UNIVERSIDADE FEDERAL DE CAMPINA GRANDE

CENTRO DE CIÊNCIAS E TECNOLOGIA - CCT

DEPARTAMENTO DE SISTEMA E COMPUTAÇÃO - DSC

CURSO DE MESTRADO EM INFORMÁTICA - COPIN

DISSERTAÇÃO DE MESTRADO

GEOLOCALIZADOR: UM SISTEMA DE REFERÊNCIA ESPAÇO-TEMPORAL INDIRETA NA WEB

EDER PAZINATTO

(MESTRANDO)

DR. CLÁUDIO DE SOUZA BAPTISTA

(ORIENTADOR)

CAMPINA GRANDE
JANEIRO – 2003

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE

CENTRO DE CIÊNCIAS E TECNOLOGIA - CCT

DEPARTAMENTO DE SISTEMA E COMPUTAÇÃO - DSC

CURSO DE MESTRADO EM INFORMÁTICA - COPIN

GEOLOCALIZADOR: UM SISTEMA DE REFERÊNCIA ESPAÇO-TEMPORAL INDIRETA NA WEB

EDER PAZINATTO

Dissertação submetida à Coordenação de Pós-Graduação em Informática do Centro de Ciências e Tecnologia da Universidade Federal de Campina Grande, como requisito parcial para obtenção do grau de Mestre em Informática.

Área de Concentração: Ciência da Computação

Linha de Pesquisa: Sistemas de Informações e Banco de Dados

DR. CLÁUDIO DE SOUZA BAPTISTA

(ORIENTADOR)

CAMPINA GRANDE
Janeiro – 2003

PAZINATTO, Eder

P348G

GeoLocalizador: Um Sistema de Referência Espaço-Temporal Indireta na Web

Dissertação (Mestrado), Universidade Federal de Campina Grande, Centro de Ciências e Tecnologia, Coordenação de Pós Graduação em Informática,

Campina Grande, Janeiro de 2003.

115 p. Il.

Orientador: Cláudio de Souza Baptista

Palavras-chaves:
1. Banco de Dados
2. Gazetteers
3. Thesaurus
4. Sistema de Informação Geográfica
5. Referência Espacial Indireta

CDU – 681.3.07B

“Só existem dois dias no ano que nada pode ser feito:
um se chama ONTEM e ou outro AMANHÃ. Portanto,
HOJE é o dia certo para amar, acreditar, fazer e,
principalmente, viver”.

(Dalai-Lama)

Dedico este trabalho aos meus
pais, Mario e Alaídes Pazinatto;
meu irmão Cleber; e a minha nona
Júlia B. Pazinatto(em memória).

Agradecimentos

Em primeiro lugar, gostaria de agradecer a Deus, por me dar forças e saúde para realizar este trabalho.

Agradecimentos especiais aos meus pais Mario e Alaídes, e ao meu irmão Cleber pela compreensão no momento em que me afastei de casa. O apoio, incentivo, e a torcida de vocês foram indispensáveis. Também agradeço, aos amigos e parentes do Rio Grande do Sul, que mesmo à distância torceram por mim.

Aos colegas de apartamento Josué, Fábio e Paulo Elyel. Muito obrigado pela coragem, pelo incentivo, confiança e ajuda em todos os momentos. Agradecimentos especiais a Laise pelo carinho, atenção e confiança que me deram forças e serenidade para continuar meu trabalho.

Sinceros agradecimentos a meu orientador Cláudio S. Baptista pela seriedade, competência, paciência e dedicação.

As amigas Ladjane e Fabiana pela alegria e amizade. Aos amigos Leandro e professor Jorge pelos momentos de descontração.

Agradecimentos ao pessoal do SINBAD (Laboratório de Sistemas de Informações e Banco de Dados) Benitz, Carlos Alexandre, Rodrigo Rebouças, Rodrigo Vilar e Rômulo pela imensa ajuda. Aos colegas do mestrado Claudivan, Philip, Plácido e Ricardo pelas inúmeras trocas de idéias.

Agradeço a todos professores e funcionários do Departamento de Sistemas e Computação da UFCG. Em especial a Aninha, Vera e Zeneide.

Sem vocês, tudo seria muito mais difícil.

Sumário

Agradecimentos v

Sumário vi

Lista de Figuras viii

Lista de Tabelas ix

Lista de Siglas e Abreviaturas x

Resumo xii

Abstract xiii

Capítulo 1 14

Introdução 14

1.1 Problemas na Recuperação de Informação Geográfica 15
1.2 Motivação 16
1.3 Objetivos 17
1.4 Organização da Dissertação 18

Capítulo 2 20

Sistemas de Recuperação de Informação 20

2.1 Ferramentas de Busca 21
2.2 Bibliotecas Digitais 22
2.2.1 Bibliotecas Digitais Georeferenciadas 24
2.3 Sistemas de Informação Espaço-Temporal 26
2.3.1 Dados Espaciais 27
2.3.1.1 Relacionamentos Espaciais 30
2.3.1.2 Operadores Espaciais 31
2.3.2 Dados Temporais 32
2.3.2.1 Relacionamentos Temporais 33
2.3.2.2 Operadores Temporais 34
2.3.3 O Padrão OpenGis 35
2.4 Thesaurus 37
2.4.1 Tipos de Relacionamentos 39
2.4.1.1 Relacionamentos de Equivalência 39
2.4.1.2 Relacionamentos Hierárquicos 40
2.4.1.3 Relacionamentos Associativos 41
2.4.2 WordNet 41

Capítulo 3 43

Sistemas de Referência Espaço-Temporal Indireta 43

3.1 Gazetteers 43
3.2 Trabalhos Relacionados 47

3.2.1 Alexandria Digital Library 47
3.2.2 Perseus 50
3.2.3 Geographic Names Information System 53
3.2.4 Getty Thesaurus of Geographic Names 55
3.2.5 Genuki 57
3.3 Considerações Finais 59

Capítulo 4 61

Modelagem do GeoLocalizador 61

4.1 Modelo Conceitual 61
4.2 Levantamento de Requisitos 63
4.2.1 Requisitos Funcionais 63
4.2.2 Requisitos Não Funcionais 66
4.3 Modelo de Casos de Uso 67
4.4 Diagrama de Classes 71
4.4.1 Classes da Camada de Apresentação 72
4.4.2 Classes da Camada Lógica 73
4.4.3 Classes da Camada de Dados 75
4.5 Modelagem do Thesaurus Geográfico 76

Capítulo 5 78

Arquitetura e Implementação do GeoLocalizador 78

5.1 Arquitetura do Sistema 78
5.1.1 Camada de Apresentação 79
5.1.2 Camada Lógica 80
5.1.3 Camada de Dados 81
5.2 Tecnologias Utilizadas 81
5.3 Thesaurus Geográfico 83
5.4 O GeoLocalizador 86
5.4.1 Interface do Usuário 86
5.4.2 Processo de Consulta 87
5.4.3 Operações Espaciais e Temporais 90
5.4.4 Mapa Geográfico 93
5.4.5 Apresentação dos resultados 94
5.5 Considerações Finais 99

Capítulo 6 101

Conclusão 101

6.1 Principais Contribuições 101
6.2 Trabalhos Futuros 103
6.3 Considerações Finais 104

Referências Bibliográficas 106

Publicações 113

Apêndice A 114

Lista de Figuras

Figura 2.1- Arquitetura de uma ferramenta de busca 22
Figura 2.2- Relacionamentos entre os conjuntos A e B, utilizando o método *4-intersection* 31
Figura 2.3- Relacionamentos temporais 34
Figura 3.1- Interface do *Gazetteer* do projeto ADL 50
Figura 3.2- Interface de consulta do projeto Perseus 52
Figura 3.3- Interface do consulta do projeto GNIS 54
Figura 3.4- Interface de consulta do TGN 57
Figura 3.5- Interface de consulta do *Gazetteer* Genuki 59
Figura 4.1- Modelo Conceitual do GeoLocalizador 62
Figura 4.2- Diagrama de Casos de Uso 68
Figura 4.3- Diagrama representando as classes da camada de Apresentação 72
Figura 4.4- Diagrama representando as classes da camada Lógica 73
Figura 4.5- Diagrama representando as classes da camada de Dados 75
Figura 4.6- Diagrama de classes do *thesaurus* geográfico 77
Figura 5.1- Arquitetura do GeoLocalizador 79
Figura 5.2- Interface do usuário 87
Figura 5.3- Interface com a forma de consulta direta 88
Figura 5.4- Interface de consulta com a árvore geográfica 88
Figura 5.5- Interface de seleção de lugares específicos 90
Figura 5.6- Interface com os operadores espaciais e temporais 92
Figura 5.7- Interface de localização do lugar geográfico 94
Figura 5.8- Interface com os resultados da consulta 95
Figura 5.9- Interface com a primeira fase da consulta 96
Figura 5.10- Interface com a expansão da consulta 97
Figura 5.11- Interface com a consulta expandida 98

Lista de Tabelas

Tabela 2.1- Principais características das Bibliotecas Digitais 23
Tabela 2.2- Processos de análise espacial 30
Tabela 3.1- Problemas na recuperação de Informação Geográfica 60
Tabela 4.1- Caso de Uso DigitarTermo 69
Tabela 4.2- Caso de Uso NavegarHierarquia 69
Tabela 4.3- Caso de Uso SelecionarTermo 69
Tabela 4.4- Caso de Uso DefinirOperadorEspacial 69
Tabela 4.5- Caso de Uso DefinirPeriodoTempo 70
Tabela 4.6- Caso de Uso DefinirOperadorTemporal 70
Tabela 4.7- Caso de Uso SubmeterConsulta 70
Tabela 4.8- Caso de Uso ExpandirConsulta 70
Tabela 4.9- Caso de Uso PesquisarRelacionamentos 71
Tabela 4.10- Caso de Uso ApresentarResultado 71
Tabela 5.1- Principias operadores do *thesaurus* 84
Tabela 5.2- Principais funções do *thesaurus* 85

Lista de Siglas e Abreviaturas

AAT- Art and Architecture Thesaurus

ACM- Association for Computing Machinery

ADEPT- Alexandria Digital Earth Prototype

ADL- Alexandria Digital Library

API- Application Programming Interfaces

BGN- Board on Geographic Names

BT- Broader Term

BTG- Broader Term Generic

BTP- Broader Term Partitive

CAD- Computer Aided-Design

CAM- Computer Aided Manufacturing

CEP- Código de Endereços Postais

CORBA- Common Object Request Broker Architecture

DCE- Distributed Computing Environment

DEM- Dimension Extended Method

DLF- Digital Library Federation

DTD- Document Type Definition

FFHS- Federation of Family History Societies

GML- Geographic Markup Language

GNIS- Geographic Names Information System

GNPS- Geographic Names Processing System

GoONS- Guild of One Name Studies

GPS- Global Positioning System

GRI- Getty Research Institute

HTML- HyperText Markup Language

HTTP- HyperText Transfer Protocol

IEEE- Institute of Electrical and Electronics Engineers

JCDL- Joint Conference on Digital Libraries

JDBC- Java DataBase Connectivity

JDK- Java Development Kit

JSP- Java Server Pages

MBR- Minimum Bounding Rectangle

NEP- Número de Endereço Postal

NGNDB- National Geographic Names Data Base

NIMA- National Imagery and Mapping Agency

NSF- National Science Foundation

NT- Narrower Term

NTG- Narrower Term Generic

NTP- Narrower Term Partitive

OGC- Open GIS Consortium

OpenGis- Open Geodata Interoperability Specification

PRO- Public Records Office

PT- Preferred Term

RDB- Reference Data Base

RT- Related Term

SGBD- Sistema de Gerência de Banco de Dados

SGML- Standard Generalized Markup Language

SIG- Sistema de Informação Geográfica

SoG- Society of Genealogists

SPIRIT- Spatially-Aware Information Retrieval on the Internet

SQL- Structured Query Language

SRI- Sistemas de Recuperação de Informação

SRL- Sistema de Referência Linear

SVG- Scalable Vector Graphics

TEI- Text Encoding Initiative

TGN- Getty Thesaurus of Geographic Names

TMNDB- Topographic Map Names Data Base

ULAN- Union List of Artist Names

UML- Unified Modeling Language

URL- Uniform Resource Locator

USGS- United States Geological Survey

UTM- Universal Transverse Mercator

XML- eXtensible Markup Language

XSLT- eXtensible Stylesheet Language Transformation

XSU- XML SQL Utility

W3C- World Wide Web Consortium

Resumo

A Internet suporta aplicações com diferentes domínios. Um dos domínios que estão sendo direcionados os sistemas na *Web* são os sistemas de informação espacial. Atualmente, um grande volume de dados espaciais está sendo capturado em diferentes formatos (*raster* ou *vector*) e armazenados na *Web*. Estes dados devem ser indexados para serem utilizados e justificarem o alto investimento que vem sendo feito para obtê-los. Por esta razão, serviços para dados espaciais, tais como *Gazetteers* e ferramentas de busca, são necessários para recuperação e tratamento destas informações. Nesta dissertação, apresentamos um *Web Gazetteer*, o qual provê, via uma interface gráfica, serviços para dados espaciais através do nome dos lugares geográficos. Utilizamos um *thesaurus* para a expansão da consulta e um banco de dados espacial para persistência dos dados e execução das consultas espaciais.

Abstract

The Internet has underpinned applications in different domains. One of the application domains that has been pushed towards web-based systems is spatial information systems. Currently, a huge volume of spatial data is being captured in different data formats (e.g. raster, vector) and stored somewhere on the Web. These data must be indexed in order to become useful and justify the high investment that has been done so far in their capture. Therefore, service for spatial data tools such as spatial searching engines and Gazetteers are demanded in order to better fulfill these requirements. In this thesis, we present a Web based Gazetteer, which aims to provide, via a GUI, services for spatial data in order to find place names. We use a thesaurus for query expansion and a spatial database for data persistence and spatial query manipulation.

Capítulo 1

Introdução

A crescente pesquisa e o desenvolvimento das tecnologias de informação e comunicação deram origem a um grande número de inovações, dentre as quais a Internet é a mais revolucionária. O avanço destas tecnologias possibilitou o desenvolvimento de novos métodos de criação e disseminação das informações.

Com a expansão da Internet, um novo paradigma de troca e recuperação de informações foi estabelecido. Simultaneamente ao crescimento do número de usuários que a utilizam, vem se verificando também um crescimento muito grande da quantidade de informações disponíveis na *Web*. Como resultado, a Internet tem-se tornado uma fonte de todos os tipos de informações e em diferentes formatos, incluindo não apenas dados textuais, mas também outras mídias, tais como imagens, vídeos e áudios. Estas características têm motivado a grande demanda de usuários em busca de informações geográficas. No entanto, na *Web* existe uma grande carência de sistemas capazes de indexar e apresentar informações espaciais.

Atualmente, as consultas na *Web* são efetuadas geralmente através da navegação exaustiva em páginas da Internet que dificilmente resulta em sucesso, devido ao grande número de dados ou através da pesquisa com palavras-chaves, utilizando as ferramentas de busca. A utilização destas ferramentas apresenta desvantagens, pois elas não possuem suporte a dados espaciais, e os resultados abrangem diferentes categorias de informações, cabendo ao usuário realizar o processo de analisar as informações que são relevantes.

Portanto, existe uma lacuna a ser preenchida com serviços para dados espaciais. Um *Gazetteer* é um destes serviços que permitem a recuperação e localização de dados espaciais usando o nome do lugar geográfico (*placename*).

Os Sistemas de Informações Geográficas – SIG também têm ganhado importância devido a sua ampla gama de aplicações, incluindo os mais diversos campos, como cartografia, planejamentos urbano e rural, meio ambiente, dentre outros. Contudo, estes sistemas não se propõem a oferecer operações para dados espaciais disponibilizados de forma independente de aplicação, como é o caso de um

Gazetteer ou mesmo das ferramentas de buscas que são capazes de indexar e disponibilizar estes dados para que possam ser analisados [Pazinatto *et al*, 2002a].

Os *Gazetteers* representam uma forma de referência indireta a informações espaciais, ou seja, são ferramentas capazes de indexar informações espaciais e não espaciais a partir do nome de um lugar geográfico.

Algumas propriedades caracterizam um dado geográfico, por isso precisamos levar em consideração alguns atributos que são fundamentais, tais como: um nome geográfico, responsável pela sua identificação; uma localização geográfica, que é expressa em coordenadas espaciais geográficas (por exemplo, latitude e longitude); e um tipo de feição (*feature type*), que determina a característica do dado geográfico (por exemplo: país, estado, município). Da mesma forma, a temporalidade das informações também deve ser considerada.

1.1 Problemas na Recuperação de Informação Geográfica

O desenvolvimento de aplicações voltadas para a Internet tem crescido muito nos últimos anos. Atualmente, existe uma série de sistemas (softwares) disponíveis na Internet que manipulam informações geográficas. Porém, a maioria desses sistemas armazena apenas informações espaciais sobre um determinado lugar geográfico, ignorando a temporalidade dos dados. Tais sistemas não conseguem manter informações sobre as mudanças geográficas e acontecimentos que ocorreram durante um período de tempo. Com isso, importantes mudanças, principalmente mudanças geopolíticas, não podem ser acessadas pelos usuários.

Além disso, muitos usuários de SIG encontram dificuldades na busca de informações e na exata localização de um determinado lugar geográfico. Estes problemas devem-se à falta de mecanismos capazes de refinar, ou seja, de excluir itens irrelevantes em uma consulta, ou de expandir uma consulta, através de relacionamentos entre os lugares geográficos e simples informações textuais.

Outro problema encontrado nos sistemas atuais de localização geográfica é a falta de disponibilidade e utilização dos dados espaciais. Este problema, que prejudica principalmente os usuários menos experientes em geoprocessamento, pode ser resolvido com a construção de sistemas com interfaces mais amigáveis. Alguns trabalhos nesta área já foram desenvolvidos, como em [Li *et al*, 1996], que dá ênfase

ao processo de desenvolvimento de interfaces *Web* para melhorar o acesso e a utilização dos dados espaciais.

1.2 Motivação

As técnicas de aquisição de dados espaciais com a utilização de satélites cada vez mais sofisticados conseguem alcançar altas resoluções e mandam *terabytes* de dados espaciais ao ano. Esses dados estão em diferentes formatos (*vector* e *raster*) e são armazenados em repositórios. Contudo, existe uma grande resistência, dos órgãos que detêm esses dados, para a disponibilização e utilização dos mesmos.

Embora exista um grande volume de informações espaciais, há uma enorme carência de sistemas na *Web* capazes de indexar, disponibilizar e permitir a manipulação destas informações. Estes sistemas devem ser capazes de recuperar dados espaciais e não espaciais sobre os lugares geográficos, com representação de mapas e mecanismos que possibilitem a execução de operações de seleção espacial sobre tais mapas geográficos.

Alguns sistemas que indexam informações espaciais são baseados na utilização das coordenadas (por exemplo: coordenadas de latitude e longitude) para fazer referência a um lugar geográfico. Porém, os usuários não estão acostumados com esta forma de referência, pois não possuem conhecimento sobre as coordenadas de um lugar geográfico. Portanto, existe a necessidade de se construir ferramentas que suportem consultas através do nome do lugar geográfico.

Problemas com a recuperação de informações espaciais podem ser resolvidos com a construção de *Gazetteers*, que provêem a recuperação e a localização de informações espaciais, usando o nome do lugar geográfico. Atualmente, existem alguns *Gazetteers* disponíveis na Internet, porém estes apresentam uma série de deficiências com relação aos dados espaciais e temporais relacionados a um lugar geográfico.

Muitos sistemas que realizam o tratamento de informações geográficas têm negligenciado a temporalidade dos dados. Muitos dados espaciais sofrem alterações com o passar do tempo, e sem a temporalidade dos dados não é possível manter um histórico sobre um dado geográfico. Sendo assim, uma determinada informação sobre um lugar pode perder relevância se ela não estiver diretamente associada a um

período de tempo. Diante disso, sistemas com suporte para as operações espaciais e temporais precisam ser desenvolvidos.

1.3 Objetivos

Como visto na seção anterior, existem diversos problemas relacionados à recuperação de informações espaciais. O principal objetivo deste trabalho é amenizar grande parte destes problemas, com a modelagem e implementação de um sistema de referência espaço-temporal indireta (*Gazetteer*), que incorpore técnicas de recuperação de informações geográficas.

Nesta dissertação, propomos a construção de um *Gazetteer*, chamado GeoLocalizador. As principais características do GeoLocalizador são as seguintes:

- Disponibilidade do sistema na *Web*: o GeoLocalizador pode ser acessado na *Web* com a utilização de um *browser*;

- Utilização de um SGBD (Sistema de Gerência de Banco de Dados) Objeto-Relacional: utilização de um banco de dados com suporte às operações espaciais (Dentro, Contém, Dentro de e Coberto por).

- Diferentes formas de pesquisa: possibilidade de os usuários optarem por duas formas de submeter uma consulta. Através da forma direta com a digitação do lugar geográfico em uma caixa de texto, ou com a navegação e seleção de um lugar em uma árvore com dados geográficos;

- Suporte à temporalidade dos dados: a temporalidade dos dados indica o intervalo de tempo válido de um determinado lugar geográfico;

- Suporte a operações espaciais e temporais: existência de ferramentas que permitem a execução de operações espaciais e temporais sobre os dados geográficos;

- Apresentação de um mapa geográfico: a posição do lugar geográfico que está sendo pesquisado é representada em um mapa;

- Diferentes formas de representar a posição de um lugar geográfico: apresentação de diferentes formas, compostas por coordenadas de latitude e longitude, que representam a geometria de um lugar geográfico Essas formas são: MBR (retângulo envolvente mínimo)- menor retângulo possível que cobre a área de um lugar geográfico; e Centróide- formado por duas coordenadas, que representam um ponto central obtido através das quatro coordenadas que formam o MBR do lugar geográfico;

- Métodos de expansão da consulta: construção de um *thesaurus* de nomes geográficos, que permite a expansão dos termos e a recuperação de informações espaciais com detalhamento dos relacionamentos de um *thesaurus.*

1.4 Organização da Dissertação

O restante desta dissertação está estruturado em mais cinco capítulos, que são descritos a seguir:

No capítulo 2, apresentamos uma visão geral sobre os sistemas de recuperação de informação. Neste capítulo, abordamos as bibliotecas digitais e a utilização das dimensões espaciais e temporais no processo de recuperação de informações geográficas. Também neste capítulo, abordamos a utilização dos *thesauri* no processo de indexação de informações nos sistemas de recuperação.

No capítulo 3, introduzimos os *Gazetteers*, que são a principal área de investigação desta dissertação. Neste capítulo, ilustramos alguns trabalhos, disponíveis na Internet, relacionados com nossa pesquisa. Dentre estes trabalhos, abordamos bibliotecas digitais, bibliotecas digitais georeferenciadas e *Gazetteers*, que realizam o tratamento de informações geográficas.

O capítulo 4 aborda os aspectos da modelagem utilizados na construção do GeoLocalizador. O modelo conceitual baseado nos principais componentes de um *Gazetteer*, bem como os requisitos e o diagrama de classes do sistema, são apresentados neste capítulo.

No capítulo 5, apresentamos a arquitetura do GeoLocalizador. Também descrevemos as tecnologias utilizadas e o modelo de implementação, com apresentação das interfaces envolvendo as diferentes fases do processo de consulta. Aspectos da construção do *thesaurus* geográfico também são discutidos neste capítulo.

Por fim, no capítulo 6, descrevemos as principais contribuições desta dissertação. Neste capítulo, são delineadas direções futuras dando continuidade aos trabalhos aqui desenvolvidos.

Capítulo 2

Sistemas de Recuperação de Informação

Os Sistemas de Recuperação de Informação (SRIs) são usados para armazenar, manter, pesquisar e recuperar informações. Essas informações podem estar em diversos formatos, tais como: textos, imagens, áudios e vídeos [Kowalski, 1997].

Os SRIs vêm sendo pesquisados e desenvolvidos há mais de três décadas, e muitos protótipos e sistemas experimentais têm sido desenvolvidos para bibliotecas, dicionários, enciclopédias, pesquisas e exibições de materiais em museus, livros de medicina e catálogos [Agosti & Smeaton, 1996].

Contudo, os SRIs têm sofrido algumas mudanças nos últimos anos. Estas mudanças devem-se, principalmente, a três motivos: a expansão da Internet, a adoção de interfaces gráficas mais funcionais e modernas, e o aumento da capacidade dos dispositivos de armazenamento e processamento.

Podemos dividir a evolução dos SRIs em três gerações distintas. Na primeira geração, os SRIs basicamente automatizaram tecnologias como os catálogos, a partir das quais eram permitidas apenas pesquisas baseadas no nome do autor e no título do documento. Na segunda geração, foram adicionadas mais funcionalidades aos SRIs, com a adição de pesquisas por temas e palavras-chave, bem como uma maior flexibilidade para construir consultas mais complexas. Na terceira geração, que está sendo desenvolvida atualmente, o principal interesse está em aperfeiçoar as interfaces gráficas, com características de hipertexto e com a utilização de uma arquitetura mais aberta, com adaptabilidade a mudanças [Baeza-Yates *et al*, 1999].

O principal objetivo de um SRI é minimizar o tempo que os usuários utilizam para localizar as informações de que necessitam [Kowalski, 1997]. Por isso, um SRI procura comparar as consultas que são submetidas, com um conjunto de documentos armazenados em uma coleção de dados, visando recuperar apenas as informações que são relevantes à consulta submetida pelo usuário. Para tanto, é importante que os SRIs possuam uma estrutura de dados consistente, ferramentas de busca rápidas e métodos

de recuperação eficazes, principalmente se o volume de dados for grande [Salton & McGill, 1983].

Neste capítulo, detalhamos diferentes tipos de SRIs. A seção 2.1 aborda o funcionamento das ferramentas de busca que são muito utilizadas na Internet. Na seção 2.2, apresentamos as principais características das bibliotecas digitais. Nesta seção, também detalhamos as bibliotecas digitais georeferenciadas, voltadas para o tratamento de informações geográficas. Na seção 2.3, são apresentados os sistemas de informação com adoção das dimensões espaciais e temporais. Também nesta seção, detalhamos alguns aspectos do padrão OpenGis. A seção 2.4 aborda *thesaurus*, uma forma de indexação muito utilizado nos SRIs.

2.1 Ferramentas de Busca

Com a expansão e popularização da Internet e com a grande demanda de usuários por busca de informações nesta rede, os SRIs baseados em textos e com consultas por palavras-chave vêm sendo muito utilizados.

Na última década, vários SRIs foram lançados no mercado, principalmente aqueles que são destinados a recuperar informações disponíveis na Internet; tais sistemas são conhecidos como ferramentas de busca [Rosenthal & Chu, 1996].

Atualmente, existem várias ferramentas de busca disponíveis na Internet em idiomas diferentes, dentre elas, podemos destacar as seguintes: Google[1], Yahoo[2] e AltaVista[3]. Estas ferramentas suportam consultas através de palavras-chave, conseguindo indexar e realizar pesquisas com enorme rapidez e recuperar várias páginas relacionadas ao item submetido na consulta.

O funcionamento de uma ferramenta de busca é bastante simples. Os usuários acessam-na através da sua URL (*Uniform Resource Locator*) e, através da interface de consulta fornecem os termos da pesquisa. As ferramentas de busca atuais possuem basicamente três componentes: interface de consulta, robôs de indexação e bancos de índices.

[1] http://www.google.com
[2] http://www.yahoo.com
[3] http://www.altavista.com

Os “robôs de indexação” percorrem a estrutura de grafo hipertexto da Internet, à procura de documentos que possam ser indexados no banco de índices. Através do componente de “interface de consulta” os usuários entram com o termo a ser pesquisado. Quando uma consulta é executada, é realizada uma busca no “banco de índices”.

Desta maneira, as ferramentas de busca realizam a comparação dos termos das consultas submetidas pelos usuários com os termos indexados nos bancos de índices e, então, retornam *links* para documentos que, possivelmente, possam atender à necessidade dos usuários. Na figura 2.1, representamos, de uma forma simples, a arquitetura de uma ferramenta de busca.

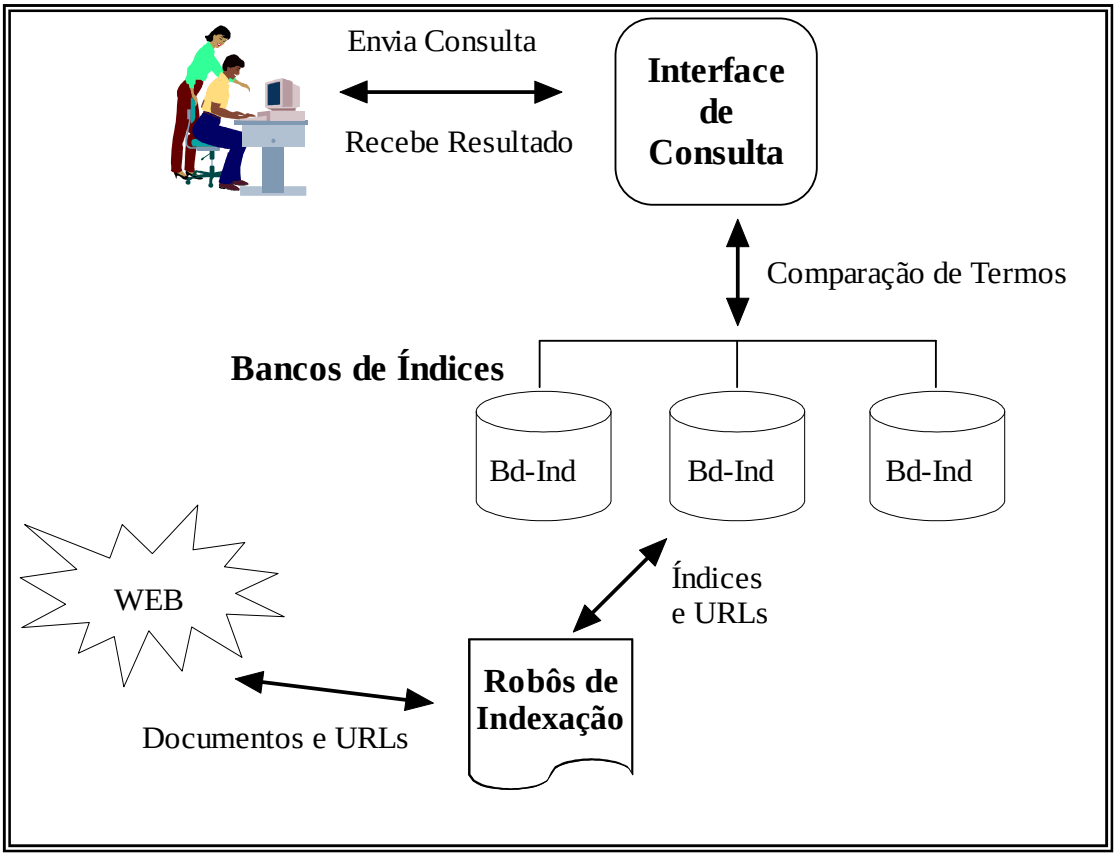


Figura 2.1- Arquitetura de uma ferramenta de busca

2.2 Bibliotecas Digitais

O surgimento e a grande circulação de documentos na forma digital impulsionaram o desenvolvimento das bibliotecas digitais (*Digital Libraries*). Uma biblioteca digital é uma coleção ordenada de informação com serviços associados, onde a informação é armazenada em formato digital e fica acessível numa rede [Arms, 1999].

