject. Nesta tabela, haveria uma chave estrangeira para a tabela GeographicObject, para relacionar o sujeito. O predicado e objeto seriam duas colunas desta tabela e armazenariam os tipos e valores das anotações.

Capítulo 4

Aspectos de Implementação

Este capítulo descreve o protótipo desenvolvido e as tecnologias utilizadas visando à validação da proposta de inserção e consulta de anotações associadas a objetos geográficos vetoriais. Dentre as operações definidas na proposta da solução, foram implementados as mais importantes para atender um cenário de uso na área de biodiversidade.

4.1 Arquitetura

A Figura 4.1 mostra a arquitetura do serviço Web implementado. Esta arquitetura é composta de 3 camadas: Interface com Usuário, Serviços e Persistência.

A camada de Interface com Usuário é responsável pela interação com o usuário. Por meio dela ocorrem a entrada e saída de dados e são invocadas as operações que o sistema disponibiliza. A arquitetura permite que o componente da camada de Interface com Usuário seja uma interface Web ou um componente de um outro sistema. No protótipo, a camada de Interface com Usuário é executada em um navegador Web.

A camada de Serviços disponibiliza as APIs do sistema, ou seja, o conjunto de interfaces de programação que define como suas funcionalidades podem ser invocadas por outros aplicativos. A camada de serviços é acessada via servidor de aplicação. Os serviços GEO estão disponíveis a partir de um servidor de mapas.

Na camada de persistência localizam-se os bancos de dados. A camada de persistência é responsável pelo armazenamento, processamento de consultas e integridade dos dados.

As tecnologias utilizadas em cada camada da arquitetura podem ser vistas na Figura 4.2. Os detalhes de cada camada serão descritos nas seções a seguir.

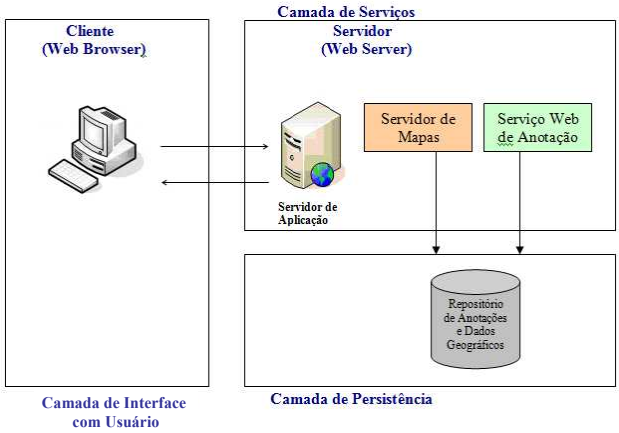


Figura 4.1: Arquitetura do serviço Web de criação e consulta de anotações.

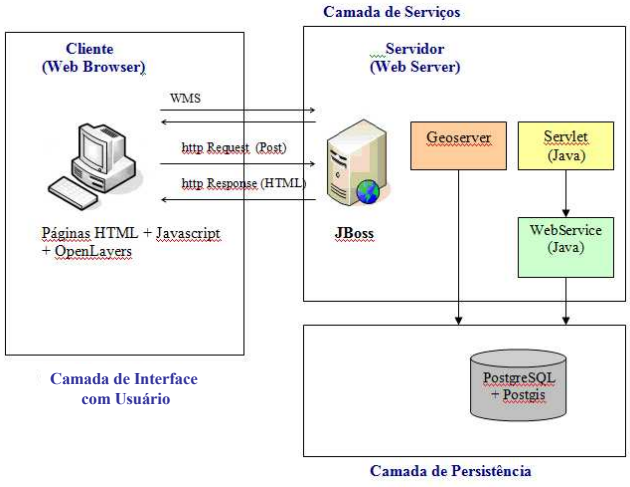


Figura 4.2: Tecnologias utilizadas na arquitetura.

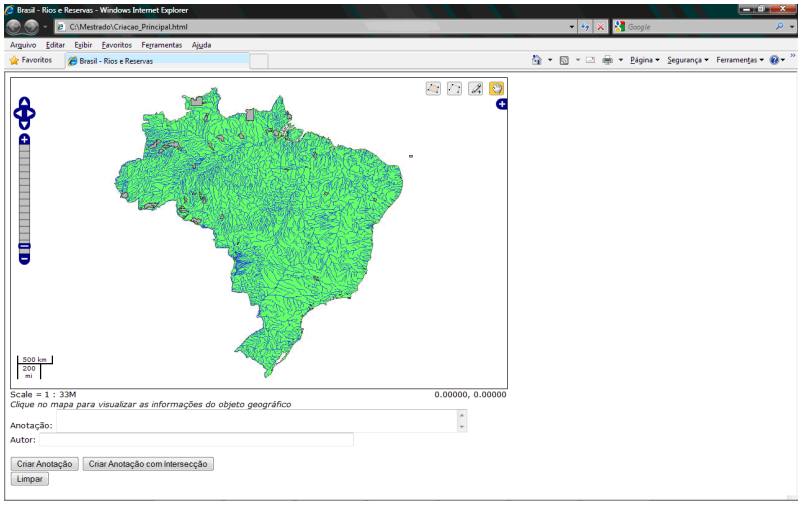


Figura 4.3: Tela de entrada da aplicação de inserção e consulta de anotação.

4.1.1 Cliente

A camada cliente da aplicação é composta de páginas HTML e Javascript executadas no Web Browser. A renderização dos mapas é realizada no servidor de mapas e a apresentação é feita utilizando-se o arcabouço OpenLayers. OpenLayers é uma biblioteca JavaScript Open Source utilizada para exibir dados espacias em páginas Web. Através da API do OpenLayers, é possível disponibilizar diversas ferramentas para interação visual com os mapas ou execução de ações em determinados eventos. A página HTML invoca um Servlet passando a operação desejada e os parâmetros necessários.

A Figura 4.3 exibe a tela principal da aplicação. A aplicação carrega o mapa do Brasil e dois tipos de objetos geográficos: rios e reservas ecológicas. A barra de ferramentas de desenho localizada no canto superior direito da tela fornece opções para que o usuário desenhe polígonos, linhas, pontos ou navegue visualmente pelo mapa (pan).

Abaixo da barra de ferramentas localiza-se o seletor de objetos geográficos, conforme Figura 4.4. Utilizando-se desse seletor é possível habilitar ou desabilitar a apresentação de um determinado tipo de objeto geográfico.

Na Figura 4.5 os objetos geográficos do tipo rios foram desabilitados, permanecendo no mapa somente os objetos do tipo reservas. Na Figura 4.6, os objetos do tipo reservas foram desabilitados, restando os objetos do tipo rios.

Do lado direito da tela, encontram-se os controles para permitir ajustar os níveis de

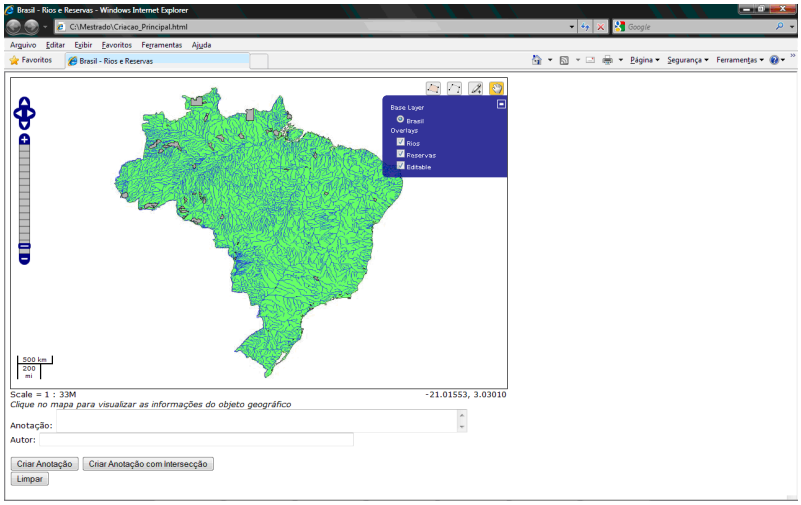


Figura 4.4: Tela com o seletor de tipo de objeto geográfico a ser apresentado.

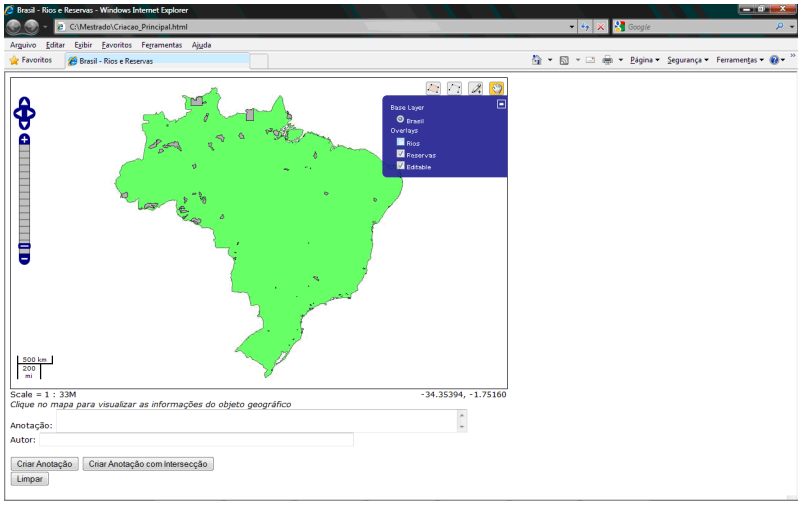


Figura 4.5: Tela com o tipo de objeto geográfico Rios desabilitado.

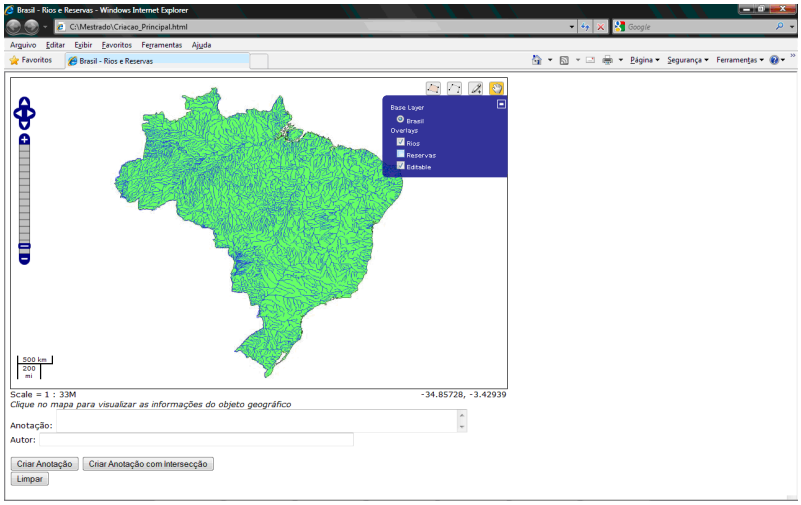


Figura 4.6: Tela com o tipo de objeto geográfico Reservas desabilitado.

zoom e clicar e arrastar o mapa. Logo abaixo do mapa, há controles para visualização da escala atual e das coordenadas do mapa referentes à posição onde encontra-se o mouse do usuário.

Para visualizar as informações alfanuméricas de um objeto geográfico, basta clicar nele. Uma tabela com as informações será apresentada abaixo do mapa. As Figuras 4.7 e 4.8 mostram respectivamente exemplos de dados alfanuméricos dos objetos Rio dos Peixes e da Reserva Chico Mendes.

