

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO – PUC-SP

ALEX MOTTA MELO DA ROCHA

**Formulação de um Modelo de Análise Epidemiológica usando
Raciocínio Baseado em Casos e Geoprocessamento**

MESTRADO EM TECNOLOGIAS DA INTELIGÊNCIA E DESIGN DIGITAL

SÃO PAULO
2012

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO – PUC-SP

ALEX MOTTA MELO DA ROCHA

**Formulação de um Modelo de Análise Epidemiológica usando
Raciocínio Baseado em Casos e Geoprocessamento**

MESTRADO EM TECNOLOGIAS DA INTELIGÊNCIA E DESIGN DIGITAL

Dissertação apresentada à Banca Examinadora da Pontifícia Universidade Católica de São Paulo, como exigência parcial para obtenção do título de Mestre em Tecnologia da Inteligência e Design Digital – área de concentração “Modelagem de Sistemas de Software” – sob a orientação do Prof. Dr. Fernando Antonio de Castro Giorno.

SÃO PAULO
2012

FICHA CATALOGRÁFICA

Rocha, Alex Motta Melo da.

Formulação de um Modelo de Análise Epidemiológica usando Raciocínio Baseado em Casos e Geoprocessamento -- São Paulo, 2012.

95 p. : il.

Dissertação (Mestrado) - Pontifícia Universidade Católica de São Paulo. Programa de Pós-Graduação em Tecnologias da Inteligência e Design Digital.

1. Inteligência Artificial. 2. Sistemas Baseados em Conhecimento. 3. Raciocínio Baseado em Casos.

4. Geoprocessamento. 5. Leptospirose Humana. I. Rocha, Alex Motta Melo da. II. Título.

BANCA EXAMINADORA

Prof. Dr. Fernando Antonio de Castro Giorno – Orientador
Pontifícia Universidade Católica de São Paulo

Profa. Dra. Ana Lúcia Manrique
Pontifícia Universidade Católica de São Paulo

Prof. Ph.D João de Fernandes Teixeira
Universidade Federal de São Carlos

Dedico esta dissertação à minha família que sempre me incentivou, apoiou e esteve comigo em todos os momentos.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus que sempre me deu força, saúde, energia e concentração para que eu pudesse lutar e alcançar mais este objetivo em minha vida que, sem dúvida, não foi fácil.

À minha família, especialmente aos meus avós, José Melo da Rocha e Margarida Maria Motta Melo da Rocha, pelos anos de inteira dedicação e incentivo para que sempre buscasse o melhor, me incentivando e apoiando a todo o momento, o que com certeza, me ajudaram a chegar até aqui. Sem esquecer-me de agradecer também aos meus tios, tias, primos e primas que são tantos, mas nenhum é menos especial que o outro.

Aos meus pais, João Paulo Mota Melo da Rocha e Ana Maria Motta Melo da Rocha, que torceram e acreditaram em mim desde o princípio, sendo de fundamental importância para eu concluir com maestria mais esta etapa da minha caminhada.

À minha noiva e, em breve, futura esposa, Adriana Aflalo Silva, minha eterna paixão, pela compreensão no decorrer desta dissertação, pelo incentivo e ajuda ao longo desses mais de cinco anos de união, que me proporcionou alcançar e conquistar mais esta vitória. Sem esquecer as inúmeras revisões, feitas por ela, em minha dissertação.

A todos meus amigos, que mesmo distantes, devido a inúmeros fatores, de maneira especial ao Luciano Santa Brígida das Neves, sempre me apoiaram e torceram pelo meu crescimento e sucesso.

Ao professor e orientador, Dr. Fernando Antonio de Castro Giorno, pelos conhecimentos transmitidos e pela resignação ao longo desta dissertação – nas inúmeras definições de temas e títulos, sobretudo, nesta fase importante em minha vida pessoal e profissional, que foi o mestrado.

Ao corpo docente do Programa de Pós-Graduação em Tecnologias da Inteligência e Design Digital (TIDD), sobretudo aos professores Dr. Ítalo Santiago Vega e Dr. Luís Carlos Petry, da Pontifícia Universidade Católica de São Paulo (PUC-SP), pelo aprendizado que me foi transferido ao longo do curso.

À secretária do TIDD, Edna Conti, por todos os conselhos, auxílios e paciência nos diversos encontros acadêmicos, ligações e trocas de e-mail.

À banca examinadora, além do meu orientador, tive a honra de contar com a presença dos professores Ph.D João de Fernandes Teixeira, Dr. Júlio Arakaki e Dra. Ana Lúcia Manrique, que contribuíram, por meio de um olhar experiente, para o sucesso desta dissertação.

Aos colegas do curso de mestrado, por todo conhecimento compartilhado, comunicação oferecida e pontos de vista singulares.

Ao Laboratório de Geoprocessamento do Instituto Evandro Chagas (LabGeo/IEC) e à Secretaria de Estado de Saúde Pública (SESPA), principalmente à Dra. Monica Fadul, pelo apoio técnico e metodológico necessários para o desenvolvimento desta dissertação, que me concederam os dados dos agravos, sendo de fundamental importância para a pesquisa.

Ao Ministério da Saúde (MS) pela disponibilização pública dos dados epidemiológicos para estudos e análises realizadas nesta dissertação, através do Sistema de Informação de Agravos de Notificação (SINAN), da Secretaria de Vigilância em Saúde (SVS).

À Empolis, empresa que forneceu, gratuitamente, o *shell* CBR-Works, que contribuiu imensamente para o desenvolvimento da técnica proposta neste trabalho, o Raciocínio Baseado em Casos (RBC).

Finalmente, a todos que direta ou indiretamente contribuíram para a constituição e realização da minha pesquisa de mestrado. Meu sincero agradecimento a todos vocês!

Allex Motta Melo da Rocha

“A mente que se abre a uma nova ideia jamais
voltará ao seu tamanho original.”

Albert Einstein

RESUMO

ROCHA, Allex Motta Melo da. **Formulação de um Modelo de Análise Epidemiológica usando Raciocínio Baseado em Casos e Geoprocessamento**. 2012. 95f. Dissertação (Mestrado em Tecnologias da Inteligência e Design Digital) – Programa de Pós-Graduação em Tecnologias da Inteligência e Design Digital, Pontifícia Universidade Católica de São Paulo, São Paulo, 2012.

A presente dissertação tem como objetivo utilizar as técnicas de Raciocínio Baseado em Casos (RBC) e Geoprocessamento, em conjunto, para identificar e caracterizar a incidência de um agravo infectocontagioso, na cidade de Belém, estado do Pará, Brasil, ocasionado pela bactéria *Leptospira interrogans*, conhecido como leptospirose humana. Questões ambientais e sanitárias, como a presença, principalmente, de excrementos de roedores, como urina e fezes influenciam o estabelecimento deste agravo. Esta pesquisa realiza um estudo do uso da técnica de RBC, com o objetivo de formular e desenvolver um modelo para identificar e auxiliar no diagnóstico da doença, por meio de dados coletados dos pacientes, tais como antecedentes epidemiológicos e clínicos, ocorridos na região, nos anos de 2008, 2009 e 2010. O sistema de RBC possibilitará que profissionais da saúde possam utilizá-lo como ferramenta de apoio na diagnose de casos da doença. Esta dissertação também aplica técnicas de Geoprocessamento para expressar visualmente e produzir análises ambientais, temporais e socioeconômicas do cenário epidemiológico da leptospirose em um bairro na cidade de Belém, nos anos citados anteriormente, devido à importância epidemiológica do mesmo. Para tal, será realizada a caracterização ambiental e socioeconômica, a fim de analisar a distribuição espacial dos principais reservatórios deste agravo e identificar os hábitos de vida das populações que moram ou trabalham nestes locais, considerando o fator ambiental para fornecer subsídios à adoção de medidas preventivas para o controle da incidência desta doença.

PALAVRAS-CHAVE: Inteligência Artificial, Sistemas Baseados em Conhecimento, Raciocínio Baseado em Casos, Geoprocessamento, Leptospirose Humana.

ABSTRACT

ROCHA, Alex Motta Melo da. **Formulação de um Modelo de Análise Epidemiológica usando Raciocínio Baseado em Casos e Geoprocessamento**. 2012. 95f. Dissertação (Mestrado em Tecnologias da Inteligência e Design Digital) – Programa de Pós-Graduação em Tecnologias da Inteligência e Design Digital, Pontifícia Universidade Católica de São Paulo, São Paulo, 2012.

This current dissertation aims to use the techniques of Case-Based Reasoning (CBR) and Geographic Information System (GIS) together to identify and characterize the impact of a contagious injury in the city of Belém, Pará state, Brazil, caused by the bacterium *Leptospira interrogans*, known as leptospirosis. Environmental and health issues, particularly as the presence of rodent droppings, urine and feces, influencing an establishment of this disease. This research conducts a study of the use of CBR technique in order to formulate and develop a model to identify and assist in the diagnosis of the disease, using data collected from patients, such as epidemiological history and clinical, in the area, in the years 2008, 2009 and 2010. The CBR system will enable health professionals can use it as a support tool in the diagnosis of cases of the disease. This dissertation also applies GIS techniques to visually express and produce environmental analysis, temporal and socioeconomic of leptospirosis epidemiology scenario in a neighborhood in the city of Belém, in the years mentioned above, due to yours epidemiological importance. So, there will be the environmental and socioeconomic characteristics, in order to analyze the spatial distribution of the main reservoirs of this disease and to identify the habits of life of people living or working in these places, considering the environmental factor to provide subsidies to adopt preventive measures to control of disease incidence.

KEYWORDS: Artificial Intelligence, Knowledge-Based Systems, Case-Based Reasoning, Geographic Information System, Human Leptospirosis.

SUMÁRIO

Lista de Figuras	12
Lista de Fórmulas	13
Lista de Gráficos.....	14
Lista de Tabelas	15
Lista de Abreviaturas ou Siglas	16
Capítulo 1: Introdução	17
1.1. Motivação.....	17
1.2. Objetivos da Pesquisa	19
1.2.1. Objetivo Geral	19
1.2.2. Objetivos Específicos	19
1.3. Resultado Esperado	20
1.4. Contribuições	20
1.5. Universo da Pesquisa	21
1.6. Método de Pesquisa	21
1.7. Organização do Trabalho	22
Capítulo 2: Fundamentação Teórica	24
2.1. Raciocínio Baseado em Casos	24
2.1.1. Histórico	25
2.1.2. Etapas de desenvolvimento de um modelo de RBC.....	27
2.1.2.1. Representação dos casos	27
2.1.2.2. Indexação dos casos	29
2.1.2.3. Recuperação dos casos	31
2.1.2.4. Adaptação dos casos.....	35
2.1.2.5. Aprendizagem	36
2.1.3. Estrutura de um sistema de RBC.....	37
2.1.3.1. Memória de casos de domínio.....	37
2.1.3.2. Mecanismo de Pesquisa	38
2.1.3.3. Descrição dos índices para diferenciar os casos.....	39
2.1.4. Vantagens e desvantagens de um sistema de RBC	39
2.1.4.1. Vantagens	39
2.1.4.2. Desvantagens.....	40
2.2. Geoprocessamento	40
2.2.1. Sistemas de Informação Geográfica.....	41
2.2.2. Equipamentos usados para entrada e saída de dados	42
2.2.2.1. Equipamentos de entrada de dados	42
2.2.2.2. Equipamentos de saída de dados	43
2.3. Epidemiologia da Leptospirose Humana	44
2.3.1. Agente etiológico	44
2.3.2. Reservatório.....	47
2.3.3. Modo de transmissão.....	47
2.3.4. Período de transmissibilidade.....	48
2.3.5. Suscetibilidade e imunidade	48
2.3.6. Aspectos clínicos e laboratoriais	48
2.3.6.1. Manifestações clínicas.....	48
2.3.6.2. Diagnóstico epidemiológico.....	49
2.3.6.3. Diagnóstico laboratorial	49
2.3.7. Incidência da Leptospirose no Pará.....	50