O consórcio de bibliotecas e agências relacionadas a tecnologias de informação eletrônica, *Digital Library Federation* (DLF)[4], define as bibliotecas digitais como organizações que proporcionam os recursos, incluindo pessoal especializado, para selecionar, estruturar, oferecer acesso intelectual, interpretar, distribuir, preservar a integridade e assegurar a persistência das coleções dos trabalhos digitais, para que estes estejam disponíveis de um modo rápido e simples para uso por uma comunidade definida ou por um grupo de comunidades [Waters, 1998].

As bibliotecas digitais suportam as principais funcionalidades das bibliotecas tradicionais. Estas funcionalidades podem ser classificadas em quatros áreas:

- coleta de dados;
- organização e representação;
- acesso e recuperação;
- análise, síntese e disseminação da informação.

As principais características de uma biblioteca digital são apresentadas na tabela 2.1.

Características das Bibliotecas Digitais
- Todas operações são efetuadas através do computador
- Os dados estão no formato digital
- O formato digital permite facilidades na manipulação e edição de novos documentos
- Um documento pode produzir um número ilimitado de cópias sem perder a qualidade
- Os documentos não se desgastam com o manuseio e com o tempo, e ocupam pouco espaço físico ao serem armazenados
- Não existe a necessidade de reservas
- Os dados podem ser distribuídos pela Internet e recuperados remotamente
- Provê rápido acesso às informações
- Não possui restrição geográfica, pode ser acessada de qualquer parte do mundo
- O número de usuários é ilimitado
- Disponível 24 horas por dia e durante 7 dias da semana

Tabela 2.1- Principais características das Bibliotecas Digitais

[4] http://www.diglib.org

A principal inovação das bibliotecas digitais, em relação às bibliotecas tradicionais é o fato de que as informações estão no formato digital. Com este formato, não existe necessidade de operações de empréstimo e reservas de materiais, também a utilização de recursos físicos como prédios ou salas é minimizada. As informações ficam armazenadas em um banco de dados e são acessadas através da Internet.

Com o grande aumento e disseminação de informações na Internet, as bibliotecas digitais precisaram incorporar funcionalidades para suportar e recuperar documentos em diferentes formatos e mídias, tais como: mapas, imagens, áudio e vídeo. As facilidades de uso e a alta qualidade na recuperação de informações em diversos formatos tornam os SRIs essenciais para o sucesso das bibliotecas digitais.

Atualmente, existem diversas bibliotecas digitais disponíveis na *Web*, com domínios diferentes. Por exemplo, as bibliotecas IEEE[5] e ACM[6] possuem artigos técnicos publicados em revistas e conferências na área de computação, enquanto as bibliotecas ADL[7] e Perseus[8] agregam informações geográficas e textos clássicos sobre Grécia e Roma.

2.2.1 Bibliotecas Digitais Georeferenciadas

A funcionalidade e a versatilidade das bibliotecas digitais proporcionaram o desenvolvimento de aplicações voltadas para o tratamento de informações geográficas, envolvendo dados espaciais e temporais. Com isso, surgiram as bibliotecas georeferenciadas, também conhecidas como *geolibraries* [Goodchild, 1999].

Uma biblioteca georeferenciada é uma biblioteca digital com geoinformações (informações georeferenciadas), na qual as pesquisas são voltadas para lugares geográficos. A geoinformação é uma informação associada a um lugar ou região geográfica na superfície da Terra [MapSciCom, 1999].

Na Internet, através das ferramentas de busca, conseguimos recuperar informações sobre diversos lugares geográficos. Porém, estas ferramentas apresentam

[5] http://www.ieee.org
[6] http://www.acm.org/dl
[7] http://www.alexandria.ucsb.edu
[8] http://www.perseus.tufts.edu

problemas quando os usuários necessitam recuperar e realizar operações sobre as informações espaciais destes lugares. As bibliotecas digitais georeferenciadas procuram suprir estas deficiências.

O objetivo das bibliotecas digitais georeferenciadas é prover novos tipos de serviços baseados em lugares geográficos. Por exemplo, o processo de recuperação de informações em bibliotecas georeferenciadas pode ser realizado através de consultas, tais como [Zhu *et al*, 1999]: 'quais dados existem nesta região geográfica?'; e 'qual região registrou maior desmatamento na última década?'.

Para responder a estas consultas, as bibliotecas digitais georeferenciadas requerem o tratamento das dimensões espaciais e temporais, com funções de indexação, métodos de pesquisa, recuperação e apresentação destes dados. Processos como o refinamento das consultas e a forma de apresentação dos dados estão diretamente ligadas às interfaces gráficas de cada biblioteca georeferenciada.

O conteúdo de uma biblioteca georeferenciada não esta limitado apenas às informações associadas a mapas ou imagens da superfície da Terra, mas também inclui qualquer informação que pode ser relacionada com uma localização geográfica. Uma entidade georeferenciada possui, geralmente, dois componentes principais: os atributos espaciais (incluindo localização com a especificação de uma geometria) e os atributos descritivos, também conhecidos como atributos não espaciais.

Esse tipo de biblioteca possui as mesmas características de funcionamento de uma biblioteca digital, porém, com o domínio espaço-temporal. Nestas bibliotecas, os usuários emitem consultas com relação à localização de determinados lugares geográficos. A localização pode ser expressa através de coordenadas espaciais, com posições de latitude e longitude, e também através de relacionamentos espaciais, considerando parâmetros como proximidade de determinada região geográfica, ou com a ocorrência de algum evento.

O tratamento de alguns serviços é essencial em uma biblioteca georeferenciada. Alguns mecanismos para representar e referenciar as características espaciais e temporais devem ser implementados na interface gráfica dessas bibliotecas, tais como:

- representação de mapas, a partir das quais os usuários podem, através de técnicas de seleção espacial, selecionar regiões;

- realização de operações de *zoom* (usado para aproximação (*zoom in*), ou distanciamento (*zoom out*) de um objeto) e *pan* (usado para o movimentação horizontal ou vertical da área de visualização de um objeto, permitindo o deslocamento do campo de visão) sobre o mapa; e
- adoção de ferramentas, com suporte à execução de consultas que envolvam operadores espaciais e temporais.

2.3 Sistemas de Informação Espaço-Temporal

Os sistemas de informação espaço-temporal surgiram diante da necessidade de integrar as dimensões espaciais e temporais em um único modelo. A forma mais antiga e ainda mais comum de processar e apresentar informações sobre entidades espaço-temporal é através de mapas [Câmara, 1996]. Os elementos de um mapa são, em geral, armazenados de forma georeferenciada, segundo um sistema de coordenadas (por exemplo: latitude, longitude e altitude).

Um sistema de informação espaço-temporal possui as características das entidades espaciais e temporais. Os relacionamentos entre estas entidades estabelecem associações espaciais caracterizadas pela dimensão tempo, como, por exemplo, um cruzamento entre dois rios varia em forma e posição ao longo do tempo. Um relacionamento espaço-temporal é capaz de manter o registro das várias formas que este cruzamento teve ao longo do tempo.

Um dos sistemas mais utilizados nesta área é conhecido como Sistemas de Informações Geográficas (SIG). Os SIGs são sistemas usados para armazenar, consultar, manipular e visualizar dados geográficos [Aronoff, 1989], ou seja, dados que representam objetos e fenômenos em que a localização geográfica é uma característica inerente à informação e indispensável para analisá-la.

Uma das principais características de um SIG é a sua capacidade de manipular dados geográficos e descritivos de forma integrada, provendo-os de uma maneira consistente para análise e consulta. Dessa forma, é possível ter acesso a informações descritivas de uma entidade geográfica a partir de sua localização e vice-versa. Os SIGs são, hoje, uma realidade em diversas empresas públicas e privadas, onde são usados em diferentes domínios [Miranda *et al*, 2002].

Existe uma grande variedade de aplicações que envolvem dados geográficos (espaço-temporal). Estas aplicações incluem campos como planejamento urbano, otimização de rotas, demografia, cartografia, cadastros urbano e rural, administração de recursos naturais, controle de queimadas e epidemias [Aronoff, 1989].

A forma de obtenção dos dados, para suprir os sistemas de informação espaço-temporal tem variado nos últimos anos. Primeiramente, a captura dos dados era feita através de fotos, digitalização de mapas e dados censitários. Atualmente, as imagens de satélites conseguem capturar com rapidez uma grande quantidade de informações com elevado grau de precisão. Existe também um grande número de aplicações onde os dados são coletados e georeferenciados diretamente, com a utilização de um GPS (*Global Positioning System*).

Nesta seção, apresentamos os conceitos e as principais características dos dados espaciais e temporais. Na seção 2.3.1, abordaremos os dados espaciais com seus relacionamentos e operadores. Na seção 2.3.2, apresentaremos os relacionamentos e operadores temporais, bem como a importância da temporalidade dos dados nos sistemas de informação.

2.3.1 Dados Espaciais

O número de aplicações que utilizam informações espaciais tem aumentado significativamente, nos últimos dez anos. Entre estas aplicações, destacam-se os SIGs, aplicações CAD-CAM e robótica.

Um dado espacial é composto pelo *footprint* (área de cobertura de um dado espacial) de um lugar geográfico. Este *footprint* consiste de geometrias, como pontos, linhas, polígonos e imagens, que representam entidades espaciais, como lagos, rios, pontes, estradas e cidades.

Os mapas são um dos métodos mais tradicionais de armazenamento, análise e apresentação de dados espaciais. Através dos mapas podemos representar a forma como pensamos o espaço, em duas dimensões com diferentes formatos e escalas [Câmara, 1996]. Os mapas podem ser classificados como temáticos (apresentam dados relativos a um tema particular) e topográficos (contêm um conjunto de dados, distribuídos em diversos temas).

Um dado espacial contém atributos não espaciais como o nome do lugar, um tipo (cidade, lago, montanha, etc) referente ao lugar geográfico, além de outras informações. Esses atributos podem representar informações textuais, numéricas e audiovisuais.

Em um SIG, todo dado espacial é georeferenciado, possibilitando a execução de análises e consultas espaciais. O georeferenciamento é usado para localizar um lugar geográfico na superfície da terra ou em um mapa. Por exemplo: Rua Silva Barbosa, 1059, Bodocongó, Campina Grande-PB, Brasil. Neste trabalho, apresentamos dois métodos para representar o espaço: a referência indireta e direta [Hill, 1999].

O método de referência indireta é o mais conhecido e, conseqüentemente, o mais utilizado pelos usuários. Neste método, os usuários referenciam o espaço através de diferentes formas, para obter características ou a localização espacial. Os *Gazetteers* são exemplos de sistemas baseados na referência indireta a dados espaciais. A seguir, detalhamos algumas formas de referência indireta, tais como:

- Nomes dos lugares geográficos: a geo-referência é feita através do nome de países, cidades, rios, monumentos. Por exemplo: Campina Grande, Paraíba, Brasil. Esta forma de referência apresenta problemas, pois existem diversos locais geográficos com o mesmo nome (por exemplo: London, Inglaterra e London, Canadá) e existem diferentes nomes para um mesmo lugar geográfico (por exemplo: Pequim e Beijing, na China);
- Endereços postais: os endereços postais são utilizados na localização de casas, empresas, universidades, etc. Por exemplo: Rua Tuiuti, Santa Bárbara do Sul, RS, Brasil;
- Código de Endereços Postais (CEP): esta forma de geo-referência é bastante utilizada na entrega de correspondências postais. Cada país possui um sistema para a denominação dos CEPs. Por exemplo: CEP 58.109-010 – Brasil. O CEP não possui referência a características naturais;
- Número de Endereço Postal (NEP): embora o NEP também não possua referência a características naturais, ele é muito utilizado como forma de referência indireta de casas e escritórios [Longley *et al*, 2001]. Existem padrões de códigos de endereço que são utilizados em muitos países, por

exemplo: de um lado da rua, códigos com números pares (280, 282, 284), do outro lado da rua são utilizados números ímpares (281, 283, 285);

- Prefixo telefônico: forma de geo-referência utilizada para denominar regiões que possuem o mesmo prefixo telefônico. Por exemplo: 051, capital e região metropolitana do Rio Grande do Sul;
- Sistema de Referência Linear (SRL): esta forma de geo-referência é utilizada para localização de determinados lugares em ferrovias e rodovias. Por exemplo: Km 80 da rodovia BR 230;
- Métodos de seleção espacial: a geo-referência é realizada através do esboço (de um retângulo, polígono ou ponto) de uma região geográfica sobre um mapa.

A referência espacial direta utiliza sistemas de coordenadas para expressar a localização do objeto espacial em relação a sua posição na superfície da Terra. A seguir, apresentamos algumas formas de referência direta a dados espaciais:

- Latitude e longitude: usando as coordenadas de latitude e longitude qualquer ponto da superfície terrestre pode ser localizado por meio de graus, minutos e segundos. Latitude é a distância angular entre um ponto qualquer da superfície terrestre e os planos paralelos (círculos do globo terrestre cujos planos são perpendiculares ao eixo dos pólos norte e sul), a partir da linha do Equador. Longitude é a distância angular entre um ponto qualquer da superfície terrestre e os meridianos (círculos do globo terrestre que passam pelos pólos norte e sul), a partir do meridiano de Greenwich [Câmara, 1996]. As coordenadas de latitude variam de -90º a 90º (ou 90ºN (norte) e 90ºS (sul)) e de longitude variam de 180º a -180º (ou 180ºW (oeste) e 180ºE (leste)). Por exemplo: Moscou (55º37'N, 36 º0'E), latitude: 55 graus e 37 minutos ao Norte do Equador; e longitude: 36 graus a leste (East) de Greenwich;
- *Universal Transverse Mercator* (UTM): as coordenadas UTM utilizam a projeção *Mercator* e dividem a terra em 60 zonas (fusos) verticais, com 6º graus de longitude. Por exemplo: *Detroit* 4.695.438 e 328.887 fuso 17.

A referência direta exige dos usuários alto grau de conhecimento em posicionamento geográfico. Porém, representa a forma mais exata de referência

espacial. Existem projetos, como o *GeoWeb Map Browser*[9], que disponibilizam na Internet ferramentas de busca baseadas em consultas espaciais com referência direta.

2.3.1.1 Relacionamentos Espaciais

No tratamento das informações espaciais, precisamos levar em consideração que os dados geográficos não existem sozinhos no espaço, tão importante quanto a identificação e a localização, é a exploração dos relacionamentos entre estes dados.

Os relacionamentos entre os objetos espaciais são realizados através dos operadores espaciais. Estes relacionamentos podem ser categorizados em [Pullar & Egenhofer, 1988]:

- Topológicos: contém, disjunto, sobrepõe, dentro_de, encontra, igual, cobre e coberto_por;
- Direcionais: acima, abaixo, esquerda, direita, norte, sul, oeste, leste, nordeste, noroeste, sudeste e sudoeste;
- Métricos: área, perímetro, comprimento, distância, longe e próximo;
- Rede: conectado, próximo, anterior e menor_caminho;
- Conjuntos: união, interseção, diferente e igual;

Com a ajuda destes relacionamentos, podemos elaborar consultas com processos de análise espacial. Na tabela 2.2 são apresentados alguns exemplos:

Análise	**Questão**	**Exemplo**
Localização	Onde...?	Qual é a região de maior produção de abacaxi?
Condição	Quanto... ?	Qual é a produção deste estado?
Roteamento	Por onde ir...?	Qual é o melhor caminho para chegar a um determinado lugar?
Tendência	O que mudou...?	Qual era a área de produção de abacaxi há 10 anos atrás?
Proximidade	Quais são...?	Quais são as 10 cidades mais próximas da capital?

Tabela 2.2- Processos de análise espacial

[9] http://wings.buffalo.edu/cgi-bin/mapbrowse

2.3.1.2 Operadores Espaciais

Diversos estudos e pesquisas foram realizados e observou-se que a maioria dos relacionamentos entre objetos espaciais são de caráter topológico [Bogorny & Iochpe, 2001]. Utilizamos, nesta dissertação, os operadores dos relacionamentos topológicos; por esse motivo, esse método será apresentado nesta seção.

Os relacionamentos topológicos determinam se dois objetos interceptam-se em um plano e qual tipo de interseção existe entre estes objetos. Existem vários métodos que descrevem diferentes relacionamentos topológicos, expressando possíveis relações entre objetos geográficos. Os métodos mais conhecidos são: *4-intersection* [Egenhofer, 1991], *9-intersection* e *Dimension Extended Method* (DEM) [Clementini, 1993].

Os principais tipos de relacionamentos obtidos da união dos métodos dos relacionamentos topológicos são: disjunto, encontra (ou toca), igual, contém, sobrepõe, cobre, dentro_de e coberto_por. A figura 2.2 apresenta os relacionamentos topológicos segundo o método *4-intersection*. Esse método baseia-se em conjuntos para representar as topologias existentes.

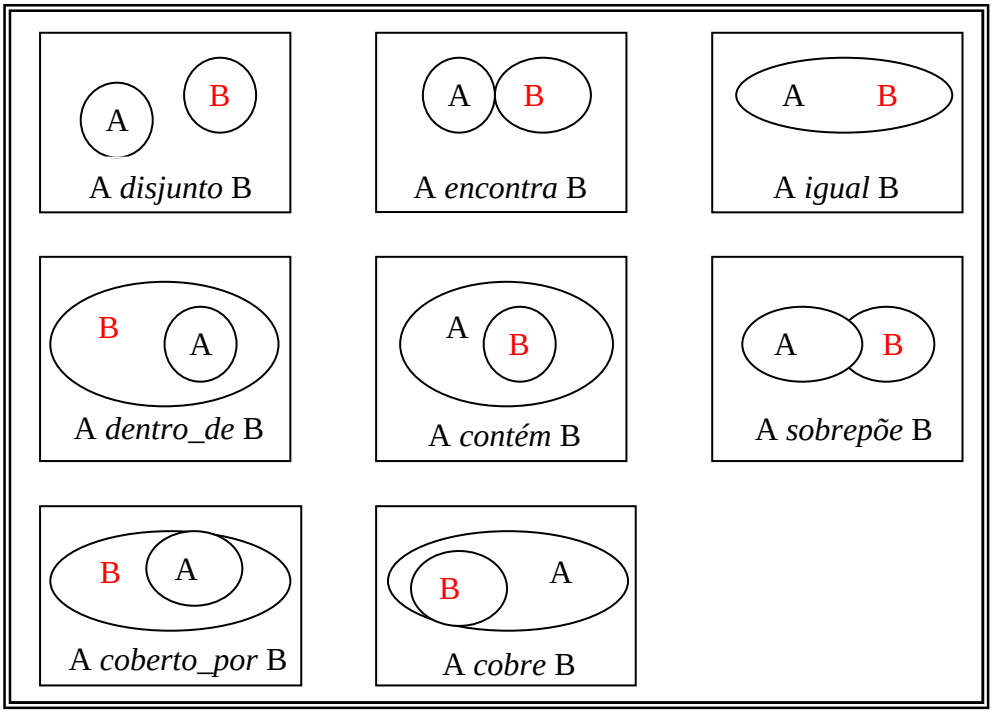


Figura 2.2- Relacionamentos entre os conjuntos A e B, utilizando o método *4-intersection*

O método *4-intersection* indica se existe ou não relacionamento topológico entre dois objetos quaisquer (A e B), pela análise dos quatros casos a seguir:

existência de interseção entre os limites A e B; entre o limite de A e o interior de B; entre o limite de B e interior de A; e entre o interior de ambos os conjuntos [Bogorny & Iochpe, 2001]. Dependendo das combinações destes quatro casos, o método *4-intersection* identifica o tipo de relacionamento existente entre os objetos A e B. A seguir, detalhamos cada relacionamento apresentado na figura 2.2:

- Disjunto: não existe nenhuma forma de interseção entre os conjuntos A e B;
- Encontra: os limites de A e B se intercepta;
- Igual: existe igualdade entre os conjuntos A e B;
- Dentro_de: existe interseção entre o conjunto A e o interior do conjunto B.
- Contém: existe interseção entre o conjunto B e o interior do conjunto A.
- Sobrepõe: existe interseção entre os conjuntos de A e B.
- Coberto_por: existe interseção entre o limite de A e o interior de B.
- Cobre: existe interseção entre o limite de B e o interior de A.

2.3.2 Dados Temporais

A informação temporal é aplicada principalmente em sistemas nos quais é importante ter acesso não apenas ao estado atual das informações, mas também às transformações sofridas por estas [Bezerra, 2000]. Para essas aplicações, a manutenção de dados históricos é um requisito fundamental.

O tempo pode ser representado através de duas formas: por um instante, que determina um único momento; e por um período, que pode ser representado por um intervalo de tempo entre duas datas diferentes [Allen, 1983].

Em um sistema de informação, a dimensão temporal modela o tempo em que os fatos ocorreram no mundo real. O tempo pode ser classificado como tempo válido ou tempo de transação [Jensen & Dyreson, 1998]. O tempo válido de um fato é o tempo no qual o fato foi verdade (válido) na realidade modelada. Um fato pode estar associado a vários instantes ou intervalos de tempo. O tempo de transação de um fato é o tempo no qual este fato tornou-se corrente no banco de dados. O uso dos tempos válido e de transação se torna importante nas aplicações em que ocorrem diferenças

significativas entre o tempo em que os fatos realmente ocorreram e o tempo de registro destas ocorrências.

Um sistema temporal pode utilizar categorias para classificar a temporalidade das informações sobre determinados dados [Baptista, 2000]. As principais categorias utilizadas são: indeterminado, não existe uma exatidão sobre a ocorrência do evento, por exemplo, 'verão de 1998' (ocorre em diferentes épocas do ano em diversos países); absoluto, possui alta precisão da ocorrência do evento, por exemplo, '19 de março de 2001'; e relativo, o tempo está associado a uma data exata, seguida por atributos de quantidade e direção, por exemplo, '30 dias depois da final da copa do mundo de 2002'.

2.3.2.1 Relacionamentos Temporais

Assim como o espaço, o tempo também possui topologia, e relacionamentos topológicos podem ser definidos para o aspecto tempo. Os relacionamentos temporais topológicos possuem alto nível de abstração e podem ajudar os usuários no entendimento e uso da dimensão temporal. Por exemplo, indicar numa relação temporal que a ocorrência de um objeto deve preceder a de outro pode ser importante para o usuário. Estes relacionamentos são semelhantes aos relacionamentos espaciais apresentados na seção 2.3.1.1.

Um dos primeiros trabalhos, voltados para a formalização do tempo através de relacionamentos, foi o de Allen [Allen, 1983]. Allen definiu um cálculo de intervalos baseados em 13 relacionamentos que podem existir entre dois intervalos. O comportamento destes 13 relacionamentos segue três regras básicas:

1- os relacionamentos são mutuamente exclusivos;
2- dois intervalos quaisquer são relacionados entre si seguindo um dos 13 relacionamentos;
3- alguns relacionamentos podem possuir comportamento transitivo, ou seja, se um intervalo A vem antes do intervalo B, e B acontece antes do intervalo C, então o intervalo A vem antes do intervalo B.

Através dos relacionamentos temporais podemos efetuar análises sobre determinadas informações. A dimensão temporal possibilita aos usuários a realização de consultas sobre eventos e acontecimentos históricos. Nos sistemas de informação

espaço-temporal, consultas envolvendo comparações de objetos espaciais em diferentes épocas e possíveis tendências podem ser previstas. Por exemplo, em um SIG, a dimensão temporal permite responder a algumas questões [Langran, 1992], tais como:

- quando as mudanças ocorreram? Por exemplo: em janeiro de 1978 o distrito passou a ser município;
- que tipos de mudanças ocorreram? Por exemplo: na década de 70, muitos municípios foram criados, alterando a geopolítica do estado;
- qual a média das mudanças? Por exemplo: na década de 70, foram emancipados em média 10 municípios por estado;
- qual a periodicidade das mudanças? Por exemplo: na década de 70, a cada ano era emancipado um município por estado.

2.3.2.2 Operadores Temporais

Os relacionamentos topológicos temporais são definidos através de 7 operadores. Na figura 2.3, apresentamos esses operadores que geram os 13 relacionamentos, segundo o cálculo de intervalos de Allen.

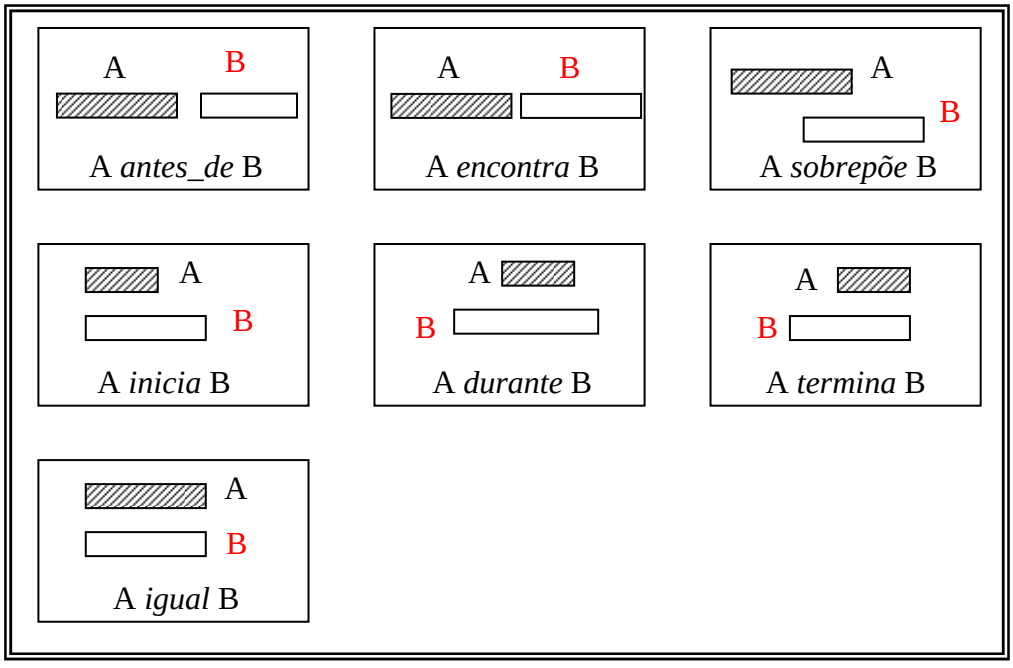


Figura 2.3- Relacionamentos temporais

A determinação dos 13 relacionamentos é feita devido à dualidade de cada relacionamento. Por exemplo, no relacionamento “Antes-de”, o intervalo A vem antes do intervalo B, indicando um relacionamento, mas o intervalo B pode ocorrer antes do A, gerando o segundo relacionamento, ainda dentro do relacionamento “Antes_de”. Esta dualidade também é representada nos outros relacionamentos, exceto no relacionamento “Igual”, originando, assim, 13 relacionamentos diferentes.

A definição de cada operador temporal da figura 2.3 é apresentada abaixo [Baptista, 2000]:

- Antes_de: o intervalo A está antes do B e eles não se sobrepõem;
- Encontra: o intervalo A está antes do B, sendo que o intervalo B inicia, quando intervalo A termina.
- Sobrepõe: o intervalo A começa antes de B e eles se sobrepõem;
- Inicia: os dois intervalos (A e B) iniciam juntos;
- Durante: o intervalo A começa e termina durante a execução do intervalo B (A contido em B);
- Termina: os dois intervalos (A e B) terminam juntos;
- Igual: os intervalos A e B iniciam e terminam juntos.

Estes operadores temporais são utilizados em diversas aplicações. Nos sistemas que envolvem informações geográficas é fundamental a existência da temporalidade dos dados, a fim de manter um histórico das mudanças que ocorreram. Em aplicações que envolvem dados multimídia com utilização de animações, vídeo e áudio, a dimensão temporal também possui grande importância.

2.3.3 O Padrão OpenGis

Muitos dados espaciais estão disponíveis em diversos repositórios e na Internet, mas estes dados geralmente são complexos, heterogêneos e não são compatíveis com outras aplicações. Diante destes problemas, em março de 1998, foi criado o padrão OpenGis (*Open Geodata Interoperability Specification*) desenvolvido pelo Consórcio OpenGis[10] (OGC).

[10] http://www.opengis.org

O padrão OpenGis foi desenvolvido pelo consórcio de empresas, universidades, órgãos governamentais e outras instituições empenhadas no desenvolvimento de especificações caracterizadas pela uniformidade de conceitos e heterogeneidade de implementações. A criação deste padrão contou principalmente com o incentivo de empresas que desenvolvem softwares em SIGs.

A principal finalidade do padrão OpenGis é o tratamento da interoperabilidade, com o objetivo de permitir o acesso transparente e seletivo de informações geográficas [OGC, 2001]. Neste sentido, o padrão procura oferecer acesso aos dados espaciais de forma independente da estrutura dos dados e do formato dos arquivos que são usados.

As especificações de implementações do OpenGIS fornecem instruções para os desenvolvedores de software construírem sistemas que sejam compatíveis com outros sistemas desenvolvidos a partir destas especificações. Algumas versões iniciais das especificações abstratas estão sendo disponibilizadas em etapas, incluindo especificações para várias plataformas de computação distribuída, como CORBA (*Common Object Request Broker Architecture*) e DCE (*Distributed Computing Environment*). Estas especificações procuram minimizar os problemas de compartilhamento de informações e da passagem dos ambientes atuais de processamento monolítico e bancos de dados proprietários para um ambiente de computação distribuída e baseada em componentes.

No padrão OpenGis, o mundo é modelado através de entidades geográficas, denominadas feições. Uma feição geográfica está associada a uma determinada localização na superfície da Terra, contendo uma lista de propriedades (nome e tipo) e geometrias (pontos, linhas e polígonos).

Para realizar o tratamento da interoperabilidade, foram criadas especificações como GML[11] (*Geographic Markup Language*). GML é uma codificação da linguagem XML[12] (*Extensible Markup Language*), com o propósito de transportar e armazenar informações, incluindo características geográficas com propriedades espaciais e não espaciais [Cox *et al*, 2001]. Como é XML, GML representa informações geográficas na forma textual, tornando fácil a realização de revisões e mudanças.

[11] http://www.opengis.net/gml/01-029/GML2.html
[12] http://www.w3.org/TR/REC-xml

As especificações do OpenGis fornecem um modelo comum para o tratamento da dimensão espacial, possibilitando a integração de informações provenientes de diversas fontes. Desse modo, estas especificações contribuem para resolver os problemas de integração nos sistemas de informação com recuperação de dados geográficos.

2.4 Thesaurus

O processo de indexação é fundamental para a recuperação das informações. Porém, um dos problemas mais comuns encontrados nos SRIs é que a terminologia empregada pelos usuários na formulação das consultas não é mesma terminologia usada quando os termos foram indexados no banco de dados [Alani, 2001].

Uma forma de indexação bastante conhecida e utilizada nos SRIs é o *thesaurus*. Um *thesaurus* é formado por um conjunto de termos dentro de um determinado vocabulário, com uma série de relacionamentos semânticos entre estes termos [Tudhope *et al*, 2001].