4.1.2 Servidor

O servidor utilizado para publicação dos dados geográficos foi o Geoserver. O GeoServer é um servidor de mapas que segue as especificações de padrões abertos do OGC, tais como Web Map Service (WMS), Web Coverage Service (WCS) e Web Feature Service (WFS-T).

A Figura 4.9 mostra algumas telas da interface administrativa do Geoserver. A primeira é a tela de entrada, na qual após fazer o login, são disponibilizados os menus com as funcionalidades do Geoserver. A segunda é a tela de conexão com o banco de dados. Nesta tela são configurados os parâmetros para conexão com o banco de dados. É possível testar se a conexão está corretamente configurada. A terceira é a tela de configuração dos *Feature Types*.

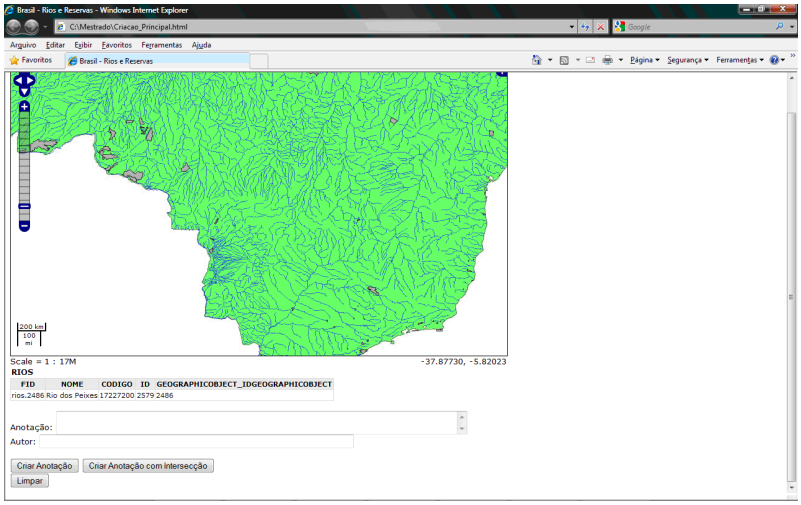


Figura 4.7: Tela com os detalhes alfanuméricos de um objeto do tipo Rio.

Após conectar com o banco, pode-se configurar o estilo dos objetos geográficos existentes no banco. Pela interface administrativa do Geoserver é possível conectar no banco PostgreSQL, configurar os *Feature Types* importados e disponibilizá-los através dos serviços OGC.

Além do Geoserver, no servidor também encontram-se o Servlet e o serviço Web que fornecem as funcionalidades para incluir e consultar anotações. Servlet é um programa em Java que estende a funcionalidade de um web server, gerando conteúdo dinâmico e interagindo com os clientes, utilizando o modelo request/response. Os Servlets não são restritos ao modelo HTTP de request/response, onde na realidade são pequenas aplicações de servidores, mas o modelo HTTP é o modelo mais comumente utilizado. O Servlet implementado recebe os parâmetros enviados na requisição HTTP e invoca o método do serviço Web correspondente à operação desejada, passando os parâmetros necessários. Como resposta, retorna uma página HTML.

O serviço Web também foi implementado na linguagem Java e utiliza a API para serviços Web, JAX-WS[1], que já vem no JDK6. O serviço Web possui métodos para inserção e consulta de anotações. Conecta-se ao banco PostgreSQL e executa funções criadas no banco para inserções das anotações. As funções utilizam outras funções espaciais disponíveis no próprio PostgreSQL. No protótipo, o servidor de aplicação JBoss[2] foi

[1]http://java.sun.com/developer/technicalArticles/J2SE/jax_ws_2
[2]http://www.jboss.org

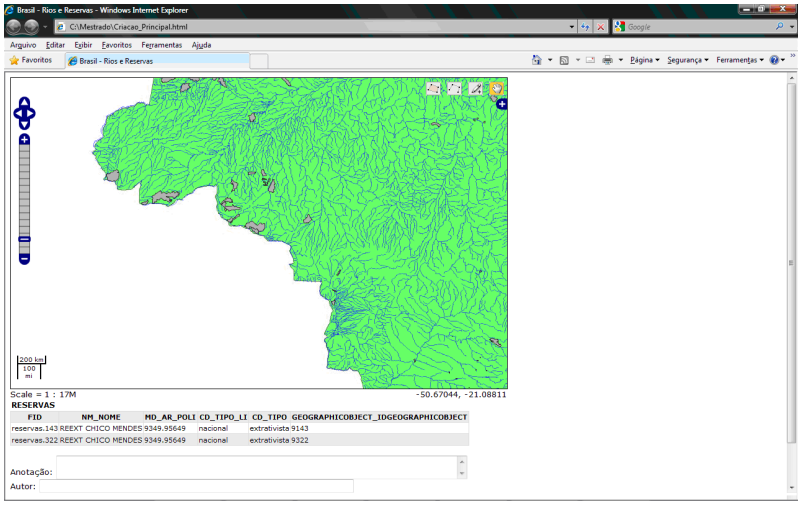


Figura 4.8: Tela com os detalhes alfanuméricos de um objeto do tipo Reserva.

utilizado para disponibilizar as operações de anotação e os serviços GEO.

4.1.3 Banco de Dados

O SGBD escolhido para ser utilizado na aplicação foi o PostgreSQL. Trata-se de um SGBD gratuito, de código aberto, bastante conhecido e que possui uma extensão para gerenciamento de dados espaciais, o PostGIS. Inicialmente o banco de dados foi carregado com dados obtidos de mapas digitais do Brasil no formato Shapefile distribuídos em [23]. Os temas referentes a Recursos Naturais e Meio Ambiente escolhidos foram: Hidrografia, Áreas de Reservas e Parques Nacionais e Estaduais.

Foram criadas tabelas para armazenamento das anotações seguindo o modelo de dados definido no Capítulo 3. Os objetos geográficos Rios, Reservas e Parques, importados dos arquivos *Shapefile*, estendem a entidade *GeographicObject*. Para isso, em cada tabela foi criada uma coluna nova chamada *geographicobject_idgeographicobject* que é uma referência para a tabela *GeographicObject*.

Para demonstrar como realizar uma integração entre a aplicação e um sistema de registro de coletas, foram criadas tabelas de registro de coletas e catálogos de espécies idênticas às do BioCORE definidas em [33], no mesmo banco de dados da aplicação. As tabelas foram populadas com dados sobre *Lepidopteras*. Além das tabelas, visões e funções também foram criadas no banco de dados para encapsular a recuperação e manipulação

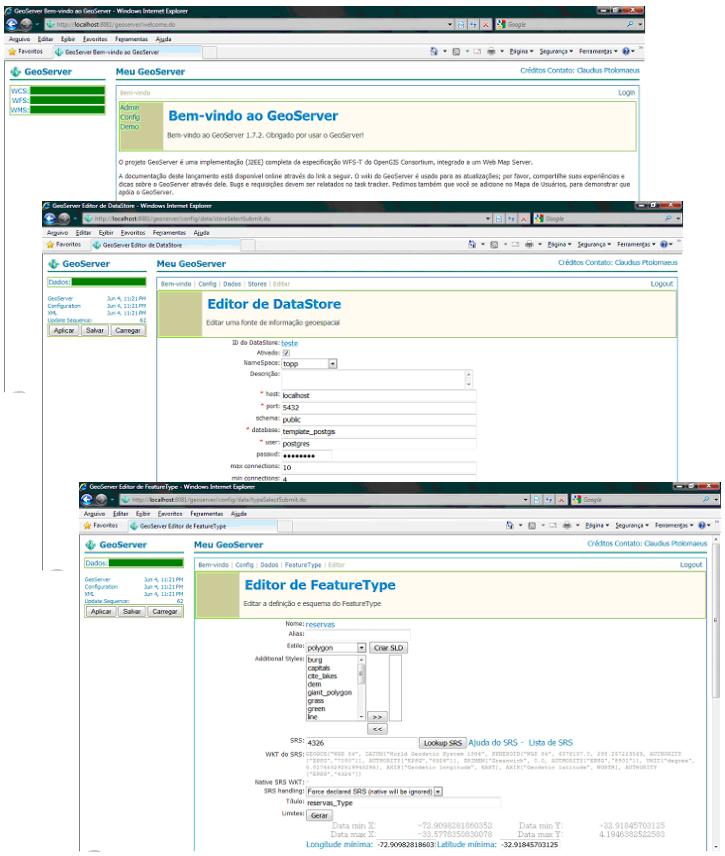


Figura 4.9: Interface administrativa do Geoserver para configuração de *Feature Types*.

de dados na camada de persistência.

A criação das tabelas de Rios, Reservas e Parques a partir de arquivos *Shapefile* ocasionou diferença em relação ao modelo de dados proposto inicialmente. Na modelagem original, os atributos *Name* e *Type* dos objetos geográficos seriam armazenados na tabela GeographicObject. Para contornar essa diferença e evitar trabalho de migração destas tabelas para a GeographicObjects, foi criada uma visão.

A visão *GeographicObjects* é uma junção de todas as tabelas geográficas (Rios, Reservas e Parques) e possui o identificador alfanumérico, o nome, a geometria, o identificador geográfico e o tipo do objeto. A visão foi criada para facilitar a implementação do serviço Web. Foram criadas funções de inserção de anotação, com diferentes parâmetros de entrada.

No protótipo foi tratado somente anotações do tipo texto. Assim, as tabelas referentes

aos demais tipos, como imagem e vídeo não foram criadas. A coluna *Type* da tabela *SuperimposedInformation* também não foi criada.

4.2 Cenário de Uso

A validação do serviço Web de anotação foi realizada considerando um potencial cenário de uso na área de biologia. Neste cenário, um biólogo cria anotações com informações sobre coletas de borboletas, consulta anotações existentes, além de consultar informações provenientes de um sistema de coletas.

Um cenário bastante comum para biólogos é fazer pesquisas e registrar informações coletadas em campo. Durante a realização de uma pesquisa, os biólogos definem a estratégia a ser adotada para a coleta de espécimes. Isso inclui estabelecer uma metodologia, um local, um período, quais serão as formas de distribuir armadilhas, entre outros. Depois de realizada a coleta, biólogos usualmente registram informações sobre espécimes, tais como classificação taxonômica, como, onde, quando e por quem foram coletadas. Um cenário de utilização da aplicação desenvolvida neste trabalho de mestrado pode ser ilustrado com o registro de informações sobre coletas de borboletas.

Em [20], foi realizado um trabalho visando descrever quantitativamente a comunidade de borboletas frugívoras da Reserva Municipal de Santa Genebra, um fragmento florestal remanescente em Campinas, SP, com ênfase na biologia populacional de *Anaea ryphea (Cramer) (Nymphalidae: Charanidae)*. As coletas realizaram-se durante 26 meses. O procedimento de coleta constituiu na distribuição de 12 armadilhas em três ambientes com diferentes níveis de perturbação: 4 na borda da mata, 4 no interior da mata e 4 próximas à trilha central que corta a mata.

A borda da mata faz limite com plantações de soja, milho e girassol, e em alguns trechos, com habitações humanas. Este habitat é caracterizado pela interrupção brusca entre a mata e esses ambientes intensamente alterados, apresentando alta insolação e luminosidade, temperaturas mais elevadas, ventos mais intensos em algumas épocas do ano e presença acentuada de espécies invasoras (capins e ervas). A trilha central é resultante de uma estrada abandonada que atravessa a reserva e reproduz parcialmente uma clareira estreita e alongada. Exibe condições intermediárias de luminosidade, temperatura e condições de vento entre aquelas observadas na borda e no interior da mata. O interior da mata apresenta menor luminosidade, menores temperaturas e maior umidade.