2.4. Conclusões	52
Capítulo 3: Estado da Arte	53
3.1. Sistema de RBC para Diagnóstico Histopatológico.....	54
3.1.1. Análise do artigo	55
3.2. Sistema de RBC para Diagnóstico de Pacientes	55
3.2.1. Análise do artigo	57
3.3. Sistematização da Assistência em Enfermagem Pediátrica aplicando RBC.....	58
3.3.1. Análise do artigo	58
3.4. Metodologia de RBC para o Diagnóstico de Gripe Suína	59
3.4.1. Análise do artigo	60
3.5. Diagnósticos de Esclerose Múltipla: Abordagem em Raciocínio Baseado em Casos....	61
3.5.1. Análise do artigo	62
3.6. Conclusões	63
Capítulo 4: Formulação do Modelo de Análise Epidemiológica	64
4.1. Implementação do sistema de RBC	64
4.1.1. Atribuição dos pesos das variáveis.....	64
4.1.1.1. Atividades de pré-processamento.....	65
4.1.1.2. Padrões identificados.....	66
4.1.2. Desenvolvimento do SisLeptos	69
4.1.2.1. Representação dos casos	70
4.1.2.2. Indexação dos casos	71
4.1.2.3. Recuperação dos casos	73
4.1.2.4. Adaptação dos casos.....	73
4.2. Geoprocessamento aplicado à Análise da Leptospirose	73
4.3. Conclusões	77
Capítulo 5: Considerações Finais.....	78
5.1. Resultados Obtidos	79
5.2. Dificuldades Encontradas.....	81
5.3. Trabalhos Futuros	82
Referências Bibliográficas.....	84
Webgrafia	87
Referências Complementares	88
Apêndice A: Telas do SisLeptos	90
Anexo A: Ficha Cadastral de Leptospirose	94

LISTA DE FIGURAS

Figura 2.1. Divisão do raciocínio em três tipos de inferência.	25
Figura 2.2. Ciclo de um sistema de RBC.	27
Figura 2.3. Modelo de representação dos casos de um sistema de RBC.....	29
Figura 2.4. Modelo de indexação dos casos de um sistema de RBC.	31
Figura 2.5. Modelo de similaridade de um sistema de RBC.	34
Figura 2.6. Modelo de adaptação de um sistema de RBC.....	35
Figura 2.7. Equipamentos de entrada de dados em um SIG.....	43
Figura 2.8. Equipamentos de saída de dados em um SIG.	43
Figura 2.9. Ciclo de vida da leptospirose.	47
Figura 2.10. Fotos da urbanização desenfreada e ausência de infraestrutura básica.....	51
Figura 4.1. Casos com os antecedentes epidemiológicos da leptospirose.....	65
Figura 4.2. Casos filtrado e padronizado utilizado no Treemap.....	66
Figura 4.3. Casos suspeitos de leptospirose agrupados por meses.....	67
Figura 4.4. Casos suspeitos de leptospirose, na área de estudo, agrupados por meses.	67
Figura 4.5. Faixa etária e gênero, na área de estudo, com maior incidência da leptospirose..	68
Figura 4.6. Casos confirmados de leptospirose agrupados por bairro.....	69
Figura 4.7. Representação dos casos do sistema SisLeptos.	70
Figura 4.8. Banco de casos do sistema SisLeptos.	71
Figura 4.9. Representação das características indexadas e não-indexadas do sistema.	72
Figura 4.10. Indexação dos casos do sistema.....	72
Figura 4.11. Mapa da área de estudo, o bairro do Guamá.....	74
Figura 4.12. Mapa da altimetria da área de estudo, o bairro do Guamá.....	75
Figura 4.13. Mapa de áreas, pontos e drenagem.	76
Figura 5.1. Tela de consulta com informações sobre um caso de leptospirose.....	79
Figura 5.2. Tela com o resultado de um caso confirmado de leptospirose.	80
Figura 5.3. Tela com o resultado de um caso descartado de leptospirose.....	81
Figura A.1. Tela de autor do SisLeptos, no idioma português.	90
Figura A.2. Tela de autor do SisLeptos, no idioma inglês.	90
Figura A.3. Tela do projeto do SisLeptos.	91
Figura A.4. Tela da metodologia do SisLeptos.	91
Figura A.5. Tela de contato do SisLeptos.	92
Figura A.6. Tela de ajuda do SisLeptos.	92
Figura A.7. Primeira página da ficha cadastral com as informações sobre leptospirose.	94
Figura A.8. Segunda página da ficha cadastral com as informações sobre leptospirose.	95

LISTA DE FÓRMULAS

Fórmula 2.1. Fórmula de similaridade pelo Vizinho mais Próximo.	34
Fórmula 3.1. Fórmula de similaridade pelo Vizinho mais Próximo para calcular a SS.	60

LISTA DE GRÁFICOS

Gráfico 2.1. Casos de leptospirose na região Norte do Brasil, do ano de 2001 a 2011.....	46
Gráfico 2.2. Casos de leptospirose no Pará, do ano de 2011 a 2011.....	50
Gráfico 2.3. Casos suspeitos de leptospirose na área de estudo: 2008 - 2010.....	51

LISTA DE TABELAS

Tabela 2.1. Casos confirmados de leptospirose no Brasil: 2001 - 2011.....	45
Tabela 2.2. Casos de leptospirose nos dez estados mais afetados do Brasil: 2001 - 2011.....	45
Tabela 2.3. Casos de leptospirose nos dez estados mais afetados do Brasil: 2001 - 2011.....	46
Tabela 2.4. Manifestações clínicas em 47 pacientes com leptospirose: 2008 - 2010.....	52
Tabela 4.1. Casos confirmados de leptospirose na área de estudo: 2008 - 2010.....	69

LISTA DE ABREVIATURAS OU SIGLAS

ASP	<i>Active Server Pages</i>
BD	Banco de Dados
CBR	<i>Case-Based Reasoning</i>
GIS	<i>Geographic Information System</i>
GPS	<i>Global Position System</i>
EM	Esclerose Múltipla
ESTs	<i>Expressed Sequence Tags</i>
EUA	Estados Unidos da América
IA	Inteligência Artificial
IBGE	Instituto Brasileiro de Geografia e Estatística
IEC	Instituto Evandro Chagas
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
IIS	<i>Internet Information Services</i>
LabGeo	Laboratório de Geoprocessamento
MOP	<i>Memory Organization Packets</i>
MS	Ministério da Saúde
NHS	<i>National Health Service</i>
POM	Pacotes de Organização de Memória
RA	Raciocínio por Analogia
RBC	Raciocínio Baseado em Casos
RNAs	Redes Neurais Artificiais
SAD	Sistema de Apoio à Decisão
SAE	Sistematização da Assistência de Enfermagem
SBC	Sistemas Baseados em Conhecimento
SESPA	Secretaria de Estado de Saúde do Pará
SFDA	<i>Swine Flu Diagnostic Assistant</i>
SGBD	Sistema de Gerenciamento de Banco de Dados
SI	Sistemas de Informação
SIG	Sistemas de Informação Geográfica
SINAN	Sistema de Informação de Agravos de Notificação
SS	<i>Similarity Score</i>
SVS	Secretaria de Vigilância em Saúde
TI	Tecnologia da Informação
XML	<i>Extensible Markup Language</i>

Capítulo 1: Introdução

Este capítulo é responsável por introduzir e descrever os fatores motivacionais desta dissertação, bem como o objetivo geral e os objetivos específicos, o resultado esperado, as contribuições, o universo da pesquisa, o método de pesquisa e a organização do trabalho.

1.1. Motivação

Ao longo das duas últimas décadas, a Ciência da Computação tem sofrido transformações produzidas, principalmente, pelo avanço da Inteligência Artificial (IA) e dos sistemas inteligentes. Essas transformações podem ser vistas como um processo de busca por soluções para problemas complexos. Neste contexto, os sistemas inteligentes sofreram diversos progressos que possibilitaram solucionar problemas das mais variadas temáticas.

Os avanços dos sistemas inteligentes se devem a evolução da IA, que conforme Rich, Knight e Nair (2010) são constituídos por um conjunto de formalismos e técnicas computacionais voltados para a simulação de processos cognitivos, cuja finalidade é solucionar problemas complexos de forma confiável.

Dentre as técnicas computacionais existentes, o Raciocínio Baseado em Casos (RBC) tem sido desenvolvido para auxiliar na resolução de problemas em diversas áreas do conhecimento. O RBC é uma técnica de representação e processamento de conhecimento, cuja finalidade é criar modelos computacionais, baseado no raciocínio humano, através de experiências passadas para solucionar problemas atuais, resultando em sistemas inteligentes. As experiências serão representadas por casos, que servirão de modelo para resolver um novo problema, a partir de comparações entre casos recuperados da base.

Os casos contidos na base do RBC são eventos que ocorreram anteriormente, os quais serão recuperados para verificar o grau de similaridade com um novo problema, por meio do cálculo de similaridade, com a intenção de recuperar casos úteis para a resolução do problema, identificando as características relevantes, classificando quanto ao grau de equivalência e sugerindo novas soluções ao problema consultado. O RBC sugere uma ou mais soluções ao problema e, em seguida, armazena-os na base de casos para ser comparado com um problema futuro.

A outra técnica aplicada para o desenvolvimento do modelo de análise epidemiológica é o Geoprocessamento, abordada na seção 2.2, que consiste em diversas tecnologias espaciais, com a finalidade de coletar, manipular e mostrar as informações de forma clara, sendo responsável por classificar, a partir de dados georreferenciados, em três níveis: baixo, médio e

alto risco de contaminação por leptospirose. O georreferenciamento realiza um diagnóstico sobre a coleta de informação no espaço geográfico, através de pontos de controle que, no caso, possibilitam realizar uma análise complexa da localização dos focos da bactéria.

Todos os anos inúmeras pessoas adoecem e, ainda pior, falecem, decorrentes de doenças infectocontagiosas. A *Leptospira interrogans* é a bactéria causadora do agravo infectocontagioso responsável por transmitir a leptospirose, presente, principalmente, na urina de roedores contaminados. A infecção também pode ser propagada por bovinos, suínos e caninos infectados com a doença. Neste cenário, o trabalho tem como foco identificar as áreas da cidade de Belém para auxiliar na diagnose do surto, através de informações fornecidas pelo Instituto Evandro Chagas (IEC) e pela Secretaria de Estado de Saúde Pública (SESPA). A escolha da área de estudo é explicada com mais fundamentos no capítulo 2, na seção 2.3.1.

No Brasil, segundo a Sociedade Brasileira de Infectologia (2009), as doenças infecciosas e parasitárias foram a terceira origem de óbitos nos hospitais do país, atrás somente das doenças causadas nos aparelhos circulatório e respiratório. Dados do Ministério da Saúde (2011) apontam a leptospirose como um dos agravos mais notificados em 2010, com 3.732 casos e 373 óbitos originados pelo agravo, além de 40 originados por outras causas. Com base neste problema de saúde pública, a seguir, serão abordados os fatores motivacionais desta dissertação.

As elevadas taxas de incidência de agravos infectocontagiosos no estado do Pará, gerou a necessidade de desenvolver uma pesquisa que possa auxiliar na diagnose epidemiológica. A partir dos recursos emergentes de Tecnologia da Informação (TI), tem se colocado como um grande desafio para a comunidade acadêmica conciliar a Medicina com os progressos tecnológicos. Os avanços da TI observados na área epidemiológica nos últimos anos proporcionaram maior concisão diagnóstica para as doenças, devido à capacidade que os Sistemas de Informação (SI) possuem em manipular amplos volumes de dados.