A palavra *thesaurus* (derivada do grego e que significa “depósito de riquezas”) ficou conhecida no começo do século XIX, quando o Dr. Peter Mark Roger publicou um dicionário com o nome “*Thesaurus of English Words and Phrases*”. Este dicionário chamou a atenção pelo fato de que as palavras não eram apresentadas pela ordem alfabética, como nos dicionários tradicionais, mas de acordo com as idéias que elas representavam.

Os *thesauri* começaram a ser aplicados por pesquisadores e bibliotecários, como uma ferramenta para indexar manualmente livros e documentos. Com o passar dos anos, novas técnicas, como a utilização de palavras-chave, indexação de temas e subtemas, além das técnicas já existentes, como pesquisa através do nome dos autores e título dos documentos passaram a ser utilizadas, agilizando ainda mais o processo de recuperação de informação.

Um *thesaurus* pode ser construído de forma manual, semi-automática ou puramente automática. A construção manual pode apresentar algumas desvantagens, tais como: necessidade de vários especialistas para a classificação das palavras; alto consumo de tempo e subjetividade, pois os especialistas podem possuir opiniões diferentes sobre determinado termo.

A construção automática de um *thesaurus* permite um tratamento mais rápido das relações semânticas, porém, com um grau de confiança menor, devido aos diferentes sentidos e usos que um termo pode assumir. A técnica de construção de *thesauri* mais confiável é a semi-automática, que mescla técnicas computacionais com o trabalho humano, usado para fazer uma revisão final nas relações semânticas entre os termos.

Os termos dos *thesauri* podem ser apresentados de duas formas: na forma hierárquica ou alfabética. Na forma hierárquica, a disposição dos termos é feita segundo a sua posição na hierarquia. Por exemplo, um termo que representa generalidade (BT - *Broader Term*) será seguido dos termos que expressam especificidade (NT - *Narrower Term*). Na forma alfabética, a disposição dos termos segue uma ordem alfabética, incluindo descrições dos termos relacionados.

Um *thesaurus* pode ser utilizado para atender, basicamente, a quatro finalidades:

- Tradução: prover um meio para traduzir a linguagem natural de autores, indexadores e usuários dentro de um vocabulário controlado, usado para indexação e recuperação;
- Consistência: manter consistência no processo de indexação de termos;
- Indicação de relacionamentos: indicar relacionamentos entre os termos;
- Recuperação: servir como uma ferramenta na recuperação de documentos.

Os sistemas de recuperação espaço-temporal utilizam *thesauri* geográficos para auxiliar os usuários nas consultas sobre lugares, provendo as variações de nomes com o decorrer do tempo, bem como nomes alternativos para uma mesma área geográfica [Harpring, 1997], além de outras formas de consultas, como lugares geográficos mais próximos, dentro, ou a uma certa distância de um determinado lugar ou característica geográfica. Um *thesaurus* pode ainda ser enriquecido através de relacionamentos de adjacência e direcionais [Jones, 1997 e Taylor *et al*, 1994].

A utilização do *thesaurus* geográfico para o tratamento de relações entre dados espaciais apresenta algumas vantagens, tais como [Pazinatto *et al*, 2002b]:

- permite a expansão de termos numa consulta, aumentando a cobertura;
- pode ser rapidamente adaptado para codificar relações espaciais, através de *Gazetteers* ou bibliotecas digitais;

- possibilita o fácil desenvolvimento de sistemas de busca, através da estrutura de dados do *thesaurus*.

2.4.1 Tipos de Relacionamentos

Com a utilização de um *thesaurus* geográfico podemos construir diversos relacionamentos e enriquecer o conteúdo que é retornado nas consultas. Os *thesauri* geralmente implementam três tipos de relacionamentos: os relacionamentos de equivalência, hierárquicos e associativos [Aitchison & Gilchrist, 1987]. Estes relacionamentos definem a estrutura de um *thesaurus* e são apresentados nas seções 2.4.1.1, 2.4.1.2 e 2.4.1.3, a seguir.

Através dos relacionamentos de equivalência, hierárquicos e associativos, métodos de expansão ou restrição de consultas podem ser implementados, provendo grande ganho de cobertura na execução de pesquisas. Novas técnicas envolvendo o melhoramento destes relacionamentos estão sendo implementadas. Em [Tudhope *et al*, 2001], discussões e estudos de casos exploram o enriquecimento dos relacionamentos para aumentar a capacidade de recuperação de informações em *thesauri*.

2.4.1.1 Relacionamentos de Equivalência

Os relacionamentos de equivalência definem todos os termos que são sinônimos dentro do *thesaurus*. Dentre estes termos, podemos definir os termos preferidos (PT - *Preferred Term*), que indicam dentre vários termos sinônimos qual é o mais utilizado ou mais atual em determinado contexto. Outros termos que representam o mesmo significado são definidos como termos não preferidos (*Non-Preferred Term*). Existem dois tipos de relacionamentos de equivalência: UF (*Use For*) e USE. O relacionamento UF faz a relação de um termo preferido com termos não preferidos (por exemplo, Estado da Paraíba UF Pb), enquanto o relacionamento USE define a relação de um termo não preferido com um termo preferido (por exemplo, Pb USE Estado da Paraíba).

Relacionamentos de equivalência podem ser usados no tratamento de informações geográficas. Por exemplo, na listagem de todos os nomes que se referem a uma mesma cidade. A cidade de Lisboa, em Portugal, possui vários nomes equivalentes, mas em idiomas diferentes, tais como: Lisboa**,** Lisbon, Lisbonne, Lissabon, Lisabon, Olissibona, Ulixbone, Luzbona, Lixbuna, Olisipo.

2.4.1.2 Relacionamentos Hierárquicos

Os relacionamentos hierárquicos classificam os termos do *thesaurus* colocando-os em uma hierarquia. Para isso, o *thesaurus* define a notação de generalidade (BT- *Broader Term*) e especificidade (NT- *Narrower Term*). Todos os termos que estão na hierarquia, em níveis abaixo de determinado termo são considerados NT deste termo. Por exemplo: Paraíba, Ceará e Pernambuco são NT do termo Brasil, enquanto que Brasil é BT de Paraíba, Ceará e Pernambuco.

Relacionamentos hierárquicos são muito utilizados em sistemas de localização geográfica e podem ser especializados em diferentes relacionamentos:

- relacionamentos genéricos: utilizam a notação BTG (*Broader Term Generic*) e NTG (*Narrower Term Generic*), é usado na mesma classe com termos de graus diferentes. Por exemplo: Sabiá (NTG) => Pássaros (BTG);
- relacionamentos todo-parte: utilizam a notação BTP (*Broader Term Partitive*) e NTP (*Narrower Term Partitive*), em situações que um conceito é inerentemente incluído em outro, de forma que os termos podem ser organizados como uma hierarquia, seguindo a relação todo-parte (BTP) e é-parte (NTP). Por exemplo: Campina Grande (NTP) => Paraíba (BTP);
- relacionamentos de instância: definem a ligação entre um simples termo, expresso por um substantivo comum, e uma instância individual deste termo que, freqüentemente, é um nome próprio. Por exemplo: Rio Amazonas (NT) => Rios (BT).

2.4.1.3 Relacionamentos Associativos

Os relacionamentos associativos definem uma relação entre os termos, ou seja, termos que são semanticamente ou conceitualmente relacionados. Estes relacionamentos utilizam a notação RT (*Related Term*) para expressar o relacionamento entre os termos.

A finalidade dos relacionamentos associativos é fazer a ligação semântica entre os termos preferidos, os quais não são hierarquicamente relacionados [Sintichakis & Constantopoulos, 1997]. Por isso podemos dizer que um relacionamento associativo é qualquer relacionamento semântico não hierárquico e sem equivalência entre um par de termos preferidos. As relações associativas de um *thesaurus* possuem simetria, ou seja, são sempre capazes de definir o inverso de um relacionamento associativo. Por exemplo: Planetas (RT) Astronomia e Astronomia (RT) Planetas.

2.4.2 WordNet

Vários *thesauri* com diferentes domínios e outros trabalhos com estruturas similares foram desenvolvidos nos últimos anos. Entre eles, podemos citar: Getty Thesaurus of Geographic Names[13]; The cook's Thesaurus[14]; e WordNet[15]. Nesta seção, abordaremos o projeto WordNet.

WordNet é um banco de dados léxico de palavras e termos em inglês, que possibilita aos usuários encontrar sinônimos, antônimos, advérbios e verbos em relação ao item submetido na pesquisa. O WordNet foi desenvolvido pelo *Cognitive Science Laboratory*, na universidade de *Princeton*, sob a direção do professor George Miller. Atualmente, está disponível na Internet a versão 1.7.1 do sistema.

No WordNet, não são apenas listadas palavras e seus significados, mas também como estas palavras ou termos se relacionam [Fellbaum, 1993]. Por exemplo, ao ser digitada a palavra "animal", não são apresentados aos usuários apenas os significados, sinônimos e antônimos, mas também relações semânticas de hiperonímia

[13] http://www.getty.edu/research/tools/vocabulary/tgn
[14] http://www.foodsubs.com
[15] http://www.cogsci.princeton.edu/~wn

(animal é uma hiperonímia de gato), hiponímia (gato é uma hiponímia de animal), meronímia (folha é uma meronímia de árvore, ou seja, é uma “parte” de um “todo”) e holonímia (o “todo” de uma relação é o holonímia).

Capítulo 3

Sistemas de Referência Espaço-Temporal Indireta

No capítulo 2, abordamos o funcionamento das ferramentas capazes de recuperar informações espaciais e temporais. Neste capítulo, apresentamos alguns conceitos e as características dos projetos que serviram de embasamento e que são relacionados com a nossa pesquisa. A seção 3.1 aborda o conceito e as principais características dos sistemas de referência indireta - *Gazetteers*. Na seção 3.2, apresentamos alguns trabalhos relacionados na área de referência indireta a dados espaciais e temporais. A arquitetura, o funcionamento e as principais limitações desses trabalhos são apresentados nesta seção.

3.1 Gazetteers

No início da década de 90 começaram a surgir as primeiras bibliotecas digitais com informações que estavam dispostas em livros, revistas e enciclopédias. Várias bibliotecas digitais foram criadas e dentre elas começaram a surgir às bibliotecas digitais georeferenciadas, com dados espaciais, conhecidas como *Geolibraries* [Goodchild, 1999]. Com a finalidade de incrementar estas bibliotecas surgiram os *Gazetteers*, ferramentas capazes de indexar informações espaciais e não espaciais.

Vários tipos de informação possuem referência a locais específicos da superfície da terra. Relatórios sobre o meio ambiente, fotografias de um determinado lugar, imagens da terra obtidas por satélites, estatísticas econômicas e demográficas, guias de cidades e planos municipais, são exemplos de informações que podem ser georeferenciadas. A forma que tem prevalecido para se fazer a geo-referência é através do nome (associado as coordenadas geográficas).

Os usuários que pretendem localizar informações relevantes sobre um determinado assunto necessitam, muitas vezes, de o fazer através da referência a uma localização específica, usualmente descrita por um nome geográfico. Por exemplo,

“Encontrar todas as informações relacionadas com a floresta amazônica”. Os usuários têm interesse em encontrar todos os itens relevantes que contenham as palavras floresta amazônica (relatórios e artigos) e também informações como fotografias aéreas, imagens de satélites, vídeos e áudios, que se referem à área geográfica da floresta amazônica. Essa forma de referência espacial, feita através do nome do lugar geográfico, é conhecida como referência indireta e é suportada pelo uso de *Gazetteers*.

Um *Gazetteer* é considerado como um dicionário de nomes geográficos, provendo nome, localização e as características dos dados geográficos [Hill, 1998]. O termo ‘nomes geográficos’ é utilizado para determinar um “lugar” ou “tipo” geográfico, tal como, cidade (tipo) de Campina Grande (lugar). A localização geográfica é normalmente indicada pela coordenadas de latitude e longitude relativas ao lugar geográfico.

Em um *Gazetteer*, a referência indireta é feita através da entrada do nome do lugar geográfico para recuperar informações, tais como, as coordenadas, imagens espaciais, tipo ou categoria e outras informações sobre o lugar geográfico de interesse.

Os *Gazetteers* geralmente são usados como índices em atlas, como enciclopédia de lugares geográficos. Por exemplo, o Columbia *Gazetteer* of the World [Cohen, 1998], como *thesauri*, tal como, o Getty Thesaurus of Geographic Names [Getty, 2002], ou como sistemas do governo que armazenam dados do território americano como GeoNet Names Server [GEOnet, 2002] e Geographic Names Information System (GNIS) [GNIS, 1998].

Em um atlas geográfico, um *Gazetteer* provê uma lista alfabética dos nomes que aparecem no atlas, com a localização espacial de cada lugar geográfico apresentada em um mapa. Em enciclopédias e *thesauri*, os *Gazetteers* definem uma estrutura hierárquica entre os lugares geográficos e fornecem informações sobre a localização com coordenadas de latitude e longitude [Hill, 2002]. Nos sistemas GeoNet Names Server e GNIS é dado ênfase na diferenciação dos nomes que um mesmo lugar geográfico pode ter, como por exemplo, Brasil e República Federativa do Brasil.

Nos *Gazetteers*, um lugar geográfico é caracterizado por um conjunto de atributos. Alguns destes atributos são indispensáveis na definição de um lugar geográfico [Hill, 2000]:

- nome: o nome determina um lugar geográfico. Porém, o atributo nome não é único, para um mesmo lugar geográfico podem existir nomes diferentes, por exemplo: *Pequim* e *Beijing* na China. Da mesma forma, vários lugares geográficos podem possuir o mesmo nome, por exemplo, nos Estados Unidos existem 18 cidades chamadas *Springfield*.
- localização: a localização determina as coordenadas (latitude e longitude) que podem ser expressas através de um ponto, um retângulo ou um polígono. Por exemplo, a localização do Brasil pode ser feita com o esboço de um retângulo, sobre um mapa geográfico, utilizando dois pares das coordenadas de latitude e longitude, que representam os dois pontos extremos do retângulo, tais como: -74°.20'O, -33°.66'S e -34°.78'L, 05°.20'N;
- tipo: o tipo determina a característica de um lugar geográfico. Por exemplo, o tipo 'país' contém lugares geográficos como Brasil, Itália e Argentina.

O atributo "tipo" pode incluir diferentes categorias como: áreas políticas (cidades, estados, países, etc); áreas turísticas (parques, museus, monumentos, etc) características naturais (montanhas, vulcões, florestas, oceanos, ilhas, etc); estruturas feitas pelo homem (pontes, canais, torres, etc); e ainda áreas imprecisas (sudeste da Paraíba, sul do Brasil, etc) [Hill & Qi, 1999a]. Uma categoria pode ser definida de forma mais específica, como lagos, ou mais genérica, como características hidrográficas, permitindo assim uma maior flexibilidade no processo de consultas ao *Gazetteer*.

Atualmente, existem *Gazetteers* com diferentes domínios [Hill *et al*, 1999b]. Por exemplo, New York Times Atlas[16] possui uma seção que pode ser usada para observar um nome geográfico e encontrar páginas (URLs) com informações do lugar pesquisado. Alguns *Gazetteers* possuem informações sobre lugares e diversos tipos de características, como por exemplo, a história do local, dados da população, dados físicos como altitude e mudanças geopolíticas que ocorreram em um período de tempo.

[16] http://geogdata.csun.edu/NYpage1.html

Existem *Gazetteers* onde as informações sobre os lugares geográficos estão disponíveis em termos de hierarquias através de *thesauri*, designados para recuperação de informações. Como exemplo de tais sistemas, podemos citar os sistemas do governo americano: U.S. Library of Congress[17] e o GeoRef Thesaurus[18] desenvolvido pela American Geological Institute. O governo americano, também possui agências que mantêm *Gazetteers* que contêm nomes oficiais de lugares geográficos (cidades, estados, países). Como exemplo, podemos citar os *Gazetteers*: U.S. Geological Survey's Geographic Names Information System (GNIS); National Imagery and Mapping Agency's Geographic Names Processing System (GNPS)[19]; e U.S. Board of Geographic Names (BGN)[20].

Vários tipos de consultas podem ser executadas em um *Gazetteer*. As principais formas de consultas são: a textual, através do nome do lugar geográfico; e a forma espacial, através da seleção de uma região geográfica sobre um mapa. A seguir, listamos alguns tipos de consultas que podem ser executadas através de um *Gazetteer*:

- o usuário entra com o "nome" de um lugar geográfico e encontra a localização espacial disposta em uma imagem, com coordenadas geográficas e outras informações;
- o usuário entra com um "tipo" ou categoria, tal como, áreas políticas, e encontra um conjunto de lugares geográficos deste "tipo", com informações específicas de cada lugar;
- o usuário esboça sobre o mapa uma região geográfica e encontra nomes e características daquela região. Esse tipo de consulta pode também ser filtrada por categorias ou "tipos", por exemplo: 'encontrar todas as características hidrográficas (lagos, pântanos e rios) de uma determinada região'. A área de pesquisa é selecionada sobre o mapa através de técnicas de seleção espacial, tais como: MBR, Centróide e Fecho Convexo (menor polígono possível que cobre a área de um lugar geográfico).

Os *Gazetteers* representam uma forma inteligente de recuperação de informação. A grande capacidade que os *Gazetteers* possuem, em relacionar representações gráficas de um objeto geográfico com nomes de lugares e localização

[17] http://www.loc.gov
[18] http://www.agiweb.org/agi/georef.html
[19] http://www.nima.mil
[20] http://mapping.usgs.gov/www/gnis/bgn.html

geográfica, tem incentivado diversas pesquisas em torno da criação de técnicas de seleção espacial mais eficientes, com suporte a consultas espaciais mais refinadas [Schlieder *et al*, 2001].

3.2 Trabalhos Relacionados

Nesta seção, apresentamos os principais trabalhos relacionados com a nossa pesquisa. Vários projetos incluindo *Gazetteers*, bibliotecas digitais e ferramentas de busca que realizam pesquisas sobre dados espaciais e temporais são apresentados. Em cada projeto detalhamos o funcionamento, a arquitetura, modelo da interface e as suas principais limitações. Discutimos, nesta seção, os seguintes projetos: Alexandria Digital Library[21]; Perseus[22]; Geographic Names Information System[23]; Getty Thesaurus of Geographic Names[24]; e Genuki[25].

3.2.1 Alexandria Digital Library

O projeto Alexandria Digital Library (ADL) começou a ser desenvolvido em 1994, na universidade da Califórnia, Santa Bárbara, Estados Unidos com o propósito inicial de construção de uma biblioteca digital. Esse projeto é financiado pela *National Science Foundation* (NSF), agência de fomento a pesquisa norte-americana [Smith & Frew, 1995].

Em 1998 o projeto passou para a segunda fase, foi renomeado para ADEPT (*Alexandria Digital Earth Prototype*) e passou a desenvolver um *Gazetteer*. ADL foi um dos primeiros projetos de pesquisa sobre o desenvolvimento de bibliotecas georeferenciadas (*Geolibraries*), produzindo vários conceitos nesta área.

O *Gazetteer* do projeto ADL tem como objetivo principal prover um amplo acesso a informações espacialmente indexadas [Gonçalves & Medeiros, 1997]. Para indexar estas informações, o *Gazetteer* utiliza dois *thesauri* do governo americano: o Getty Thesaurus of Geographic Names, e o GeoRef Thesaurus [Goodman, 2000].

[21] http://www.alexandria.ucsb.edu
[22] http://www.perseus.tufts.edu
[23] http://geonames.usgs.gov/gnishome.html
[24] http://www.getty.edu/research/tools/vocabulary/tgn
[25] http://www.genuki.org.uk

Esses *thesauri* fazem o mapeamento dos nomes dos lugares geográficos, de forma que possam ser relacionados e categorizados dentro do *Gazetteer*.

Diversos estudos e pesquisas vêm sendo desenvolvidos e milhões de dólares têm sido investidos ao longo dos anos neste projeto. Embora o *Gazetteer* do projeto ADL ainda apresente algumas limitações, até os dias atuais representa um ponto de referência na Internet para a pesquisa de lugares geográficos. Com uma interface simples e funcional facilita, aos usuários, a submissão de consultas.

Arquitetura

O projeto ADL utiliza uma arquitetura em três camadas: um *browser Web*, responsável pela formulação das consultas e apresentação dos resultados; um servidor de *middleware*, responsável pelo acesso a base de dados e recuperação das informações; e um servidor de banco de dados, responsável pela persistência dos dados. Esta arquitetura possui quatro componentes: coleções, catálogos, interface gráfica e um componente de aquisição [Frew *et al*, 1998].

As coleções contêm o esquema de armazenamento das informações, com coleções de itens indexados espacialmente. Esse componente possui procedimentos para interpretar e retornar os dados da consulta. O projeto ADL emprega um sistema de armazenamento hierárquico distribuído para armazenar suas próprias coleções.

O componente de catálogo provê agilidade na execução das consultas submetidas pelos usuários, com o uso de várias estruturas de índices e metadados de informações.

O componente de interface gráfica é formado por um conjunto de páginas HTML (*HyperText Markup Language*) que implementam três funções: visualizador de mapas (definição de regiões no mapa), consulta via *Gazetteer* (definição de um lugar geográfico e seu tipo de feição) e consultas ao catálogo de metadados (forma de consulta genérica traduzida para SQL - *Structured Query Language*).

O objetivo do componente de aquisição é estender o esquema de metadados, incorporando novas informações ao conjunto de dados existentes até o momento, bem como criar novos registros de metadados para os itens selecionados na consulta.

Interface do Usuário

A interface gráfica do projeto ADL é baseada em uma coleção de páginas HTML, com duas formas de consulta: espacial e textual. A interface gráfica contém um mapa mundi, que pode ser usado para submeter uma consulta espacial. Os usuários podem selecionar uma determinada região sobre o mapa, através do esboço de um MBR. O *Gazetteer* permite a visualização espacial dos resultados da consulta, ou seja, os resultados da consulta são apresentados no mapa através de pontos que podem ser selecionados disparando páginas com metadados daquela informação. Na forma de consulta textual, os usuários podem digitar o nome de um lugar geográfico e então especificar um tipo de feição ou categoria, uma operação espacial e outros atributos. Essas consultas podem utilizar dois operadores espaciais: *Overlap* e *Contains*. O resultado da consulta apresenta várias informações como coordenadas geográficas, formato e origem dos dados, relacionamentos hierárquicos e apresentação de um mapa com a localização do lugar geográfico submetido inicialmente na consulta. Um *Gazetteer* associado com dois *thesauri*, que dispõem de uma hierarquia de lugares geográficos permite a recuperação das informações relacionadas a um lugar.

A interface gráfica do *Gazetteer* do projeto ADL pode ser visualizada na figura 3.1. Neste exemplo, realizamos uma consulta sobre o lugar geográfico *New York*, com a seleção da categoria aeroportos. Conforme a figura 3.1, podemos perceber que o sistema retornou quatro aeroportos, e então, selecionamos o aeroporto municipal de *New York*. As informações (nome, tipo de feição, referência espacial, coordenadas, etc) sobre este aeroporto são mostradas no frame da página inferior à esquerda da figura 3.1.

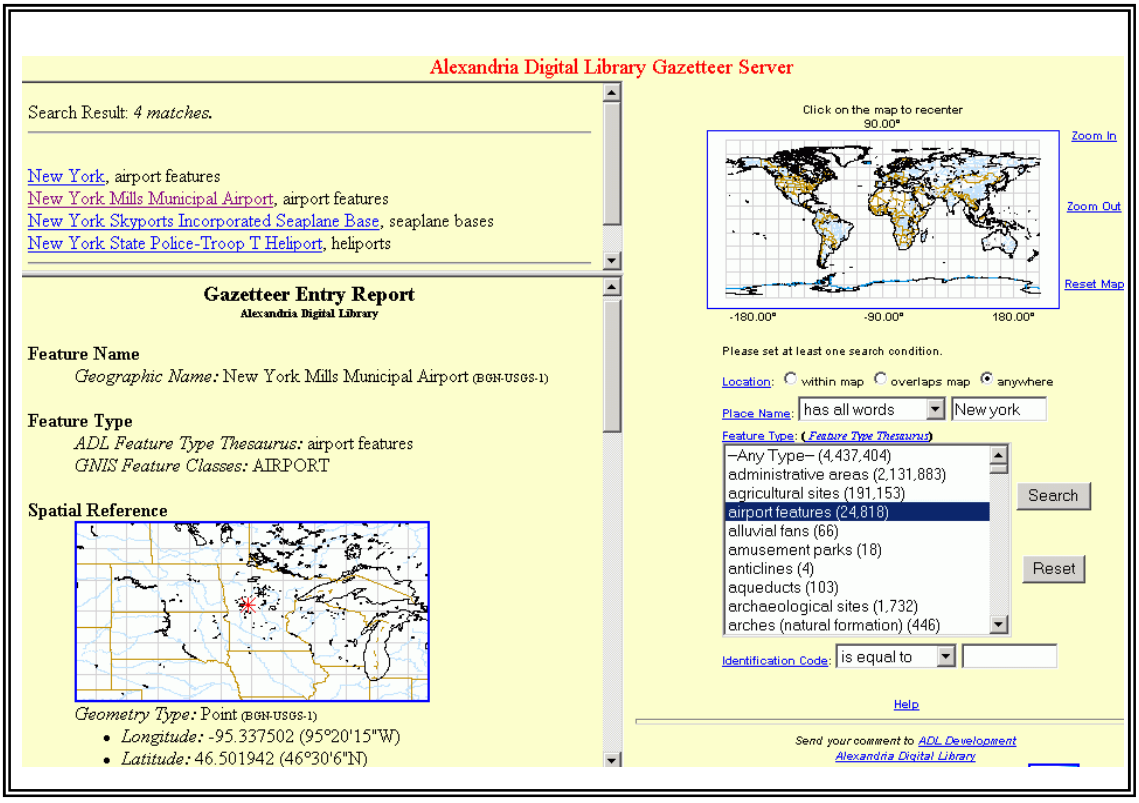


Figura 3.1- Interface do *Gazetteer* do projeto ADL

Limitações

O *Gazetteer* do projeto ADL apresenta algumas limitações, tais como: a falta de informações sobre a temporalidade dos dados que impede a manutenção de um histórico, com a data das informações e de operações que envolvem operadores temporais; possui suporte limitado às operações espaciais, pois utiliza apenas os operadores *Overlap* e *Contains*; limitação nas consultas espaciais realizadas sobre o mapa; apresenta problemas com relação à interoperabilidade, pois o *Gazetteer* do projeto ADL não consegue trocar informações com outros *Gazetteers*.

3.2.2 Perseus

O projeto Perseus é uma biblioteca digital com informações textuais e gráficas dispostas em diversos idiomas sobre diferentes regiões do planeta, sobretudo dados da América do Norte e da Europa [Crane, 1996]. Este projeto iniciou-se em 1985, na universidade de *Tufts*, *Massachusets*, Estados Unidos. Porém, a disponibilização da primeira versão da biblioteca digital ocorreu apenas em julho de 1987.

O projeto Perseus reúne informações sobre textos gregos, peças de artes e materiais arqueológicos da história antiga, disponibilizando-os através da biblioteca digital. Associado a estas informações existe um *Gazetteer*, que auxilia no processo de pesquisa. Esse *Gazetteer* mantém relacionamentos entre os lugares geográficos e a origem da informação. O principal objetivo da biblioteca digital Perseus é servir de estudo e fonte de discussão em escolas e órgãos que possuem interesse sobre a história da Grécia antiga [Crane, 1998].

O Perseus gerencia uma biblioteca digital que contém mais de 65 milhões de termos, boa parte destes termos estão disponíveis em quatro idiomas: inglês, latim, italiano e alemão. O projeto Perseus, também contém mais de 30.000 imagens divididas em 1.000 mapas diferentes para a representação dos lugares geográficos.

Arquitetura

A arquitetura do projeto Perseus é composta por um conjunto de coleções com imagens e textos em SGML (*Standard Generalized Markup Language*) [Maler & Andaloussi, 1995] e XML [Bray *et al*, 2000]. A estrutura desses documentos tem permitido a disponibilidade do conteúdo destes textos em uma variedade de plataformas. Esses documentos estão formatados de acordo com as diretrizes estabelecidas pela *Text Encoding Initiative* (TEI)[26] [Ide & Veronis, 1995]. Algumas ferramentas para gerenciar documentos SGML e XML conforme DTDs (*Document Type Definition*) foram construídas permitindo suporte a documentos com diferentes estruturas [Mahoney *et al*, 2000]. Essas ferramentas extraem a estrutura e a descrição de metadados entregando documentos bem formados ao sistema de exibição de textos. A persistência destes dados foi feita em um SGBD relacional.

Interface do Usuário

A interface gráfica da biblioteca digital do projeto Perseus é formada com páginas HTML que contêm uma ferramenta de busca. Através desta interface, os usuários podem digitar o nome do lugar geográfico e associá-lo com coleções de informações envolvendo publicações de textos com especificação de autores e títulos, materiais arqueológicos, peças de arte, informações sobre guerra civis que ocorreram em determinados países e outras coleções. A interface também provê um mapa mundi

[26] http://www.tei-c.org

com a localização do lugar geográfico relativo ao termo que foi submetido na consulta. Neste mapa os usuários podem executar operações de *zoom in* e *zoom out* e realizar consultas espaciais clicando sobre determinadas regiões no mapa com dados da escala e coordenadas de latitude e longitude. Um *Gazetteer* realiza o processo de busca dos lugares geográficos. A interface de consulta da biblioteca digital Perseus é apresentada na figura 3.2. No Perseus, os usuários realizam as consultas através da entrada de um nome geográfico. No exemplo da figura 3.2, apresentamos a primeira etapa do processo de uma consulta para recuperar informações relacionadas com o lugar geográfico *New York*.

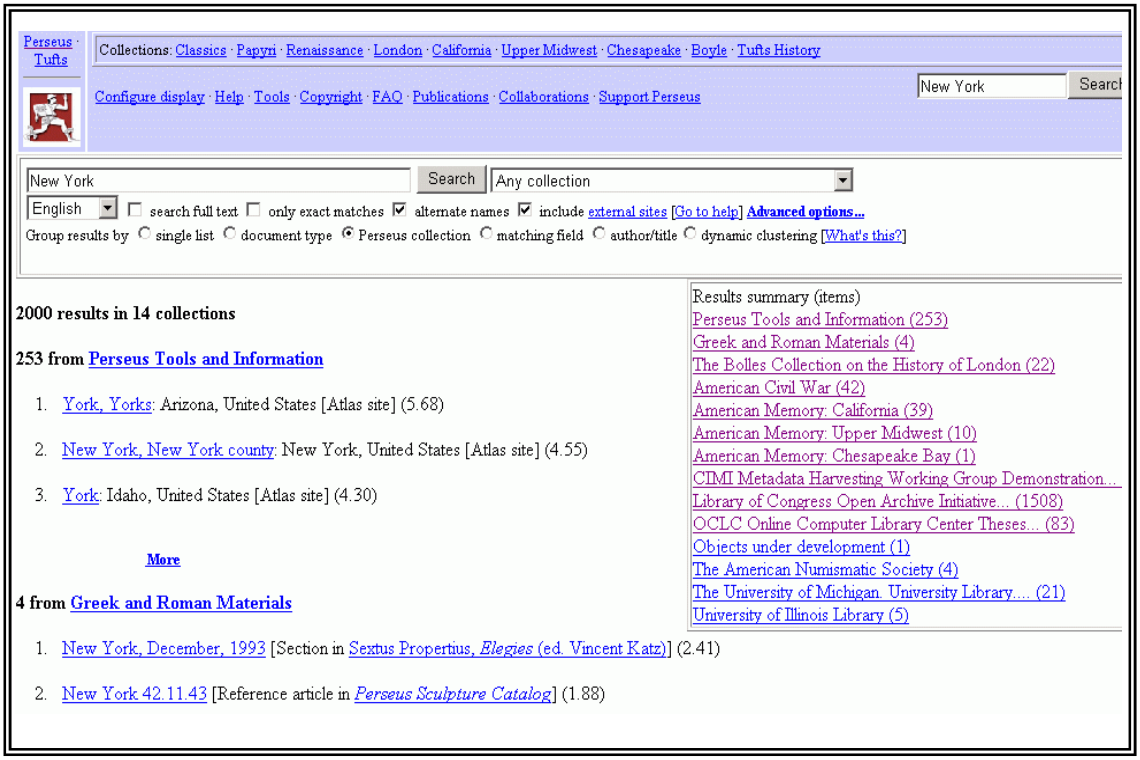


Figura 3.2- Interface de consulta do projeto Perseus

Limitações

O projeto Perseus apresenta algumas limitações, tais como: pequeno suporte sobre a temporalidade das informações; as consultas espaciais são limitadas; inexistência de operações temporais e espaciais; resultados com informações textuais e espaciais são apresentados em diferentes níveis na interface gráfica; várias informações irrelevantes sobre o termo da pesquisa são recuperadas, isso exige que o usuário exerça um processo de filtragem das informações.