As armadilhas foram penduradas a alturas entre 1,5 a 3,5 m. Foram realizadas 4 amostragens por mês em intervalos de aproximadamente 7 dias, totalizando 104 amostragens. As armadilhas foram abertas em horários centrais do dia (entre 11:00 e 13:00). Os indivíduos foram identificados e liberados em seguida. Espécimes testemunhos de todas as espécies amostradas foram depositados no Museu de História Natural da Unicamp.

O biólogo capturou 4117 indivíduos, pertencentes a 44 espécies em 15 gêneros de *Nymphalidae*, distribuídas em 5 subfamílias: *Brassolinae* (3 espécies), *Charaxinae* (10), *Morphinae* (1), *Nymphalinae* (23) e *Satyrinae* (7). *Anaea ryphea* foi a espécie mais abundante, com 21,5% dos indivíduos capturados, e as 10 espécies mais abundantes somaram 81% das capturas. A maior parte das espécies capturadas pode ser classificada como espécies comuns. Seis espécies foram capturadas em apenas um dos habitats, nove compartilharam pelo menos dois habitats e 29 foram capturadas nos três ambientes. Um exemplo de levantamento de espécies capturadas em cada habitat pode ser visto na Figura 4.10. Foram encontrados 887 indivíduos da espécie *Anaea ryphea*, por exemplo. A sigla BTI indica que a espécie foi encontrada nos três habitats, na borda da mata, na trilha no interior da mata e no interior da mata. As colunas Borda, Trilha e Interior mostram a quantidade encontrada em cada habitat.

Habitat	Subfamília	Espécie	Autor	Borda	Trilha	Interior	Total
BTI	Charaxinae	*Anaea ryphea*	(Cramer, 1775)	270	501	116	887
BTI	Nymphalinae	*Biblis hyperia*	(Cramer, 1780)	372	241	6	619
BTI	Nymphalinae	*Smyrna blomfildia*	(Fabricius, 1782)	89	223	119	431
BTI	Nymphalinae	*Hamadryas februa*	(Hübner, 1823)	232	83	22	337
BTI	Nymphalinae	*Hamadryas epinome*	(Felder & Felder, 1867)	23	180	103	306
BTI	Nymphalinae	*Myscelia orsis*	(Drury, 1782)	2	122	90	214
BTI	Charaxinae	*Memphis morvus*	(Prittwitz, 1865)	43	99	55	197
BTI	Charaxinae	*Memphis arginussa*	(Geyer, 1832)	27	90	22	139
BTI	Satyrinae	*Taygetis laches*	(Fabricius, 1793)	18	31	56	105
BTI	Nymphalinae	*Hamadryas feronia*	(Linnaeus, 1758)	87	12	2	101
BTI	Nymphalinae	*Hamadryas arete*	(Doubleday, 1847)	29	52	3	84
BTI	Nymphalinae	*Hamadryas amphinome*	(Linnaeus, 1767)	35	35	4	74
BT	Charaxinae	*Memphis appias*	(Hübner, 1825)	18	55	0	73

Figura 4.10: Exemplo de dados de coletas sobre espécies de *Lepidoptera* capturadas em três habitats (B = Borda, T = Trilha e I = Interior da Mata) [20].

Com os dados das coletas, o biólogo avaliou a diversidade e abundância das espécies da região, assim como sua relação com plantas hospedeiras, parasitas, etc.

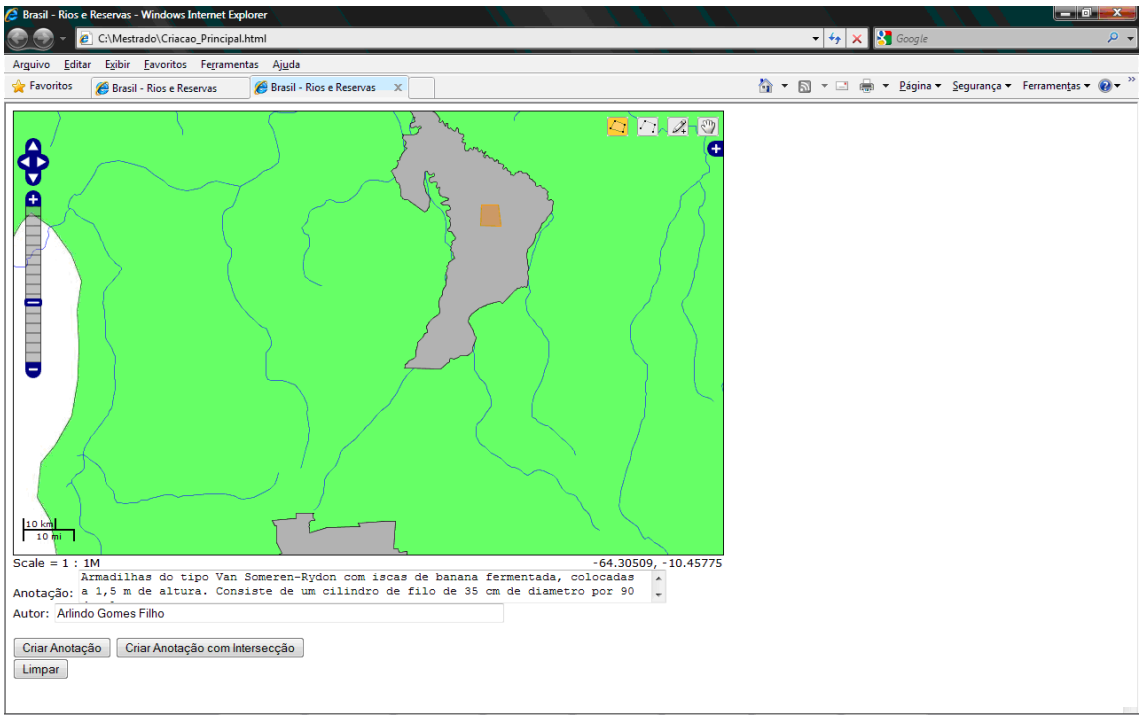


Figura 4.11: Exemplo de criação de anotação associada a uma região de uma reserva ecológica.

4.3 Exemplo de Utilização do Serviço Web

No cenário descrito acima é importante associar as informações de coletas com os tipos de habitats. Muitas vezes, os habitats estão localizados em diferentes posições geográficas e possuem características relevantes para o estudo.

Inicialmente o biólogo poderá utilizar a operação de criação de anotação para registrar onde as armadilhas estão localizadas. Na barra de ferramentas de desenho, há opções para desenhar polígonos, linhas ou pontos. A localização das armadilhas pode ser representada com diferentes geometrias. A anotação pode estar associada a uma única geometria ou a um conjunto delas. Após desenhar a geometria e preencher os dados da anotação, o biólogo invoca a operação utilizando o botão "Criar Anotação". Nesse caso, a anotação é armazenada como um texto livre, sem seguir nenhuma estrutura.

A tela da aplicação que exemplifica a criação de uma anotação é apresentada na Fi-

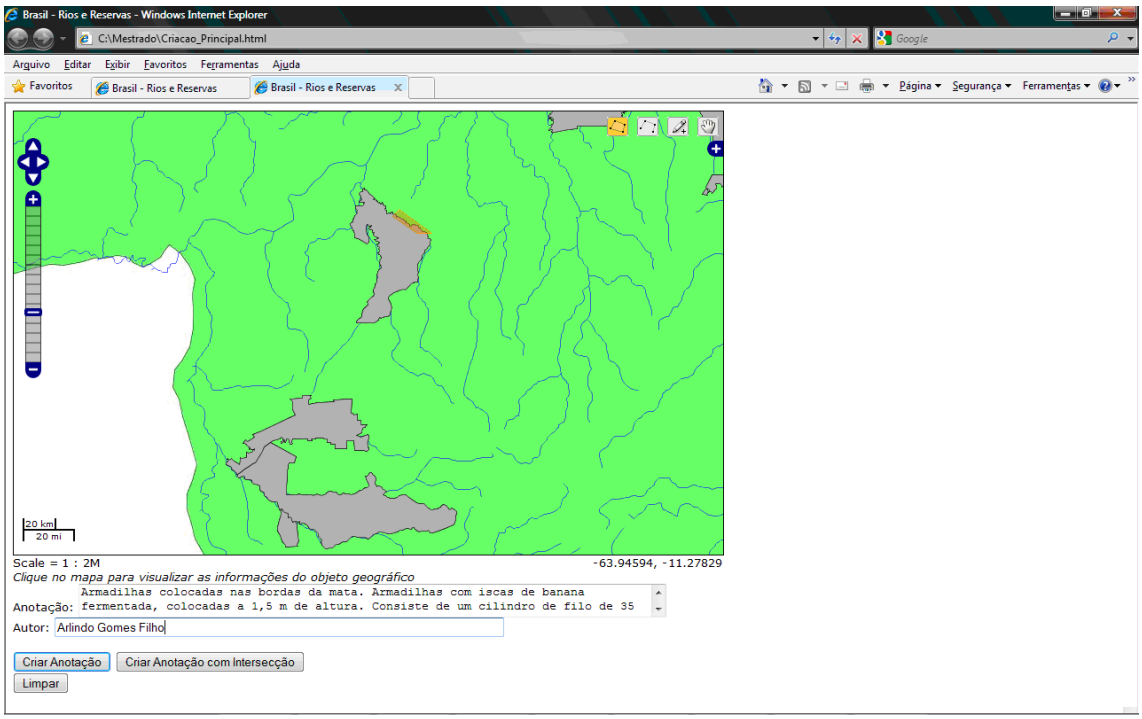


Figura 4.12: Exemplo de criação de anotação próxima a fronteira de uma reserva ecológica.

gura 4.11. No exemplo, um polígono foi desenhado sobre a área de uma reserva ecológica, os dados da anotação foram preenchidos e a operação "Criar Anotação" foi executada. Este exemplo mostra como o biólogo poderia registrar as armadilhas localizadas no interior da mata. O polígono desenhado está totalmente contido dentro da área da reserva ecológica.

O biólogo também pode registrar as armadilhas localizadas nas bordas da mata. A tela da aplicação que exemplifica esse tipo de anotação é apresentada na Figura 4.12. Neste exemplo, o polígono foi desenhado próximo à fronteira e está parcialmente contido na área da reserva ecológica.

A aplicação dá ao usuário possibilidade de associar informações sobrepostas a objetos geográficos. No cenário, o objeto geográfico é a reserva ecológica e as informações sobrepostas são os dados observados durante as coletas. Quando uma região, uma linha ou um ponto são demarcados no mapa para criar uma anotação, é feita uma associação entre a reserva e a anotação, assim, além dos dados já existentes, como nome, área, etc, a reserva

passa a ter novas informações.

É possível associar a anotação com dados de coletas registrados em um sistema, ou até mesmo integrar catálogos de espécies à aplicação de anotação. No cenário ilustrado, a aplicação foi integrada ao banco de dados do BioCORE [4], dando possibilidade ao usuário de associar sua anotação com registros do Museu de História Natural da Unicamp. Os detalhes da modelagem do banco de dados do BioCORE podem ser visto em [33]. A associação entre as anotações e o catálogo de espécies foi feita diretamente em banco de dados. Foi criada uma tabela de ligação entre a tabela *SuperimposedInformation* e a tabela *Catalog*, onde são armazenados os identificadores de cada tabela. A tabela *Catalog* pertence ao BioCORE e armazena o código da instituição, o código da coleção a que pertence a espécie, o número do catálogo que foi atribuído ao lote de espécies, a via de conservação, o número de campo que os biólogos atribuem a um espaço específico onde realizam a coleta, entre outros. Para que o biólogo possa visualizar as informações do catálogo deve invocar a operação "Consultar Catálogo". A Figura 4.13 apresenta a tela com os dados do catálogo do BioCORE. Do lado esquerdo da tela, o parâmetro de busca informado foi parte do texto da anotação ("Armadilhas do tipo"). Do lado direito da tela, foram retornados os dados do registro do catálogo associado a anotação encontrada. O registro do catálogo encontrado é da instituição ZUEC, tem número 1583 e data de 2009.