O serviço de saúde, na região amazônica, devido às grandes distâncias, falta de infraestrutura básica e profissional capacitado, está cada vez mais precário. Conforme Brasileiro (2008, *apud* SANTANA, 2008, p. 1), presidente do Sindicato dos Enfermeiros de Sergipe, “há sobrecarga de trabalho, equipamentos sem manutenção, quantidade de macas insuficiente para a demanda de pacientes”.

Com o alto número de pacientes atendidos em hospitais, postos de saúde e prontos-socorros, gera-se uma sobrecarga no atendimento, em decorrência ao número aquém de funcionários, sem contar com a falta de preparo de alguns empregados e poucos equipamentos disponíveis para a realização de exames.

Com a tentativa de combater tais limitações, propôs-se a formulação de um modelo de RBC, tendo como apoio a utilização do Geoprocessamento para identificar e classificar a incidência da leptospirose humana nos principais bairros, de Belém do Pará, afetados com a incidência da doença. Todas as informações de pacientes são verídicas e autorizadas para uso em fins acadêmicos. Esta pesquisa conta com a colaboração do IEC e da SESPA, órgãos de pesquisa científica do Ministério da Saúde (MS), que cederam os dados dos pacientes para a realização desta dissertação.

1.2. Objetivos da Pesquisa

A seguir, é descrito o objetivo geral e os objetivos específicos desta pesquisa.

1.2.1. Objetivo Geral

Esta pesquisa tem como objetivo formular um modelo de análise epidemiológica usando, em conjunto, as técnicas de RBC e Geoprocessamento, para identificar casos de leptospirose humana, levando em consideração questões espaciais e temporais. O modelo será testado com casos reais da doença, de uma área pré-determinada, colhidos ao longo deste trabalho.

1.2.2. Objetivos Específicos

Os objetivos específicos desta dissertação são:

- Modelar e implementar, como prova de conceito, um sistema inteligente, utilizando RBC, para identificar casos da leptospirose humana de pacientes, em uma área pré-definida;
- Proporcionar apoio para médicos e especialistas responsáveis por produzir o diagnóstico de leptospirose humana, a partir do desenvolvimento de um Sistema de Apoio à Decisão (SAD) que gere automaticamente uma análise confiável, esperando prover uma solução rápida, economizando tempo e recursos financeiros governamentais, por meio de informações coletadas de casos dos pacientes, como antecedentes epidemiológicos e dados clínicos;
- Expressar visualmente, com técnicas de Geoprocessamento, as características do cenário epidemiológico do agravo na área pré-definida, com o intuito de auxiliar na diagnose do surto.

1.3. Resultado Esperado

O resultado esperado é um modelo capaz de identificar e analisar dados que se expressam qualitativa e quantitativamente de forma diferenciada em um determinado espaço geográfico, os quais poderão ser avaliados de forma probabilística, somando esforços para a consolidação de novas técnicas para análises epidemiológicas. Para tal, serão utilizadas técnicas de Geoprocessamento para visualizar as áreas de riscos, a fim de compreender o cenário ecoepidemiológico estabelecido na região estudada, como questões ambientais e socioeconômicas.

1.4. Contribuições

Esta pesquisa procura contribuir com estudos exploratórios a respeito da formulação de um modelo, fundamentado nas técnicas de RBC e Geoprocessamento, para a compreensão do cenário epidemiológico da leptospirose humana, devido a doença possuir índices de morbidade¹ expressivos, estar associada a grande volume de dados epidemiológicos e demandar análises complexas para o seu entendimento.

Considerando a natureza diferenciada dos dados epidemiológicos, ambientais, socioeconômicos e culturais que interferem na incidência de uma doença infectocontagiosa, como a leptospirose humana, o presente trabalho pretende investigar se a utilização, em conjunto, das técnicas de RBC e Geoprocessamento, é eficiente para caracterizar cenários ecoepidemiológicos deste agravo. Esta dissertação tem como estudo de caso a leptospirose humana na cidade de Belém, estado do Pará, devido diversas características, como estar localizada em uma área de várzea, possuir níveis de altimetria diferenciada em seu território, expressiva taxa de umidade e pluviosidade, apresentar gradientes de exclusão social, ter um elevado número de pessoas expostas a áreas de riscos e carecer de serviços básicos de educação e saúde.

O sistema prospectivo busca manipular dados que se expressam qualitativa e quantitativamente de forma diferenciada em um determinado espaço geográfico, os quais possam ser avaliados de forma probabilística, somando esforços para a consolidação de novas técnicas para análises epidemiológicas. Neste contexto, com este trabalho, também se almeja contribuir em medidas preventivas para o controle da incidência da doença, face às variáveis que a determinam, a partir de análises da distribuição espaço-temporal da incidência da leptospirose humana.

¹ Taxa de portadores da doença em relação à população total analisada.

A utilização do Geoprocessamento será fundamental, em virtude de a técnica promover a integração de dados ambientais, socioeconômicos e de saúde, que ao serem georreferenciados permitem identificar padrões da distribuição de doenças, ou dados relacionados a estas, tornando-se, desta forma, uma importante ferramenta de manipulação e análise de dados. Desta forma, as técnicas de Geoprocessamento serão utilizadas por permitirem a visualização do resultado das análises geradas pelo RBC, o que facilita a compreensão do cenário epidemiológico estabelecido na área de estudo.

1.5. Universo da Pesquisa

A pesquisa desenvolvida tem como universo a população exposta e contaminada por leptospirose humana, nos bairros do Guamá, Jurunas e Montese. Estes bairros foram escolhidos através da semelhança entre suas características socioeconômicas, ambientais e epidemiológicas, que influenciaram no cenário epidemiológico estabelecido. Sendo que, a utilização da técnica de Geoprocessamento foi restrita ao bairro do Guamá, por decorrência na limitação dos recursos tecnológicos disponíveis, os quais geraram dados ambientais e socioeconômicos, como características relacionadas aos tipos de habitação, infraestrutura e drenagem.

1.6. Método de Pesquisa

Para o desenvolvimento deste trabalho, faz-se necessário a concretização dos seis passos apresentados a seguir:

O primeiro passo consiste no levantamento do referencial teórico, em fontes primárias, secundárias e terciárias de informações, sendo que, as fontes primárias se constituirão dos dados ambientais, socioeconômicos e epidemiológicos coletados em campo relacionados à leptospirose humana, da área de estudo. As fontes secundárias se comporão de livros, artigos, monografias, dissertações e teses, além dos Sistemas de Informação oficiais do Ministério da Saúde, bem como o Sistema de Informação de Agravos de Notificação (SINAN) e a Secretaria de Estado de Saúde do Pará (SESPA), da Secretaria de Vigilância em Saúde (SVS). As fontes de dados terciárias se formarão dos dados disponibilizados através da Web, como por exemplo, as bibliotecas digitais.

O segundo passo incide na coleta de dados da leptospirose humana, através dos Sistemas de Informação do Ministério da Saúde, como o SINAN e a SESPA, a fim de utilizá-

los posteriormente para a comprovação do modelo de análise epidemiológica proposto nesta dissertação.

O terceiro passo é a elaboração de técnicas de depuração dos dados para que possa gerar um formato único de dados que comporão a base de casos do sistema que será desenvolvido utilizando a técnica de RBC, levando em consideração a natureza diferenciada dos dados coletados e levantados em campo ao longo da pesquisa.

O quarto passo é a formulação de um modelo de análise epidemiológica, utilizando as técnicas de RBC e Geoprocessamento, para auxiliar especialistas responsáveis por produzir o diagnóstico de leptospirose humana, e fornecer subsídios à adoção de medidas preventivas para o controle da incidência da doença, por parte de órgãos governamentais.

O quinto passo consiste na identificação e classificação da incidência da leptospirose humana no bairro do Guamá, com o uso das técnicas de Geoprocessamento, para expressar visualmente e produzir análises ambientais, temporais e socioeconômicas do cenário epidemiológico do agravo, nos anos de 2008, 2009 e 2010, a partir da utilização dos dados georreferenciados da área.

O sexto e último passo é a avaliação dos resultados obtidos, levando em consideração a utilização, em conjunto, das técnicas de RBC e Geoprocessamento, para caracterizar a leptospirose humana nas áreas de estudo, nos anos de 2008, 2009 e 2010, ao analisar questões socioeconômicas e ambientais, a fim de compreender as relações que influenciaram a incidência da doença na região.

1.7. Organização do Trabalho

De maneira sucinta, a seguir são expostos, em ordem cronológica, os cinco capítulos desta dissertação, com o intuito de descrever o desenvolvimento da pesquisa:

O primeiro capítulo, a Introdução, apresenta o tema, descrevendo a problemática e o fator motivacional desta dissertação. Posteriormente, aponta o objetivo geral e os objetivos específicos da pesquisa. Adicionalmente, apresenta o resultado esperado, as contribuições, o universo da pesquisa, o método de pesquisa e a organização do trabalho.

O segundo capítulo, aborda a Fundamentação Teórica, fornece base para esta pesquisa. Este capítulo descreve a técnica do Raciocínio Baseado em Casos, juntamente com o histórico, etapas de desenvolvimento, estrutura, além das vantagens e desvantagens dos sistemas de RBC. Em seguida, trata sobre o Geoprocessamento, junto com os conceitos fundamentais, histórico, as técnicas de Sensoriamento Remoto, Digitalização de Dados e Sistemas de Informação Geográfica (SIG). Posteriormente, apresenta a bactéria *Leptospira*

interrogans, causadora da leptospirose. Entre os assuntos, serão abordados o agente etiológico, o reservatório, o modo de transmissão, o período de incubação, o período de transmissibilidade, susceptibilidade e imunidade, os aspectos clínicos e laboratoriais, bem como as manifestações clínicas e o diagnóstico laboratorial. Por fim, descreve a incidência da leptospirose humana no estado do Pará, referenciando os bairros do Guamá, Jurunas e Montese.

O terceiro capítulo, o Estado da Arte, apresenta as pesquisas mais recentes sobre os Sistemas Baseados em Conhecimento (SBC), com foco em Raciocínio Baseado em Casos, mostrando projetos de pesquisas atuais na área e trabalhos científicos, especialmente artigos, relacionados ao tema de estudo desta dissertação.

O quarto capítulo, a Formulação do Modelo de Análise Epidemiológica, apresenta a análise do surto de leptospirose humana, objeto de estudo desta dissertação. Neste capítulo, é descrito o desenvolvimento do modelo de RBC detalhadamente, desde a atribuição dos pesos das variáveis, as etapas de desenvolvimento, até a arquitetura do sistema, levando em consideração as informações coletadas dos pacientes, como antecedentes epidemiológicos e dados clínicos, a fim de identificar a incidência da leptospirose nas áreas propostas. Neste capítulo, também são detalhadas as fases que compõem a etapa de determinação dos pesos utilizados para a indexação dos casos, com o objetivo de garantir que o sistema gere resultados confiáveis e, ainda, cada etapa da construção do modelo. Em seguida, mostra-se o uso das técnicas de Geoprocessamento para identificar e classificar a incidência da leptospirose no bairro do Guamá, nos anos de 2008, 2009 e 2010, a partir da utilização dos dados georreferenciados desta área.

Por fim, as Considerações Finais, no quinto capítulo, descrevem algumas conclusões em relação às técnicas aplicadas neste trabalho, os resultados obtidos pelo sistema, desde a construção do modelo, junto com as taxas de acertos, até o produto final, dificuldades encontradas, recomendações e extensões para trabalhos futuros.

Capítulo 2: Fundamentação Teórica

Este capítulo é responsável por abordar a Fundamentação Teórica que norteia e embasa esta pesquisa, com o enfoque nas técnicas de Raciocínio Baseado em Casos e Geoprocessamento. A bactéria *Leptospira interrogans*, causadora da leptospirose, também é abordada.