3.2.3 Geographic Names Information System

O projeto Geographic Names Information System (GNIS) foi desenvolvido pela U.S. Geological Survey (USGS) em cooperação com U.S. Board on Geographic Names (BGN), ambos órgãos do governo americano de pesquisa geológica e mapeamento geográfico. As pesquisas iniciais sobre o GNIS iniciaram em 1968, quando os dados de nomes geográficos do estado americano de *Massachusetts* foram coletados e armazenados na forma digital.

O projeto GNIS contém informações de todo território americano com áreas administrativas, coordenadas espaciais e apresentação de mapas. O nome oficial de cada lugar geográfico é descrito na base de dados com referência as características dos estados e municípios americanos. O GNIS é o repositório oficial com informações dos nomes dos lugares geográficos dos Estados Unidos.

Os lugares geográficos são recuperados através do *Gazetteer* da National Imagery and Mapping Agency (NIMA) desenvolvido pela BGN, que contém nomes associados com lugares e características geográficas de vários países [Nowakowski, 1995].

Arquitetura

Atualmente, a base de dados do projeto GNIS é composta por três bancos de dados: National Geographic Names Data Base (NGNDB); USGS Topographic Map Names Data Base (TMNDB); e Reference Data Base (RDB).

O NGNDB é maior banco de dados utilizado pelo GNIS. Ele contém listas de nomes identificados com códigos, e com a localização de mais 2 milhões de características geográficas do território americano. O TMNDB e o RDB foram desenvolvidos para incrementar o NGNDB. O TMNDB possui a relação de todos os mapas topográficos. O RDB contém as informações sobre a origem dos dados utilizados no NGNDB.

Interface do Usuário

A interface gráfica do projeto GNIS é composta por páginas HTML. O processo de consulta é realizado sobre lugares geográficos (nome de estados ou municípios americanos) com a ajuda de um *Gazetteer*. Algumas ferramentas para auxiliar o processo de consulta são oferecidas aos usuários tais como: pesquisar por

nomes geográficos relacionados a uma determinada categoria como estados ou cidades; selecionar o nome ou o código específico de um estado (por exemplo, New York, 36 NY); especificar informações sobre um determinado município através do seu nome; e selecionar apenas informações que estejam dentro uma determinada faixa de códigos, pois cada lugar geográfico é identificado através de um código. O resultado da consulta apresenta várias informações, tais como: coordenadas de latitude, longitude e altitude; e categoria (tipo) do lugar geográfico. O lugar pode ser visualizado em mapas em diferentes formatos: no formato *raster* obtido de mapas topográficos; através de fotografias aéreas; com imagens de satélites; ou através de mapas produzidos por centros de pesquisas. Na figura 3.3 submetemos uma consulta sobre o estado americano de *New York*, com a categoria *Lakes* (lagos). A consulta retornará apenas informações (nome, coordenadas geográficas, altitude, município e mapas com a localização espacial) dos lagos que estão relacionados com o estado de *New York*. A interface de consulta do projeto GNIS pode ser visualizada na figura 3.3.

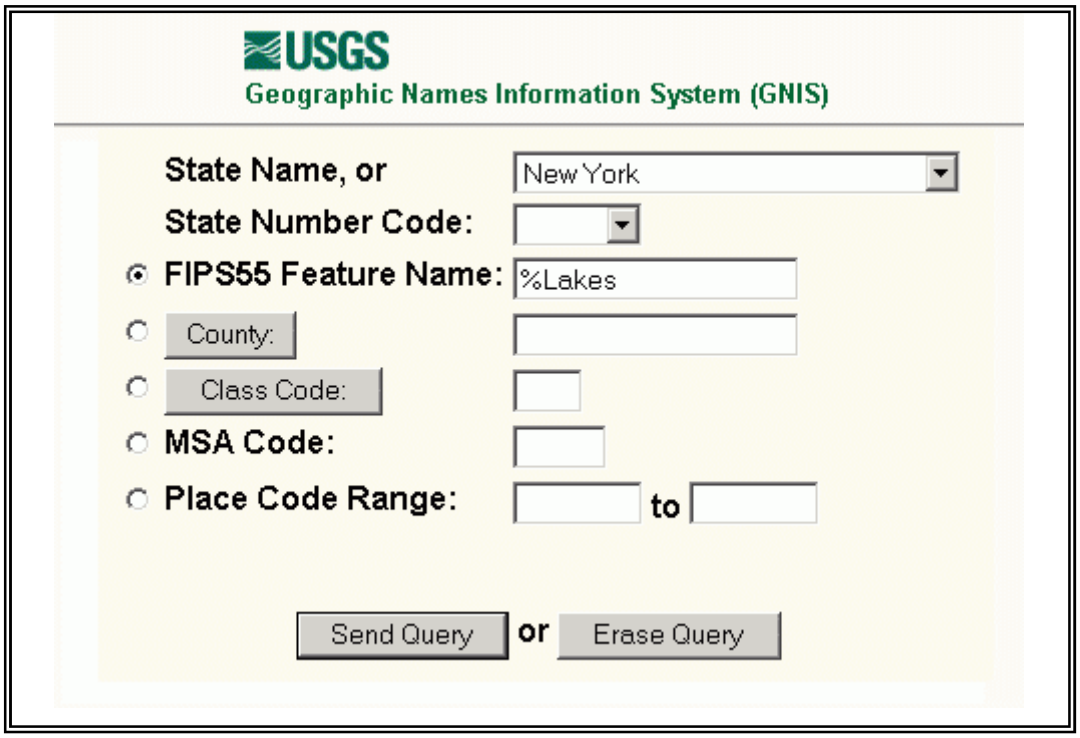


Figura 3.3- Interface do consulta do projeto GNIS

Limitações

O sistema de consulta do projeto GNIS apresenta algumas limitações, tais como: inexistência da temporalidade dos dados; falta de operações espaciais e temporais; não existe suporte a consultas espaciais; na execução de determinadas

consultas, existe a necessidade de uma faixa de códigos (por exemplo: “recuperar apenas os lugares geográficos com código acima de 1.290 e abaixo 2.800”), para delimitar o número de registros que são retornados, pois o sistema não permite a recuperação de mais de 2.000 registros em uma consulta.

3.2.4 Getty Thesaurus of Geographic Names

O Getty Thesaurus of Geographic Names (TGN) foi desenvolvido pelo programa de pesquisa Getty Research Institute (GRI) mantido pelo governo americano, com sede em *Los Angeles*, Estados Unidos. Os trabalhos no centro de pesquisa e desenvolvimento GRI iniciaram em 1997, porém a conclusão e disponibilização na Internet ocorreram apenas no ano de 2000.

O TGN é um *thesaurus* que contém um vocabulário estruturado de nomes geográficos, com informações de todos os continentes do planeta. O TGN contém cerca de 1 milhão nomes sobre lugares geográficos incluindo características físicas e áreas administrativas, tais como: cidades, estados e países.

O principal objetivo do TGN é a criação de um ambiente diferente e inovador para o desenvolvimento da pesquisa, com investigações críticas e discussões sobre as áreas das artes e humanidades [Fink, 1999]. O TGN pode ser usado como uma fonte de pesquisa, com informações que podem ser acessadas em qualquer parte do mundo através da Internet.

O instituto de pesquisa GRI possui outros dois vocabulários: Art & Architecture Thesaurus (AAT)[27]; e o Union List of Artist Names (ULAN)[28]. Estes dois vocabulários podem ser usados em conjunto com o TGN [Getty, 1999]. Com o uso conjunto destes três vocabulários o GRI pode prover acesso a informações como objetos, conceitos, autoridades e lugares geográficos que estão relacionados a estas informações. Os vocabulários AAT, ULAN e TGN são muito utilizados em museus, bibliotecas de artes, projetos bibliográficos relacionados a materiais históricos e por estudantes em busca de informações.

O TGN representa um ponto de referência na Internet na pesquisa de lugares geográficos. No resultado de uma consulta, os lugares geográficos são apresentados

[27] http://www.getty.edu/research/tools/vocabulary/aat
[28] http://www.getty.edu/research/tools/vocabulary/ulan

em forma hierárquica, proporcionando um maior aprendizado aos usuários. As informações contidas no *thesaurus* TGN são utilizadas em outros projetos, como no *Gazetteer* do projeto ADL.

Arquitetura

O banco de dados do GRI possui os vocabulários de termos do AAT, ULAN e do TGN com informações sobre artes, arquitetura e lugares geográficos, respectivamente. Esses vocabulários foram construídos e são mantidos pelo Getty Vocabulary Program, que uma divisão do GRI. Os três vocabulários estão disponíveis na Internet sendo acessados via *browser*.

As informações do TGN estão dispostas no banco de dados, de forma que cada registro é um lugar geográfico representado por um ID (identificador) único [Harpring, 1997]. Juntamente com um ID para cada lugar geográfico, existe um registro indicando a posição deste lugar na hierarquia de lugares geográficos e outras informações como coordenadas, fonte das informações e um tipo de feição que descreve o tipo do lugar geográfico, tais como: município, estado, país, etc. Informações sobre o nome de um lugar geográfico podem incluir nomes históricos e traduções em diferentes idiomas. Entre esses nomes, o *thesaurus* sinaliza um termo como preferido ou *Preferred Term* indicando qual dos termos é o mais atual ou o nome oficial do lugar geográfico. Cada lugar possui três atributos: um nome (por exemplo: Paraíba); um tipo de feição (por exemplo: estado); e coordenadas (por exemplo: latitude: -34.53 e longitude: -07.06).

O TGN trabalha em conjunto com outros projetos. Quando algumas informações sobre determinados lugares geográficos estão em outras fontes, como no *Gazetteer* GNIS da USGS, existem referências para o projeto que detém a informação.

Interface do Usuário

A interface gráfica do TGN é composta por páginas HTML. A ferramenta de pesquisa suporta apenas a entrada de nomes de lugares geográficos. Na primeira fase da consulta, uma série de lugares que estão relacionados com o lugar geográfico submetido na consulta são retornados. Estes lugares são *links* que possuem informações mais específicas de cada lugar. Assim que o usuário clica no *link* de um nome geográfico uma série de informações são apresentadas aos usuários, tais como:

tipo de feição; coordenadas de latitude e longitude; posição do lugar, em uma hierarquia de lugares geográficos; origem da informação; e designação de outros nomes para este lugar geográfico. Na figura 3.4, apresentamos a interface de consulta do TGN, executando uma consulta sobre o lugar geográfico *New York*.

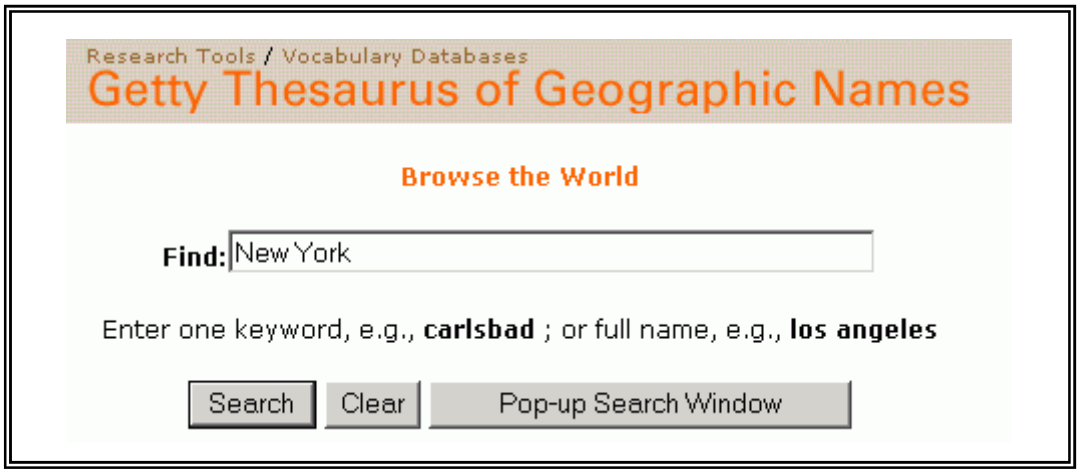


Figura 3.4- Interface de consulta do TGN

Limitações

Embora o TGN possua um amplo vocabulário de nomes geográficos, algumas limitações são apresentadas: existe apenas uma forma de submeter uma consulta, através do nome do lugar geográfico; não possui informações sobre a temporalidade dos dados; as coordenadas espaciais são representadas por apenas dois pontos, fornecendo uma pequena noção da localização geográfica aos usuários; não suporta operações temporais e espaciais; não possui imagens ou mapas informando a localização espacial do lugar; poucas informações sobre um lugar geográfico.

3.2.5 Genuki

O projeto Genuki envolve os países do Reino Unido (Inglaterra, Irlanda do Norte, País de Gales e Escócia), a República da Irlanda e as ilhas de *Channel* e de *Man*. O projeto Genuki é mantido desde 1995, como um trabalho conjunto das universidades de *Manchester*, *Oxford*, *Colchester*, *Newcastle*, *St Andrews* e *Dublin*.

A finalidade do projeto Genuki é servir como uma biblioteca digital com informações sobre a genealogia (estudo da origem das famílias ou descendência dos

indivíduos) pertinente aos países do Reino Unido e da Irlanda [Austen *et al*, 1995]. O projeto Genuki conta o ajuda de vários colaboradores de diversas universidades e principalmente com o apoio da federação de genealogia do Reino Unido - *Federation of Family History Societies*, localizada na Inglaterra.

O projeto Genuki possui um *Gazetteer*, que possibilita encontrar informações que estão relacionadas com determinados lugares geográficos [Tippey, 1998]. No processo de consulta, esse *Gazetteer* acessa informações em diversas bases de dados, inclusive outros *Gazetteer* que incluem informações geográficas da Irlanda e de cada país que compõe o Reino Unido.

Arquitetura

O projeto Genuki acessa informações de diversos bancos de dados. As informações são recuperadas de vários servidores [Hicks, 2001], dentre eles: o *Public Records Office* (PRO), *Society of Genealogists* (SoG)[29], *Federation of Family History Societies* (FFHS)[30] e o *Guild of One Name Studies* (GoONS)[31].

Para organizar o grande número de informações, o projeto Genuki adota uma estrutura hierárquica que consiste em quatro níveis. No nível mais alto estão as informações de todo o Reino Unido e Irlanda. Nos três níveis inferiores estão as informações das regiões (*county*), municípios e cidades.

Interface do Usuário

A interface gráfica do projeto Genuki é formada por uma coleção de páginas HTML, que possuem uma ferramenta de busca. Essa ferramenta permite a pesquisa por nomes geográficos acessando um *Gazetteer*. A partir de um lugar geográfico, os usuários podem escolher visualizar apenas as informações relacionadas a lugares que estão a uma determinada distância do lugar submetido inicialmente na consulta. Um mapa com a localização do lugar geográfico é apresentado aos usuários. Na figura 3.5, apresentamos a interface de consulta do *Gazetteer* Genuki. No exemplo da figura 3.5, uma consulta sobre o lugar geográfico *Canterbury*, na Inglaterra, é executada.

[29] http://www.sog.org.uk
[30] http://www.ffhs.org.uk
[31] http://www.one-name.org

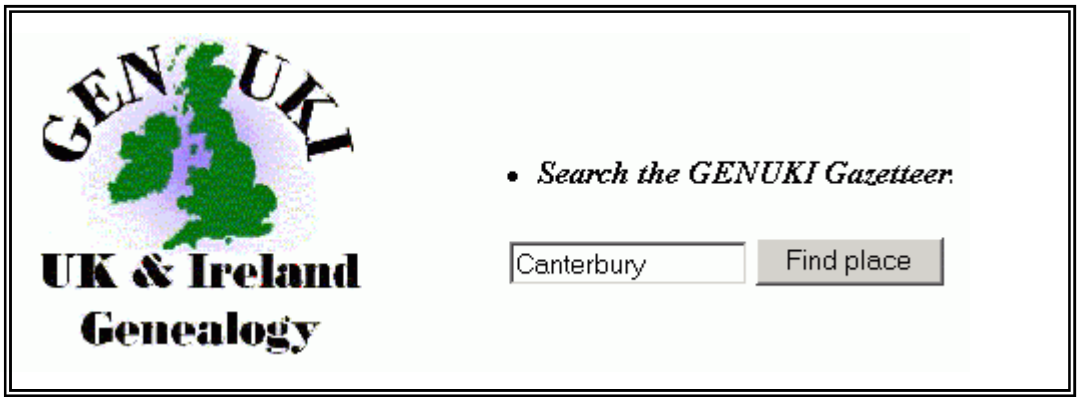


Figura 3.5- Interface de consulta do *Gazetteer* Genuki

Limitações

O *Gazetteer* do projeto Genuki apresenta algumas limitações: não possui informações espaciais; não possui suporte a temporalidade dos dados; apresenta apenas a localização (através de um mapa) de determinado lugar geográfico, sem informações adicionais; as informações contidas no *Gazetteer* não cobrem todo Reino Unido e Irlanda.

3.3 Considerações Finais

Neste capítulo, apresentamos o conceito e as principais características dos *Gazetteers*. Um *Gazetteer* é uma ferramenta que utiliza referência espacial indireta, para recuperar informações espaciais e não espaciais sobre um lugar geográfico. A principal forma de geo-referência utilizada pelos *Gazetteers* é através do nome de um lugar.

Alguns projetos que incluem o desenvolvimento de *Gazetteers*, bibliotecas digitais georeferenciadas e ferramentas de busca que realizam pesquisas sobre dados espaciais e temporais também foram apresentados neste capítulo. Dentre os projetos apresentados estão os seguintes: Alexandria Digital Library (ADL), Perseus, Geographic Names Information System (GNIS), Getty Thesaurus of Geographic Names (TGN) e Genuki.

Na apresentação e detalhamento desses projetos percebemos algumas deficiências encontradas no processo de recuperação das informações. As principais deficiências são elencadas na tabela 3.1.

Principais Problemas	Trabalhos Relacionados
Falta da temporalidade dos dados	• Projeto ADL: não possui informação sobre a temporalidade dos dados; • Projeto Perseus: pequeno suporte sobre a temporalidade dos dados; • Projeto GNIS: não possui informação sobre a temporalidade dos dados; • Projeto TGN: não possui informação sobre a temporalidade dos dados; • Projeto Genuki: não apresenta a temporalidade dos dados.
Mecanismos para refinar as consultas	• Projeto ADL: suporte limitado às operações espaciais (operadores *Overlap* e *Contains*); • Projeto Perseus: inexistência de operações espaciais e temporais e mecanismos de refinamento; • Projeto GNIS: não possui suporte para operações espaciais e temporais; • Projeto TGN: não suporta operações espaciais e temporais; • Projeto Genuki: não possui mecanismos para o refinamento das consultas
Sistemas com interfaces gráficas amigáveis e simples	• Projeto ADL: interface gráfica amigável e simples; • Projeto Perseus: resultados com informações espaciais e textuais são apresentados em níveis diferentes na interface gráfica; • Projeto GNIS: interface gráfica complexa, com utilização de códigos. • Projeto TGN: apresenta poucas informações espaciais e não espaciais sobre um lugar geográfico • Projeto Genuki: apresenta apenas a localização do lugar geográfico em um mapa, sem informações adicionais.

Tabela 3.1- Problemas na recuperação de Informação Geográfica

Na tabela 3.1 estão relacionados, de forma comparativa, os principais problemas referentes a recuperação de informação geográfica e os sistemas disponíveis na *Web* que realizam a recuperação dessas informações. As deficiências dos sistemas atuais são notórias, por isso, existe a necessidade de se construir um sistema, tal como o GeoLocalizador, para suprir tais deficiências.

Capítulo 4

Modelagem do GeoLocalizador

Nos capítulos 2 e 3 vimos os principais conceitos e características dos SRIs e dos trabalhos relacionados com nossa pesquisa. Neste capítulo, apresentamos a modelagem para a construção do *Gazetteer* GeoLocalizador, apresentando um modelo conceitual com a integração dos dados espaciais e temporais relacionados a lugares geográficos. Como visto no capítulo anterior, são notórias as deficiências apresentadas pelos sistemas atuais de referência indireta a dados espaço-temporais. Por este motivo, propomos neste capítulo um modelo capaz de suprir as principais deficiências, com a descrição das funcionalidades implementadas pelo GeoLocalizador.

O modelo conceitual do sistema será explorado na seção 4.1, onde destacamos as principais entidades e relacionamentos de um *Gazetteer*. Na seção 4.2, abordamos o levantamento dos requisitos funcionais e não funcionais, realizado a partir da análise de algumas funcionalidades do sistema e dos problemas encontrados nos projetos apresentados no capítulo 3. Na seção 4.3, apresentamos o modelo de "casos de uso", gerado pelo processo de realização de uma consulta. Na seção 4.4, expomos o projeto de diagrama de classes do sistema, onde são descritos para cada classe, os principais métodos e suas responsabilidades. A seção 4.5 aborda a modelagem de um *thesaurus* geográfico, com a implementação de seus principais relacionamentos.

4.1 Modelo Conceitual

Nesta seção apresentamos o modelo conceitual do GeoLocalizador. Neste modelo realizamos uma visão de alto nível do sistema. O GeoLocalizador é modelado com a composição de uma hierarquia de lugares geográficos expressada pelo padrão *Composite* [Gamma *et al*, 1995]. Essa composição permite a modelagem dos relacionamentos "é-parte" e "parte-todo", os quais são muito úteis na implementação

de *Gazetteers* [Baptista & Kemp, 2000]. O modelo conceitual do GeoLocalizador é apresentado na figura 4.1.

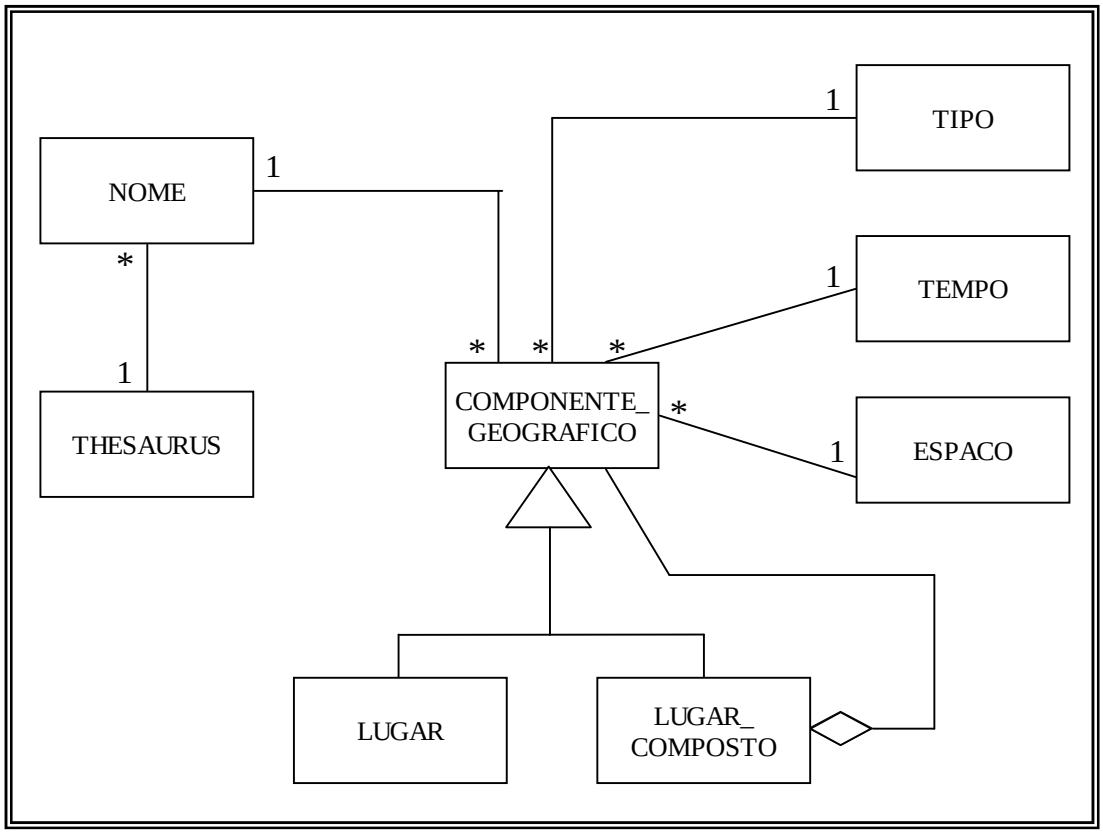


Figura 4.1- Modelo Conceitual do GeoLocalizador

Conforme a figura 4.1, a entidade Componente Geográfico representa o elemento principal do *Gazetteer*. As principais funções de cada entidade do modelo que interage com o Componente Geográfico são apresentadas abaixo:

- Lugar: esta entidade representa um lugar geográfico simples. Esse lugar não possui relacionamentos com outros elementos geográficos.
- Lugar Composto: nesta entidade é possível estabelecer relações entre elementos geográficos, com a ajuda do padrão *Composite*. Um objeto espacial é denominado composto, quando ele está diretamente relacionado, através dos relacionamentos “é-parte” e “todo-parte”, com outro objeto espacial.
- Tipo: esta entidade indica o tipo de feição ou característica de um lugar geográfico. Exemplos de tipos incluem: cidades, estados, rios, lagos, montanhas, etc.

- Tempo: esta entidade registra o intervalo de tempo válido de um determinado lugar geográfico.
- Espaco: esta entidade faz referência ao espaço geográfico de determinado objeto espacial. Este “espaco” é representado através de um MBR ou Centróide, os quais são definidos por coordenadas de latitude e longitude.
- Nome: esta entidade é responsável pela identificação de um lugar geográfico.
- *Thesaurus*: nesta entidade são modelados diversos relacionamentos, em níveis semântico e conceitual, entre lugares geográficos. O propósito destes relacionamentos é enriquecer o conteúdo das informações que envolvem um lugar geográfico.

4.2 Levantamento de Requisitos

Na modelagem do GeoLocalizador foram levantados diversos requisitos funcionais e não funcionais que o sistema deve atender. Nesta seção detalhamos estes requisitos.

4.2.1 Requisitos Funcionais

Os requisitos funcionais representam os aspectos de funcionalidade do sistema, determinando “o que” o sistema faz. No capítulo 3, descrevemos alguns sistemas disponíveis na *Web* que manipulam dados geográficos. Esses sistemas apresentam limitações, por isso, nesta seção propomos alguns requisitos funcionais que o GeoLocalizador deve atender. A seguir, apresentamos esses requisitos:

RF1- Disponibilizar o sistema de consultas na Web

Atualmente existe a necessidade de aplicações que estejam disponíveis o maior tempo possível, que sejam simples e que funcionem com eficiência. Assim, qualquer usuário pode acessar o sistema e realizar operações a qualquer momento.

O sistema de consultas deve estar disponível na *Web* permitindo que os usuários, através de um simples *browser*, possam acessar o sistema e realizar as consultas.

RF2- Possibilitar diferentes formas de submeter às consultas

A recuperação de dados geográficos, geralmente envolve informações espaciais e temporais. Por isso, existe a importância de prover diferentes métodos de acesso a esses dados. Uma das principais funcionalidades do sistema é definir a forma de submeter uma consulta.

No GeoLocalizador, as consultas devem ser submetidas através de duas formas: na forma direta através digitação do termo da pesquisa ou através da navegação em uma árvore de lugares geográficos até encontrar o lugar desejado para a pesquisa. Portanto, o sistema deverá ter um esquema que incorpore a definição de diferentes formas de submeter às consultas.

RF3- Disponibilizar a utilização de operadores espaciais e temporais

A partir do momento que o sistema manipula dados espaciais e temporais, é importante a existência de uma ferramenta que suporte operações sobre tais dados, considerando os requisitos de tempo e espaço.

O sistema deve ser capaz de permitir diferentes operações sobre os dados espaciais, com operadores que explorem as características de um determinado objeto espacial. Do mesmo modo, diferentes operações temporais com a definição de datas e intervalos de tempo, bem como a utilização de operadores temporais. Tais funcionalidades permitem aos usuários o refinamento dos dados que são retornados em uma consulta.

RF4- Definir a estrutura da hierarquia de lugares geográficos

As informações sobre os lugares geográficos devem estar organizadas em uma árvore seguindo uma hierarquia geográfica, tais como: continentes, países, estados e municípios. Essa organização facilita permite que os usuários possam navegar e a localizar um lugar específico na árvore com dados geográficos.

O sistema deve possuir um esquema para estruturar e manter a hierarquia geográfica. Novos lugares devem ser adicionados na base de dados e possíveis mudanças na árvore de lugares geográficos devem ser permitidas.

RF5- Expandir a consulta

Nos sistemas de recuperação de informação, uma consulta pode gerar resultados em diferentes etapas. Na geração destes resultados, existem métodos capazes de expandir os resultados de uma consulta, a fim de encontrar informações mais específicas sobre determinado item.

O sistema deve permitir a expansão dos termos que são retornados durante o processo de apresentação dos resultados da pesquisa. Com a expansão da consulta, diversos relacionamentos podem ser visualizados.

RF6- Definir relacionamentos entre os termos

A expansão de uma consulta gera uma série de informações que estão de alguma forma relacionadas com o item que foi expandido na consulta. Essas informações enriquecem o conteúdo dos resultados que são apresentados aos usuários.

O sistema deve definir relacionamentos entre os termos que são armazenados. Estes relacionamentos, podem indicar o posicionamento de um determinado termo em relação a outros termos em uma hierarquia, ou ainda relações de equivalência e associação.

RF7- Integração com outros sistemas

A disponibilização do sistema na *Web* permite a fácil integração com outras aplicações para troca e compartilhamento de informações. O sistema deve possuir um esquema para o tratamento da interoperabilidade, permitindo a troca de serviços, para o compartilhamento de informações entre *Gazetteers*. A troca de serviços deve ser feita com a utilização de um modelo para a padronização de entradas e invocações de funções no *Gazetteer*.