Semanalmente, o biólogo precisa registrar as informações referentes a coleta realizada em cada armadilha. Assim, ele poderá utilizar a operação de consulta para encontrar as coordenadas geográficas das armadilhas. O biólogo também pode criar anotações em outras coordenadas. Há várias formas de consultar anotações, utilizando parte do texto da anotação, pelo nome do autor, pela data, pelo nome do objeto geográfico ou selecionando uma região no mapa.

A Figura 4.14 ilustra a realização de uma consulta na qual uma região é demarcada no mapa a fim de recuperar todas as anotações nela contidas. A operação "Consultar Anotação - Região" foi executada. O resultado é retornado na forma textual, na parte inferior da tela, e contém a anotação, o autor, os nomes dos objetos geográficos associados à anotação e a data.

A Figura 4.15 ilustra como um biólogo pode consultar anotações cujas marcas estejam totalmente contidas dentro de uma reserva ecológica. O parâmetro de entrada da consulta é o nome do objeto geográfico, que foi preenchido no campo correspondente da tela. No exemplo, o objeto geográfico consultado foi a reserva REEXTE DO RIO JACI-PARANA. Foi passado somente parte do nome (jaci) no parâmetro de consulta. A operação "Consultar Anotação Contida em Objeto" foi executada. Esta consulta auxilia-o a descobrir as anotações localizadas no interior da mata.

A Figura 4.16 ilustra como um biólogo consulta anotações que estejam associadas a um determinado objeto geográfico. Neste caso, serão retornadas as anotações localizadas

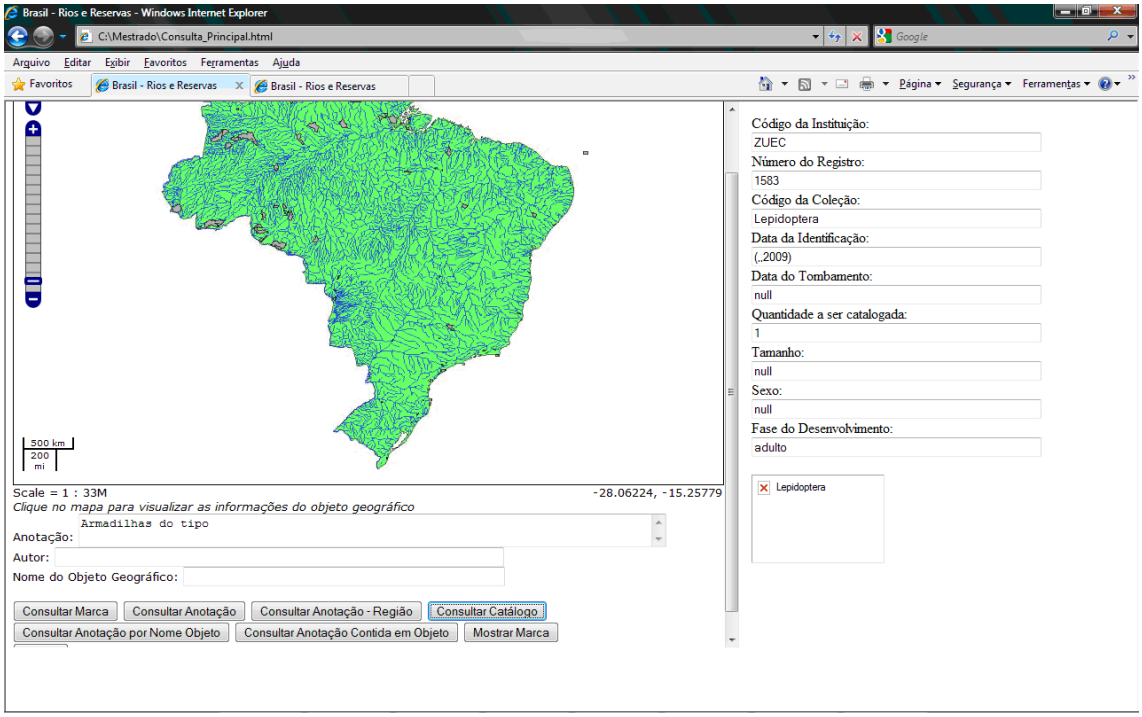


Figura 4.13: Exemplo de consulta de informações de catálogo de espécies do BioCORE associadas a anotação.

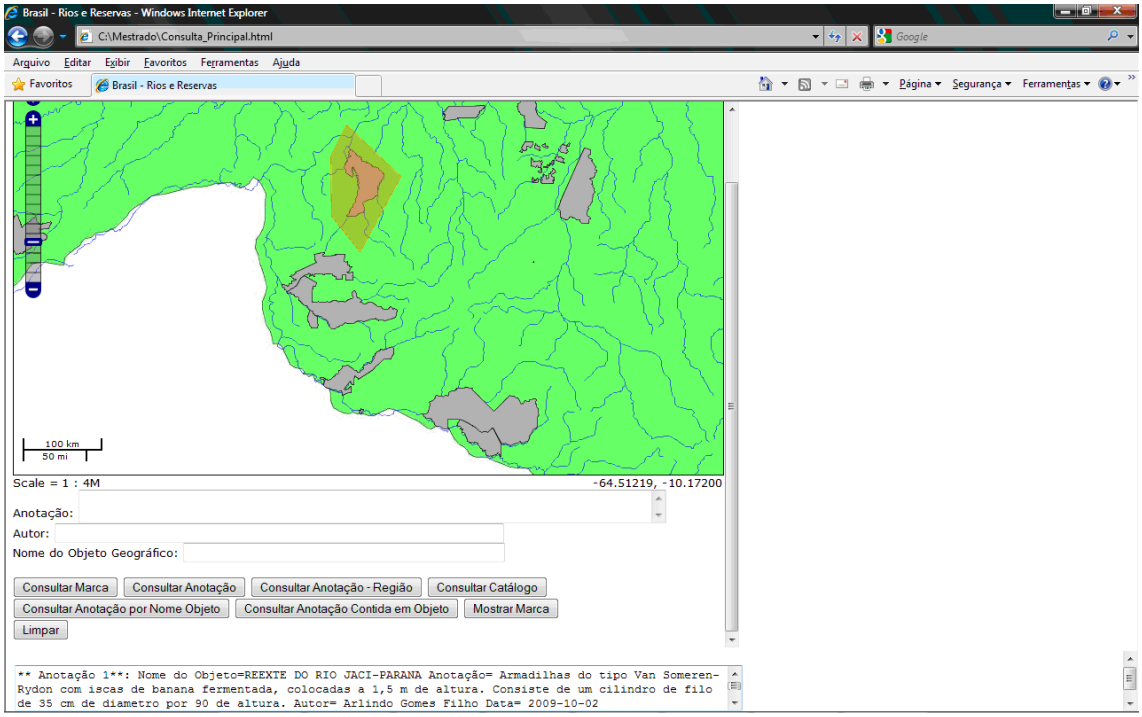


Figura 4.14: Exemplo de consulta de anotação a partir de uma região selecionada no mapa.

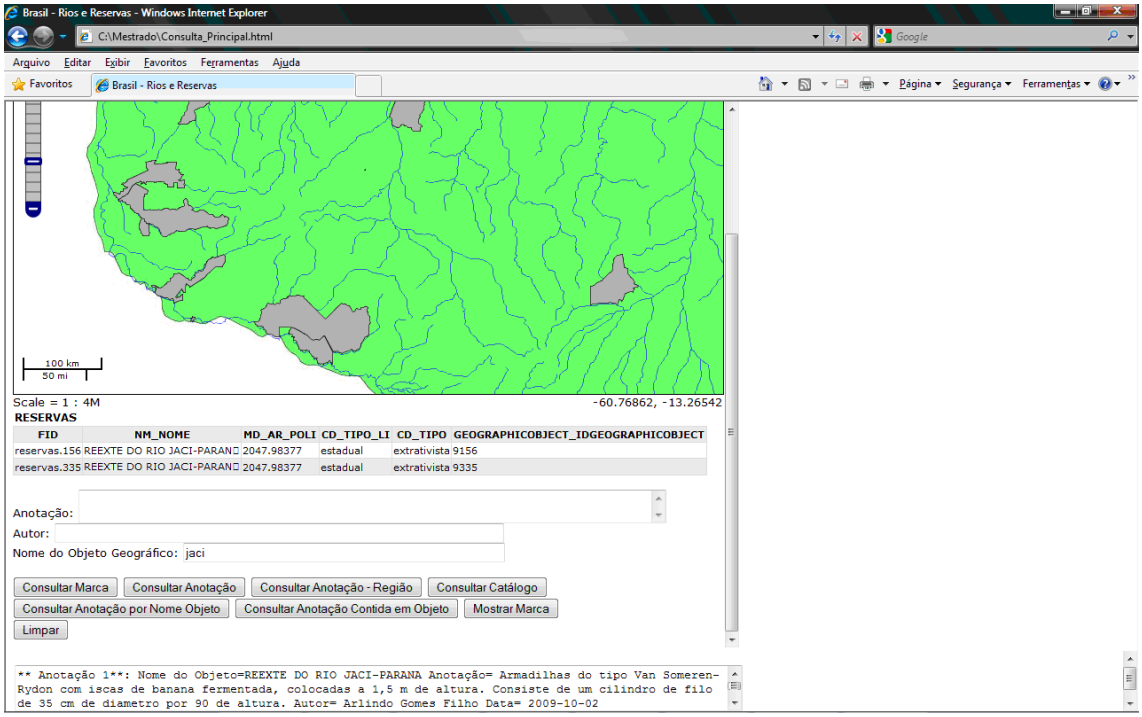


Figura 4.15: Exemplo de consulta de anotação contida no interior da área de um objeto geográfico.

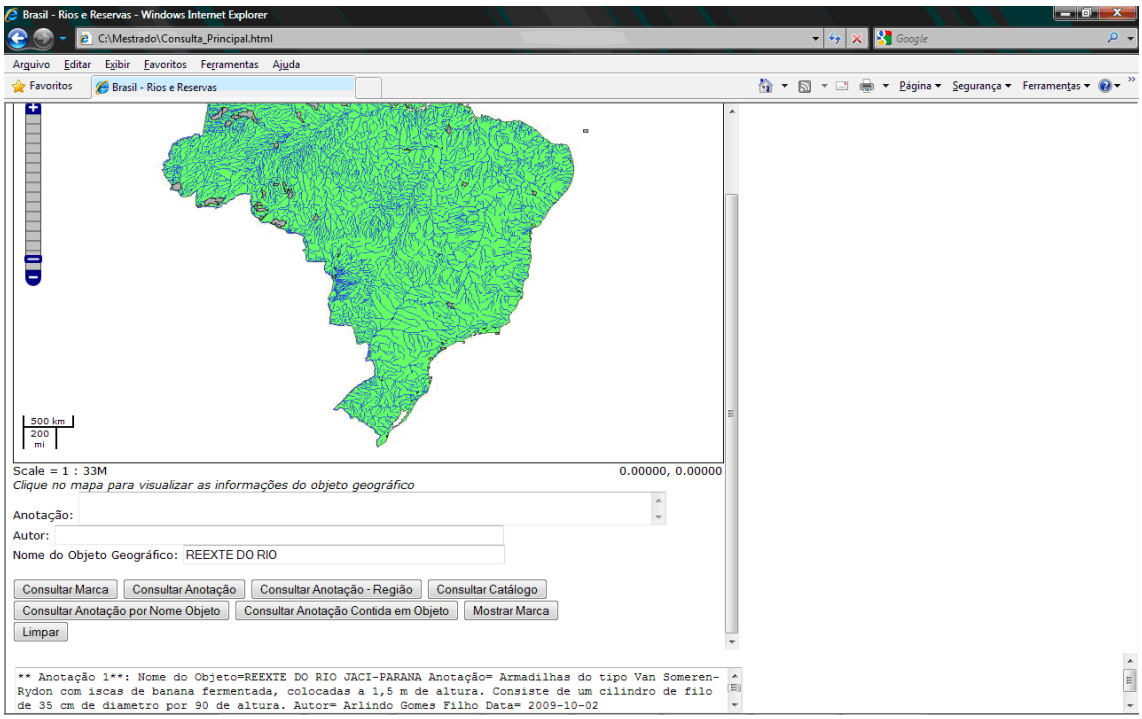


Figura 4.16: Exemplo de consulta de anotação associada a um objeto geográfico.

tanto no interior, quanto na borda da mata.