2.1. Raciocínio Baseado em Casos

Raciocínio Baseado em Casos (RBC) é uma ferramenta de raciocínio da Inteligência Artificial. A filosofia básica desta técnica é a de buscar a solução para uma situação atual através da comparação com uma experiência passada semelhante (FERNANDES, 2005, p. 27).

O Raciocínio Baseado em Casos (em inglês, *Case-Based Reasoning* – CBR) é uma técnica de representação e processamento de conhecimento, estudada pela área de Inteligência Artificial. De acordo com Fernandes (2005), o RBC teve sua origem relacionada a estudos de aspectos do aprendizado humano, sendo característica do seu funcionamento a simulação de atividades do ser humano, no qual utiliza experiências armazenadas na memória para resolver novos problemas, identificando afinidades entre as mesmas.

Esta técnica incorpora conhecimento de um especialista para resolver problemas das mais diversas temáticas. Conforme Riesbeck e Schank (1989, *apud* FERNANDES, 2005), um sistema de RBC resolve problemas por adaptar soluções que foram utilizadas para resolver problemas anteriores. No processo de adaptar e recordar uma situação semelhante quando comparado a uma nova, sistemas de RBC simulam o raciocínio do ser humano, buscando soluções resolvidas anteriormente para um problema parecido. Portanto, o RBC utiliza casos passados na busca da resolução de novos casos.

Marcos e Dias (2005) afirmam que o raciocínio pode ser dividido em três tipos de inferência: abdução, dedução e indução, conforme ilustra a figura 2.1. O raciocínio abdutivo ou hipotético, ligado à percepção estética, consiste em criar uma teoria para explicá-la posteriormente. O raciocínio indutivo ou estético, ligado à valores e crenças, é caracterizado por abstrair regras gerais por meio de observações individuais, ou seja, através de certo conhecimento ou experiência adquirida obtém-se determinada conclusão. Finalmente, no raciocínio dedutivo, ligado ao conhecimento lógico, a partir de determinadas informações deduz-se algo. O RBC faz uso, principalmente, do terceiro tipo de inferência, o raciocínio dedutivo, para a resolução de novos problemas.

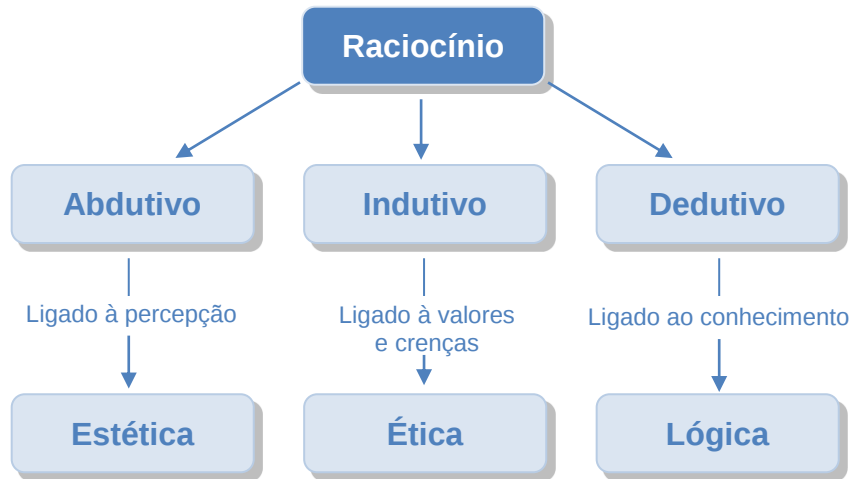


Figura 2.1. Divisão do raciocínio em três tipos de inferência.

Teixeira (2009, p. 11 e 12) afirma que “antes se achava que a inteligência era a capacidade de raciocinar. Agora, inteligência é o poder computacional”. Atualmente, com o avanço da tecnologia e com computadores cada vez mais poderosos, a inteligência é substituída pela força bruta, fazendo com que a máquina pareça pensar.

2.1.1. Histórico

De acordo com Gresse von Wangenheim e von Wangenheim (2003), o Raciocínio Baseado em Casos foi inspirado nos estudos de Schank e Abelson sobre Memória Dinâmica, iniciado em 1977. O objetivo da pesquisa era simular os modelos computacionais baseando-se em experiências do ser humano para realizar a tomada de decisão, através de experiências anteriores. Esses conhecimentos são gravados em forma de roteiros, que permitem realizar soluções de novos problemas.

Para Gresse von Wangenheim e von Wangenheim (2003), outras linhas de pesquisa de RBC incidiram do Raciocínio por Analogia (RA), que verifica as situações disponíveis em um domínio existente e procura soluções satisfatórias, com o intuito de resolver um problema de outro campo. Embora os primeiros trabalhos em RBC e em RA tenham se desenvolvido de forma independente, o RA contribuiu consideravelmente para a teoria de RBC, visto que ambos utilizam o raciocínio aplicável a determinado problema para tentar solucioná-lo. Além disso, os autores descrevem que o RBC pode ser considerado uma especialização do RA, por ter um estudo mais amplo do raciocínio computacional com base em experiências passadas.

Segundo Gresse von Wangenheim e von Wangenheim (2003, p. 31), “o primeiro sistema que pode ser chamado de raciocinador baseado em casos foi o sistema CYRUS [...]”. Conforme os autores, o CYRUS foi desenvolvido pela pesquisadora Janet Kolodner, no ano

de 1983, utilizando o modelo de Memória Dinâmica proposto por Schank. Esse modelo cria estratégias para elaborar soluções de problemas e de aprendizagem, verificando a melhor resposta existente na memória. Este sistema tem o objetivo de selecionar perguntas e possíveis respostas que correspondem às informações de reuniões passadas do ex-secretário dos Estados Unidos da América (EUA), Cyrus Vance. Assim, é possível encontrar informações semelhantes e propor a melhor solução, a partir de reuniões diplomáticas anteriores do ex-secretário.

A partir do CYRUS foram desenvolvidos diversos sistemas, nos EUA e Europa, com base na Memória Dinâmica, tais como (GRESSE VON WANGENHEIM; VON WANGENHEIM, 2003):

- MEDIATOR: desenvolvido por Simpson, no ano de 1985. O objetivo deste modelo é realizar a intervenção para negociações em um sistema de planejamento;
- PROTOS: desenvolvido por Porter e Baresiss, no ano de 1986. O sistema tem como finalidade suportar o diagnóstico em doenças auditivas, através da descrição de sintomas e histórico de pacientes e resultados de exames;
- PERSUADER: desenvolvido por Sycara, no ano de 1988. É um sistema para realizar negociações sindicais, que verifica o motivo de o empregado não estar satisfeito com a empresa ou vice-versa;
- CHEF: desenvolvido por Hammond, no ano de 1989. É um sistema que tem como objetivo criar novas receitas, a partir de receitas existentes;
- CASEY: desenvolvido por Koton, no ano de 1989. A finalidade do sistema é realizar o diagnóstico de problemas cardíacos;
- HYPO: desenvolvido por Ashley, no início da década de 1990. O sistema soluciona casos na área da legislação de marcas e patentes. O objetivo central é resolver casos de violação de patentes, como por exemplo, o vazamento de um segredo industrial;
- PATDEX: desenvolvido por Stefan Wess, no ano de 1991. O sistema é voltado para a identificação de problemas e estratégias de solução, sobretudo no diagnóstico de problemas técnicos, na área do diagnóstico de erros de inoperância em máquinas de produção.

2.1.2. Etapas de desenvolvimento de um modelo de RBC

Segundo Lee (1998), é fundamental o cumprimento de quatro etapas para a construção de um modelo de Raciocínio Baseado em Casos, como recuperar, reutilizar, revisar e reter. O ciclo do funcionamento de um sistema de RBC utiliza as premissas dos 4R, como se apresenta a seguir:

- Recuperar casos com características similares ao problema;
- Reutilizar casos equivalentes ao problema;
- Revisar a solução do problema em questão;
- Reter os casos na base para que sejam reutilizados.

A construção de um modelo de RBC corresponde às premissas dos 4R, mencionadas anteriormente, como representa a figura 2.2:

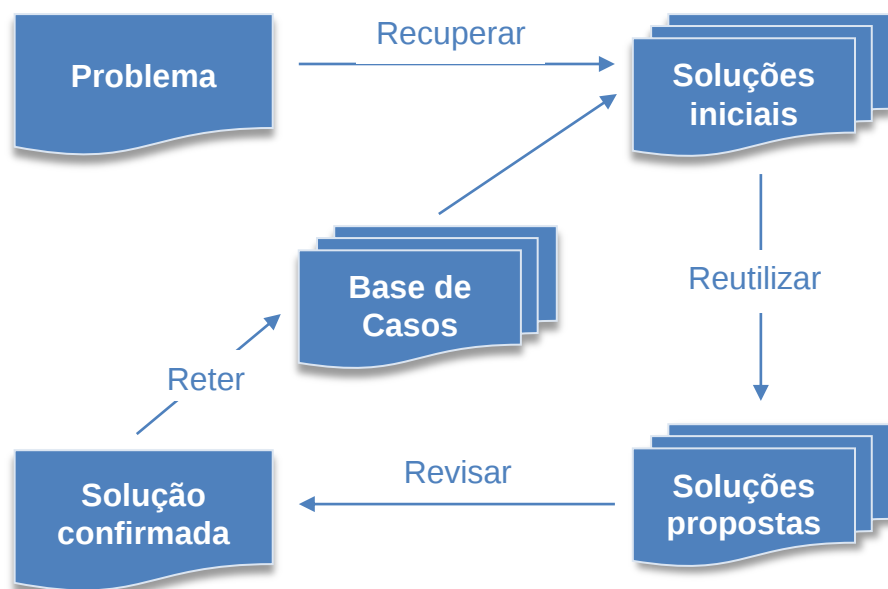


Figura 2.2. Ciclo de um sistema de RBC.
Fonte: Adaptado de Lee (1998).

2.1.2.1. Representação dos casos

A representação de casos é uma etapa fundamental para o sucesso de um sistema de RBC. Conforme Gresse von Wangenheim e von Wangenheim (2003), para desenvolver um sistema de RBC é necessário escolher a melhor maneira de representar o conhecimento, haja vista que sua representação é constituída sob a forma de casos.

De acordo com Fernandes (2005, p. 32), “caso é a forma de conhecimento contextualizado representando uma experiência que ensina uma lição útil”. Os casos são conhecimentos específicos representados em uma determinada situação ou episódio ocorrido.

Isto é, os casos podem ser vistos de formas diferentes e de tamanho variado, logo têm em comum o registro de experiências, a partir das características relevantes, para que, no futuro, esta situação seja usada com o propósito de solucionar um problema. Os casos abordam dois aspectos importantes:

- a) O primeiro aspecto descreve como as experiências vividas ou casos abstratos podem ensinar o sistema a encontrar eventos similares ao problema atual, através de atributos com maior grau de relevância. Com isso, os casos devem estar bem definidos e organizados para que o sistema consiga compreendê-los;
- b) O segundo aspecto faz referência a como a solução deve ser inserida no contexto. Neste caso, são selecionados casos semelhantes ao problema em questão, juntos com alternativas que poderão ser aproveitadas e, em seguida, adaptadas para que o sistema apresente a solução adequada ao problema.

Ao construir o RBC é recomendado verificar o objetivo do sistema no qual se realiza o raciocínio, pois, conforme Castoldi e Santos (2002), é imprescindível conter o objetivo a ser alcançado, bem como as características do problema e semelhanças entre suas partes. Deste modo, a representação deve incluir ações que possibilitam ao sistema encontrar soluções adequadas ao problema, através dos componentes mencionados. Contudo, pode haver restrições impostas a esse objetivo. Logo, deve ser atribuída outra ação que se adeque a situação, evitando que o sistema apresente um resultado precipitado.