RF8- Disponibilizar graficamente os resultados das consultas na Web

Os resultados obtidos na consulta devem ser disponibilizados graficamente na *Web*. A publicação das informações com a ilustração de mapas, localização dos lugares geográficos e *links* para páginas na Internet permite uma maior satisfação dos usuários do sistema.

4.2.2 Requisitos Não Funcionais

Através do auxilio dos requisitos não funcionais os projetistas procuram determinar "como" o sistema executa determinada ação. Para isso, esses requisitos caracterizam o sistema dentro de categorias, tais como: facilidade de uso, tempo de resposta, perfil dos usuários, etc. Alguns requisitos não funcionais do GeoLocalizador são:

RNF1- Tipo de interface

Principalmente em ambientes *Web*, a qualidade da apresentação dos dados e a funcionalidade da interface das aplicações são indispensáveis. Como visto no capítulo 3, muitas vezes interfaces gráficas complexas dificultam o processo de recuperação de informações. Para facilitar o acesso e a realização de pesquisas no sistema, um tipo de interface amigável, deve ser definido.

RNF2- Perfil dos usuários do sistema

A utilização do sistema poderá ser feita por usuários da Internet. Os usuários com maior afinidade na área de geoprocessamento e informações geográficas poderão explorar as principais características do sistema. Do mesmo modo, usuários leigos e sem afinidades com essa área podem buscar informações com a ajuda das ferramentas de consulta.

RNF3- Independência de plataforma

Com a difusão de diferentes plataformas e o crescimento do número de usuários, se faz necessário o desenvolvimento de aplicações que sejam portáveis. As aplicações que são executadas via *browser* favorecem os usuários que utilizam diferentes plataformas. Porém, no desenvolvimento em três camadas deve-se procurar utilizar apenas comandos padrões do SQL, para execução de consultas no banco de dados. Isso deve ser feito para não afetar a portabilidade da aplicação. O sistema deve ser implementado de forma que suporte a independência de plataformas.

RNF4- Tempo de resposta

Uma das características mais importantes para uma aplicação é a garantia que ela tenha um bom desempenho. O tempo de resposta para uma requisição feita pelos usuários deve ser o menor possível. Para que um tempo de resposta aceitável seja alcançado, o processo de recuperação de informações deve, sempre que possível, acessar apenas os dados relevantes para cada consulta requerida pelo usuário.

4.3 Modelo de Casos de Uso

A modelagem do diagrama de casos de uso é uma técnica usada para descrever e definir os requisitos funcionais de um sistema [Larman, 1998]. O modelo de casos de uso é formado por um diagrama usado para identificar como o sistema se comporta

nas várias situações que podem ocorrer durante as suas operações. Os componentes deste diagrama são os “atores” (representam qualquer entidade que interage com o sistema) e os “casos de uso” (representam todas as situações possíveis de utilização do sistema).

Na figura 4.2, apresentamos o diagrama de casos de uso para o processo de realização de uma consulta no GeoLocalizador. Nesse diagrama, são descritos os principais eventos de iteração entre o sistema e um “ator”. No diagrama da figura 4.2, o “ator” representa o usuário do sistema.

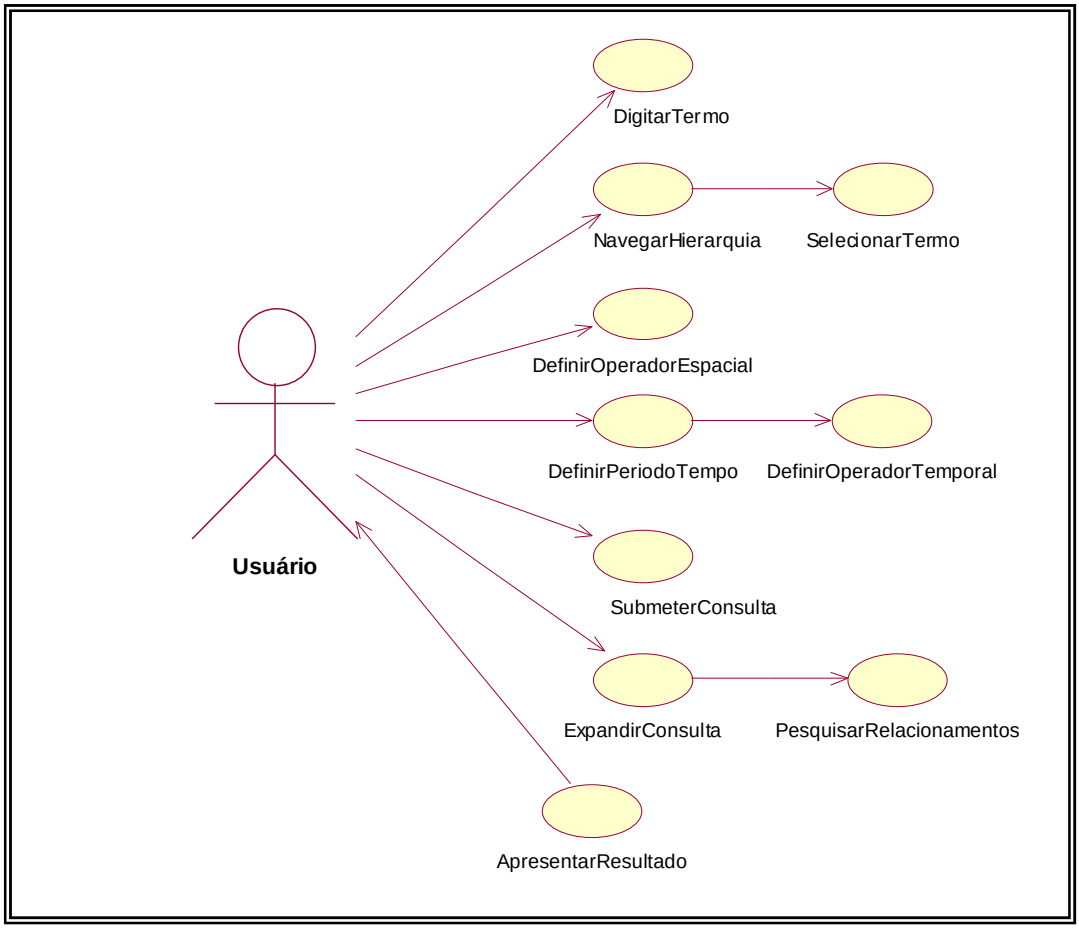


Figura 4.2- Diagrama de Casos de Uso

Conforme a figura 4.2, percebemos que o usuário (ator) acessa o sistema e inicia o processo de realização de uma consulta. Esse processo envolve diversos “casos de uso”. Nas tabelas 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9 e 4.10 são descritas as principais características de cada “caso de uso” do digrama da figura 4.2. Essas características incluem: o nome do “caso de uso”; um resumo com sua funcionalidade; e a referência para os requisitos funcionais.

Caso de Uso	DigitarTermo
Resumo	O usuário acessa o sistema via *Web* e digita em uma caixa de texto o termo a ser pesquisado.
Referência	RF1 e RF2

Tabela 4.1- Caso de Uso DigitarTermo

Caso de Uso	NavegarHierarquia
Resumo	O usuário acessa o sistema via *Web* e navega em uma árvore formada por uma hierarquia de lugares geográficos.
Referência	RF1, RF2 e RF4

Tabela 4.2- Caso de Uso NavegarHierarquia

Caso de Uso	SelecionarTermo
Resumo	O usuário navega na árvore com a hierarquia de lugares e seleciona o lugar geográfico a ser pesquisado.
Referência	RF2

Tabela 4.3- Caso de Uso SelecionarTermo

Caso de Uso	DefinirOperadorEspacial
Resumo	Com a definição do lugar geográfico a ser pesquisado, o usuário seleciona o operador espacial que será utilizado na consulta.
Referência	RF3

Tabela 4.4- Caso de Uso DefinirOperadorEspacial

Caso de Uso	DefinirPeriodoTempo
Resumo	O usuário define um período de tempo que será utilizado no processo de consulta.
Referência	RF3

Tabela 4.5- Caso de Uso DefinirPeriodoTempo

Caso de Uso	DefinirOperadorTemporal
Resumo	Com a definição do período de tempo, o usuário seleciona o operador temporal que será utilizado.
Referência	RF3

Tabela 4.6- Caso de Uso DefinirOperadorTemporal

Caso de Uso	SubmeterConsulta
Resumo	Depois de montar a consulta, com a definição do lugar geográfico e com a seleção dos operadores espacial e temporal a consulta é submetida.
Referência	RF1, RF2, RF3 e RF7

Tabela 4.7- Caso de Uso SubmeterConsulta

Caso de Uso	ExpandirConsulta
Resumo	Os itens que são retornados na primeira fase da consulta podem ser expandidos, desta forma, obtendo os relacionamentos implementados em um *thesaurus* geográfico.
Referência	RF5

Tabela 4.8- Caso de Uso ExpandirConsulta

Caso de Uso	PesquisarRelacionamentos
Resumo	Com a expansão da consulta, vários relacionamentos são estabelecidos. Através de *links* para páginas *Web*, o usuário poderá realizar pesquisas sobre os itens que são relacionados com o termo expandido.
Referência	RF5 e RF6

Tabela 4.9- Caso de Uso PesquisarRelacionamentos

Caso de Uso	ApresentarResultado
Resumo	Depois da consulta ser processada na base de dados, todos os resultados são apresentados ao usuário.
Referência	RF8

Tabela 4.10- Caso de Uso ApresentarResultado

4.4 Diagrama de Classes

Com a definição do diagrama de "casos de uso", na seção 4.3, podemos definir o diagrama de classes do sistema. O diagrama de classes demonstra a estrutura estática das classes do sistema. No processo de análise geramos um modelo para entender o domínio do problema e montamos uma possível solução para atender os requisitos do sistema.

Nesta seção, apresentamos o diagrama de classes do GeoLocalizador. A notação usada nos digramas das diversas fases deste projeto é a *Unified Modeling Language* (UML) [Rumbaugh *et al*, 1999 e Booch, 1999]. Para um melhor entendimento o diagrama de classes foi dividido em três partes. A seção 4.4.1 apresenta o digrama com as classes da camada de Apresentação. Na seção 4.4.2, abordamos o diagrama com as classes da camada Lógica. As classes responsáveis pelo tratamento dos dados são apresentadas na seção 4.4.3. O diagrama de classes completo, com a integração de todas as classes das três camadas do sistema pode ser visualizado no apêndice A.

4.4.1 Classes da Camada de Apresentação

As classes da camada de Apresentação representam a funcionalidade e a estrutura da interface do GeoLocalizador. Na figura 4.3, apresentamos o digrama envolvendo essas classes.

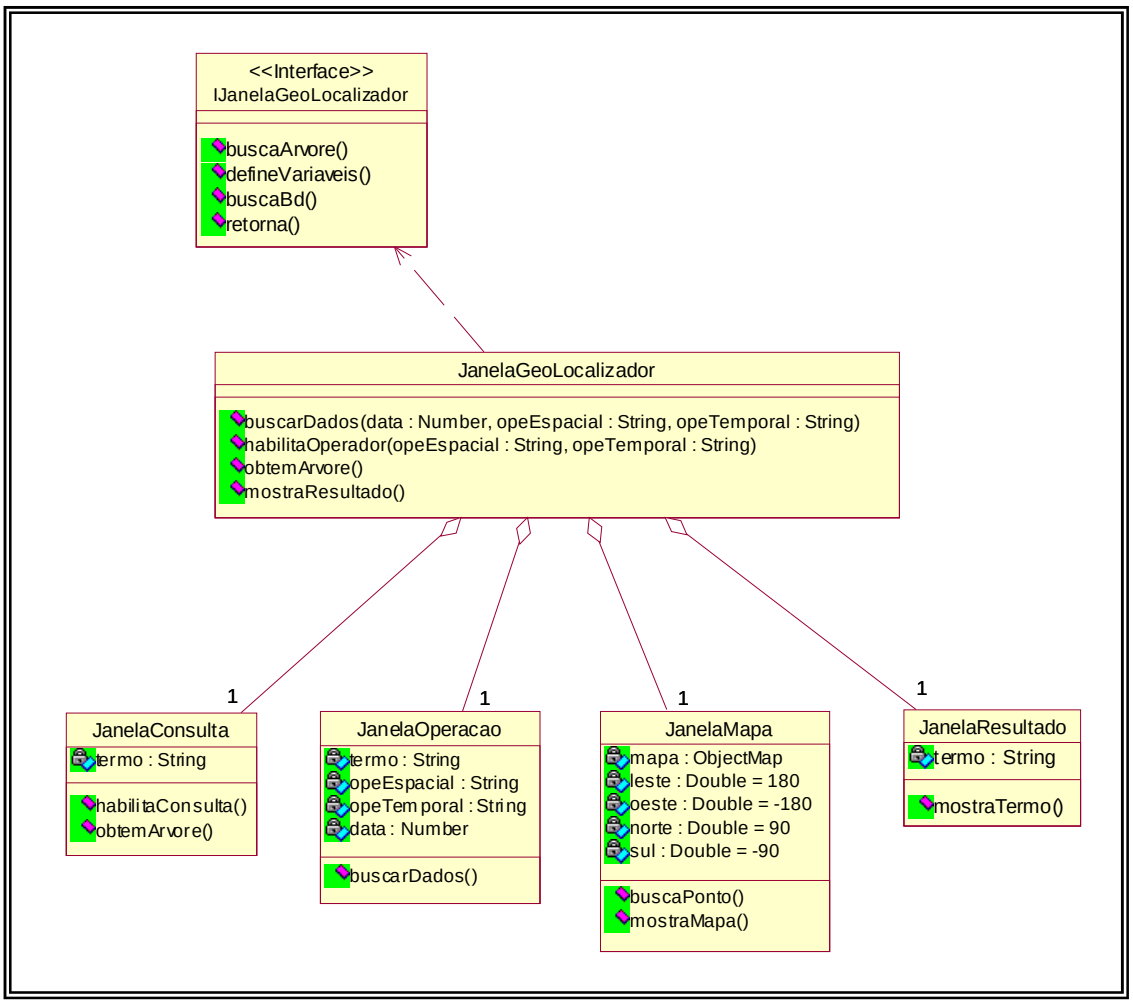


Figura 4.3- Diagrama representando as classes da camada de Apresentação

Na figura 4.3, podemos perceber a utilização da interface chamada **IJanelaGeoLocalizador**. O processo de modelagem com utilização de interfaces permitem um fraco acoplamento entre as classes da camada de Apresentação com as outras classes utilizadas na modelagem do GeoLocalizador. Da mesma forma, as interfaces possibilitam uma maior flexibilidade e a possibilidade de reusar a funcionalidade de uma classe.

As principais operações relacionadas com a interface gráfica do GeoLocalizador estão dispostas na classe **JanelaGeoLocalizador**. Essa classe possui o método *buscarDados()*, que contém todas as informações adicionadas pelo usuário

na interface do sistema. O método *habilitaOperador()* define os operadores espaciais e temporais relacionados com a consulta. O método *obtemArvore()* exerce a função de montar e atualizar a hierarquia de lugares geográficos, no momento que a interface do sistema é inicializada. O método *mostraResultado()* apresenta os resultados da consulta ao usuário.

A classe **JanelaGeoLocalizador** é composta por mais quatro classes responsáveis pela interface gráfica do GeoLocalizador. A classe **JanelaConsulta** obtém a árvore com a hierarquia de lugares geográficos e habilita o processo de consulta. A classe **JanelaOperacao** captura os dados, entre eles a definição dos operadores espaciais e temporais, fornecidos pelo usuário para formular e submeter a consulta. Na classe **JanelaMapa**, é modelado o mapa mundi com os pontos das coordenadas de latitude e longitude. Essa classe, também é responsável pela localização de um lugar geográfico no mapa. A classe **JanelaResultado** possui o método *mostraTermo()*, responsável pela apresentação dos resultados.

4.4.2 Classes da Camada Lógica

As classes que definem a camada Lógica do sistema são descritas no digrama da figura 4.4. Essas classes representam os relacionamentos, os atributos e métodos envolvendo os lugares geográficos que compõem o *Gazetteer*.

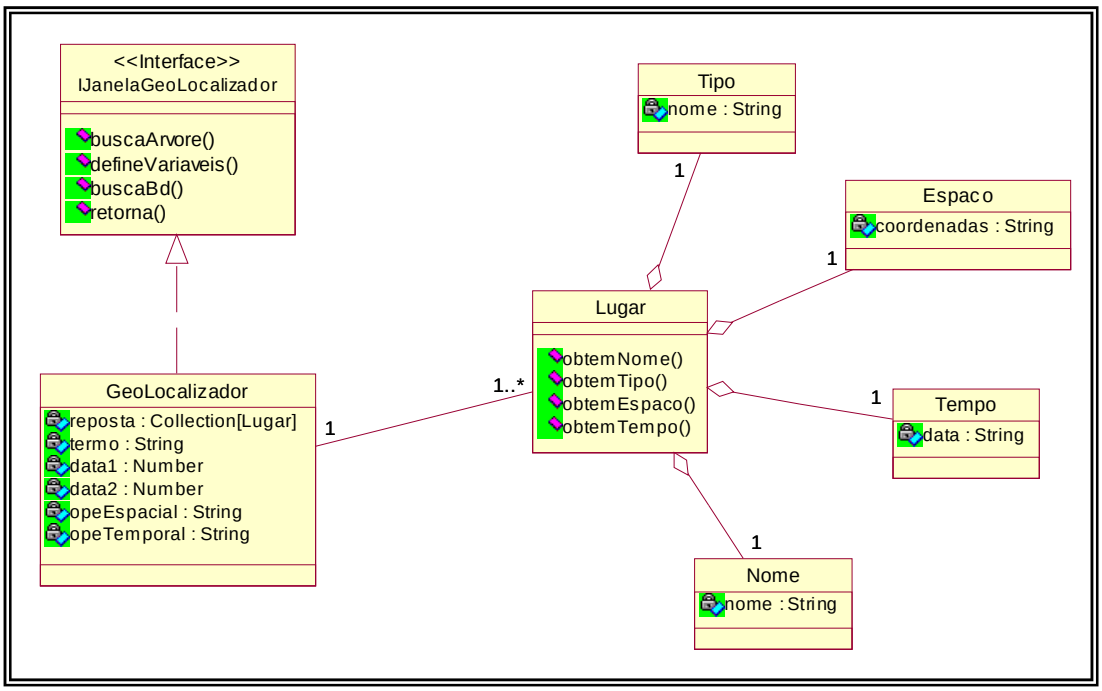


Figura 4.4- Diagrama representando as classes da camada Lógica

A classe **GeoLocalizador** é a principal classe do sistema. Essa classe centraliza todas as informações envolvendo as camadas de Apresentação, Lógica e de Dados (ver apêndice A). Essa classe implementa os métodos da interface **IJanelaGeoLocalizador**.

O método *buscaArvore()*, na classe **IJanelaGeoLocalizador**, busca os dados na base de dados para a formação da árvore com a hierarquia de lugares geográficos. Os dados para a formação da árvore geográfica são obtidos em uma tabela no banco de dados, e transformados em um documento XML. Esse documento XML possui a estrutura da árvore com a hierarquia de lugares geográficos, que é apresentada aos usuários. O método *defineVariaveis()* recebe os dados fornecidos pelo usuário para a formulação da consulta. O método *buscaBd()* é responsável pela emissão da consulta, formulada nas classes da camada de Apresentação e enviada para as classes da camada de Dados, para a execução na base dados. O método *retorna()* contém o resultado da execução da consulta.

A seguir, definimos os atributos da classe **GeoLocalizador** responsáveis pela formulação e armazenagem dos resultados da consulta. Destacamos cada atributo separadamente:

- termo: este atributo representa o nome do lugar geográfico submetido pelo usuário na formulação da consulta, por exemplo: Paraíba;
- data1: o atributo “data1” define a primeira data informada pelo usuário na composição de um período de tempo.
- data2: um período de tempo pode ser formado por uma data (neste caso, um instante de tempo) ou através de duas datas (neste caso, um intervalo de tempo). O atributo “data2” define a segunda data do intervalo de tempo utilizado na consulta.
- opeEspacial: este atributo armazena o operador espacial utilizado na consulta.
- opeTemporal: este atributo possui o operador temporal definido pelo usuário, para a formulação da consulta;
- reposta: este atributo armazena uma coleção de lugares geográficos retornados em uma consulta, por exemplo: na seguinte consulta – “o estado da Paraíba contém quais municípios?”. Nesta consulta, uma coleção

de lugares geográficos será retornada, tais como: Campina Grande, João Pessoa, Patos, Cajazeiras, etc.

A classe **Lugar** exerce uma função importante no sistema. Nessa classe estão as informações que definem um lugar geográfico. Conforme abordado no modelo conceitual da figura 4.1, um lugar geográfico é composto por quatro atributos: nome, tipo, espaço e tempo. Na modelagem do GeoLocalizador, esses atributos são definidos através das classes **Nome**, **Tipo**, **Espaco** e **Tempo**.

Os métodos da classe **Lugar** são responsáveis pela definição das características que definem um lugar geográfico. O método *obtemNome()* obtém o nome do lugar geográfico. O método *obtemTipo()* recebe o tipo de um lugar (por exemplo: para o "lugar" Campina Grande, o "tipo" é Cidade). O método *obtemEspaco()* obtém as coordenadas espaciais que definem um lugar geográfico. O método *obtemTempo()* define o período de tempo válido de um determinado um lugar geográfico.

4.4.3 Classes da Camada de Dados

As classes da camada de Dados modelam a conexão do sistema com a base de dados para a recuperação das informações. As classes desta camada são apresentadas no diagrama da figura 4.5.

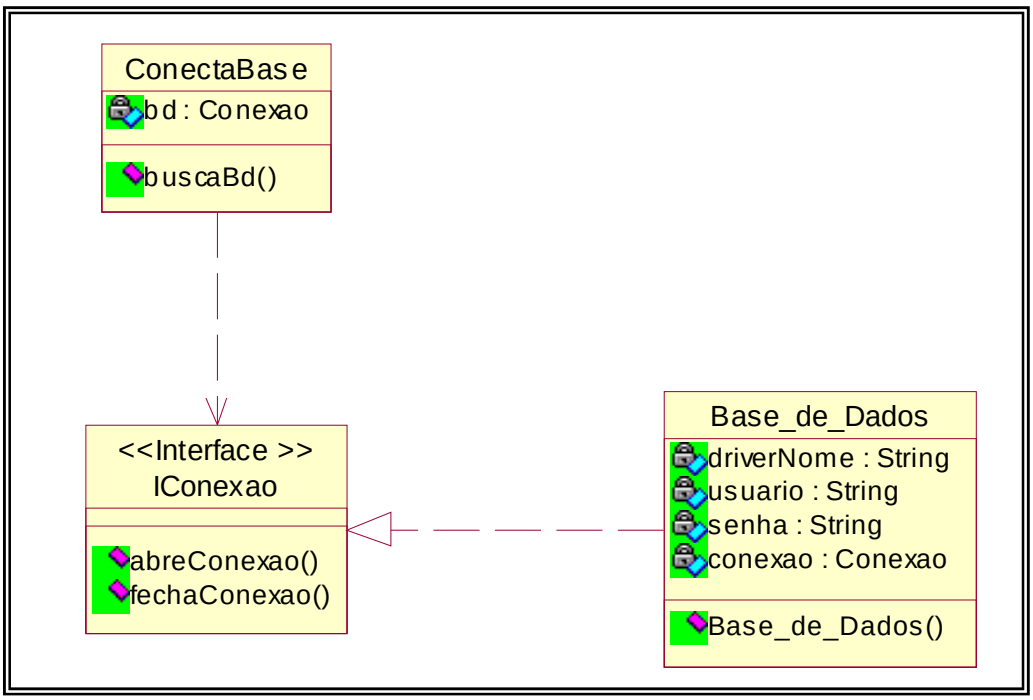


Figura 4.5- Diagrama representando as classes da camada de Dados

Conforme o diagrama da figura 4.5, na classe **ConectaBase**, o método *buscaBd()* recebe a consulta formulada pelo usuário e instância a base de dados para a execução da consulta. A interface **IConexão** implementa os métodos responsáveis pelo início e encerramento da conexão com o banco de dados. Esta interface permite o tratamento de diferentes formas de abrir e fechar a conexão, através dos métodos *abreConexao()* e *fechaConexao()*. Com isso, diferentes bancos de dados podem ser usados com métodos distintos de acesso.

Na classe **Base_de_Dados**, as operações sobre o banco de dados são executadas. No método *Base_de_Dados()* é criado um objeto que representa o banco de dados sendo utilizado. Na utilização de outros bancos de dados, novos objetos serão criados para cada banco de dados. A seguir, detalhamos os atributos da classe **Base_de_Dados**:

- driverNome: este atributo possui o nome do driver utilizado na conexão com o banco de dados;
- usuario: este atributo é usado para identificação do usuário;
- senha: o atributo "senha" é utilizado para validar a identificação do usuário;
- conexao: este atributo possui as informações necessárias para realizar a conexão com o banco de dados.

4.5 Modelagem do Thesaurus Geográfico

Como mencionado no capítulo 2, um *thesaurus* é formado por um conjunto de termos com uma série de relacionamentos semânticos entre esses termos [Tudhope *et al*, 2001]. No sistema GeoLocalizador modelamos um *thesaurus* geográfico para expandir os termos que são retornados em uma consulta.

Primeiramente, precisamos definir os relacionamentos que serão modelados no *thesaurus* geográfico. Diante das funcionalidades do GeoLocalizador, definimos três relacionamentos: os relacionamentos de equivalência, hierárquico e associativo. Na figura 4.6, apresentamos o diagrama de classes para modelagem destes relacionamentos:

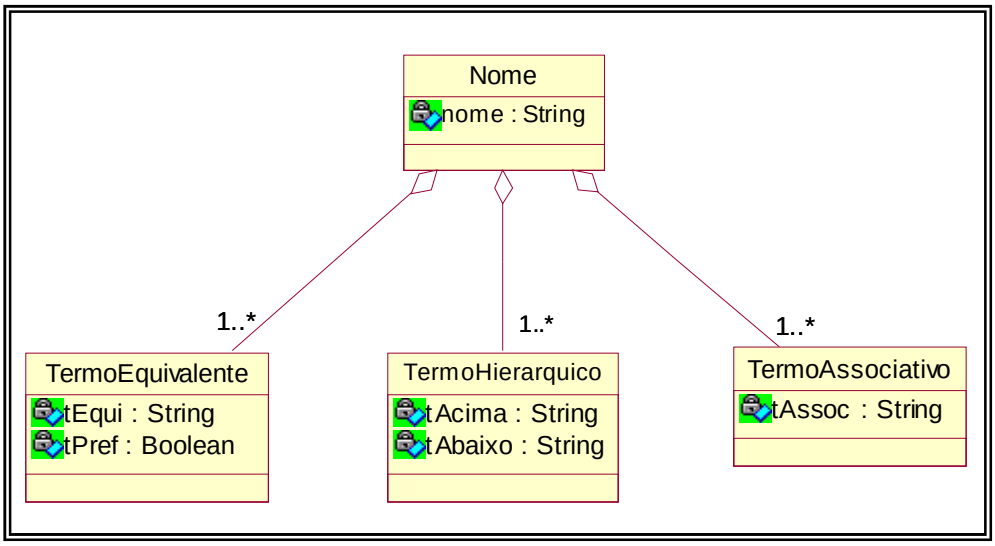


Figura 4.6- Diagrama de classes do *thesaurus* geográfico

Conforme o diagrama de classes da figura 4.6, cada classe que representa um relacionamento do *thesaurus* está diretamente ligada com a classe **Nome**, que é responsável pela identificação de um lugar geográfico. O atributo “nome” desta classe pode possuir um ou vários termos equivalentes, hierárquicos e associativos. A seguir, descrevemos a funcionalidade de cada classe que se relaciona com a classe **Nome**:

Classe **TermoEquivalente**: esta classe define todos os termos que são sinônimos de um determinado lugar geográfico. Em um domínio geográfico, termos equivalentes (atributo “tEqui”) podem expressar diferentes designações que foram ou que são dadas para uma mesma região geográfica. Nesta classe, também definimos se um determinado termo é preferido (atributo “tPref”), em relação a outros termos equivalentes. Um termo é denominado preferido, quando ele é o mais utilizado ou mais atual em determinado contexto.

Classe **TermoHierarquico**: esta classe define a posição de determinado termo em uma hierarquia de lugares geográficos. Os atributos “tAcima” e “tAbaixo” definem os termos que estão em níveis acima (tAcima) ou abaixo (tAbaixo) de determinado termo na hierarquia de lugares.

Classe **TermoAssociativo**: esta classe define os termos (atributo “tAssoc”) que estão de alguma forma relacionados com o termo em questão. Os termos associativos representam associações semânticas e conceituais, entre diferentes termos.

Capítulo 5

Arquitetura e Implementação do GeoLocalizador

No capítulo 4 especificamos o modelo conceitual, juntamente com os requisitos e a análise do sistema. Os casos de uso e o diagrama de classes do sistema foram detalhados, a fim explicitar a funcionalidade e as principais características do GeoLocalizador. Neste capítulo, apresentamos de forma detalhada a arquitetura proposta, bem como o funcionamento do GeoLocalizador, que foi desenvolvido e validado conforme as idéias e a modelagem apresentada nos capítulos anteriores.

Este capítulo está estruturado da seguinte forma: na seção 5.1, detalhamos a arquitetura do sistema. A seção 5.2 aborda as principais tecnologias utilizadas no processo de implementação. Na seção 5.3, detalhamos os operadores e funções utilizadas no desenvolvimento do *thesaurus* geográfico. Na seção 5.4, ilustramos os aspectos da interface e funcionamento do GeoLocalizador.

5.1 Arquitetura do Sistema

O principal propósito da arquitetura do GeoLocalizador é prover uma interface simples e funcional que proporcione aos usuários o fácil acesso às informações geográficas. Uma característica importante é que essa arquitetura deve prover aos usuários, a possibilidade de acessar o sistema de diferentes regiões do planeta. Para isso, a arquitetura deve ser dividida em camadas e a aplicação deve estar disponível na Internet.

Nos últimos anos, a arquitetura das aplicações distribuídas tem variado em modelos que prevêem a divisão do programa em camadas. Essa divisão pode incluir modelos com duas-camadas, três-camadas ou n-camadas. O modelo de três-camadas apresenta algumas vantagens, tais como: utilização de *browser*; facilidade na adesão de novas camadas; e a excelente separação entre as camadas de Apresentação e Lógica que proporcionam facilidades em casos de mudanças na camada de Dados, tal como, o acesso a múltiplos servidores de banco de dados [Edwards, 1999]. Estas

vantagens e outras características determinaram a adoção do modelo de três-camadas na composição da arquitetura do GeoLocalizador.

Na figura 5.1, apresentamos a arquitetura do GeoLocalizador com a divisão das três-camadas: camada de Apresentação; a camada Lógica; e a camada de Dados. Nessa figura, podemos verificar as fases do processo de consulta até a busca das informações na base de dados.