A Figura 4.17 apresenta um trecho da especificação WSDL da operação de Consulta por Nome do Objeto Geogrráfico. As seções *input* e *output* definem respectivamente os parâmetros enviados e recebidos por mensagens SOAP na invocação desta operação.

As operações de consulta normalmente retornam as informações textuais das anotação e seus metadados. Para consultar a localização geográfica da anotação, é preciso consultar a geometria de sua marca. Para isso, o biólogo deve invocar a operação "Consultar Marca". Esta consulta retorna a geometria da marca textualmente. Para exibir a geometria no mapa, deve-se invocar a operação "Mostrar Marca". As Figuras 4.18 e 4.19 mostram respectivamente exemplos de consulta e exibição da geometria da marca de uma anotação.

Ao final de todos os registros, o usuário poderá utilizar as informações armazenadas para calcular indicadores, distribuições, variações, etc. Caso estejam em formato de texto livre, será necessário extrair a informação necessária para os cálculos antecipadamente ou

```
<definitions xmlns:soap = "http://schemas.xmlsoap.org/wsdl/soap/" xmlns:tns = "teste" xmlns:xsd =
"http://www.w3.org/2001/XMLSchema" xmlns = "http://schemas.xmlsoap.org/wsdl/" targetNamespace =
"teste" name = "Annotation_QueryService">
    <types>
        <xsd:schema>
            <xsd:import namespace = "teste" schemaLocation =
                    "http://localhost:8080/queryByObjName?xsd=1"/>
        </xsd:schema>
    </types>
    <message name = "queryByObjName">
        <part name = "parameters" element = "tns:queryByObjName"/>
    </message>
    <message name = "queryByObjNameResponse">
        <part name = "parameters" element = "tns:queryByObjNameResponse"/>
    </message>
    <portType name = "Annotation_Query">
        <operation name = "queryByObjName">
            <input message = "tns:queryByObjName"/>
            <output message = "tns:queryByObjNameResponse"/>
        </operation>
    </portType>
    <binding name = "Annotation_QueryPortBinding" type = "tns:Annotation_Query">
        <soap:binding transport = "http://schemas.xmlsoap.org/soap/http" style = "document"/>
        <operation name = "queryByObjName">
            <soap:operation soapAction = ""/>
            <input>
                <soap:body use = "literal"/>
            </input>
            <output>
                <soap:body use = "literal"/>
            </output>
        </operation>
    </binding>
    <service name = "Annotation_QueryService">
        <port name = "Annotation_QueryPort" binding = "tns:Annotation_QueryPortBinding">
            <soap:address location = "http://localhost:8080/queryByObjName"/>
        </port>
    </service>
</definitions>
```

Figura 4.17: Exemplo de WSDL de consulta de anotação associada a um objeto geográfico.

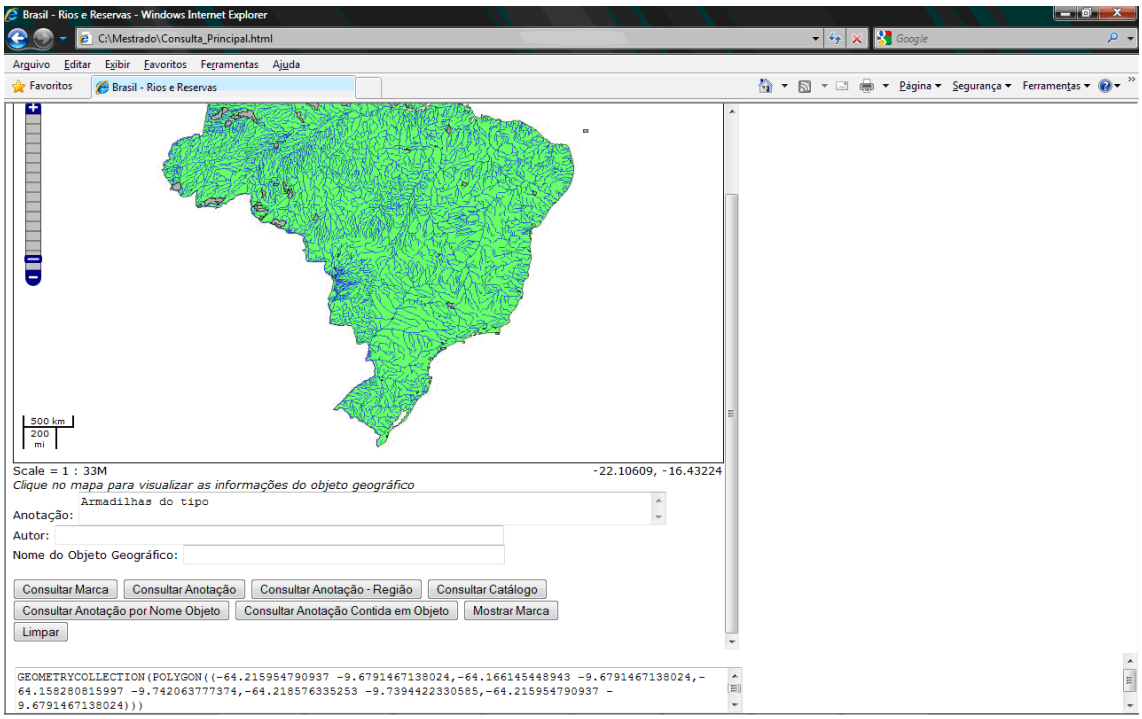


Figura 4.18: Exemplo de consulta de geometria da marca de uma anotação.

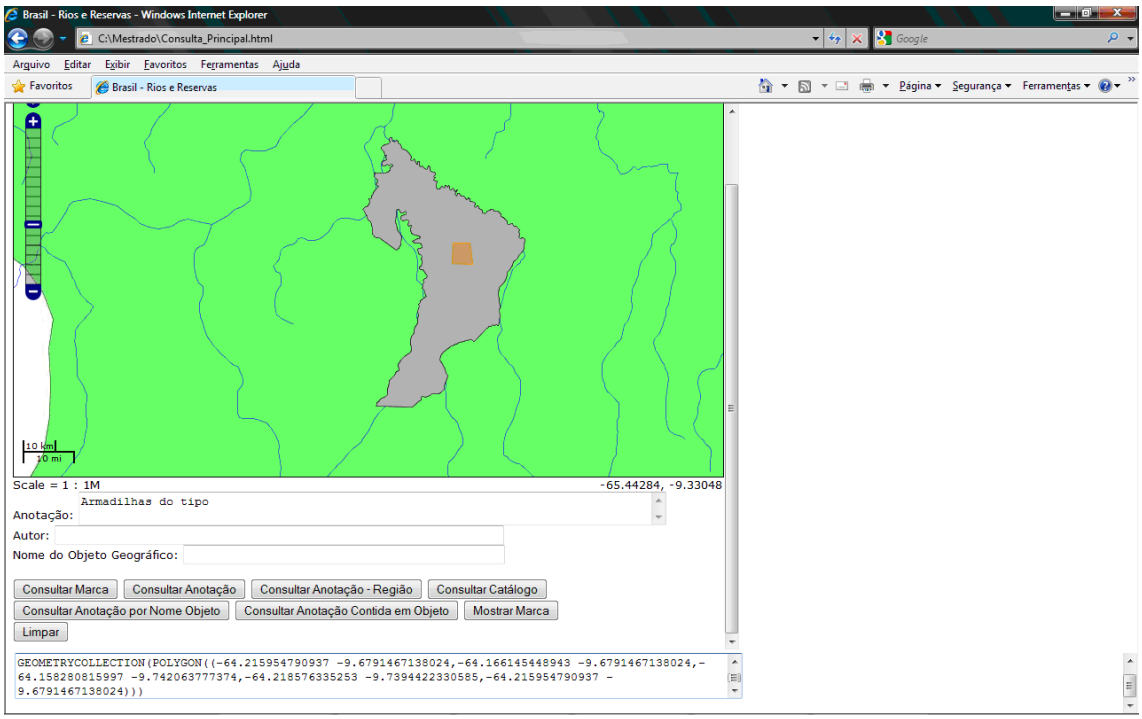


Figura 4.19: Exemplo de exibição da marca de uma anotação no mapa.

criar consultas mais específicas para cada tipo cálculo.

Um usuário mais avançado da aplicação poderia utilizar a operação de consulta de anotações passando como parâmetro uma consulta, que pode ser geográfica ou não. No cenário da coleta de borboletas, o biólogo poderia consultar as anotações que estão na borda da mata, invocando a operação que tem como parâmetro uma consulta geográfica, passando a instrução que faz a recuperação dos dados das anotações que estejam próximas aos limites da reserva, ou que não estejam totalmente contidas na reserva.

4.4 Operações Implementadas

Para validar a proposta de arquitetura do serviço Web e atender ao cenário de uso de biodiversidade de coletas de borboletas, foram escolhidas as seguintes operações para serem implementadas:

- **boolean** *InsertAnnotation* (**Text** Anotacao, **Geometry** Regiao, **Metadata** Metadados, **Context** Contexto)

- **List** {**Text, Metadata** } *QueryAnnotation* (**List** {**Text** PalavraChave})

- **List** {**Text, Metadata** } *QueryAnnotation* (**Geometry** Regiao)

- **List** {**Text, Metadata** } *QueryAnnotation* (**String** NomeObj)

- **MarkGeometry** *QueryMark* (**Text** Anotacao)

As operações foram escolhidas por serem relevantes para permitir o registro e consulta dos dados do cenário de biodiversidade. Além de implementar algumas das operações definidas nas interfaces de inserção e consulta, foi implementada também uma operação de consulta a informações do catálogo. Esta operação recupera as informações de catálogo do BioCORE associadas à anotação. O parâmetro de entrada da operação é uma lista de palavras que devem coincidir com a anotação.

Catalog *QueryCatalog* (**List** {**Text** PalavraChave})

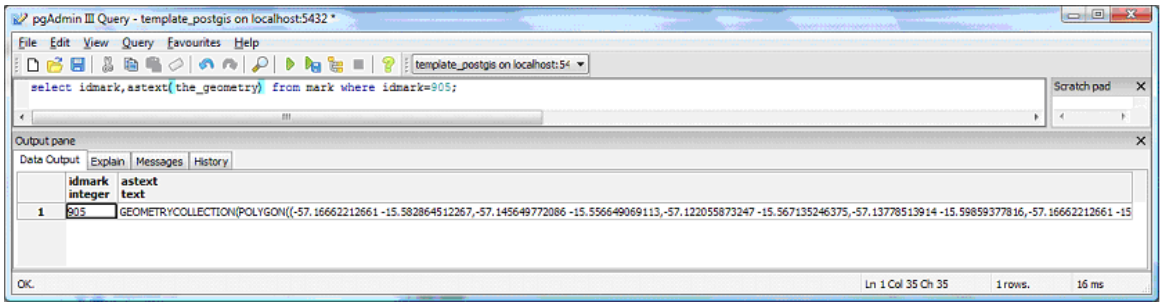


Figura 4.20: Informações armazenadas na tabela Mark.