Existem dois componentes básicos que o sistema deve adotar em sua estrutura para representar os casos. O primeiro refere-se à descrição do problema, onde uma situação é compreendida dentro do domínio para facilitar o raciocinador a encontrar seu objetivo. Com isso, ajuda a recuperar em particular casos que são equivalentes, a partir da descrição do problema para que as soluções sejam armazenadas na memória de dados do sistema, com o intuito de recuperar alternativas adequadas. O segundo componente aborda a descrição da solução, a qual compreende um conjunto de alternativas que apresentam ao raciocinador passos das soluções encontradas, podendo ou não associá-las ao problema, dependendo do que foi encontrado. Os casos podem incluir outros componentes opcionais, tais como a visualização dos efeitos da solução aplicada e a justificativa, mostrando com detalhe os resultados obtidos e indicando-os, independente de alcançar sucesso ou fracasso.

Para elucidar melhor um modelo de RBC, a seguir é descrito, de maneira simples, um caso em um sistema de suporte a manutenção de impressoras (figura 2.3):

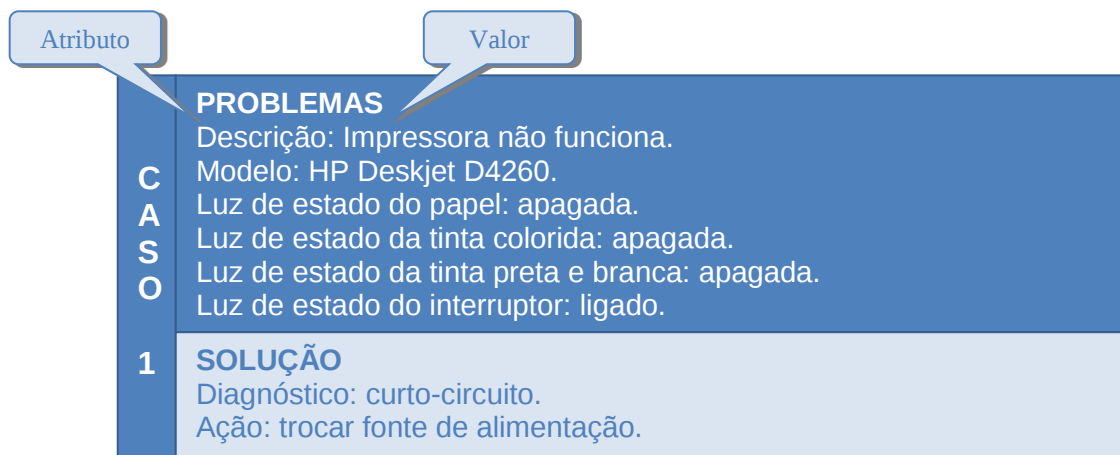


Figura 2.3. Modelo de representação dos casos de um sistema de RBC.

Fonte: Adaptado de Gresse von Wangenheim e von Wangenheim (2003).

Este modelo apresenta os atributos e valores selecionados para representar um problema do sistema, como a descrição, modelo, as luzes de estado do papel e das tintas, o estado do interruptor e as soluções para o caso, bem como o diagnóstico e a ação.

2.1.2.2. Indexação dos casos

Segundo Heinrich (2001), a etapa mais importante do RBC é a indexação, onde se definem os caminhos que ajudam a recuperar soluções úteis, com o intuito de resolver problemas futuros. Isto é, a indexação é feita de maneira a identificar, principalmente, o grau de importância que o índice tem ao caso corrente. O sistema deve analisar quais componentes serão utilizados, a fim de recuperar os casos com eficiência e rapidez contidos na base. Para tal, é necessário aplicar o cálculo de similaridade. Por meio desse cálculo, podem-se avaliar os casos que serão utilizados pelo sistema na resolução de novos problemas.

A indexação não é uma etapa trivial. Logo, a análise dos atributos que podem ser associados como índices devem ser analisados com cuidado, pois é necessário conhecer os requisitos para que o atributo seja considerado um bom índice. A associação de índice é responsável por recuperar os casos com maior confiabilidade, através dos atributos que são de vital importância para a etapa de recuperação em sistema de RBC.

Os componentes indexados têm um valor de entrada específico que serve como base para encontrar outras referências. A partir da descrição do problema, os índices devem apresentar características específicas dos casos para que sejam comparados com os demais armazenados na base, verificando os casos sugeridos pelo sistema, isto é, aqueles que serão utilizados para solucionar o problema.

Segundo Melchior (1999), a indexação é dividida em duas partes:

- a) A primeira parte indica rótulos para os casos no instante que são armazenados na base, a fim de garantir que sejam recuperados no momento adequado;
- b) O segundo envolve a definição da organização dos casos para facilitar a recuperação e encontrar casos que satisfaçam o problema em questão com eficiência.

Kolodner (1993) propõe adotar os seguintes princípios para que os índices sejam atribuídos sob a forma adequada:

- Interpretar prematuramente as informações contidas nos casos para aproveitá-los em outros problemas;
- Indicar a forma que os casos podem ser utilizados;
- Permitir que os casos sejam úteis em situações diferentes;
- Possibilitar que o caso seja recuperado facilmente em problemas futuros.

Conforme Melchior (1999), além desses princípios, os índices podem ser escolhidos a partir de métodos manuais e automáticos, a fim de auxiliar na etapa de indexação dos componentes. O primeiro analisa o caso como um todo, com a finalidade de identificar que componentes serão úteis para os casos. As informações contidas nos atributos devem ser representadas de forma sólida para que o sistema seja capaz de classificar e organizar os casos de maneira apropriada. Portanto, esse método possibilita analisar os componentes de modo a garantir a confiabilidade dos dados, desta forma, reconhecendo um maior número de soluções. O segundo método é oferecido em diversas ferramentas disponíveis no mercado. Dentre as principais técnicas disponíveis, podem-se citar os métodos baseados em *checklist*, diferenças e explicação.

Esta etapa é importante para a construção de um sistema de RBC. Portanto, os pesos devem ser atribuídos com o auxílio de um especialista ou a partir da análise dos índices, isto é, através de requisitos para determinar a escolha e os valores dos índices. A figura 2.4 mostra os atributos e seus respectivos pesos. Neste sentido, os atributos são responsáveis por determinar o conhecimento do problema e a escolha dos pesos serve para definir os atributos mais relevantes para recuperar os casos armazenados na base. Ainda é possível observar a atribuição para o diagnóstico e ação da solução igual a zero, devido aos atributos não serem discriminantes², ou seja, não serem utilizados na recuperação dos casos.

² Atributos que são relevantes para a comparação dos casos.

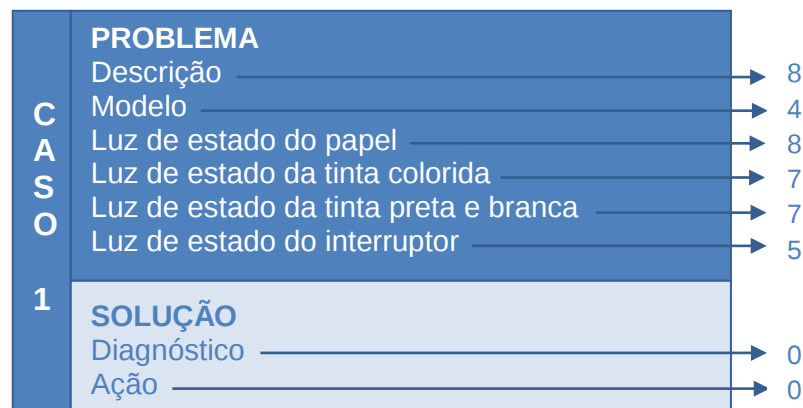


Figura 2.4. Modelo de indexação dos casos de um sistema de RBC.

2.1.2.3. Recuperação dos casos

A etapa de recuperação é responsável por encontrar um conjunto de casos semelhantes ao problema. Esta etapa contém tarefas importantes para que o sistema obtenha casos confiáveis, tais como:

- Identificar características do problema atual, com a finalidade de facilitar a recuperação dos casos;
- Recuperar um conjunto de casos contidos na base, a partir da descrição do problema;
- Buscar os casos armazenados na base, a fim de selecionar o melhor candidato, por meio dos casos recuperados;
- Selecionar e ordenar os casos, de acordo com o maior grau de similaridade, por meio de um dos cálculos de similaridade existentes, apresentado nesta seção (fórmula 2.1).

No RBC, a similaridade tem a função de classificar a percepção da semelhança existente entre dois casos. Conforme Júlio (2005), as métricas de similaridade podem ser consideradas de diferentes formas, como:

- Similaridade local e global: é classificada como similaridade global a semelhança que ocorre entre dois objetos de maneira intuitiva. Usa-se a similaridade local quando se busca uma avaliação mais específica das características entre um caso e um novo problema (GRESSE VON WANGENHEIM, 1999);
- Similaridade sintática e semântica: a similaridade sintática é a forma de avaliação que desconsidera elementos contextuais (AAMODT, 1994, *apud*

JÚLIO, 2005). A similaridade semântica é o tipo de avaliação onde os atributos são analisados e comparados em termos de semelhança sintática (JÚLIO, 2005);

- Similaridade estrutural, organizacional e pragmática: a similaridade estrutural demanda consistência estrutural entre os casos e em um dos nós conceitos. A similaridade organizacional tem relação à similaridade imposta aos casos gravados em locais próximos à base. A similaridade pragmática acontece quando duas partes exercem papéis parecidos em suas respectivas situações. (JÚLIO, 2005);
- Medidas cardinais e medidas ordinais: as medidas cardinais resultam em números que representam uma distância, enquanto as medidas ordinais definem uma ordem, como um *ranking* (JÚLIO, 2005).

A recuperação dos casos é um item essencial na construção de um modelo de RBC, cuja etapa tem como objetivo encontrar casos equivalentes ao problema em questão, por meio da atribuição correta dos componentes indexados e dos algoritmos de similaridade. Desta forma, as buscas são facilitadas, os casos são selecionados e classificados quanto a sua confiabilidade, resultando em soluções potencialmente úteis ao sistema.

Com a indexação, pode-se identificar características relevantes ao problema, através dos descritores³ informados, os quais são importantes por recuperar casos semelhantes à situação em questão. Segundo Gresse von Wangenheim e von Wangenheim (2003), quanto maior o número de componentes presentes no domínio⁴, maior é a quantidade de fatores a serem considerados nesta etapa, aumentando, assim, a complexidade no momento de recuperar novos casos. Esses fatores são descritos a partir da análise dos componentes, com a finalidade de encontrar padrões para facilitar o cálculo de similaridade.

Abel (1996) afirma que recuperar casos é uma tarefa complexa, desde a análise do julgamento até o processo de avaliação dos critérios utilizados para a recuperação de novas soluções. Segundo este autor, a utilização de um banco de casos adequado ao domínio, que contenha diversas experiências, permite recuperar casos, com eficiência, precisão e flexibilidade, de maneira satisfatória. Desta forma, somente os casos úteis ao problema em questão serão retornados, podendo ser adaptados às diversas situações. Recuperar casos relevantes ao objetivo do sistema é importante para a próxima etapa: a adaptação.

³ Atributo-valor utilizado na descrição de um caso (FERNANDES, 2005).

⁴ Conjuntos de componentes que descrevem um problema específico.