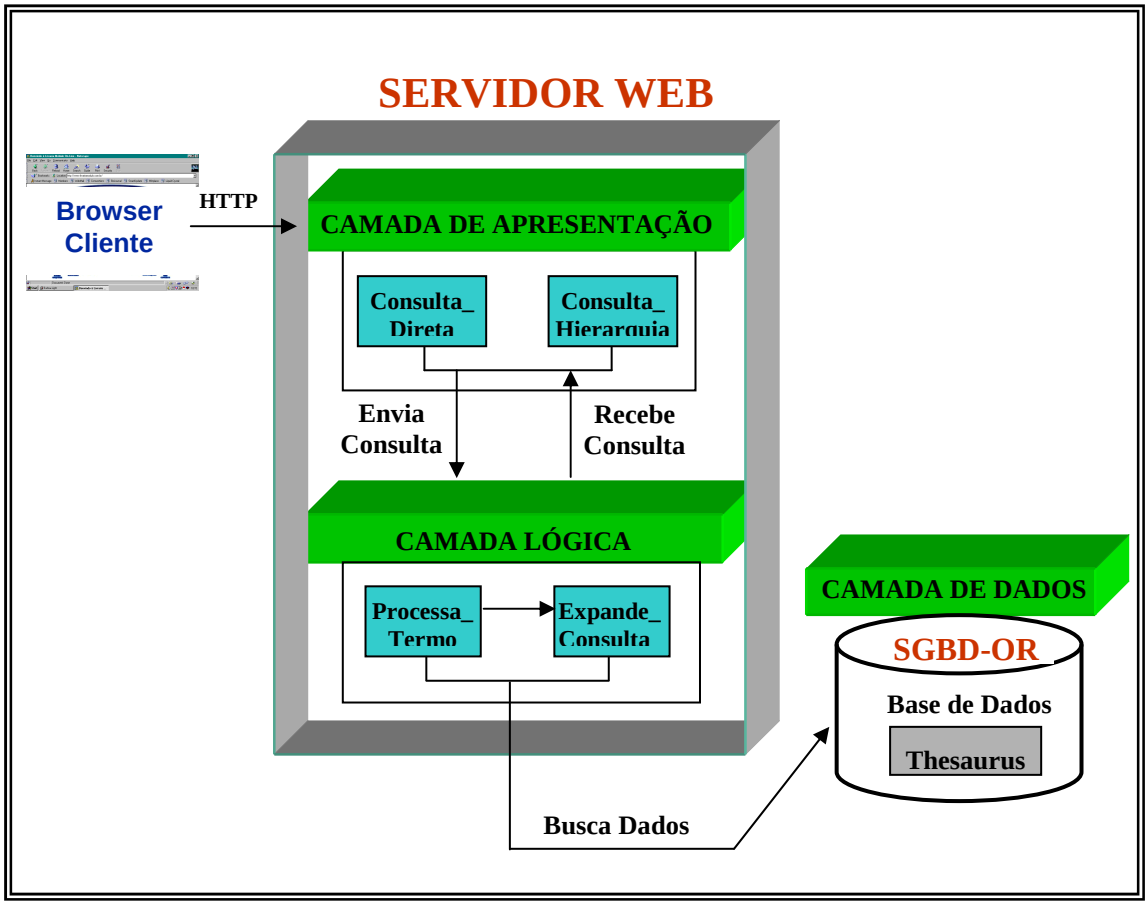


Figura 5.1- Arquitetura do GeoLocalizador

5.1.1 Camada de Apresentação

Na arquitetura organizada em três-camadas, a interface do usuário é modelada na camada de Apresentação. Nesta camada, nos preocupamos em desenvolver interfaces simples e funcionais, já que esta parte do programa é acessada pelos usuários.

Conforme a figura 5.1, com o *browser Web* via o protocolo HTTP (*HyperText Transfer Protocol*), o usuário (cliente) acessa a camada de Apresentação, onde está localizada a interface para a emissão das consultas. As classes dessa camada são executadas no Servidor *Web*, o qual dá o suporte essencial para disponibilidade dos

componentes da interface, que são apresentadas no *browser* do usuário. Essas classes definem os componentes "Consulta_Direta" e "Consulta_Hierarquia", responsáveis pelas duas formas distintas de submeter uma consulta. Esses componentes, também são responsáveis pela apresentação final dos resultados da consulta. A seguir, apresentamos as funcionalidades de cada componente desta camada:

- Consulta_Direta: recebe os dados digitados pelo usuário, para a formulação da consulta, e os envia para a camada Lógica.
- Consulta_Hierarquia: executa um processo similar, com o recebimento dos dados do usuário e envio para a camada Lógica. Porém, esse componente também é responsável pela organização e apresentação da árvore com a hierarquia de lugares geográficos.

No momento em que o sistema de consulta é inicializado, o componente "Consulta_Hierarquia" interpreta um documento XML com uma hierarquia de lugares geográficos que estão no banco de dados. Esse documento é interpretado, e uma árvore com dados geográficos é dinamicamente montada na interface do usuário. As informações que estão no documento XML são buscadas em uma tabela no banco de dados.

O processo de transformação dos dados da tabela para o documento XML é realizado através da ferramenta XSU[32] (*XML SQL Utility*) do *Oracle8i* e com a utilização da linguagem XSLT (*eXtensible Stylesheet Language Transformation*) [Kay, 2000]. Com XSU mapeamos o conteúdo de uma tabela do banco de dados para um documento XML. Esse documento precisa ser reestruturado, através da linguagem de transformação XSLT, para que o *parser* consiga interpretá-lo e gerar a hierarquia com lugares geográficos na interface do usuário.

5.1.2 Camada Lógica

Com a definição da consulta na camada de Apresentação, essa é enviada para a camada Lógica, que processa as operações e redireciona a consulta para a base de dados. Na camada Lógica está inserida toda a funcionalidade do sistema. Sua função é servir a camada de Apresentação, executando todos os processos em função das

[32] http://otn.oracle.com/docs/tech/xml/oracle_xsu/doc_library/adx04xsu.html

requisições dos usuários. Todo e qualquer acesso a base de dados deve ser feito por intermédio dessa camada.

A seguir, detalhamos cada componente desta camada da arquitetura:

- Processa_Termo: este componente é responsável pela execução da primeira fase da consulta. Nesta fase, as operações espaciais, que envolvem os MBRs dos lugares geográficos, e as operações temporais são executadas. Esse processo resulta na seleção de diversos lugares geográficos, conforme a consulta formulada pelo usuário.
- Expande_Consulta: com o resultado da primeira fase da consulta, o componente "Expande_Consulta" realiza o processo de expansão da consulta. Esse componente recebe o lugar geográfico selecionado pelo usuário e acessa o *thesaurus* localizado na base de dados.

5.1.3 Camada de Dados

Na camada de Dados são encontrados os dados usados nos experimentos, como os lugares geográficos com suas propriedades, tais como: o nome, espaço, tempo e tipo. Nesta camada está implementado o *thesaurus* geográfico utilizado no processo de expansão dos termos da consulta. Esse *thesaurus* possui os termos e os relacionamentos existentes entre eles.

A camada de Dados é implementada usando o SGBD objeto-relacional *Oracle8i*[33]. Essa camada recebe as requisições emitidas nas consultas, e então, executa as operações e envia o resultado de volta para a camada Lógica.

5.2 Tecnologias Utilizadas

O processo de desenvolvimento do sistema GeoLocalizador envolveu diferentes tecnologias. A utilização dessas tecnologias permitiu o melhor desempenho na realização das operações, bem como a construção de uma interface mais funcional para o sistema.

[33] http://www.oracle.com

A seguir, apresentamos as principais tecnologias empregadas no desenvolvimento do GeoLocalizador:

- Java [Gosling *et al*, 2000]: Java é uma linguagem de programação orientada a objeto, desenvolvida em 1995, pela *Sun Microsystems*[34]. Desde seu surgimento, esta linguagem vem sendo utilizada em larga escala em universidades e empresas. Suas principais características são a independência de plataforma, suporte a sistemas distribuídos, construção de aplicações para a Internet e facilidade no acesso a banco de dados. Para compilação e interpretação da linguagem Java foi utilizado o JDK[35] (*Java Development Kit*) 1.3.1_03.

- JSP [Patzer *et al*, 1999]: JSP (*Java Server Pages*) é uma linguagem oriunda da tecnologia Java utilizada para o desenvolvimento de aplicações *Web*. Da mesma forma que o Java, JSP tem a vantagem da portabilidade de plataforma. No GeoLocalizador, as classes JSP são utilizados nas páginas que compõem a camada de Apresentação, da arquitetura apresentada na seção 5.1. Para utilizar está linguagem é necessário a instalação de um servidor que ofereça suporte JSP, para tal, utilizamos o servidor *Web* do projeto Jakarta, o *Apache Tomcat* 4.0.1[36].

- XML [Bray *et al*, 2000]: XML é um novo padrão adotado pelo W3C[37] (*World Wide Web Consortium*), que provê um formato para descrever dados estruturados e semi-estruturados na *Web*. Essa linguagem difere de HTML[38] no sentido que ela foi projetada para descrever o conteúdo, em vez da apresentação dos documentos [Abiteboul *et al*, 1999]. No GeoLocalizador, utilizamos um documento XML para definir e estruturar a hierarquia de lugares geográficos.

- SVG [Ferraiolo, 2001]: SVG (*Scalable Vector Graphics*) é uma implementação XML, desenvolvida pela W3C. Essa linguagem é utilizada

[34] http://www.sun.com
[35] http://java.sun.com/j2se/1.3
[36] http://jakarta.apache.org/tomcat
[37] http://www.w3.org
[38] http://www.w3.org/MarkUp

para a descrição de elementos gráficos na forma vetorial em duas dimensões. Com o formato baseado em texto e não proprietário, o SVG permite a fácil edição e manipulação de imagens e mapas, sendo compatível com a linguagem Java. No GeoLocalizador, utilizamos SVG na formatação do mapa geográfico, com a definição das coordenadas de latitude e longitude do globo terrestre.

- O processo de consulta com o servidor de banco dados é feito via comandos SQL (*Structured Query Language*) [Eisenberg & Melton, 2000], com a utilização do JDBC[39] 2.0 (*Java DataBase Connectivity*) – uma API (*Application Programming Interfaces*) simples para acesso a banco de dados que permite que a aplicação Java possa estabelecer a conexão, e enviar comandos SQL para serem processados [Mukhar *et al*, 2001].

- A persistência dos dados é feita no SGBD Objeto-Relacional *Oracle8i* [Oracle, 2002], com a utilização dos cartuchos: *Oracle Spatial Cartridge*, para o tratamento das operações espaciais; e *Oracle ConText Cartridge*, para a implementação do *thesaurus* geográfico. Esse SGBD está rodando em um servidor, com o sistema operacional *Windows 2000 Server*.

5.3 Thesaurus Geográfico

Nesta seção, apresentamos os principais operadores e funções utilizadas no desenvolvimento do *thesaurus* geográfico utilizado pelo GeoLocalizador.

A implementação de um *thesaurus* geográfico possibilita o enriquecimento dos resultados da consulta, com apresentação de uma série de relacionamentos envolvendo o lugar geográfico expandido com os termos que estão no *thesaurus*. O processo de expansão da consulta é discutido na seção 5.4.

O *thesaurus* geográfico foi implementado no SGBD *Oracle8i* com a utilização do pacote *CTX_THES* do *ConText Cartridge*. Esse pacote possui uma série de operadores e funções necessárias à construção e gerenciamento das informações no

[39] http://java.sun.com/products/jdbc

thesaurus. Na tabela 5.1, apresentamos os principais operadores utilizados no desenvolvimento do *thesaurus* geográfico.

Operador	**Descrição**
BT	Define os relacionamentos *Broader Terms* (“todo-parte”).
CREATE_PHARSE	Utilizado na adição de termos ao *thesaurus*.
CREATE_THESAURUS	Operador utilizado para a criação de um *thesaurus*.
NT	Define os relacionamentos *Narrower Terms* (“é-parte”).
PT	Define os *Preferred Terms* (“termos preferidos”).
RT	Define os termos que são *Related Terms* (“termos relacionados”).
SYN	Define os termos que são sinônimos no *thesaurus*.

Tabela 5.1- Principias operadores do *thesaurus*

A definição de cada relacionamento de um *thesaurus* foi detalhada no capítulo 2 desta dissertação. Na tabela 5.2, descrevemos as principais funções do pacote *CTX_THES* utilizadas para gerenciar e emitir consultas no *thesaurus*.

Função	Descrição
ctx_thes.create_thesaurus ('thes_GeoLocal');	Função responsável pela criação do *thesaurus* “thes_GeoLocal”.
ctx_thes.create_phrase ('thes_GeoLocal','termo');	Função utilizada para a adição de termos ao *thesaurus*.
ctx_thes.drop_pharse('thes_GeoLocal', 'termo');	Função utilizada para remoção de um termo do *thesaurus*.
ctx_thes.create_relation('thes_GeoLocal', 'termo1','OPERADOR','termo2');	Função utilizada para a criação de relacionamentos entre os termos do *thesaurus*.
ctx_thes.create_relation('thes_GeoLocal', 'termo1','SYN','termo2');	Função utilizada para definir que o “termo1” é sinônimo do “termo2”.
ctx_thes.create_relation('thes_GeoLocal', 'termo1','BTP','termo');	Função utilizada para definir que o “termo1” está um (1) nível acima do “termo2” em uma hierarquia.
ctx_thes.create_relation('thes_GeoLocal’ ,'termo1','USE','termo2');	Função utilizada para definir que “termo2” é PT (“Preferred Term”) do “termo1”.
Ctx_thes.pt(termo,' thes_GeoLocal');	Função utilizada para retornar o *Preferred Term* do “termo”.
Ctx_thes.btp(termo,2,' thes_GeoLocal');	Função utilizada para recuperar todos os termos que estão até a dois (2) níveis abaixo do “termo”, na hierarquia de lugares geográficos.
Ctx_thes.ntp(termo,2,' thes_GeoLocal');	Função utilizada para recuperar todos os termos que estão até a dois (2) níveis acima do “termo”, na hierarquia de lugares.
Ctx_thes.rt(termo,' thes_GeoLocal');	Função utilizada para retornar todos os RT (“*Related Terms*”) com o “termo”.

Tabela 5.2- Principais funções do *thesaurus*

5.4 O GeoLocalizador

No GeoLocalizador, o usuário pode acessar o sistema através de um *browser* e realizar suas consultas. O *Gazetteer* implementado no GeoLocalizador permite a recuperação de informações geográficas através do nome de um lugar geográfico. Com a definição de um lugar geográfico, os usuários podem realizar consultas com suporte a operações espaciais e temporais sobre este lugar. Da mesma forma, o GeoLocalizador apresenta aos usuários a posição, em um mapa mundi, do lugar que está sendo pesquisado, bem como a possibilidade de expansão dos termos da consulta, através dos relacionamentos de um *thesaurus* geográfico.

Nesta seção, particularizamos alguns detalhes da implementação e funcionamento do GeoLocalizador. Detalhamos a interface desenvolvida, apresentando de uma maneira geral, as principais características de seu funcionamento. As janelas de submissão das consultas, com a apresentação dos resultados e todas as etapas que envolvem o processo de uma consulta são descritas.

5.4.1 Interface do Usuário

A interface inicial do GeoLocalizador é acessada através de um *browser*. Conforme descrito no capítulo 3, existem problemas em relação ao modo em que alguns *Gazetteers* e bibliotecas georeferencidas disponíveis na *Web* apresentam as informações aos usuários. Em virtude desses problemas, procuramos projetar uma interface gráfica mais amigável, baseada no projeto ADL, com a divisão da interface em páginas (frames). Primeiramente, precisamos destacar que a interface do sistema é dividida em quatro páginas, cada qual apresenta uma etapa no processo de consulta. A seguir, detalhamos o conteúdo de cada página da interface:

1. Página_1: na primeira página, localizada no frame superior à esquerda, são visualizadas as duas formas de submissão de uma consulta;
2. Página_2: na segunda página, localizada no frame inferior à esquerda, são apresentados os operadores espaciais e temporais, bem como os componentes para definição de um período de tempo. O período de tempo é utilizado pelos operadores temporais.

3. Página_3: na terceira página, localizada no frame superior à direita, é apresentado um mapa mundi. Durante o processo de consulta, a posição do lugar geográfico que está sendo pesquisado é marcada no mapa, a fim de mostrar ao usuário a localização geográfica do termo pesquisado.
4. Página_4: na quarta página, localizada no frame inferior à direita, são apresentados os resultados da consulta. Também neste frame, o usuário pode realizar a expansão da consulta.

A interface inicial do GeoLocalizador, composta com as quatro páginas detalhadas anteriormente é apresentada na figura 5.2.

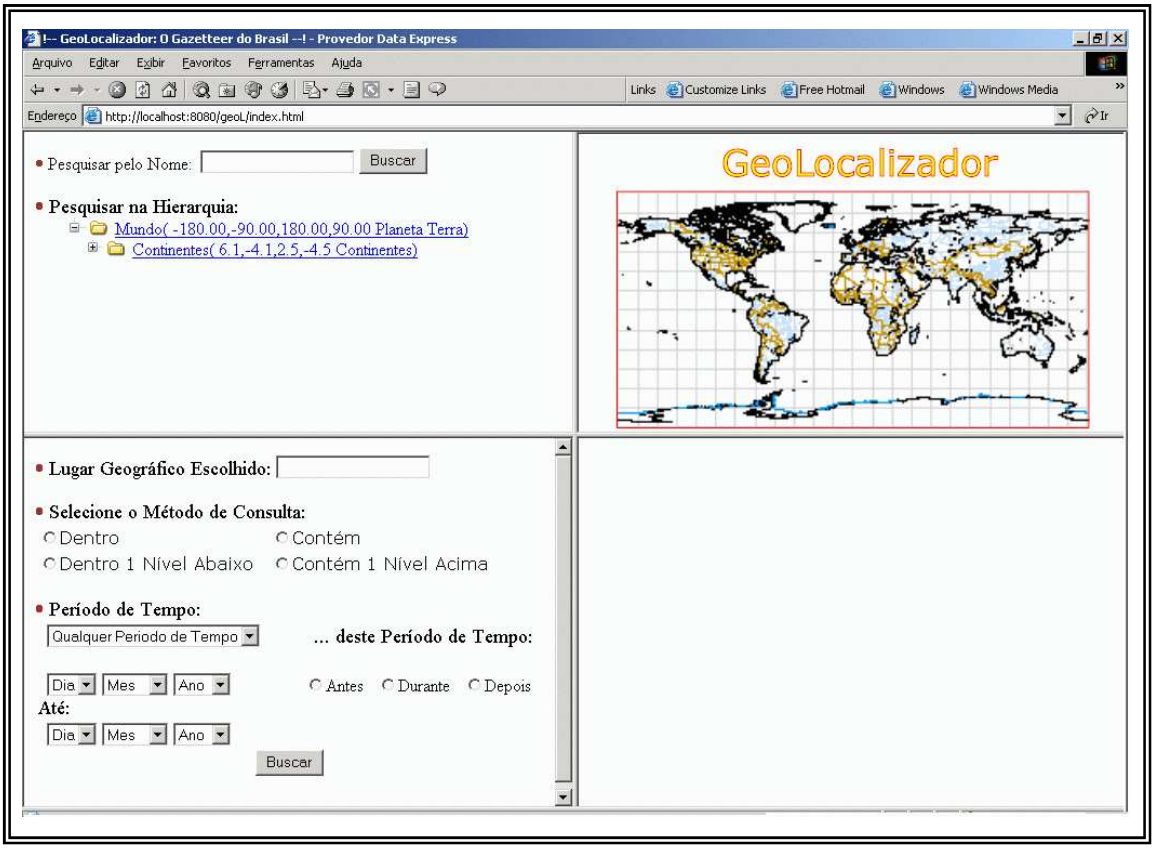


Figura 5.2- Interface do usuário

O conteúdo e o funcionamento de cada página apresentada na figura 5.2 são detalhados nas subseções a seguir.

5.4.2 Processo de Consulta

Conforme apresentado na seção 5.4.1, assim que o usuário acessa a interface do GeoLocalizador, são apresentadas duas formas de submeter uma consulta: a forma

direta, através da digitação do lugar geográfico a ser pesquisado em uma caixa de texto; ou através da navegação na árvore de lugares geográficos até encontrar o lugar a ser pesquisado.

Na forma direta, o usuário precisa apenas digitar o lugar a ser pesquisado em uma caixa de texto, no campo 'Pesquisar pelo Nome:'. Na figura 5.3, apresentamos um exemplo desta forma de submeter uma consulta. Neste exemplo, uma consulta sobre o lugar geográfico 'Brasil' é realizada.

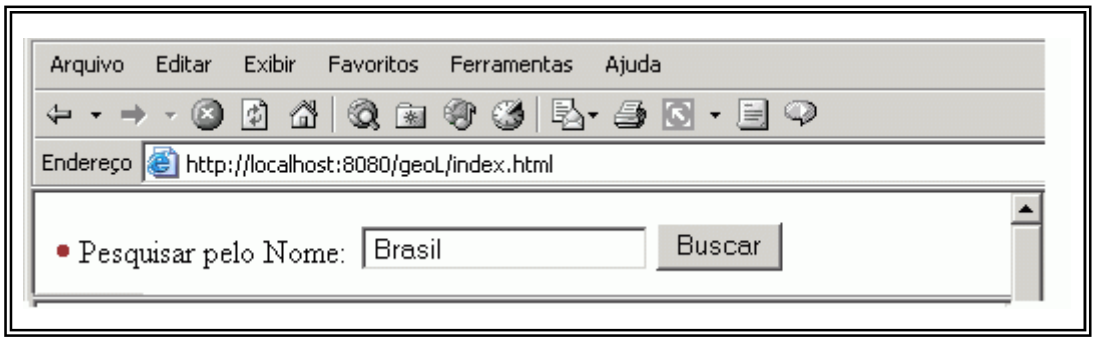


Figura 5.3- Interface com a forma de consulta direta

A forma de consulta apresentada na figura 5.3 é bastante comum, pois é utilizada principalmente nas ferramentas de buscas encontradas na Internet. Contudo, problemas como a digitação incorreta do lugar a ser pesquisado afeta o resultado da consulta, muitas vezes, são recuperadas informações irrelevantes com relação ao termo pesquisado.

Na segunda forma de consulta, com a utilização da árvore com a hierarquia de lugares geográficos, o usuário pode navegar na árvore e localizar o lugar a ser pesquisado. Um exemplo desta forma de consulta é apresentado na figura 5.4.

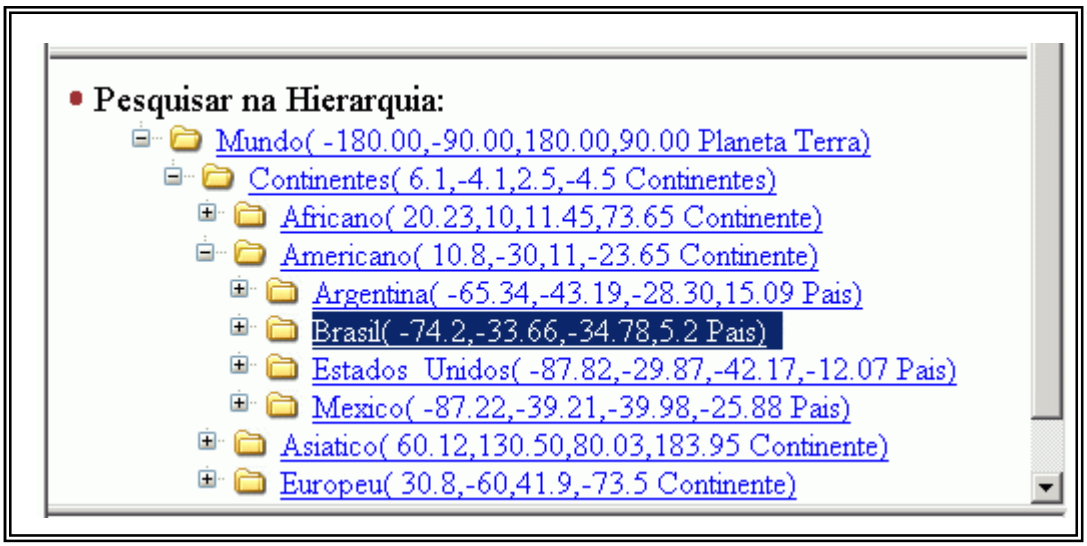


Figura 5.4- Interface de consulta com a árvore geográfica

No exemplo da figura 5.4, percebemos que o usuário navegou na árvore até encontrar o lugar geográfico ‘Brasil’. A apresentação dos lugares geográficos com o detalhamento dos níveis hierárquicos auxilia o usuário identificar rapidamente a posição do item de pesquisa na árvore geográfica. Da mesma forma, outras informações como o MBR e o Tipo do lugar são apresentados. Por exemplo, na figura 5.4 o usuário selecionou o lugar (Brasil), com o MBR (-74.2,-33.66,-34.78,5.2) e com o Tipo (País).

Conforme descrito na seção 5.1, os dados que formam a árvore geográfica, apresentada na figura 5.4, são recuperadas da base de dados. Muitas vezes, informações sobre um lugar geográfico podem estar registradas várias vezes na base de dados, com diferentes períodos de tempo. Este fato ocorre, devido às mudanças geopolíticas que originam, geralmente, novos lugares geográficos. Por exemplo, o estado de Goiás pode estar registrado duas vezes na base de dados: um registro com as informações antes da criação do estado de Tocantins; e outro registro com informações a partir da criação de Tocantins. Precisamos destacar, que o estado de Tocantins foi criado em 1988 e gerou a divisão da área geográfica do estado de Goiás.

Em casos onde existem diversas informações sobre um lugar, regras são aplicadas na organização da árvore geográfica. Por exemplo, a árvore registra apenas uma vez um lugar geográfico na hierarquia, com as informações mais recentes deste lugar. Neste sentido, o estado de Goiás pode estar registrado duas vezes na base de dados, com informações anteriores e posteriores a criação do estado de Tocantins. Porém, na árvore geográfica serão apresentadas apenas as informações (especificamente o MBR) mais recentes sobre Goiás, ou seja, informações de Goiás após a criação de Tocantins.

Em uma consulta, assim que o usuário digita um lugar na caixa de texto (campo ‘Pesquisar pelo Nome:’) ou seleciona um lugar na árvore com dados geográficos (campo ‘Pesquisar na Hierarquia:’), que está registrado mais do que uma vez na base de dados, uma nova página será apresentada ao usuário. Nesta página, o usuário deverá selecionar o lugar geográfico baseando-se no Tipo e no ano de validade das informações. Por exemplo, suponha que o usuário seleciona o estado de Goiás na árvore com dados geográficos. Neste caso, uma nova página será apresentada (parte inferior da figura 5.5) solicitando que o usuário selecione o lugar geográfico que deseja realizar a consulta. Na figura 5.5, ilustramos um exemplo com a seleção do lugar geográfico ‘Goiás’.

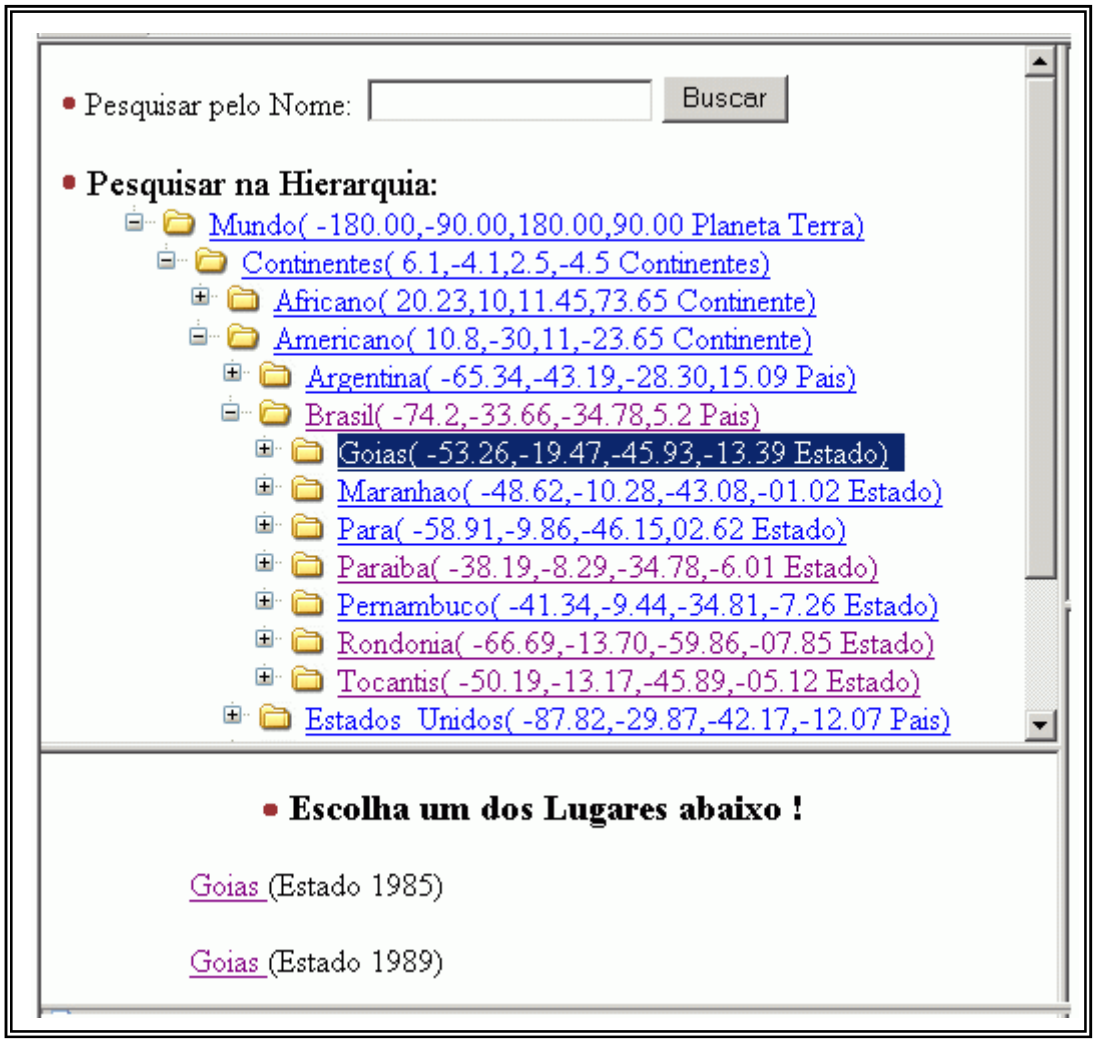


Figura 5.5- Interface de seleção de lugares específicos

Assim que o usuário define o termo a ser pesquisado, conforme uma das formas de consulta apresentadas nas figuras 5.3 e 5.4 (5.5), a interface para a seleção dos operadores espacial e temporal é habilitada. Aspectos desta interface são detalhados na subseção 5.4.3.

5.4.3 Operações Espaciais e Temporais

Com a determinação do termo a ser pesquisado, o usuário pode definir os operadores espaciais e temporais que serão utilizados na formulação da consulta.

No GeoLocalizador as operações espaciais são executadas com a utilização do MBR de cada lugar geográfico. Usando o MBR podemos realizar diversas consultas, tais como: ‘encontrar todos os estados (ou cidades) que estão dentro da área geográfica coberta pelo MBR de um determinado lugar geográfico’.