Foi implementado uma operação para exibição da marca consultada no mapa. Essa operação utiliza a API do OpenLayers para exibir a geometria recuperada pela operação de consulta de marcas.

A operação de consulta a anotação contida em objeto geográfico é uma variação da operação: **List** {**Text, Metadata** } *QueryAnnotation* (**String** NomeObj). Além de considerar o nome do objeto, a operação utiliza o relacionamento topológico "contains" para verificar se a anotação está contida na reserva.

4.4.1 Exemplos de Informações Armazenadas pela Operação de Inserção de Anotação

As operações de inserção de anotações manipulam quase todas as tabelas do modelo de dados. A seguir são apresentados exemplos de informações armazenadas durante a execução da operação de inserção de anotações. Neste exemplo, a anotação foi associada a um polígono desenhado no interior de uma reserva ecológica.

A Figura 4.20 demonstra os valores armazenados na tabela Mark após a inserção de uma anotação. A primeira coluna armazena o identificador da marca e a segunda armazena a geometria da marca, no caso, um polígono. O polígono é composto por quatro pontos. O último ponto coincide com o primeiro, pois é um polígono fechado. Como o usuário pode desenhar várias geometrias associadas a uma anotação, a operação de inserção sempre persiste uma coleção de geometrias.

A tabela SuperimposedInformation faz o relacionamento da marca com o texto da

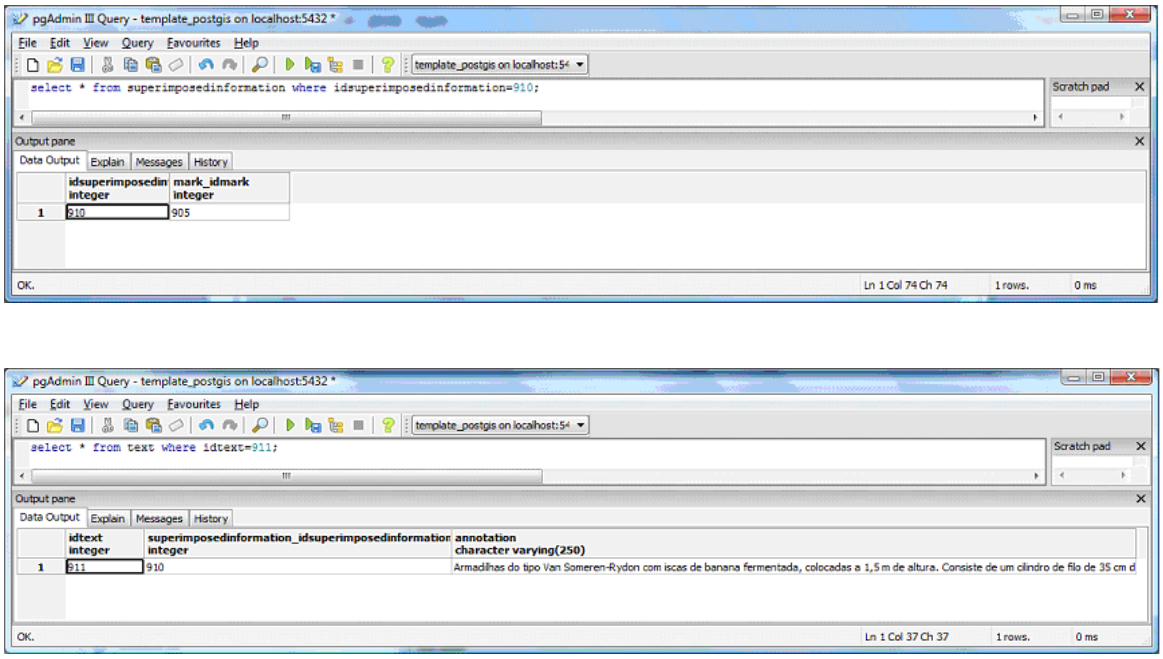


Figura 4.21: Informações armazenadas nas tabelas SuperimposedInformation e Text.

anotação, na tabela Text. A Figura 4.21 demonstra os valores armazenados nas tabelas SuperimposedInformation e Text. Na primeira coluna da tabela SuperimposedInformation está armazenado o identificador da informação sobreposta e na segunda o identificador da marca. A coluna Type não foi criada pois o protótipo trata somente o formato texto. Na tabela Text estão armazenados o identificador da anotação, o identificador da informação sobreposta e o texto da anotação.

A Figura 4.22 demonstra os valores armazenados na tabela Metadata. A tabela Metadata está relacionada com a tabela Mark e possui o identificador do metadado, o identificador da marca e os metadados da anotação, como nome do autor, data e idioma.

A tabela Excerpt faz o relacionamento da marca com o(s) objeto(s) geográfico(s) anotado(s) na tabela GeographicObject_Has_Excerpt. A Figura 4.23 demonstra os valores armazenados nas tabelas Excerpt e GeographicObject_Has_Excerpt. Na primeira coluna da tabela Excerpt está armazenado o identificador do conteúdo (excerpt) da anotação e na segunda o identificador da marca. Na tabela GeographicObject_Has_Excerpt estão

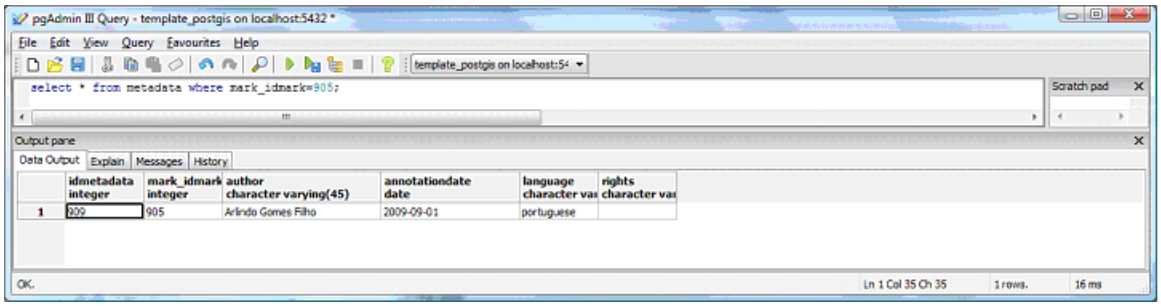


Figura 4.22: Informações armazenadas na tabela Metadata.

armazenados o identificador da tabela, o identificador do excerpt, o identificador do objeto geográfico anotado e a geometria anotada.

O usuário pode desenhar várias geometrias associadas a uma anotação. Assim, a operação de inserção de anotação recupera, para cada geometria desenhada, todos os objetos geográficos que estejam totalmente ou parcialmente contidos na geometria e os associa com o conteúdo da anotação, conforme trecho de código a seguir.

A quantidade de geometrias existentes na coleção de geometrias recebida como parâmetro pela operação de inserção de anotação é recuperada e armazenada em um contador (linha 1). Enquanto existir geometria, ou seja, enquanto o contador de geometrias for maior que zero, ocorre uma iteração (linha 2). A geometria correspondente à posição indicada pelo contador é recuperada da coleção (linha 3). Para cada objeto geográfico que tiver algum ponto em comum com a geometria recuperada (linha 4), será inserida uma linha na tabela que relaciona objeto geográfico com conteúdo de marca (linhas 8 e 9). A cada iteração o contador é decrementado (linha 17).

Não são persistidos dados na tabela Reservas na inserção de anotações. Os valores armazenados nesta tabela são provenientes de carga de dados. A tabela Reservas armazena o identificador alfanumérico, o nome, a área em metros quadrados, o tipo de responsável, o tipo da reserva, a geometria e o identificador geográfico. Com a visão GeographicObjects é possível descobrir de que tipo (Rio, Reserva ou Parque) é um objeto geográfico anotado. Com o tipo do objeto, é possível recuperar os demais dados do objeto diretamente na tabela correspondente, no caso Reservas. A Figura 4.24 demonstra os valores armazenados na visão GeographicObjects e na tabela Reservas.

```
1 SELECT NumGeometries(GeomFromText(collection)) INTO contador;
2 WHILE contador > 0 LOOP
3    SELECT GeometryN(GeomFromText(collection),contador) INTO geometria;
4    FOR objid3 in SELECT * FROM geographic_objects
5               WHERE not disjoint(the_geom,geometria) = true
6    LOOP
7       BEGIN
8          SELECT nextval('geral') INTO id;
9          INSERT INTO geographicobject_has_excerpt(
10                                        idgeographicobject_has_excerpt,
11                                        excerpt_idexcerpt,
12                                        geographicobject_idgeographicobject,
13                                        the_geometry)
14         VALUES (id, idexcerpt2, objid3.idgeographicobject, geometria);
15      END;
16   END LOOP;
17 contador := contador - 1;
18 END LOOP;
```

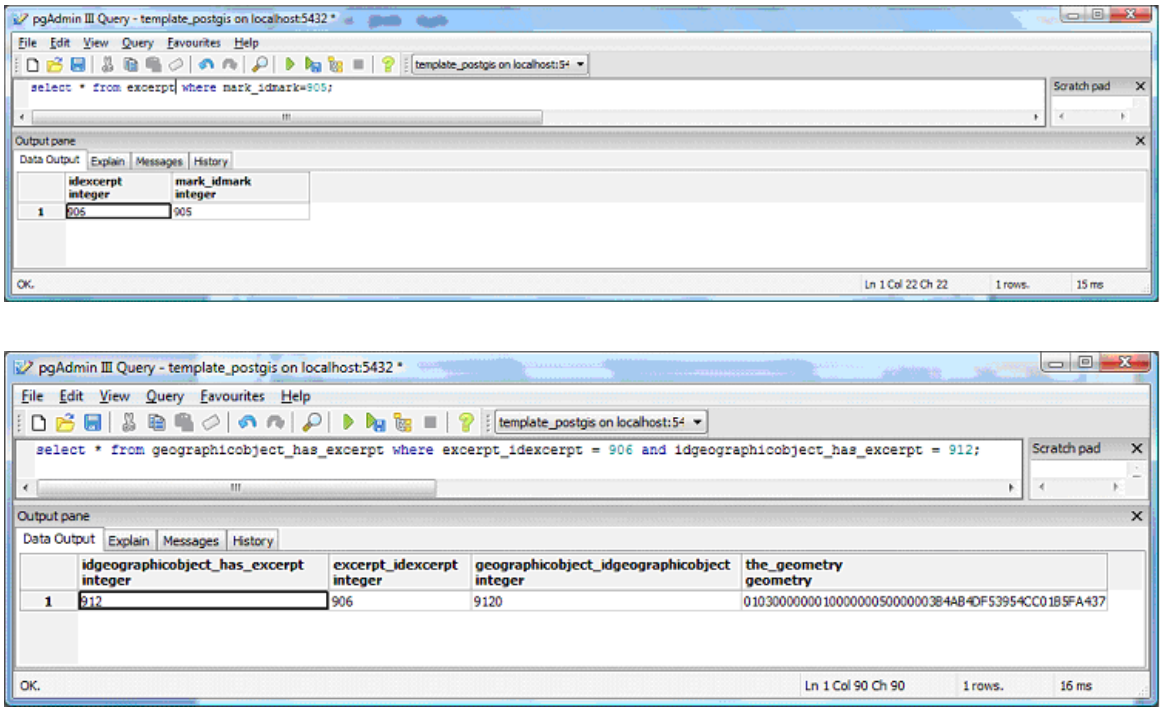


Figura 4.23: Informações armazenadas nas tabelas Excerpt e GeographicObject_Has_Excerpt.