Essas análises são feitas através de métodos que otimizam a recuperação das experiências passadas com o intuito de reter em sua memória casos que possibilitem ser empregados para resolver problemas, através do cálculo de similaridade que organiza quais casos serão utilizados pelo sistema. Existem diversos métodos para realizar a recuperação dos casos, tais como:

- a) *Matching* e *Ranking*: para Kolodner (1993), o método *Matching* se propõe a realizar a comparação de casos e do problema em questão, determinando o grau de equivalência entre eles. E o método *Ranking* tem como finalidade classificar os casos conforme sua utilidade, verificando os melhores candidatos a partir da semelhança do problema;
- b) Vizinho mais Próximo: segundo Gresse von Wangenheim e von Wangenheim (2003), o objetivo desse método é visitar, como o próprio nome sugere, o vizinho mais próximo. Para tal, são utilizados algoritmos que medem a distância entre os casos visitados na base, aumentando, assim, o desempenho do sistema, tendo em vista que não realizará a comparação de todos os casos da base, mas sim aqueles que serão visitados, a partir da distância estabelecida pelo algoritmo.

De acordo com Fernandes (2005), existem diferentes algoritmos para o algoritmo de recuperação pelo Vizinho mais Próximo, como:

- Distância Euclidiana: mede a distância geométrica em um espaço multidimensional;
- Distância Euclidiana Quadrada: é o quadrado da distância euclidiana padrão;
- Distância *City-block* (Manhattan): serve para medir a distância entre as dimensões;
- Distância Percentual: medida muito útil para dados em que a dimensão analisada for natural.

O uso de um dos métodos de recuperação apresentados possibilita encontrar, na base de casos, uma solução semelhante ao problema corrente. Para tal, o sistema verifica o grau de similaridade dos casos existentes na base, como ilustra a figura 2.5, através da aplicação da fórmula de similaridade pelo Vizinho mais Próximo (fórmula 2.1).

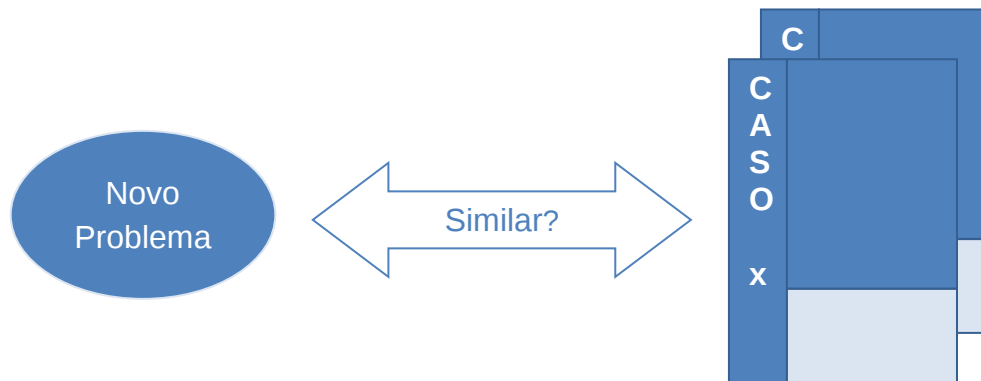


Figura 2.5. Modelo de similaridade de um sistema de RBC.

Um dos cálculos mais aplicados para encontrar a similaridade entre os casos é a fórmula de similaridade pelo Vizinho mais Próximo:

$$Similaridade(N, F) = \sum_{i=1}^n f(N_i, F_i) * w_i$$

Fórmula 0.1. Fórmula de similaridade pelo Vizinho mais Próximo.
Fonte: Fernandes (2005).

Conforme Fernandes (2005), a fórmula de similaridade pelo Vizinho mais Próximo tem como variáveis:

N = Novo caso;

F = Casos existentes na memória de casos;

n = Número de atributos;

i = Atributo individual variando de 1 a n;

f = Função de similaridade para o atributo i nos casos N e F;

w = Peso relativo ao atributo i.

Esta fórmula é repetida para cada caso do banco com o intuito de obter-se um *ranking*. As similaridades são comumente normalizadas em um intervalo entre zero e um (zero quando não há similaridade e um quando a similaridade é exata). No entanto, existe certa dificuldade na determinação dos pesos referentes às características. A limitação deste método é a convergência para a solução correta e a quantidade de casos recuperados. No geral, o tempo de recuperação aumenta proporcionalmente com o número de casos (KOSLOSKY, 1999).

2.1.2.4. Adaptação dos casos

A etapa de adaptação de casos utiliza métodos que verificam e analisam a melhor solução para o problema, a partir dos mecanismos que precedem esta fase. Desta forma, os mecanismos definem estratégias para adaptar a solução, através dos casos recuperados potencialmente úteis.

Quando um caso é fornecido ao sistema, o algoritmo de recuperação encontra eventos pertinentes (semelhantes) ao problema. Logo, devem-se analisar os índices dos casos, a fim de verificar as diferenças entre as características, conforme os mecanismos utilizados para a adaptação e, com isso, aplicar a melhor regra para que o sistema de RBC seja capaz de fornecer uma boa solução. Portanto, quando houver a necessidade de reformular parte de uma solução antiga, de acordo com a regra utilizada, é necessário adaptar os casos a partir dos atributos indexados, os quais servirão como base para a modificação de casos existentes. Desta forma, o processo de adaptação dos casos deve ser confiável para ser capaz de indicar sucesso ou insucesso de uma solução gerada pelo sistema.

Segundo Vergara (1995), os sistemas de RBC apresentam dois tipos de adaptação generalizados:

- a) Adaptação Estrutural: este tipo de adaptação é executado uma única vez pelo sistema, e seu funcionamento consiste na substituição de regras dos atributos e transformação de casos antigos selecionados pelo sistema;
- b) Adaptação Derivacional: semelhante à adaptação estrutural. Entretanto, esta adaptação exige maior compreensão do problema, fazendo com que o sistema a execute duas vezes, a partir das regras e mecanismos que são utilizados com a finalidade de gerar novas soluções.

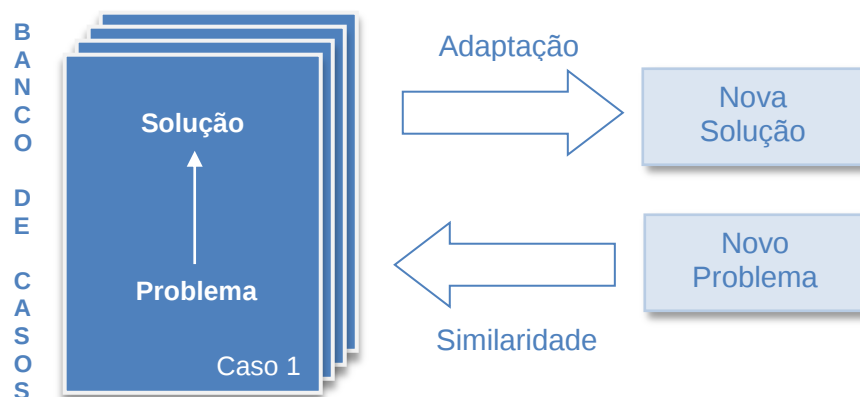


Figura 2.6. Modelo de adaptação de um sistema de RBC.

A figura 2.6 demonstra a adaptação de um sistema de RBC. Neste contexto, a adaptação tem como objetivo modificar casos existentes, se necessário, para solucionar novos problemas, levando em consideração as diferenças entre os casos recuperados (de acordo com o grau de similaridade) e a situação em questão.

Vergara (1995) também informa que existem quatro técnicas específicas para adaptação de um caso, desde as mais simples até as mais complexas. As técnicas são:

- a) Adaptação nula ou estrutural: esta técnica é nula, portanto apenas aplica a solução do caso mais similar recuperado à nova situação;
- b) Soluções parametrizadas: é a comparação específica do caso e do problema, utilizando as diferenças encontradas para modificar a solução dos parâmetros no caso recuperado;
- c) Abstração e Especialização: é uma técnica estrutural que realiza simples adaptações para gerar soluções de casos complexos;
- d) Reinstalação: é uma técnica derivacional que não opera na solução original, mas sim nos métodos utilizados para produzir a solução.

2.1.2.5. Aprendizagem

Conforme Lagemann (1998), a etapa de aprendizagem tem como característica a retenção de casos na memória e indexação correta dos problemas. Em outras palavras, é o acúmulo de novas experiências relativamente úteis ao objetivo do sistema, a partir da atribuição correta aos índices dos casos. O autor também informa que um sistema de RBC apenas será eficiente quando for capaz de aprender através de experiências passadas e indexar corretamente os problemas.

De acordo com Fernandes (2005), para facilitar a aprendizagem, o sistema deve ser competente em identificar as informações corretas e processá-las assim que disponíveis, a fim de realizar uma análise dos componentes indexados e dos casos que obtiveram sucesso ao longo do tempo, auxiliando na aprendizagem do sistema com a inclusão de casos importantes. Logo, o processo de aprendizagem deve especificar o que deseja aprender para aumentar as possibilidades do sistema obter sucesso, ou seja, encontrar uma solução confiável ao problema.

Com o aumento no número de casos adicionados na base, é necessária a validação desses novos conhecimentos, por meio de um especialista, ou até mesmo por um administrador responsável pelo sistema, a fim de armazenar apenas as soluções confiáveis para que posteriormente, sejam utilizadas em situações similares.

Para Koslosky (1999), os casos passados fazem com que o sistema de RBC tome decisões e aprendam sob três maneiras distintas:

- a) Generalização e especialização: a generalização é uma regra aplicada para analisar ou resolver novos problemas e na especialização da base de conhecimento;
- b) Pesquisa restringida: esta pesquisa ajuda a recuperar soluções potencialmente úteis ao objetivo do sistema através dos casos indexados na base. Um modelo de RBC apresenta índices de entrada para recuperar uma solução relevante e, posteriormente, adaptá-la para solucionar um problema;
- c) Avaliação comparativa: um problema pode ser analisado e comparado por meio das características dos casos existentes na base. Desta forma, o sistema aprende por indução, isto é, abstrai regras gerais por meio de conhecimentos específicos.

2.1.3. Estrutura de um sistema de RBC

A estrutura de um sistema de RBC precede de três itens fundamentais: memória de casos de domínio, mecanismos de pesquisas e descrição dos índices para diferenciar os casos.

2.1.3.1. Memória de casos de domínio

A base de casos é composta por um conjunto de acontecimentos que representam as experiências passadas de um especialista, sendo que o sistema aprende com as informações concedidas em forma de conhecimento. A memória compreende os casos e mecanismos de acesso ao conhecimento (KOLODNER, 1993).

Fernandes (2005) ainda informa que os modelos de memória nada mais são que a estrutura de organização dos casos. Assim, os modelos têm o intuito de organizar experiências passadas que obtiveram sucesso e insucesso ao longo do tempo, para proporcionar eficiência no processo computacional, por meio dos mecanismos utilizados para recuperar soluções úteis aos problemas.

Lee (1998) propõe dois aspectos que devem ser considerados para a escolha da modelagem de memória. O primeiro analisa o tipo de filosofia que reproduz a simulação de um determinado sistema, podendo ser rede semântica, memória episódica ou memória dinâmica. E o segundo trata do tipo de organização que será adotada para os casos, como a estruturação da memória de casos.

O mesmo autor também menciona que existem diversos tipos de modelo de memória utilizada para representação do conhecimento: redes semânticas, frames, representação em árvores, grafos conceituais, entre outras. Em meio aos mais utilizados, estão:

- a) Memória Dinâmica: utiliza uma estrutura hierárquica de Pacotes de Organização de Memória (POM, em inglês, *Memory Organization Packets* – MOP) para organizar os casos. Dalfovo *et al.* (1998) preconizam que os POMs são compostos de frames que constituem uma unidade básica de Memória Dinâmica;
- b) Memória Prototípica: também conhecida como memória de categoria e exemplares, sendo que este tipo de memória é recomendado, principalmente, aos sistemas que fazem classificação. Conforme Fernandes (2005), a memória prototípica classifica os casos armazenados em categorias, proporcionando um aumento na velocidade de recuperação dos casos associados a uma mesma categoria. Outro benefício deste modelo é possibilitar a construção de uma memória que compreenda todo um subdomínio.