As relações espaciais no GeoLocalizador são implementadas através de operadores espaciais do *Spatial Cartridge* do SGBD *Oracle8i* e com os relacionamentos de um *thesaurus* geográfico. A seguir, apresentamos a definição das quatro operações espaciais implementadas no GeoLocalizador:

- Dentro: esta operação espacial recupera todos os lugares geográficos que estão dentro da área coberta pelo MBR do lugar submetido inicialmente na consulta. Esta operação utiliza os operadores espaciais "*CoveredBy*" e "*Inside*" do *Oracle8i*;
- Contém: esta operação recupera todos os lugares cujo MBR cobre a região geográfica do lugar submetido na consulta. Esta operação utiliza os operadores espaciais "*Covers*" e "*Contains*" do *Oracle8i*;
- Dentro 1 Nível Abaixo: esta operação é similar a operação "Dentro", porém recupera apenas lugares geográficos que estão um nível abaixo na hierarquia de lugares, utilizando o Tipo de cada lugar geográfico. Por exemplo, se o lugar submetido na consulta é um país, esta operação deve ser executada obedecendo duas condições: 1) recuperar apenas os lugares que as áreas geográficas estão dentro do MBR do lugar que está sendo pesquisado; e 2) selecionar apenas os lugares cujo Tipo seja estado, não recuperando lugares geográficos que estão dentro do MBR sendo pesquisado, mas que são do Tipo: município, cidade, bairro, distrito, etc. Esta operação utiliza o operador BT ("*Broader Term*") do *thesaurus* geográfico;
- Contém 1 Nível Acima: esta operação possui funcionamento similar com a operação "Dentro 1 Nível Abaixo", porém recupera apenas os lugares que estão 'Contidos' um nível acima na hierarquia de lugares geográficos. Por exemplo, em uma pesquisa realizada sobre um município está operação recupera apenas os lugares cujo MBR cobre o MBR deste município e que o Tipo seja estado. Esta operação utiliza o operador NT ("*Narrower Term*") do *thesaurus* geográfico.

Do mesmo modo que os operadores espaciais, as operações temporais também auxiliam no refinamento das consultas. No entanto, é preciso que o usuário especifique um período de tempo, para que os operadores possam ser utilizados.

A definição de um período de tempo pode ser formulada através de duas formas: com um intervalo de tempo, que é estipulado com a definição de duas datas, uma data inicial e a outra indicando a data final do intervalo; e um instante de tempo, definido por uma data específica.

No GeoLocalizador, a definição de um instante de tempo é feita com a especificação de uma simples data, tal como: 23-03-2002. No entanto, um intervalo de tempo pode ser estipulado através de quatro maneiras:

1. selecionar um ano, por exemplo: 1990. Neste caso, o intervalo abrange todos os meses deste ano;
2. definir uma década, por exemplo: 1980d, a qual compreende as informações do ano de 1980 até o ano de 1989;
3. definir um século, por exemplo: século XIX;
4. definir duas datas compondo um intervalo, por exemplo: entre as datas de 26-02-1978 até 21-03-2002.

Após a definição de um período de tempo, o usuário pode selecionar a operação temporal que será aplicada na consulta. O GeoLocalizador, implementa três operadores: “Antes”; “Durante”; e “Depois”. Entretanto, caso nenhum período de tempo seja estipulado a seleção de um dos operadores temporais será invalidada, e a consulta será feita em qualquer período de tempo. Na figura 5.6, apresentamos um exemplo com a definição dos operadores espaciais e temporais em uma consulta.

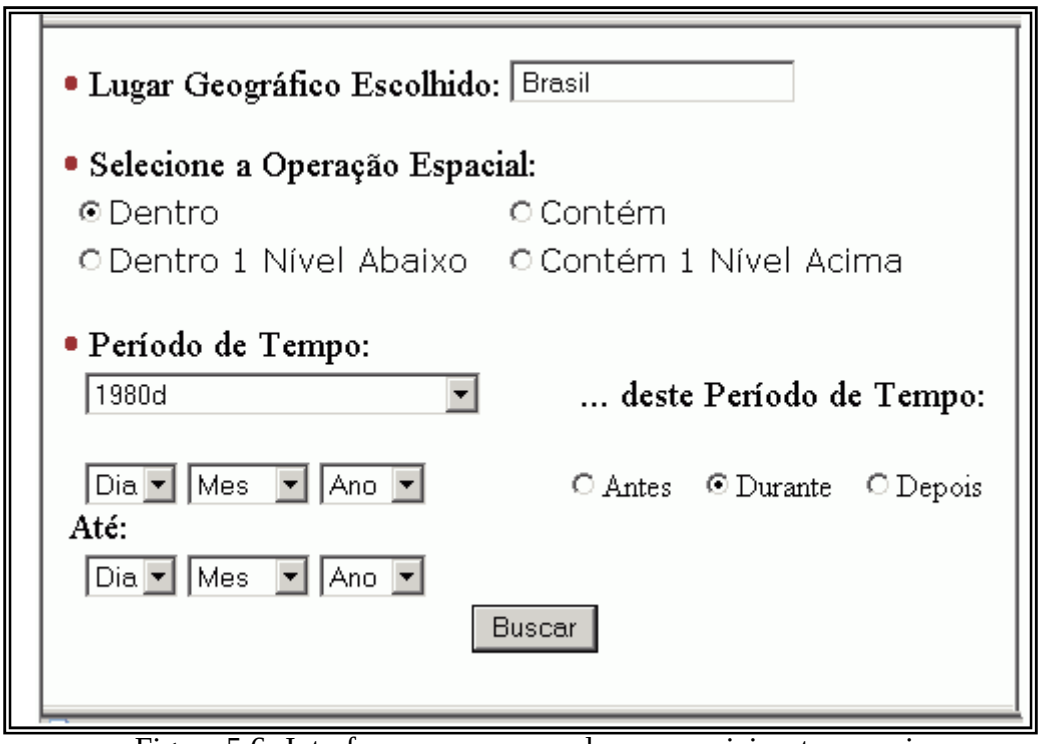


Figura 5.6- Interface com os operadores espaciais e temporais

A figura 5.6 ilustra o exemplo de uma consulta realizada sobre o lugar geográfico ‘Brasil’. Essa consulta foi formulada com as seguintes definições:

- seleção do operador espacial “Dentro”;
- definição do período de tempo, com o intervalo “década de 80”;
- seleção do operador temporal “Durante”.

As definições estipuladas na figura 5.6 implicam na seguinte consulta: “selecionar todos os lugares que estão dentro (‘Dentro’) da região geográfica coberta pelo MBR do país ‘Brasil’, cujo intervalo de tempo válido esteja ‘Durante’ a ‘década de 1980’”.

Na figura 5.6, o usuário seleciona um período de tempo. Essa seleção pode ser feita de diversas maneiras. No *combobox* ‘Período de Tempo:’, o usuário pode selecionar intervalos de tempo, tais como: determinados anos (por exemplo: 1990); décadas (por exemplo: 1990d); e séculos (por exemplo: XXI). Este *combobox*, também é utilizado para o usuário definir um intervalo de tempo, composto por duas datas ou através de um instante de tempo.

Para definir um intervalo de tempo com duas datas específicas (por exemplo: 27-09-2002 e 15-12-2002), o usuário precisa selecionar a opção “Definir Data” no *combobox* ‘Período de Tempo:’, e então utilizar os *combobox* ‘Dia’, ‘Mes’ e ‘Ano’ para a definição das duas datas. Da mesma forma, é realizado o processo para especificar um instante de tempo, porém, neste caso apenas uma data é especificada.

Com a apresentação das seções 5.4.2 e 5.4.3 encerramos o processo de formulação de uma consulta. Nas seções 5.4.4 e 5.4.5, detalhamos as interfaces responsáveis pela apresentação dos resultados desta consulta.

5.4.4 Mapa Geográfico

Conforme visto nas seções 5.4.2 e 5.4.3, com a escolha de um lugar geográfico e com a seleção dos operadores espacial e temporal, o usuário pode submeter a consulta. Assim que a consulta é submetida, um mapa geográfico é apresentado com a ilustração de um ponto indicando a localização geográfica do lugar submetido na

consulta. O mapa mundi com o ponto representando a posição de um lugar geográfico pode ser visualizado no exemplo da figura 5.7.

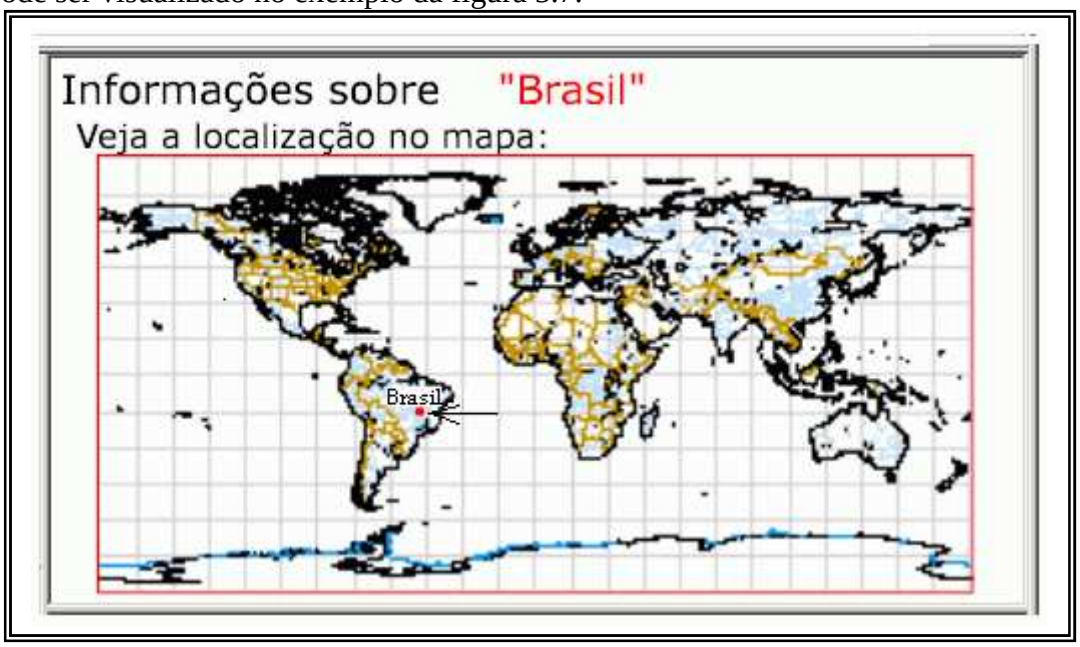


Figura 5.7- Interface de localização do lugar geográfico

Na figura 5.7, podemos perceber a existência de um ponto indicando localização do lugar geográfico ‘Brasil’. Esse ponto representa as coordenadas de latitude e longitude do Centróide do Brasil.

O mapa apresentado na figura 5.7 proporciona aos usuários um melhor esclarecimento das informações que estão sendo obtidas no processo de consulta. A indicação com a localização do lugar geográfico que está sendo pesquisado possibilita ao usuário fazer projeções com relação à distância e posicionamento com outras regiões do mundo. Na seção 5.4.5, abordamos a interface responsável pela apresentação dos resultados da consulta na forma textual, bem como o processo de expansão.

5.4.5 Apresentação dos resultados

Nesta seção, detalhamos a interface de apresentação dos resultados da consulta. A seguir, detalhamos as duas fases do processo de consulta.

Na primeira fase, são apresentados os lugares geográficos que satisfazem a consulta formulada pelo usuário. Esses resultados apresentados estão de acordo com o nome do lugar geográfico, período de tempo e as operações espaciais e temporais

definidas pelo usuário. A seguir, apresentamos o processo completo da primeira fase de uma consulta:

- primeiramente, definimos o nome do lugar geográfico a ser pesquisado, por exemplo: 'Brasil'. Esse processo é realizado conforme a descrição na seção 5.4.2;
- com a informação do lugar geográfico a ser pesquisado o usuário pode fazer a seleção de um operador espacial (por exemplo: Dentro), de um período de tempo (por exemplo: Década de 1980) e de um operador temporal (por exemplo: Durante), assim finalizando o processo de formulação da consulta. Essas operações são discutidas na seção 5.4.3;
- assim que a consulta é formulada e submetida, a localização do lugar geográfico sendo pesquisado é marcada com o ponto Centróide em um mapa mundi. Esse processo é ilustrado na seção 5.4.4;
- ao mesmo tempo em que a localização do lugar é marcada no mapa geográfico, os resultados da consulta contendo os lugares geográficos que satisfazem a condição estabelecida conforme o operador espacial, período de tempo e operador temporal são apresentados em uma nova página.

A interface com a apresentação dos resultados da primeira fase, conforme a consulta: "selecionar todos os lugares que estão dentro da região geográfica coberta pelo MBR do país Brasil, e que o intervalo de tempo válido destes lugares esteja Durante a década de 1980"; é apresentada na figura 5.8.

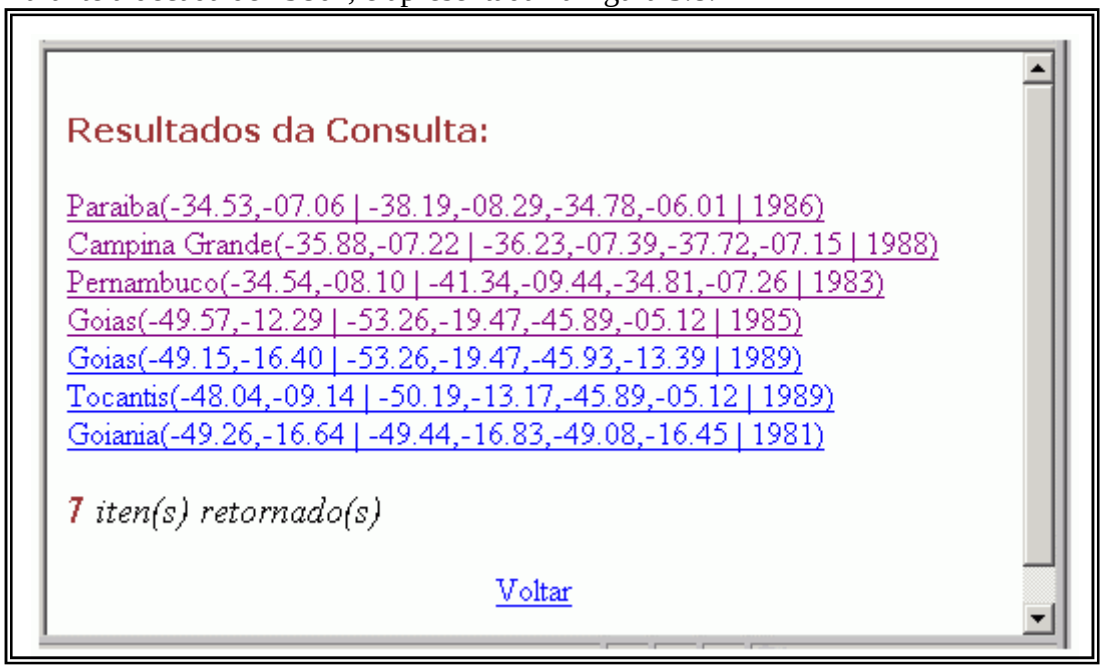


Figura 5.8- Interface com os resultados da consulta

Conforme a figura 5.8, podemos perceber que juntamente com o nome dos lugares geográficos retornados na consulta, também são apresentadas as coordenadas do ponto Centróide, do MBR e o período de tempo válido para estes lugares geográficos. Essas informações proporcionam um maior esclarecimento sobre os lugares geográficos retornados na consulta. A figura 5.9 apresenta a interface do sistema, com as quatro páginas envolvidas na primeira fase da consulta.

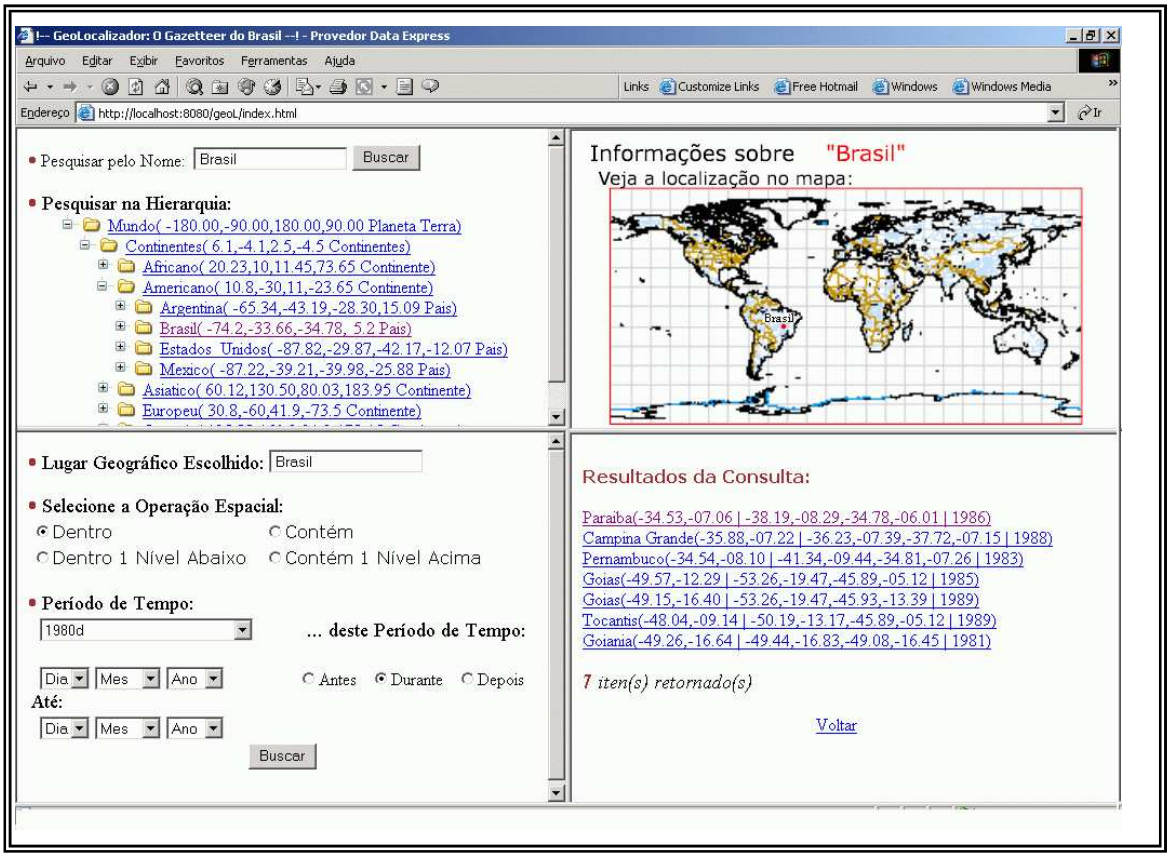


Figura 5.9- Interface com a primeira fase da consulta

A segunda fase do processo de consulta é realizada com a expansão de um dos lugares geográficos que foram retornados na primeira fase da consulta apresentada na figura 5.8. O processo de expansão da consulta é realizado através de um *thesaurus* geográfico implementado no SGBD *Oracle8i*. Alguns detalhes do processo de implementação do *thesaurus* geográfico são discutidos na seção 5.3.

Para expandir a consulta, basta o usuário selecionar um dos lugares geográficos que foi retornado na primeira fase da consulta (conforme apresentado na figura 5.8). Neste exemplo, selecionamos o lugar geográfico ‘Paraíba’. Assim que este lugar é selecionado, uma nova marcação com a localização do ponto Centróide do estado da ‘Paraíba’ é ilustrada no mapa mundi. Ao mesmo tempo em que um mapa geográfico é apresentado ao usuário, o processo de expansão é processado pelo

sistema buscando diversas informações que estão relacionadas com o lugar geográfico expandido. A figura 5.10 apresenta o resultado da expansão do lugar geográfico ‘Paraíba’.

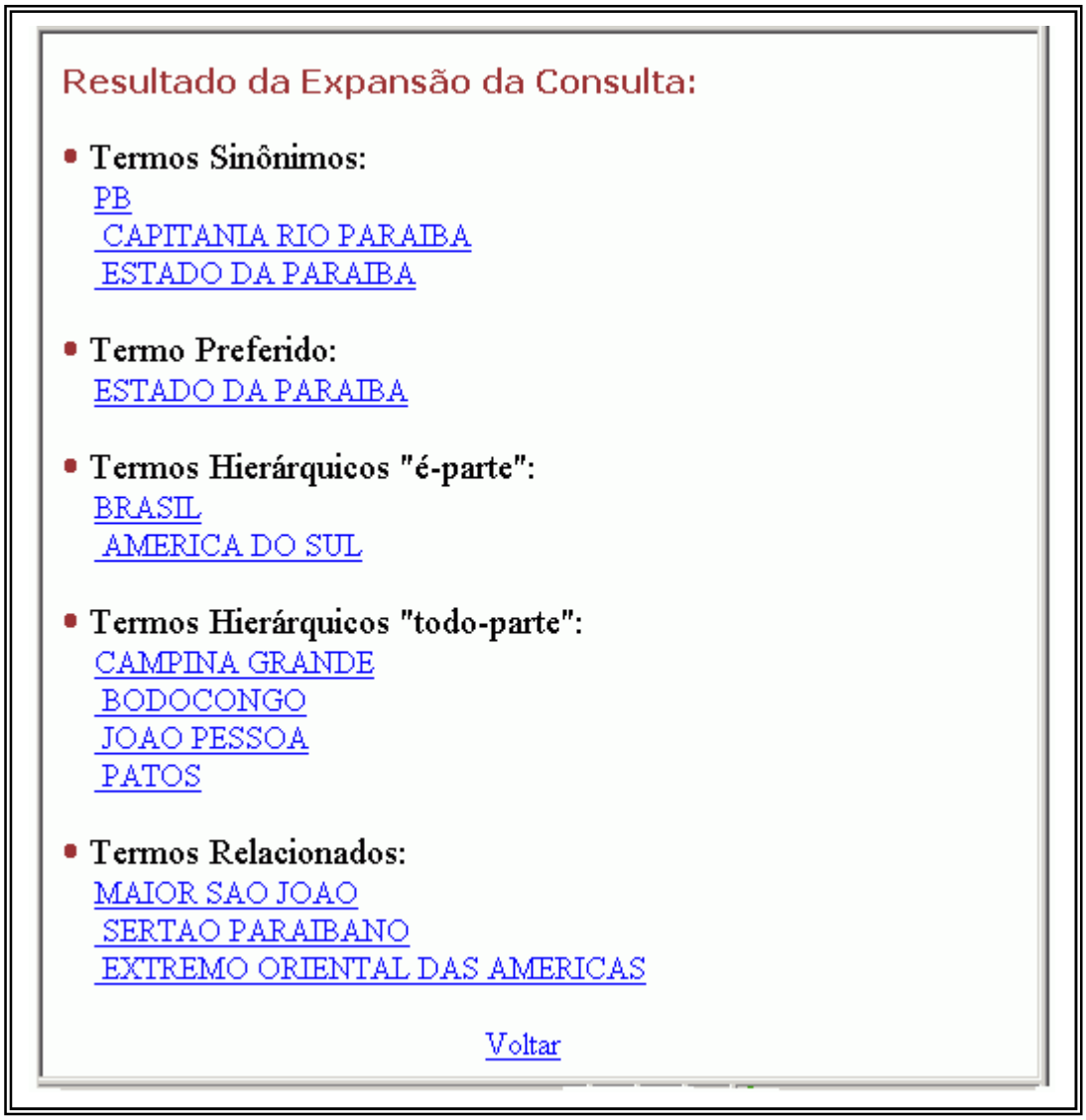


Figura 5.10- Interface com a expansão da consulta

A expansão da consulta é feita com a utilização do *thesaurus* geográfico, possibilitando a seleção de termos sinônimos, formação de hierarquias e outros relacionamentos, tais como, evidenciar que o termo gaúcho está relacionado com o estado do Rio Grande do Sul, ou que a floresta amazônica está relacionada com o Brasil.

Conforme apresentado no exemplo da figura 5.10, o *thesaurus* encontrou uma série de relacionamentos com o termo geográfico ‘Paraíba’. O relacionamento de equivalência representa todos os termos que são sinônimos. Entre esses termos, existe

o Termo Preferido (*Preferred Term*), que indica qual dos termos é o mais atual ou o termo oficial para o lugar geográfico.

O relacionamento hierárquico estabelece as relações “todo-parte” e “é-parte”, que denota a posição do termo na hierarquia. Em uma hierarquia geográfica, os termos “todo-parte” representam todos os termos que estão em níveis abaixo do termo expandido. Os termos “é-parte” representam os termos que estão em níveis acima, do termo expandido, na hierarquia.

O relacionamento associativo representa todos os termos que estão semanticamente ou conceitualmente relacionados com o termo expandido. No exemplo da figura 5.10, percebemos que informações como ‘o maior São João do mundo’, ‘sertão paraibano’ e ‘extremo oriental das Américas’ estão relacionadas com o lugar geográfico Paraíba.

Na figura 5.11, apresentamos a interface completa do sistema GeoLocalizador, após o processo de expansão da consulta.

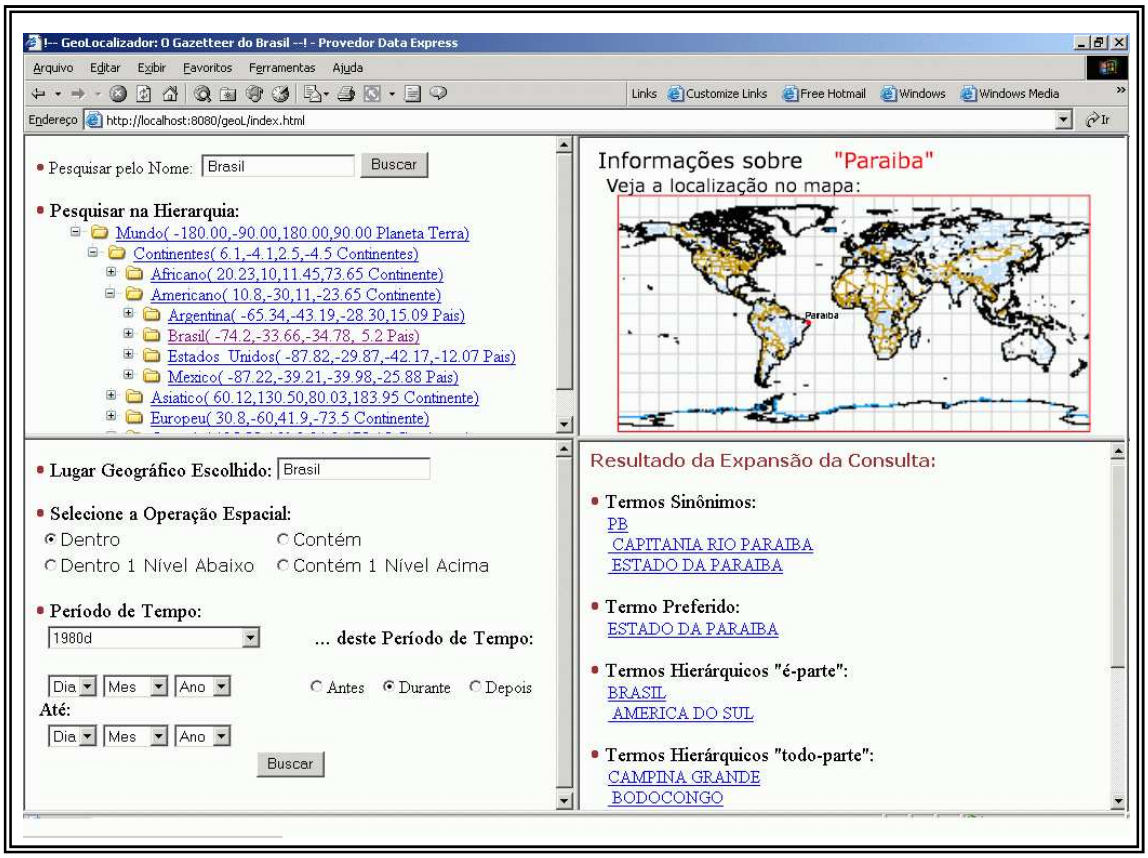


Figura 5.11- Interface com a consulta expandida

Na figura 5.11, percebemos que os termos encontrados com a expansão da consulta são também *links* com URLs para páginas *Web*. Esses *links* estão armazenados dentro do banco de dados, e apontam para páginas da Internet, a partir

das quais são encontradas informações individuais de cada item selecionado pelo usuário. Por exemplo, na figura 5.11 o termo ‘Estado da Paraíba’ é sinônimo e termo preferido em relação ao termo expandido (Paraíba). O *link* desse termo aponta para a página, cuja URL é http://www.pm.pb.gov.br/historiadaparaiba.html, onde são encontradas diversas informações sobre a história política, turismo e as características do estado da Paraíba.

5.5 Considerações Finais

Neste capítulo apresentamos os detalhes da implementação e funcionamento do GeoLocalizador. Conforme apresentado no capítulo 3, os sistemas disponíveis na *Web*, tais como os *Gazetteers*, bibliotecas georeferenciadas e ferramentas de busca que realizam pesquisas sobre dados geográficos apresentam problemas. Neste trabalho alguns requisitos foram elencados para procurar suprir esses problemas (seção 4.2). A seguir, apresentamos quais requisitos funcionais foram atendidos:

- Disponibilizar o sistema de consultas na *Web* (RF1): o sistema está disponível na *Web*, podendo ser acessado através de um *browser*. Esse requisito foi atendido conforme apresentado na seção 5.4.1.
- Possibilitar diferentes formas de submeter às consultas (RF2): no GeoLocalizador o usuário pode submeter uma consulta através de duas formas: (1) consulta direta, com a digitação do nome do lugar geográfico em uma caixa de texto e (2) navegação e seleção de um lugar em uma árvore com dados geográficos. Esse requisito funcional foi atendido na seção 5.4.2.
- Disponibilizar a utilização de operadores espaciais e temporais (RF3): na formulação de uma consulta, o usuário pode utilizar operações espaciais e temporais sobre os dados geográficos. Esse requisito foi atendido conforme apresentado na seção 5.4.3.
- Definir a estrutura da hierarquia de lugares geográficos (RF4): a estrutura da árvore com dados geográficos segue uma hierarquia com áreas políticas permitindo os relacionamentos “é-parte” e “parte-todo”. Esse requisito foi atendido conforme detalhado na seção 5.4.2.

- Expandir a consulta (RF5): o usuário pode expandir a consulta, a fim de encontrar informações mais específicas de determinado item. Esse requisito foi atendido conforme apresentado na seção 5.4.5.
- Definir relacionamentos entre os termos (RF6): os relacionamentos entre os termos da consulta são definidos em um thesaurus geográfico. Esse requisito foi atendido na seção 5.4.5.
- Disponibilizar graficamente os resultados das consultas na *Web* (RF8): os resultados da consulta são disponibilizados graficamente na *Web*. Esse requisito foi atendido conforme apresentado nas seções 5.4.4 e 5.4.5.

Capítulo 6

Conclusão

Os sistemas de informação espaço-temporais estão ganhando cada vez mais importância em órgãos governamentais e não governamentais. Os *Gazetteers* têm papel importante em tais sistemas, pois se prestam como uma ferramenta de apoio que auxilia os usuários, que não sejam especialistas na área de geoprocessamento, ou que não possuam conhecimento sobre as coordenadas de um lugar geográfico.