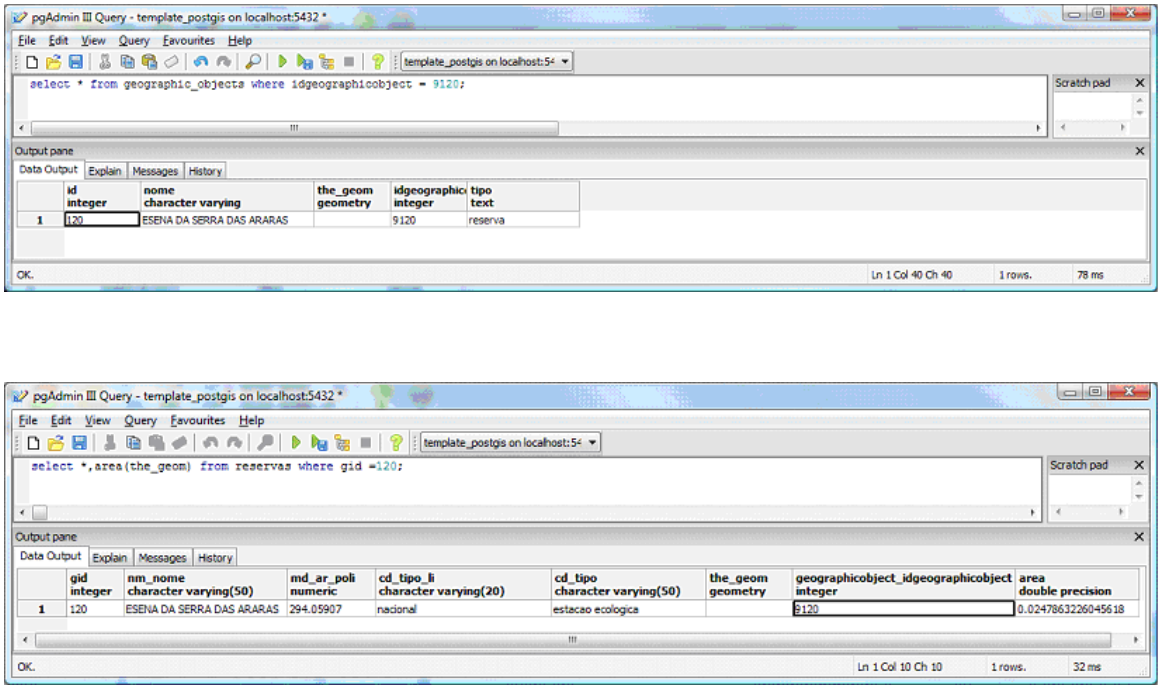


Figura 4.24: Informações armazenadas na visão GeographicObjects e na tabela Reservas.

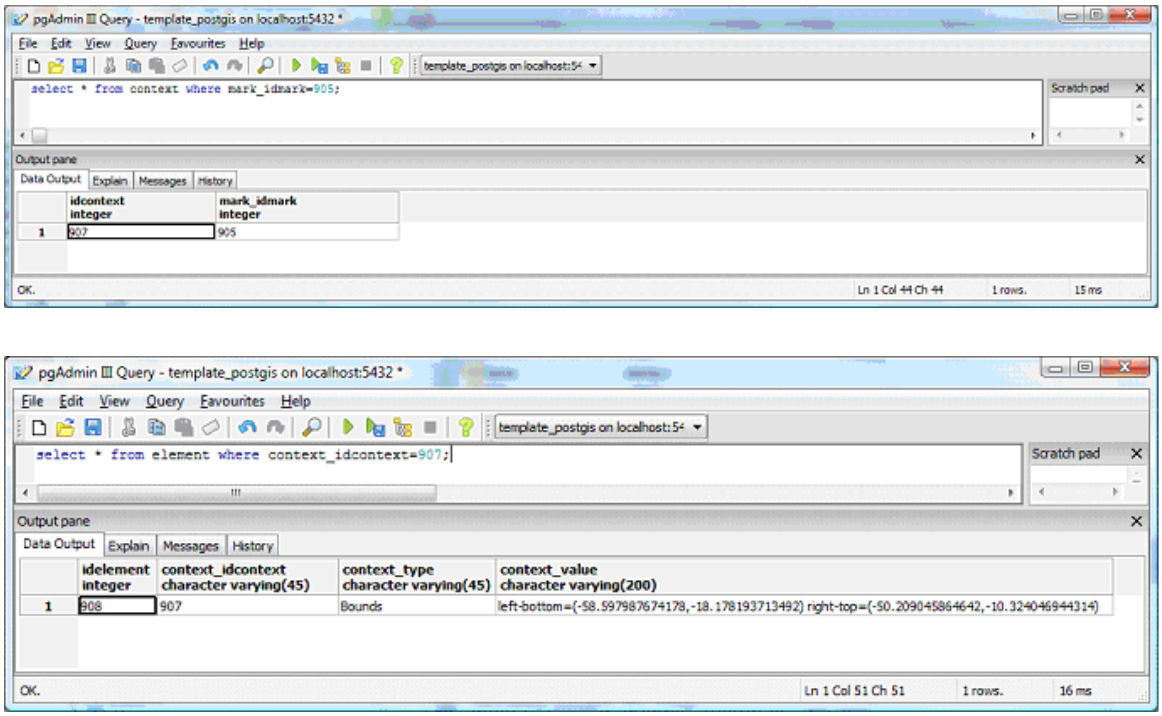


Figura 4.25: Informações armazenadas nas tabelas Context e Element.

A tabela Context faz o relacionamento da marca com os elementos de contexto da anotação. Os elementos de contexto são definidos no formato tipo/valor e podem assumir diferentes valores dependendo do tipo de aplicação. A Figura 4.25 demonstra os valores armazenados nas tabelas Context e Element. Na primeira coluna da tabela Context está armazenado o identificador do contexto e na segunda o identificador da marca. Na tabela Element estão armazenados o identificador do elemento, o identificador do contexto, o tipo e valor do contexto. No caso em tipo e valor do contexto estão armazenadas as coordenadas que representam o nível de zoom que o usuário estava visualizando quando criou a anotação.

4.4.2 Exemplo de Consulta

Na operação de consulta de anotações de uma região, o usuário desenha uma geometria no mapa e invoca o servlet. O servlet invoca a respectiva operação que atenderá a requisição

```
1 SELECT
2    GEOGRAPHIC_OBJECTS.NOME,
3    TEXT.ANNOTATION,
4    METADATA.AUTHOR,
5    METADATA.ANNOTATIONDATE
6 FROM
7    GEOGRAPHICOBJECT_HAS_EXCERPT,
8    GEOGRAPHIC_OBJECTS,
9    EXCERPT,
10   SUPERIMPOSEDINFORMATION,
11   TEXT,
12   METADATA
13 WHERE
14   NOT DISJOINT(GEOGRAPHIC_OBJECTS.THE_GEOM, <COLLECTION>) = TRUE AND
15   GEOGRAPHIC_OBJECTS.IDGEOGRAPHICOBJECT =
16   GEOGRAPHICOBJECT_HAS_EXCERPT.GEOGRAPHICOBJECT_IDGEOGRAPHICOBJECT AND
17   GEOGRAPHICOBJECT_HAS_EXCERPT.EXCERPT_IDEXCERPT = EXCERPT.IDEXCERPT AND
18   EXCERPT.MARK_IDMARK = SUPERIMPOSEDINFORMATION.MARK_IDMARK AND
19   SUPERIMPOSEDINFORMATION.IDSUPERIMPOSEDINFORMATION =
20   TEXT.SUPERIMPOSEDINFORMATION_IDSUPERIMPOSEDINFORMATION AND
21   METADATA.MARK_IDMARK = SUPERIMPOSEDINFORMATION.MARK_IDMARK;
```

do usuário. Da mesma forma que a inserção, a consulta sempre recebe como parâmetro uma coleção de geometrias. O parâmetro enviado para a operação segue o padrão:

```
GEOMETRYCOLLECTION(POLYGON((
 -64.215954790937 -9.6791467138024,
 -64.166145448943 -9.6791467138024,
 -64.158280815997 -9.742063777374,
 -64.218576335253 -9.7394422330585,
 -64.215954790937 -9.6791467138024)))
```

Neste caso, somente um polígono faz parte da coleção. Outros tipos de elementos, como pontos ou linhas também são permitidos. Um relacionamento topológico (NOT DISJOINT) é utilizado para recuperar os objetos geográficos que estejam contidos na região informada e que possuem associação com alguma anotação (linha 14). A consulta retorna o(s) nome(s) do(s) objeto(s) geográfico(s), a(s) anotação(s), o(s) autor(es) e data(s) (linhas 2, 3, 4 e 5).

A partir do objeto geográfico que tem algum ponto em comum com a geometria informada (linha 14), junções são feitas com as tabelas de conteúdo de marca (linha 15 e 17), marca (linha 18), informação sobreposta (linhas 19 e 20) e metadados (linha 21).

O retorno da operação de consulta é um texto, que na verdade, contém o código da página HTML que será exibida no navegador Web como resposta.

Capítulo 5

Conclusões e Trabalhos Futuros

5.1 Conclusões

O trabalho apresentado nesta dissertação consiste na especificação e implementação de um serviço Web para gerenciamento de informações sobrepostas associadas a dados geográficos vetoriais para sistemas de informação de Biodiversidade. O serviço trata informações sobrepostas no formato texto, referenciadas como anotações. O serviço faz uso de um repositório que também foi modelado neste trabalho. As entidades do modelo de dados seguiram o conceito de *Superimposed Information*.

Uma interface genérica para manipulação de anotações foi definida e dela foram derivadas algumas interfaces mais específicas para facilitar a utilização do serviço.

Para validar a solução, foi desenvolvido um protótipo para a área de biodiversidade considerando um potencial cenário de uso. Neste protótipo uma parte das interfaces definidas foi implementada a fim de validar a arquitetura dos serviços.

Em resumo, as principais contribuições deste projeto são:

- Especificação e implementação de um modelo de dados para armazenamento de informações sobrepostas;
- Especificação e implementação parcial de serviço Web para anotação de dados geográficos vetoriais que utiliza operadores e relacionamentos espaciais;
- Construção de um protótipo voltado para a área de biodiversidade que utilize dados reais;

O protótipo desenvolvido utiliza padrões abertos e software livre. Uma das dificuldades ocorridas no desenvolvimento deste trabalho foi a falta de suporte e documentação sobre alguns dos softwares livres utilizados, como o OpenLayers. A documentação ainda é um ponto considerado com menor relevância nas comunidades de software livre.

5.2 Trabalhos Futuros

Algumas possíveis extensões deste trabalho, tanto do ponto de vista de pesquisa quanto de implementação, são listadas a seguir.