Para que um sistema de RBC seja bem sucedido, deve-se analisar qual memória é mais indicada ao objetivo do sistema, pois a escolha adequada da memória proporciona melhor rotulação e organização dos casos contidos na base.

2.1.3.2. Mecanismo de Pesquisa

Pesquisar é fundamentar-se com informações obtidas através de bibliografias. Desta forma, o conhecimento é aprimorado conforme o aumento do número de assuntos consultados. Dalfovo *et al.* (1998) informam que as pesquisas são classificadas sob duas formas, dependendo do tipo de informações coletadas e das análises feitas com a obtenção desses dados:

- a) Pesquisa Quantitativa: é responsável por descobrir quantos casos contidos numa mesma base compartilham uma ou mais características;
- b) Pesquisa Qualitativa: esta pesquisa tem como base explorar as características dos casos para explicitar suas recuperações em relação às informações disponibilizadas pelo sistema.

2.1.3.3. Descrição dos índices para diferenciar os casos

Segundo Dalfovo *et al.* (1998), para construir um modelo de RBC é necessário, inicialmente, identificar os índices ou as características dos problemas; em seguida, selecionar os casos similares ao problema atual para que, posteriormente, adapte a solução eleita e utilize conforme a realidade da situação presente.

Os índices auxiliam o RBC a encontrar situações similares, ou seja, recuperar casos que possuem as mesmas características do problema em questão, através do cálculo de similaridade que tem a finalidade de verificar a equivalência entre os problemas e os casos existentes na base. Além disso, os índices também são responsáveis por caracterizar os casos para que sejam organizados corretamente na memória, a fim de facilitar na recuperação de casos úteis ao objetivo do sistema. Os índices dos casos são tão importantes quanto; por exemplo, os índices utilizados para identificar os livros em uma biblioteca.

2.1.4. Vantagens e desvantagens de um sistema de RBC

Em seguida, serão apresentadas as principais vantagens e desvantagens dos sistemas de Raciocínio Baseado em Casos.

2.1.4.1. Vantagens

Os indivíduos são conhecidos como seres pensantes, logo são capazes de resolver problemas que, muitas vezes, parecem não ter solução. O RBC simula a inteligência humana, sendo que quanto maior o número de experiências adquiridas, melhor seu desempenho. Assim, Fernandes (2005) aponta os seguintes motivos para se utilizar um sistema de RBC:

- Aquisição do conhecimento: o conhecimento de um sistema de RBC fica armazenado na própria base de casos;
- Manutenção do conhecimento: um usuário pode ser eleito para administrar os casos na base, desta forma, sem a necessidade da intervenção de especialistas;
- Eficiência crescente na resolução de problemas: o RBC armazena as soluções que obtiveram insucessos para evitar que futuras soluções não sejam bem sucedidas, assim como guarda as soluções bem sucedidas, evitando, assim, possíveis insucessos.

2.1.4.2. Desvantagens

Da mesma forma que o RBC possui vantagens para sua implementação por simular o raciocínio do homem, esta técnica apresenta também limitações semelhantes ao do ser humano, como conhecimento limitado e incerto, que necessita ser adquirido ao longo do tempo. Algumas dessas desvantagens são caracterizadas a seguir:

- Casos inconscientes: limitação em obter dados disponíveis e confiáveis;
- Ausência de conhecimento: os casos disponíveis não cobrem todo o domínio;
- Identificação dos índices: dificuldade em atribuir índices de acordo com a realidade do problema.

Esta seção abordou a técnica de Raciocínio Baseado em Casos, junto com o histórico, etapas de desenvolvimento, estrutura, vantagens e desvantagens dos sistemas de RBC. A seção seguinte tratará sobre o Geoprocessamento, desde os conceitos fundamentais até as técnicas de Sensoriamento Remoto, Digitalização de Dados e Sistemas de Informação Geográfica (SIG).

2.2. Geoprocessamento

Ao longo das últimas décadas, o Geoprocessamento tem sido bastante difundido em relação ao auxílio na tomada de decisão, e também na realização de análises complexas e verificação de fenômenos de natureza física e humana. Conforme Moura (2005), o termo Geoprocessamento surgiu com a intenção de processar dados georreferenciados, e aplicar um processo que apresente melhorias para representação dos espaços geográficos.

Para cumprir com o objetivo que o Geoprocessamento propõe, faz-se necessário uma atualização constante das informações geográficas. Logo, deve-se realizar a captura de imagens através de sensores acoplados em satélites de detecção remota, por meio de técnicas de Sensoriamento Remoto. Esperando uma melhor compreensão do conteúdo trabalhado, a seguir são descritos alguns conceitos e fundamentos básicos acerca do assunto.

Segundo Carvalho Filho *et al.* (2000), Geoprocessamento é o conjunto de tecnologias computacionais relacionadas com a coleta, tratamento, manipulação, análise e apresentação de dados espaciais e geográficos, bem como as técnicas de Sensoriamento Remoto, Digitalização de Dados e Sistemas de Informação Geográfica (SIG). Esta técnica é capaz de realizar um complexo processamento digital de imagens, por meio de mapas e Banco de Dados (BD) georreferenciados, com o intuito de reunir tudo em um SIG.

Desta forma, entende-se Geoprocessamento como um conjunto de técnicas voltadas para a manipulação de dados georreferenciados, os quais serão utilizados em sistemas específicos, onde a informação da localização geográfica será relacionada a atributos dentro de um BD.

De acordo com Miranda (2005), o georreferenciamento é uma atividade do Geoprocessamento, com o objetivo de localizar um determinado espaço geográfico através do Sistema de Posicionamento Global (*Global Position System* – GPS), a partir da transmissão precisa de informações de posicionamento, permitindo, desta forma, a localização de pontos na superfície terrestre. Por meio das informações obtidas, é possível realizar a manipulação de mapas ou imagem geográfica, através do ponto referência de localização, denominado como pontos de controle. Portanto, nota-se que o GPS é imprescindível para a área de Geoprocessamento, tendo em vista que, dificilmente, um projeto não necessita realizar a coleta de dados em campo.

Os pontos de controle são responsáveis por informar ao sistema a localização do espaço para realizar a manipulação das informações apresentadas. A obtenção das coordenadas dos pontos de controle pode ser obtida em campo, a partir de levantamentos topográficos, com o auxílio de GPS, ou ainda por meio de mesas digitalizadoras, outras imagens e mapas (em papel ou digitais).

2.2.1. Sistemas de Informação Geográfica

É o termo dado ao Sistema de Informação Geográfica (SIG, em inglês, *Geographic Information Systems* – GIS) para o software que trata a geoinformação. O SIG possibilita a confecção de mapas, o georreferenciamento, processamento e visualização das informações (FURTADO, 2002).

Miranda (2005) afirma que o conceito de Sistemas de Informação Geográfica mudou consideravelmente ao longo das últimas décadas, conforme o uso destes sistemas nas diversas áreas do conhecimento humano. Logo, geraram-se muitas definições, um tanto quanto controversas, o qual resultou em um problema para a comunidade de SIG.

Neste contexto, busca-se descrever as definições de SIG relacionadas à área da computação com o estudo de Geoprocessamento no Brasil. A seguir são descritos alguns destes fundamentos básicos:

“[...] SIG como sistema de processamento (análise) e visualização de mapas. No processamento de mapas, cada arquivo representa um mapa (também chamado um plano de informação, tema ou cobertura). Os mapas são usualmente armazenados em um formato padrão e manipulados por uma

função que pode adicionar, subtrair ou pesquisar por padrões. A saída destas operações é outro mapa” (MAGUIRE, 1991, *apud* MIRANDA, 2005, p. 26).

Uma definição mais adequada à área computacional conceitua SIG como, segundo Miranda (2005, p. 25), “um sistema de computador que coleciona, edita, integra e analisa informação relacionada à superfície da Terra”.

Finalmente, chega-se a outra definição que parece ser a mais completa e abrangente sobre este assunto. Ramirez (1994) define SIG como um conjunto de ferramentas computacionais e procedimentos desenvolvidos para dar suporte à aquisição, gerenciamento da persistência (inserção, consulta, alteração, remoção) de dados geoespaciais, que descrevem pontos na superfície da Terra. Nota-se, então, que os SIGs caracterizam-se pela captura, modelagem, manipulação, recuperação, análise e apresentação de dados georreferenciados.

Borghetti (2006) ressalta que a rápida expansão dos SIGs está ligada aos avanços computacionais e seu potencial para solucionar problemas das mais diversas temáticas, como estudos de uso da Terra, topografia, geologia, análise ambiental, clima, solos, entre outros. Com isso, houve a necessidade de construir SIGs voltados para a resolução de problemas de outras áreas, denominados como SIGs genéricos.

Entende-se por SIG genérico como uma classe ou tipo de sistemas de informação. Este tipo de sistema trata de informações referenciadas espacialmente à Terra. Existem também sistemas de informação espaciais aplicados; por exemplo, à área de Medicina que não devem ser confundidos com os SIGs.

2.2.2. Equipamentos usados para entrada e saída de dados

A seguir, serão expostos os principais procedimentos de entrada de dados e as formas de apresentação dos dados de saída utilizados em um SIG.

2.2.2.1. Equipamentos de entrada de dados

Conforme Miranda (2005), os dispositivos de entrada de dados servem para converter dados analógicos em mídia eletrônica ou digital para serem utilizados no SIG, entre os equipamentos mais comuns, observados na figura 2.7, estão:

- Mesas digitalizadoras: permitem transformações de informações analógicas para digitais através de processo manual;
- Rastreadores óticos (*scanners*): tem a mesma função das mesas digitalizadoras, entretanto são mais modernos e exigem pouca interação com o usuário.

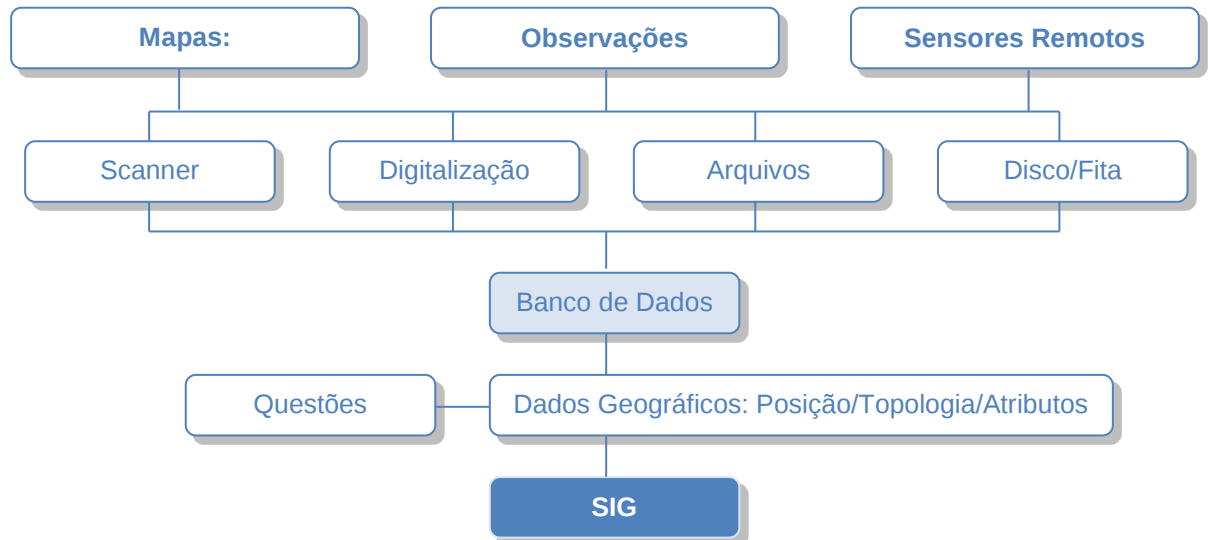


Figura 2.7. Equipamentos de entrada de dados em um SIG.
Fonte: Adaptado de Silva (2000).