O *Gazetteer* desenvolvido nesta dissertação procura suprir as principais deficiências que são encontradas em outros *Gazetteers* disponíveis na *Web*. A implementação de algumas funcionalidades, dentre elas o tratamento da dimensão temporal, representação da árvore com dados geográficos e suporte às operações espaciais e temporais, habilita o GeoLocalizador a realizar consultas através da referência indireta a dados espaço-temporais.

O processo de expansão da consulta, com a utilização de um *thesaurus* geográfico, implementado no GeoLocalizador, possibilita a correlação dos termos que foram adicionados ao *thesaurus*, com o item que está sendo pesquisado. Essa técnica de recuperação de informação permite uma maior cobertura nos resultados que são apresentados para os usuários.

6.1 Principais Contribuições

A implementação de algumas características diferenciam o GeoLocalizador dos outros *Gazetteers* disponíveis na *Web*. A seguir, descrevemos as suas principais contribuições:

- Diferentes formas de submeter uma consulta: as duas formas distintas de submeter uma consulta no GeoLocalizador permitem uma maior flexibilidade no processo de formulação desta. A apresentação de uma árvore com dados geográficos, onde o usuário pode navegar e encontrar

um lugar a ser pesquisado, proporciona aos usuários um novo método de submissão de consultas. Ao mesmo tempo em que este método é inovador nos sistemas de consulta na *Web*, a navegação na árvore oferece também um maior aprendizado, pois permite que as informações sejam apresentadas em níveis, com a clara separação entre os lugares que representam continentes, países, estados e municípios;

- Suporte à temporalidade dos dados: com a informação sobre a temporalidade dos dados é possível recuperar diversas informações sobre um mesmo lugar geográfico, em diferentes períodos de tempo. Com o passar do tempo, novas fronteiras são formadas em decorrência de mudanças geopolíticas, como a queda do muro de Berlim, na Alemanha, em 1989, que resultou na união da Alemanha Ocidental, bem como a divisão da União Soviética, em 1991, originando novos países, como a Rússia e a Ucrânia. Estes acontecimentos mudaram a geopolítica mundial e novas fronteiras entre os países foram formadas em épocas distintas;

- Utilização de operadores espaciais e temporais: a utilização de operadores espaciais e temporais permite a execução de processos de refinamento das consultas. Com a utilização de operadores espaciais, podemos executar operações com o envolvimento dos MBRs dos lugares geográficos. Diante disso, podemos recuperar informações que estão relacionadas apenas a determinadas regiões geográficas, da mesma forma que informações sobre certas regiões podem ser descartadas. Os operadores temporais permitem a execução de consultas sobre determinados períodos de tempo. Por exemplo, apenas eventos ou acontecimentos que ocorreram ‘antes’, ‘durante’ ou ‘depois’ de determinado período de tempo podem ser recuperados;

- *Thesaurus* geográfico: a utilização de um *thesaurus* geográfico permite a expansão da consulta e, conseqüentemente, torna-se possível obter uma classificação para as informações recuperadas, mais apurada quanto à relevância destas informações. A definição de relacionamentos entre os termos do *thesaurus* e o termo expandido da consulta proporciona uma

série de informações que estão semanticamente ou conceitualmente relacionadas.

6.2 Trabalhos Futuros

As pesquisas sobre os *Gazetteers* e novas funcionalidades que podem ser implementadas em tais sistemas dão margem a uma série de trabalhos. A partir dos resultados obtidos com o desenvolvimento do GeoLocalizador, pode-se identificar algumas questões importantes para o aperfeiçoamento e extensão deste trabalho. A seguir, apresentamos algumas funcionalidades que podem ser adicionadas ao GeoLocalizador em trabalhos futuros:

- Tratamento da interoperabilidade: atualmente, os *Gazetteers* enfrentam problemas em relação à interoperabilidade para o compartilhamento das informações. Vários estudos estão sendo realizados sobre o protocolo do projeto ADL [Janée, 2002] para padronização de um formato para a troca de serviços entre *Gazetteers*. A estrutura do GeoLocalizador possui pequenas diferenças em relação ao protocolo que está sendo proposto pelo projeto ADL. Dessa forma, uma pequena adaptação será necessária para a integração do GeoLocalizador com outros *Gazetteers*;

- Adição de novos *thesauri*: o processo de expansão da consulta do GeoLocalizador é realizado com utilização de um *thesaurus* geográfico. A estrutura do GeoLocalizador permite o fácil acoplamento de novos *thesauri*. Com isto, *thesauri* com diferentes domínios podem ser utilizados, aumentando o número de itens relevantes que são retornados no processo de uma consulta;

- Novos operadores espaciais: adição de novos operadores proporcionam diferentes operações espaciais. Essa funcionalidade permite maior capacidade de refinamento das consultas no GeoLocalizador. Por exemplo, a implementação de um operador espacial pode permitir recuperar todos os lugares geográficos que possuem, ao menos, um ponto de suas fronteiras

(norte, sul, leste, oeste) em comum, ou seja, que são vizinhos, tais como: Paraíba e Pernambuco;

- Automatizar o processo de carga de dados: até a presente implementação, a carga de dados no GeoLocalizador está sendo feita de forma manual. Trabalhos futuros podem ser direcionados no sentido de automatizar o processo de inserção de dados no GeoLocalizador;

- Interface gráfica: a interface gráfica do GeoLocalizador pode ser melhorada com a utilização de ferramentas que permitem *zoom in*, *zoom out* e *pan* sobre o mapa geográfico. A implementação destas ferramentas possibilitará uma melhora no processo de visualização com aproximação, através de *zoom*, de pequenas regiões geográficas;

- Implementação de técnicas de seleção espacial: técnicas de seleção espacial como esboço de uma região sobre mapas geográficos podem ser implementadas, para melhorar o processo de consulta. As seguintes técnicas podem ser utilizadas: MBR, Centróide, e Fecho Convexo.

6.3 Considerações Finais

Na medida em que aumenta o volume de informações na Internet, as ferramentas de busca precisam ser incrementadas, e novas técnicas e métodos de recuperação devem ser implementados nos SRIs. Com o crescimento do uso de aparelhos de georefereciamento, como os GPSs, começa a surgir uma grande demanda de serviços espaciais com a finalidade de oferecer aos usuários informações sobre posicionamento geográfico.

Da mesma forma, prefeituras e governos estaduais têm investido pesado em sistemas de informações geográficas, com a finalidade de divulgar e impulsionar o turismo em diferentes regiões. Um *Gazetteer* pode ser utilizado para prover a localização espacial destes lugares geográficos.

Nos últimos anos, as pesquisas sobre *Gazetteers* vêm aumentando de forma surpreendente. Primeiramente, pesquisas e projetos de desenvolvimento de *Gazetteer*

concentravam-se principalmente nos Estados Unidos e no Canadá. Atualmente, vários países da Europa e da Ásia, estão desenvolvendo grandes projetos voltados à construção de *Web Gazetteers*.

Todo o interesse sobre o desenvolvimento de *Gazetteers* foi evidenciado no JCDL'2002 (*Joint Conference on Digital Libraries*), no *Workshop* sobre *Gazetteers* Digitais, ocorrido em julho de 2002, em *Portland*, *Oregon*, Estados Unidos. Nesta conferência, foram apresentados projetos e pesquisas sobre *Gazetteers* em diversos países, tais como: Estados Unidos, Canadá, Alemanha, Reino Unido e China.

Na Europa, projetos envolvendo vários países, com altos investimentos, estão sendo desenvolvidos. Um exemplo é o projeto SPIRIT[40] (*Spatially-Aware Information Retrieval on the Internet*), que envolve pesquisadores de grandes universidades espalhadas em diversos países, tais como: Suíça, França, Inglaterra, Alemanha e Holanda. A finalidade deste projeto é a construção de uma ferramenta de busca capaz de reconhecer dados geográficos. Enfim, nos próximos anos, deve ocorrer uma proliferação de *Gazetteers* na Internet, com informações de várias partes do planeta.

[40] http://www.cs.cf.ac.uk/department/posts/SPIRITSummary.pdf

Referências Bibliográficas

[Abiteboul et al, 1999] Abiteboul, S.; Suciu, D.; Buneman, P.: "*Data on the web: From Relations to Semistructured Data and XML*". Morgan Kaufmann Publishers, 1999.

[Agosti et al, 1996] Agosti, M.; Smeaton, A.: "*Information Retrieval and Hypertext*". Kluwer Academic Publishers, 1996.

[Aitchison et al, 1987] Aitchison, J.; Gilchrist, A.: "*Thesaurus Construction: a practical manual*", Aslib London, 1987.

[Alani, 2001] Alani, H: "Spatial and Thematic Ontology in Cultural Heritage Information Systems". Tese de Doutorado, University of Glamorgan/Prifysgol Morgannwg, Maio 2001.

[Allen, 1983] Allen, J.: "Maintaining knowledge about Temporal Intervals". *Comunications of ACM*, 26(11), pp. 832-843, (1983).

[Arms, 1999] Arms, W.: "*Digital Libraries*". MIT Press, 1999.

[Aronoff, 1989] Aronoff, S.: "*Geographic Information System a Management Perspective*". Ottawa: WDL Publications, 1989.

[Austen et al, 1995] Austen, M.; Dunstan, V.; Randell, B.; Stanier, A.; Stringer, P.; Woodgate, J.: "An Information Service for United Kingdom & Ireland Genealogy based on the Internet's World Wide Web". *Computers in Genealogy*, September (1995).

[Baeza-Yates et al, 1999] Baeza-Yates, R.; Ribeiro-Neto, B.: "*Modern Information Retrieval*". Addison Wesley Longman, Wokingham, UK., 1999.

[Baptista, 2000] Baptista, S., C.: "STepLib: A Digital Library for Spatial-Temporal and Multimedia Data". Tese de Doutorado, University Kent at Canterbury; Setembro 2000.

[Baptista et al, 2000] Baptista, S., C.; Kemp, Z.: "An Integrated Metamodel for Knowledge Representation in Geolibraries". In *Proceedings of the 7th IEEE Conference on Advances in Digital Libraries*, IEEE Computer Society Press, Bethesda, MD, USA, pp. 151-159, May (2000).

[Bezerra, 2000] Bezerra, P. E.; "Representação e Manipulação de Objetos em Banco de Dados Temporais Indeterminados". Dissertação de Doutorado, UFPB, Agosto, 2000.

[Bogorny et al, 2001] Bogorny, V.; Iochpe, C.: "Estendendo o Modelo Relacional do OpenGIS para Garantir Restrições Espaciais Topológicas". *XVI Simpósio Brasileiro de Banco de Dados SBBD*, pp. 25-39, (2001).

[Booch, 1999] Booch, G.: "*The Unified Modeling Language User Guide*". Addison-Wesley, 1999.

[Bray et al, 2000] Bray, T.; Paoli, J.; Sperberg-McQueen, M., C.; Maler E.: "Extensible Markup Language (XML) 1.0 (Second Edition)". W3C Recommendation, Outubro, 2000. http://www.w3.org/TR/REC-xml. Acessado em novembro de 2002.

[Câmara, 1996] Camara, G.: "Anatomia de Sistemas de Informação Geográfica". *IX Anais do SIBGRAPI'96* (1996).

[Clementini, 1993] Clementini, E.: "A Small Set of Formal Topological Relationships for End-User Interaction". *Third International Symposion, SSD*, Singapore, (1993).

[Cohen, 1998] Cohen, S., B.: "*The Columbia Gazetteer of the World*". Columbia University Press, 1998.

[Cox et al, 2001] Cox, S.; Cuthbert, A.; Lake, R.; Martell, R.: "Geography Markup Language (GML) 2.0 (Second Edition)". OpenGIS® Implementation Specification, OGC: n. 01-029, February 2001. Disponível em http://www.opengis.net/gml/01-029/GML2.html**.** Acessado em outubro de 2002.

[Crane, 1996] Crane, G.: "Building a digital library: The Perseus Project as a case study in the humanities". *Proceedings of the First ACM International Conference on Digital Libraries*, pp. 3-10 (1996).

[Crane, 1998] Crane, G.: "The Perseus Project and Beyond How Building a Digital Library Challenges the Humanities and Technology". Technical Report, *D-Lib Magazine*, January (1998).

[Edwards, 1999] Edwards, J.: "*3-Tier Server/Client at Work*". John Wiley & Sons, 1999.

[Egenhofer, 1991] Egenhofer, M.: "A Framework for the Definition of Topological Relationships and Algebraic Approach to Spatial Reasoning Within this Framework". *NCGIA*, (1991).

[Eisenberg et al, 2000] Eisenberg, A.; Melton, J.: "SQL Standardization: The Next Steps". ACM Sigmod, 2000. Disponível em: http://www.acm.org/sigmod/record/issues/0003/standards.pdf. Acessado em novembro de 2002.

[Fellbaum, 1993] Fellbaum, C.: "*English Verbs as a Semantic Net*", Technical Report, Agosto (1993).

[Ferraiolo, 2001] Ferraiolo, J. (ed.): "Scalable Vector Graphics (SVG) 1.0 Specification". W3C Recommendation, Setembro, 2001. Disponível em: http://www.w3.org/TR/SVG. Acessado em setembro de 2002.

[Fink, 1999] Fink, E.: "The Getty Information Institute A Retrospective". *D-Lib Magazine*, v.5, March (1999).

[Frew et al, 1998] Frew, J.; Freeston, M.; Freitas, N.; Hill, L.; Janee, G.; Lovette, K.; Nideffer, R.; Smith, T.; Zheng, Q.: "The Alexandria Digital Library architecture". *Proceedings of the Second European Conference on Research and Advanced Technology for Digital Libraries (ECDL'98)*, pp. 61-73. Heraklion, Crete, Greece, September (1998).

[Gamma et al, 1995] Gamma, E.; Helm, R.; Johnson, R.; Vlissides, J.: "*Design Patterns, Elements of Reusable Object-Oriented Software*". Addison-Wesley, 1995.

[GEOnet, 2002] "GEOnet Names Server". Disponível em: http://www.nima.mil/gns/html. Acessado em outubro de 2002.

[Getty, 1999] "*Guidelines for forming language equivalents: a model based on the art & architecture thesaurus*". International Terminology Working Group - Getty Information Institute (1999).

[Getty, 2002] "Getty Thesaurus of Geographic Names". Disponível em: http://www.getty.edu/research/tools/vocabulary/tgn. Acessado em novembro de 2002.

[GNIS, 1998] U.S.Geological Survey (1998). Geographic Names Information System (GNIS). Disponível em: http://geonames.usgs.gov/gnishome.html. Acessado em novembro de 2002.

[Gonçalves et al, 1997] Gonçalves, A., M.; Medeiros, B., M., C.: "Bibliotecas Digitais para Dados Geográficos". *GISBRASIL 97 – III Congresso e Feira para Usuários de Geoprocessamento*, Maio (1997).

[Goodchild, 1999] Goodchild, M.: "*Distributed Geolibraries: Spatial Information Resources*". National Academy Press, USA, 1999.

[Goodman, 2000] Goodman, A., B.: "*GeoRef Thesaurus 9th edition*". American Geological Institute, 2000.

[Gosling et al, 2000] Gosling, J.; Joy, B.; Steele, G.; Bracha, G.: "*Java(TM) Language Specification (2nd Edition)*", Addison-Wesley Pub Co, 2000.

[Harpring, 1997] Harpring, P.: "The Limits of the World: Theoretical and Practical Issues in the Construction of the Getty Thesaurus of Geographic Names". *In Proceedings of the 4th International Conference on Hypermedia and Interactivity in Museums, ICHIM' 97*, Le Louvre, Paris, France, pp. 237-251, (1997).

[Hicks, 2001] Hicks, G.: "GENUKI". *Cardiganshire FHS's journal*, 2(7), February (2001).

[Hill, 1998] Hill, L.: "Building Georeferenced Collections Gazetteer Services". *Taxonomy Authority file Workshop*, Washington, DC; June (1998).

[Hill, 1999] Hill, L.: "Gazetteer and collection-level metadata developments". *In R. T. Kaser & V. C. Kaser (Eds.), Metadiversity*. The Grand Challenge for Biodiversity Information Management through Metadata. *Proceedings of the Symposium sponsored by the U.S. Geological Survey Biological Resources Div. & the National Federation of Abstracting & Information Services (NFAIS)*, pp. 141-145 (1999).

[Hill, 2000] Hill, L.: "Core Elements of Digital Gazetteers: Placenames, Categories, and Footprints". *Proceedings of the 4th European Conference - ECDL*, pp.280-290 (2000).

[Hill, 2002] Hill, L.: "New Protocols for Gazetteer and Thesaurus Services". *D-Lib Magazine*, 8(3) (2002).

[Hill et al, 1999a] Hill, L.; Qi, Z.: "Indirect Geospatial Referencing through Place Names in the Digital Library: Alexandria Digital Library Experience with Developing and Implementing Gazetteers". *Proceedings of the American Society for information Science Annual Meeting*, Washington, pp. 57-69, November (1999).

[Hill et al, 1999b] Hill, L.; Frew, J.; Zheng, Q.: "Geographic Names: The Implementation of a Gazetteer in a Georeferenced Digital Library". *D-Lib Magazine*, January (1999).

[Ide et al, 1995] Ide, N.; Veronis, J.: "Text *Encoding Initiative: Background and Context*", Kluwer Academic Publishers, October, 1995.

[Janée, 2002] Janée, G: "Gazetteer and thesaurus service protocols". *Joint Conference on Digital Libraries – Workshop Digital Gazetteers Integration into distributed digital library services*, Portland, USA, (2002).

[Jensen et al, 1998] Jensen, C.; Dyreson, E., C.: "A Consensus Glossary of Temporal Database Concepts - February 1998 Version". *Springer-Verlag*, pp. 367-405, (1998).

[Jones, 1997] Jones, B., C.: "GIS and Museums". *MDA Information*, 2(3), pp. 21-26, (1997).

[Kay, 2000] Kay, M.: "*XSLT Programmer's Refererence*". Wrox Press Inc, 2000.

[Kowalski, 1997] Kowalski, G.: "*Information Retrieval Systems: Theory and Implementation*". Kluwer Academic Publishers, USA, 1997.

[Langran, 1992] Langran, G.: "*Time in Geographic Information Systems*". Taylor & Francis, 1992.

[Larman, 1998] Larman, C.: "*Applying UML and Patterns – An Introduction to Object-Oriented Analysis and Design*". Prentice-Hall, 1998.

[Li et al, 1996] Li, S., C.; Bree, D.; Moss, A.; Petch, J.: "Developing Internet-based user interfaces for spatial data access and usability". *NCGIA 3rd International Conference/Workshop on Integrating Geographical Information Systems and Environmental Modelling*, pp. 21-25, Santa Fe, New Mexico: USA, january (1996).

[Longley et al, 2001] Longley, A., P.; Goodchild, F., M.; Maguire, J., D.; Rhind, W., D.: "*Geographical Information Systems and Science*". John Wiley & Sons, 2001.

[Mahoney et al, 2000] Mahoney, A.; Rydberg-Cox, A., J.; Smith, A., D.; Wulfman, E., C.: "Generalizing the Perseus XML Document Manager". *Workshop Web-Based Language Documentation and Description*, Philadelphia, USA, December (2000).

[Maler et al, 1995] Maler, E.; Andaloussi, E., J.: "*Developing SGML DTDs: From Text to Model to Markup*". Prentice Hall PTR, December, 1995.

[MapSciCom, 1999] Mapping Science Committee (1999). "Distributed Geolibraries: Spatial Information Resources". Summary of a Workshop: Panel on Distributed Geolibraries, Mapping Science Committee (National Research Council [US]). Washington,D.C.: National Academy Press. Disponível em: http://www.nap.edu/html/geolibraries/. Acessado em dezembro de 2002.

[Miranda et al, 2002] Miranda, A., V., R.; Baptista, S., C.; Almeida, R., R.; Catão, B.; Pazinatto, E.: "IGIS: um Framework para Sistemas de Informações Geográficas em N-Camadas usando um SGDB Objeto-Relacional". *GeoInfo'2002 – IV Simpósio Brasileiro de Geoinformática*, pp. 155-162, Caxambu, MG: Brasil, Dezembro (2002).

[Mukhar et al, 2001] Mukhar, K.; Lauinger, T.; Carnell, J.: "*Beginning Java Databases: JDBC, SQL, J2EE, EJB, JSP, XML*". Wrox Press Inc, 2001.

[Nowakowski, 1995] Nowakowski, J.: "New products and services: GNIS online". *U. S. Geological Survey yearbook* (1995).

[OGC, 2001] OpenGis Consortium. "Abstract Specification Overview". topic 0, version 4. Disponível em: http://www.opengis.org/techno/abstract/99-100r1.pdf. Acessado em dezembro 2002.

[Oracle, 2002] Oracle Technology Network: "*Oracle Documentation*". Disponível em: http://technet.oracle.com/docs/content.html. Acessado em dezembro de 2002.

[Patzer et al, 1999] Patzer, A.; Li, S.; Houle, P.; Wilcox, M.; Phillips, R.; Ayers, D.; Bergsten, H.; Diamond, J.; Bogovich, M.; Ferris, M.; Fleury, M.; Halberstadt, A.; Mohseni, P.; Vedati, K.; Zeiger, S.: "*Professional Java Server Programming: with Servlets, Java Server Pages (JSP), XML, Enterprise Java Beans (EJB), JNDI, CORBA, Jini and Javaspaces*". Wrox Press Inc, 1999.

[Pazinatto et al, 2002a] Pazinatto, E.; Baptista, S., C.; Miranda, A., V., R.: "GeoLocalizador: um Sistema de Referência Espaço-Temporal Indireta utilizando um SGBD Objeto-Relacional". *GeoInfo'2002 – IV Simpósio Brasileiro de Geoinformática*, pp. 49-56, Caxambu, MG: Brasil, Dezembro (2002).

[Pazinatto et al, 2002b] Pazinatto, E.; Baptista, S., C; Miranda, V., A., R.: "Modelagem e Implementação de um Sistema de Referência Espaço-Temporal Indireta utilizando um SGBD Objeto-Relacional". *In Proceedings International Information Technology Symposium - I2TS'2002*, pp. 19, Florianópolis, Outubro (2002).

[Pullar et al, 1988] Pullar, D.; Egenhofer, M.: "Towards Formal Definitions of Topological Relations Among Spatial Objects". *In Proceedings of the 3rd International Symposium on Spatial Data Handling*, Marble, Sydney, Australia, pp. 225-241, (1988).

[Rosenthal et al, 1996] Rosenthal, M; Chu, H.: “Search Engines for the World Wide Web: A comparative study and evaluation methodology”. *Proceedings of the ASIS Annual Conference*, Baltimore, pp.127-135, October 19-24 (1996).

[Rumbaugh et al, 1999] Rumbaugh, J.; Booch, G.; Jacobson, I.: “*The Unified Modeling Language Reference Manual*”. Addison-Wesley, 1999.

[Salton et al, 1983] Salton, G.; McGill, M., J.: “Introduction to Modern Information Retrieval”. McGraw- Hill, 1983.

[Schlieder et al, 2001] Schlieder, C.; Vögele, T.; Visser, U.: “Qualitative Spatial Representation for Information Retrieval by Gazetteers”. *Conference of Spatial Information Theory -COSIT*, California: Springer (2001).

[Sintichakis et al, 1997] Sintichakis, M.; Constantopoulos, P.: “A Method for Monolingual Thesauri Merging”. *In Proceedings of the 20th Annual International ACM SIGR*, pp. 129-138, (1997).

[Smith et al, 1995] Smith, T.; Frew, J.: “Alexandria Digital Library”. *Communications of the ACM*, 38(4): pp. 61-62, April (1995).

[Taylor et al, 1994] Taylor, C.; Tudhope, D.; Beynon-Davies P.: “Representation and Manipulation of Conceptual, Temporal and Geographical Knowledge in a Museum Hypermedia System”. In *Proceedings of the European Conference on Hypermedia Technology, ECHT '9*4, pp. 239-244, (1994).

[Tippey, 1998] Tippey, D.: “GENUKI -- The UK and Ireland Genealogical Information Service on the Internet”. *Family History Monthly*, May (1998).

[Tudhope et al, 2001] Tudhope, D.; Alani, H.; Jones, C.: “Augmenting thesaurus relationships: possibilities for retrieval”, *The Journal of Digital Information*, 1(8), (2001).

[Waters, 1998] Waters, J., D.: “What Are Digital Libraries”. *CLIR* (*Council Library and Information Resources*), ed. 4, july/august (1998).

[Zhu et al, 1999] Zhu, B.; Ramsey, M.; Ng, T.; Chen, H.; Schatz, B.: “Creating a Large-Scale Digital Library for Georeferenced Information”. *D-Lib Magazine*, july (1999).

Publicações

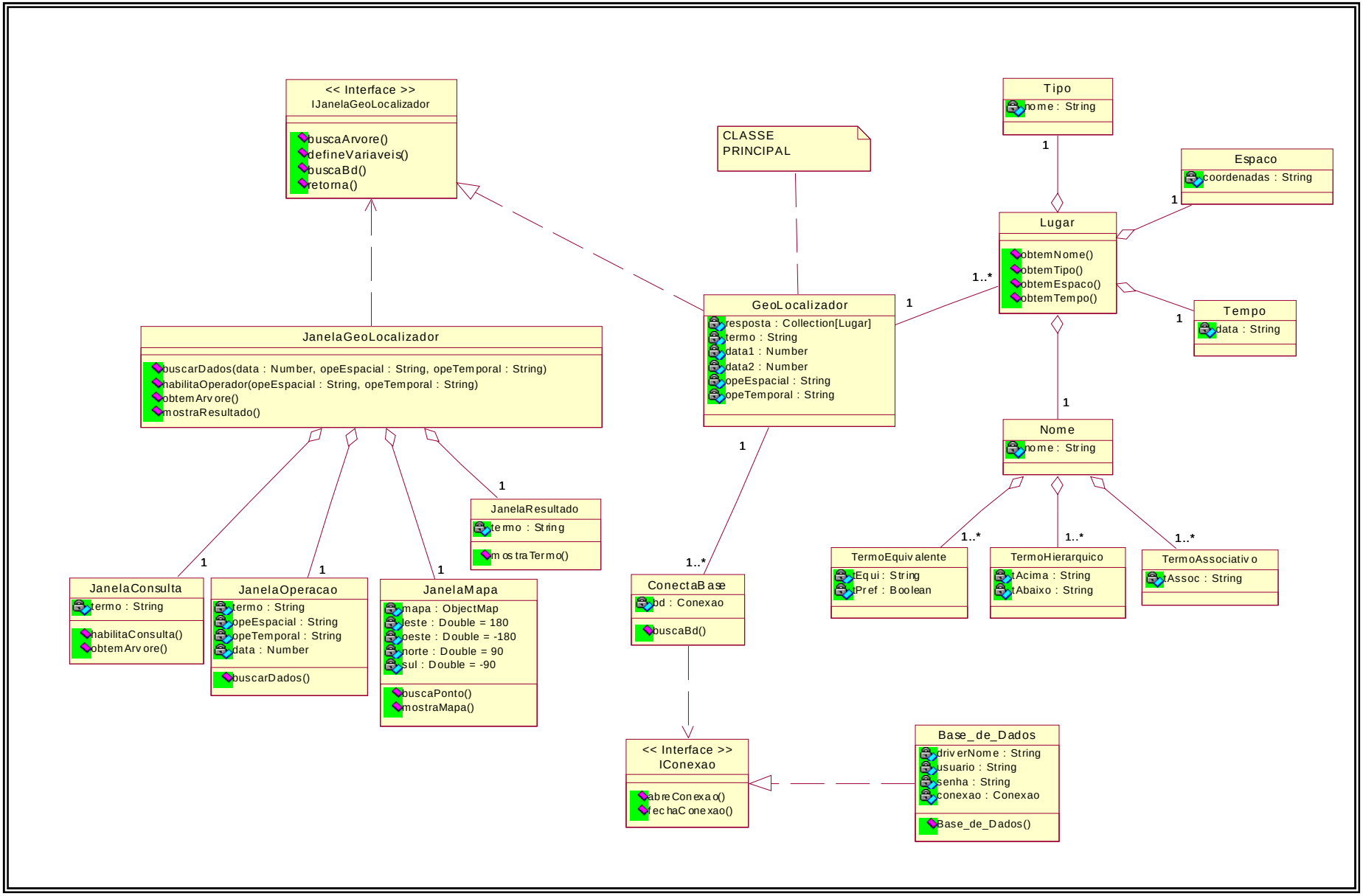
Pazinatto, E.; Baptista, S., C.; Miranda, A., V., R.: “GeoLocalizador: um Sistema de Referência Espaço-Temporal Indireta utilizando um SGBD Objeto-Relacional”. *GeoInfo’2002 – IV Simpósio Brasileiro de Geoinformática*, pp. 49-56, Caxambu, MG: Brasil, Dezembro (2002).

Pazinatto, E.; Baptista, S., C; Miranda, V., A., R.: “Modelagem e Implementação de um Sistema de Referência Espaço-Temporal Indireta utilizando um SGBD Objeto-Relacional”. *In Proceedings International Information Technology Symposium - I2TS’2002*, pp. 19, Florianópolis, Outubro (2002).

Miranda, A., V., R.; Baptista, S., C.; Almeida, R., R.; Catão, B.; Pazinatto, E.: “IGIS: um Framework para Sistemas de Informações Geográficas em N-Camadas usando um SGDB Objeto-Relacional”. *GeoInfo’2002 – IV Simpósio Brasileiro de Geoinformática*, pp. 155-162, Caxambu, MG: Brasil, Dezembro (2002).

Apêndice A

Diagrama de classes do GeoLocalizador


<< Interface >>
IJanelaGeoLocalizador
buscaArvore()
defineVariaveis()
buscaBd()
retorna()
CLASSE
PRINCIPAL
Tipo
nome : String
Espaco
coordenadas : String
Lugar
obtemNome()
obtemTipo()
obtemEspaco()
obtemTempo()
Tempo
data : String
GeoLocalizador
resposta : Collection[Lugar]
termo : String
data1 : Number
data2 : Number
opeEspacial : String
opeTemporal : String
JanelaGeoLocalizador
buscarDados(data : Number, opeEspacial : String, opeTemporal : String)
habilitaOperador(opeEspacial : String, opeTemporal : String)
obtemArvore()
mostraResultado()
Nome
nome : String
JanelaResultado
termo : String
mostraTermo()
JanelaConsulta
termo : String
habilitaConsulta()
obtemArvore()
JanelaOperacao
termo : String
opeEspacial : String
opeTemporal : String
data : Number
buscarDados()
JanelaMapa
mapa : ObjectMap
leste : Double = 180
oeste : Double = -180
norte : Double = 90
sul : Double = -90
buscaPonto()
mostraMapa()
ConectaBase
bd : Conexao
buscaBd()
TermoEquivalente
tEqui : String
tPref : Boolean
TermoHierarquico
tAcima : String
tAbaixo : String
TermoAssociativo
tAssoc : String
<< Interface >>
IConexao
abreConexao()
fechaConexao()
Base_de_Dados
driverNome : String
usuario : String
senha : String
conexao : Conexao
Base_de_Dados()
1
1..*