5.2.1 Extensões na Implementação

- Implementação das demais operações definidas na seção 3.2, que não foram utilizadas no protótipo;
- Especificação e implementação de interfaces de serviços de gerenciamento de anotações que tratem novos formatos como imagens, vídeos, etc. Alterações no modelo de dados para armazenamento de novos tipos de dados;
- Integração dos serviços de gerenciamento de anotação com algum serviço de ontologia. Esta integração visa proporcionar o uso de termos do domínio de interesse no texto da anotação, facilitando o processo de consulta;

5.2.2 Extensões para Outros Domínios

- Especificação e adaptações no repositório de dados e nas interfaces de serviços de gerenciamento de anotações para permitir sua utilização em outros domínios;
- Integração do repositório de dados e dos serviços aos sistemas BioCORE e SuperIDR;

5.2.3 Controle de Acesso

- Especificação e implementação um sistema de controle de acesso para garantir a segurança da aplicação. O sistema de controle de acesso poderia permitir cadastrar variados perfis com diferentes níveis de acesso por funcionalidade. Por exemplo, este sistema poderia dar permissão a biólogos para inserirem ou removerem anotações e a alunos para somente consultarem anotações;

5.2.4 Testes de Desempenho: Tempo de Processamento e Custos de Armazenamento

- Validação do uso do sistema em um ambiente concorrente. Análise do tempo de resposta da aplicação quando muitos usuários inserem ou consultam anotações ao mesmo tempo;
- Análise de desempenho dos serviços definidos.

- Análise do espaço de armazenamento e gerenciamento de memória em situações em que uma anotação está associada a um conjunto muito grande de objetos ou a objetos que possuem muitas coordenadas.

Referências Bibliográficas

[1] M. Agosti and N. Ferro. A Formal Model of Annotations of Digital Content. *ACM Transactions on Information Systems*, 26(1):3, 2007.

[2] D. W. Archer, L. Delcambre, F. Corubolo, L. Cassel, S. Price, U. Murthy, D. Maier, E. A. Fox, S. Murthy, J. Mccall, K. Kuchibhotla, and R. Suryavanshi. Superimposed Information Architecture for Digital Libraries. In *ECDL '08: Proceedings of the 12th European conference on Research and Advanced Technology for Digital Libraries*, pages 88–99, Berlin, Heidelberg, 2008. Springer-Verlag.

[3] V. H. M. Azevedo, M. S. P. Meirelles, R. P. D. Ferraz, and A. R. Filho. Interoperabilidade entre Objetos Geográficos Heterogêneos. *Simpósio Brasileiro de Geoinformática - GEOINFO*, pages 235–245, 2006.

[4] BioCORE. Tools, models and techniques to support research in biodiversity. http://www.lis.ic.unicamp.br/projects/biocore/. Acessado em 17/10/2009.

[5] F. Brandão and J. A. Ribeiro. Estudo do XML, GML, SVG e WebServices (WMS e WFS) para formatação e divulgação de informações geográficas. *Simpósio Brasileiro de Sensoriamento Remoto*, pages 5611–5617, 2007.

[6] D. Brickley and R. V. Guha. RDF Vocabulary Description Language 1.0: RDF Schema. *Technical report, RDF Core Working Group, World Wide Web Consortium – W3C*, 2004.

[7] V. P. Canhos. Informática para biodiversidade: padrões, protocolos e ferramentas. *Ciência e Cultura*, 55(2):45–47, 2003.

[8] G. Câmara, M. A. Casanova, G. C. Magalhães, and C. M. B. Medeiros. *Anatomia de Sistemas de Informação Geográfica*. 10a Escola de Computação, Instituto de Computação, Unicamp, Campinas, 1996.

[9] H. A. da Rocha. Metadados para workflows científicos no apoio ao planejamento ambiental. Master's thesis, Unicamp, 2003.

[10] A. da S. Fagundes. Projeto e implementação de um banco de metadados para o sistemas de informação de biodiversidade do estado de são paulo. Master's thesis, Unicamp, 1999.

[11] R. da S. Torres, C. M. B. Medeiros, M. A. Gonçalves, and E. A. Fox. A Digital Library Framework for Biodiversity Information System. *International Journal on Digital Libraries*, 6(1):3–17, 2006.

[12] J. Daltio. Aondê: Um serviço web de ontologias para interoperabilidade em sistemas de biodiversidade. Master's thesis, Unicamp, 2007.

[13] J. Daltio, C. B. Medeiros, L. Gomes Jr., and T. M. Lewinsohn. A framework to process complex biodiversity queries. In *SAC '08: Proceedings of the 2008 ACM symposium on Applied computing*, pages 2293–2297, New York, NY, USA, 2008. ACM.

[14] S. R. de Sousa. A semantic approach to describe geospatial resources. *LNCS – ER 2009 Workshops*, 5833:327–336, 2009.

[15] Deegree - Free Software for Spatial Data Infrastructures. http://www.deegree.org/. Acessado em 17/10/2009.

[16] L. Delcambre and D. Maier. Models for Superimposed Information. *Lecture Notes in Computer Science, Advances in Conceptual Modeling ER '99*, 1727:264–280, 1999.

[17] L. Delcambre, D. Maier, S. Bowers, M. Weaver, L. Deng, P. Gorman, J. Ash, M. Lavelle, and J. Lyman. Bundles in Captivity: An Application of Superimposed Information. *Proceedings of the 17th International Conference on Data Engineering*, pages 111–120, 2001.

[18] Dublin Core Metadata Initiative. http://dublincore.org. Acessado em 18/05/2008.

[19] ESRI Shapefile Technical Description. http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf. Technical report, ESRI, 1998.

[20] A. G. Filho. A Comunidade de Borboletas Frugívoras da Reserva de Santa Genebra, Campinas, SP, Com Ênfase na Flutuação Populacional de Anaea Ryphea (Cramer) (Nymphalidae: Charanidae) e Sua Relação com as Plantas Hospedeiras. Master's thesis, Unicamp, 2003.

[21] GeoServer. http://geoserver.org/display/geos/welcome. Acessado em 17/10/2009.

[22] GeoTools. http://geotools.codehaus.org/. Acessado em 17/10/2009.

[23] GISMAPS. http://www.gismaps.com.br/. Acessado em 17/10/2009.

[24] T. Gruber. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal of Human-Computer Studies*, 43(5–6):907–928, 1995.

[25] R. H. Güting. An introduction to spatial database systems. *The VLDB Journal*, 3(4):357–399, 1994.

[26] P. Hall. Biodiversity E-tools to Protect our Natural World. *Converging Sciences Conference*, 2004.

[27] G. N. Hess, C. Iochpe, and S. Castano. Towards a Geographic Ontology Reference Model for Matching Purposes. *Simpósio Brasileiro de Geoinformática - GEOINFO*, pages 25–28, 2007.

[28] Introduction to SVG. http://www.w3schools.com/svg/svg_intro.asp. Acessado em 17/10/2009.

[29] L. C. Gomes Jr. Uma arquitetura para consultas a repositórios de biodiversidade na web. Master's thesis, Unicamp, 2007.

[30] C. G. N. Macario, S. R. Sousa, and C. M. B. Medeiros. Annotating geospatial data based on its semantics. In *17th ACM SIGSPATIAL Conference*, New York, NY, USA, 2009. ACM.

[31] C. G. N. Macário and C. B. Medeiros. Specification of a framework for semantic annotation of geospatial data on the web. *SIGSPATIAL Special*, 1(1):27–32, 2009.

[32] D. Maier and L. Delcambre. Superimposed Information for the Internet. In *Proceedings of the ACM SIGMOD Work-shop on the Web and Databases*, pages 1–9, 1999.

[33] J. E. G. Malaverri. Um serviço de gerenciamento de coletas para sistemas de biodiversidade. Master's thesis, Unicamp, 2009.

[34] J. G. Malaverri, B. S. C. M. Vilar, and C. B. Medeiros. A tool based on web services to query biodiversity information. In *WEBIST*, pages 305–310, 2009.

[35] MapServer. http://mapserver.org/. Acessado em 17/10/2009.

[36] S. P. Matias. Processamento de Consultas ao Banco de Dados de Biodiversidade do BIOTA. Master's thesis, Unicamp, 2000.

[37] S. Murthy, D. Maier, and L. Delcambre. Mash-o-matic. In *DocEng '06: Proceedings of the 2006 ACM Symposium on Document Engineering*, pages 205–214, 2006.

[38] S. Murthy, D. Maier, L. Delcambre, and S. Bowers. Superimposed Applications using SPARCE. In *20th International Conference on Data Engineering*, page 861, 2004.

[39] U. Murthy, K. Ahuja, S. Murthy, and E. A. Fox. SIMPEL: A Superimposed Multimedia Presentation Editor and Player. In *6th Joint Conference on Digital Libraries*, pages 11–15, 2006.

[40] U. Murthy, R. da S. Torres, and E. A. Fox. Integrated Support for Superimposed Applications in Biomedical Information Systems. 2006. Virginia Tech.

[41] U. Murthy, R. da S. Torres, and E. A. Fox. SIERRA - A Superimposed Application for Enhanced Image Description and Retrieval. *European Conference on Digital Libraries*, pages 540–543, 2006.

[42] U. Murthy, E. A. Fox, Y. Chen, E. M. Hallerm, and T. R. C. Falcão R. da S. Torres, E. J. Ramos. Superimposed image description and retrieval for fish species identification. In *ECDL '09: Proc. of the 13th European conference on Research and Advanced Technology for Digital Libraries*, 2009.

[43] U. Murthy, D. Maier, L. Delcambre, and S. Bowers. Putting Integrated Information into Context: Superimposing Conceptual Models with SPARCE. In *First Asia-Pacific Conference of Conceptual Modeling*, pages 71–80, 2004.

[44] U. Murthy, R. Richardson, and E. A. Fox. Enhancing Concept Mapping Tools Below And Above To Facilitate The Use Of Superimposed Information. *Proceedings of the Second International Conference on Concept Mapping*, pages 5–8, 2006.

[45] J. Nogueras-Iso, F. J. Zarazaga-Soria, J. Lacasta, R. Béjar, and P. R. Muro-Medrano. Metadata Standard Interoperability: Application in the Geographic Information Domain. *Computers, environment and urban systems*, 28(6):611–634, 2003.

[46] O que é biodiversidade? http://www.wwf.org.br/informacoes/questoes_ambientais/biodiversidade/index.cfm. Acessado em 17/10/2009.

[47] O que é um Tablet PC? . http://www.microsoft.com/brasil/windowsxp/tabletpc/evaluation/about.mspx. Acessado em 17/10/2009.

[48] OpenGIS® Standards and Related OGC documents. http://www.opengeospatial.org/standards. Acessado em 17/10/2009.

[49] OWL_Working_Group. http://www.w3.org/2007/owl/wiki/owl_working_group. Acessado em 17/10/2009.

[50] G. R. Queiroz and K. R. Ferreira. Tutorial sobre Bancos de Dados Geográficos. Technical report, Instituto Nacional de Pesquisas Espaciais, 2006, Disponível em http://www.dpi.inpe.br/TutorialBdGeo_GeoBrasil2006.pdf, Acessado em 17/10/2009.

[51] Scalable Vector Graphics (SVG) – XML Graphics for the Web. http://www.w3.org/graphics/svg/. Acessado em 17/10/2009.

[52] H. N. Uchoa and P. R. Ferreira. Geoprocessamento com Software Livre, 2004, Disponível em http://www.geolivre.org.br, Acessado em 17/10/2009.

[53] R. Veigas and V. Soares. Querying a geographic database using an ontology-based methodology. *Simpósio Brasileiro de Geoinformática - GEOINFO*, pages 19–22, 2006.

[54] L. R. Venancio, R. Fileto, and C. B. Medeiros. Aplicando Ontologias de Objetos Geográficos para Facilitar Navegação em GIS. *Simpósio Brasileiro de Geoinformática - GEOINFO*, pages 1–1, 2003.

[55] B. S. C. M. Vilar. Processamento de consultas baseado em ontologias para sistemas de biodiversidade. Master's thesis, Unicamp.

[56] XHTML 1.0 – DTDs. http://www.w3.org/tr/xhtml1/dtds.html. Acessado em 06/10/2009.