2.2.2.2. Equipamentos de saída de dados

Miranda (2005) ainda cita que os dispositivos de saída de dados são responsáveis por exibir dados e informações geradas pelo SIG, entre os tipos mais comuns, vistos na figura 2.8, estão:

- Monitores: permitem a visualização de dados;
- Impressoras: imprimem os resultados processados pelos SIG, como por exemplo, gráficos, mapas, relatórios e etc.;
- Traçadores gráficos (*plotters*): produzem cópias de dados geográficos (em papel ou filme) de alta qualidade.

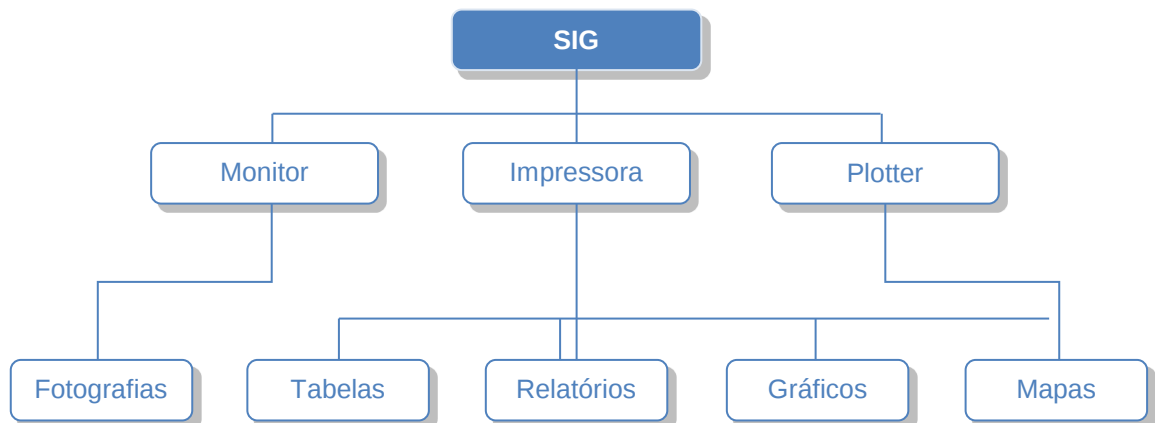


Figura 2.8. Equipamentos de saída de dados em um SIG.
Fonte: Adaptado de Silva (2000).

Após a abordagem dos conceitos fundamentais de Geoprocessamento, bem como os Sistemas de Informação Geográfica e seus respectivos equipamentos de entradas e saídas de dados, o tópico subsequente tratará sobre epidemiologia da leptospirose humana.

2.3. Epidemiologia da Leptospirose Humana

Segundo Almeida Filho e Rouquayrol (1990), a epidemiologia é a ciência que estuda o a distribuição dos fenômenos de saúde-doença⁵ na sociedade. Esta ciência tem como método básico a análise de frequência, distribuição e determinantes das enfermidades e de seus agravos à saúde coletiva.

Nesta seção, apresenta-se a bactéria *Leptospira interrogans*, responsável por causar a leptospirose. Dentre os assuntos oferecidos, serão vistos o agente etiológico, reservatório, modo de transmissão, período de incubação, período de transmissibilidade, susceptibilidade e imunidade, aspectos clínicos e laboratoriais, como manifestações clínicas e diagnósticos laboratoriais.

2.3.1. Agente etiológico

Segundo Paula (2005), a leptospirose, também denominada como síndrome de Weil ou popularmente conhecida como “doença do xixi do rato”, é uma antroponose, isto é, doença transmissível ao homem de forma secundária. Trata-se de uma doença infecciosa febril, potencialmente grave, causada por microorganismos pertencentes à espécie *Leptospira*, cujo período de incubação varia de 1 a 30 dias, com tempo médio de uma a duas semanas após a infecção.

Este agravo é encontrado tipicamente em países subdesenvolvidos ou em desenvolvimento, em continentes como Ásia, América do Sul e Central, cujos sistemas de infraestrutura e saneamento básico se encontram em situação precária ou inexistente. No Brasil, levando em consideração o tamanho populacional, as áreas mais favorecidas para transmissão concentram-se na região Amazônica.

No entanto, de acordo com dados obtidos do Ministério da Saúde (2011), as regiões sul e sudeste do Brasil, principalmente, devido à densidade demográfica, têm maior incidência nos casos diagnosticados de leptospirose. Nos últimos dez anos do levantamento, foram registrados 38.161 casos da doença no Brasil, como demonstra a tabela 2.1. Apesar do

⁵ “Saúde-doença é uma expressão que indica a extensão que separa o estado de vida de completo bem-estar – físico, mental e social – e o de doença, intermediados por todas as variações possíveis entre um e outro” (BELLUSCI, 1995, p. 25).

número alarmante, Martins e Castiñeiras (2009) informam que os casos notificados, provavelmente, representam somente 10% do total de casos existentes no país.

Casos confirmados de leptospirose no Brasil: 2001 – 2011

Região	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011*	Total
Norte	142	227	248	224	272	752	247	314	358	260	295	3.339
Nordeste	651	638	514	807	746	679	542	611	922	713	492	7.315
Sudeste	1.222	957	999	1.318	1.363	1.693	1.235	1.019	1.483	1.491	1.180	13.960
Sul	1.649	907	1.192	673	1.088	1.175	1.263	1.553	1.081	1.219	1.205	13.005
Centro-oeste	44	40	52	74	65	70	33	52	45	49	18	542
Total	3.708	2.769	3.005	3.096	3.534	4.369	3.320	3.549	3.889	3.732	3.190	38.161

* dados preliminares sujeito à revisão

Tabela 2.1. Casos confirmados de leptospirose no Brasil: 2001 - 2011.

Fonte: SINAN/SVS/MS, 2011.

O estado brasileiro que tem o maior índice de casos confirmados (8.271) e óbitos (1.080) por leptospirose é São Paulo, sobretudo por concentrar a maior parte da população do país. Em seguida, os estados com maiores notificações confirmados são Rio Grande do Sul e Santa Catarina, com 5.481 e 4.546 casos, respectivamente. No entanto, o Rio de Janeiro é o segundo em número de óbitos por este agravo, com 574. Enquanto isso, o Pará foi o nono estado com maior número de casos no Brasil e o sétimo que mais vitimou pessoas, conforme são apresentados na tabela 2.2 (MINISTÉRIO DA SAÚDE, 2011).

Casos de leptospirose nos dez estados mais afetados do Brasil: 2001 - 2011

Situação	SP	RS	SC	PR	RJ	PE	ES	BA	PA	CE	Total
Confirmado	8.271	5.481	4.546	3.008	2.972	2.785	1.831	1.546	1.347	1.000	32.787
Cura	7.191	5.148	4.336	2.618	2.398	2.434	1.744	1.296	1.157	902	29.224
Óbito	1.080	333	210	390	574	351	87	250	190	98	3.563
Total	9.351	5.814	4.756	3.398	3.546	3.136	1.918	1.796	1.537	1.098	36.350

Tabela 2.2. Casos de leptospirose nos dez estados mais afetados do Brasil: 2001 - 2011.

Fonte: SINAN/SVS/MS, 2011.

De acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE, 2011), considerando a proporção entre os casos confirmados de leptospirose e o número de habitantes nos dez estados mais afetados do Brasil (vide tabela 2.3), o Pará divide a sexta posição com Rio de Janeiro e São Paulo, atrás apenas de Santa Catarina, Espírito Santos, Rio Grande do Sul, Paraná e Pernambuco, respectivamente, entre os estados com maior relato de contaminação da doença. Este cenário foi um dos fatores determinantes para a escolha da área de estudo.

Leptospirose nos dez estados mais afetados do Brasil			
Estados	População	População (%)	Casos
SP	41.262.199	0,02	8.271
RS	10.693.929	0,05	5.481
SC	6.248.436	0,07	4.546
PR	10.444.526	0,03	3.008
RJ	15.989.929	0,02	2.972
PE	8.796.448	0,03	2.785
ES	3.514.952	0,05	1.831
BA	14.016.906	0,01	1.546
PA	7.581.051	0,02	1.347
CE	8.452.381	0,01	1.000

Tabela 2.3. Casos de leptospirose nos dez estados mais afetados do Brasil: 2001 - 2011.
Fonte: IBGE e SINAN/SVS/MS, 2011.

Na região Norte do Brasil houve estabilidade dos casos confirmados de leptospirose entre os anos de 2001 a 2005, com um elevado aumento em 2006, voltando a ficar estáveis nos anos de 2007 e 2008. Contudo, com as fortes chuvas e, conseqüentemente, alagamentos, devido ao lixo que a própria população deposita nas ruas, entre outros fatores, esse número não foi erradicado, como mostra o gráfico 2.1. Desta forma, esta pesquisa teve como exemplo norteador os casos de leptospirose humana dos anos de 2008, 2009 e 2010 nos bairros do Guamá, Jurunas e Montese, situados na capital do estado do Pará.

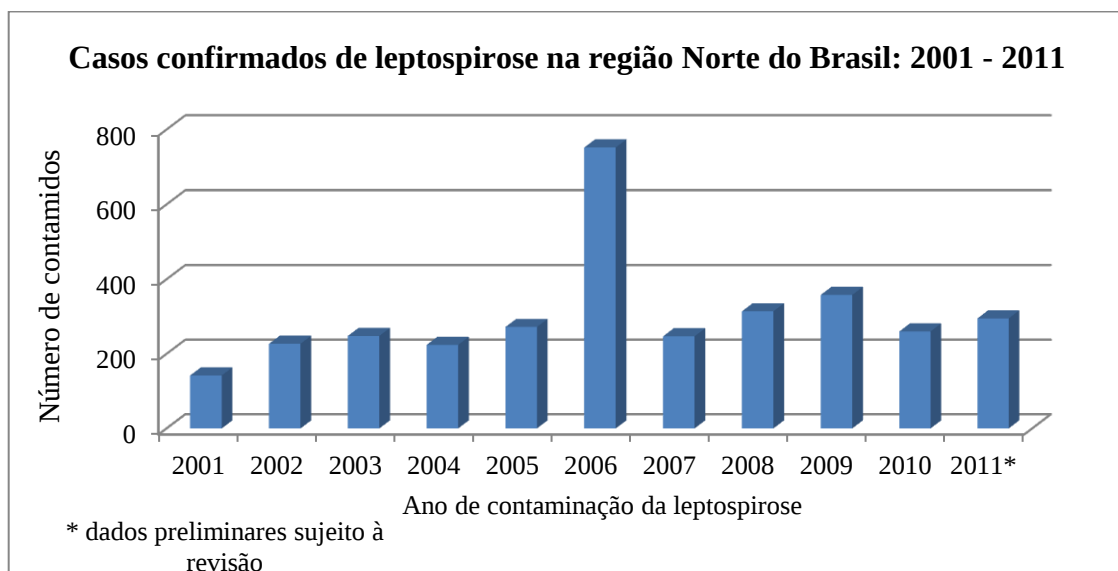


Gráfico 2.1. Casos de leptospirose na região Norte do Brasil, do ano de 2001 a 2011.
Fonte: SINAN/SVS/MS, 2011.

2.3.2. Reservatório

Animais domésticos e silvestres são os reservatórios fundamentais para existência dos focos da doença. Neste caso, o ser humano é apenas hospedeiro acidental e terminal dentro da cadeia de transmissão, como representa a figura 2.9 (DEPARTAMENTO DE VIGILÂNCIA EPIDEMIOLÓGICA, 2005).

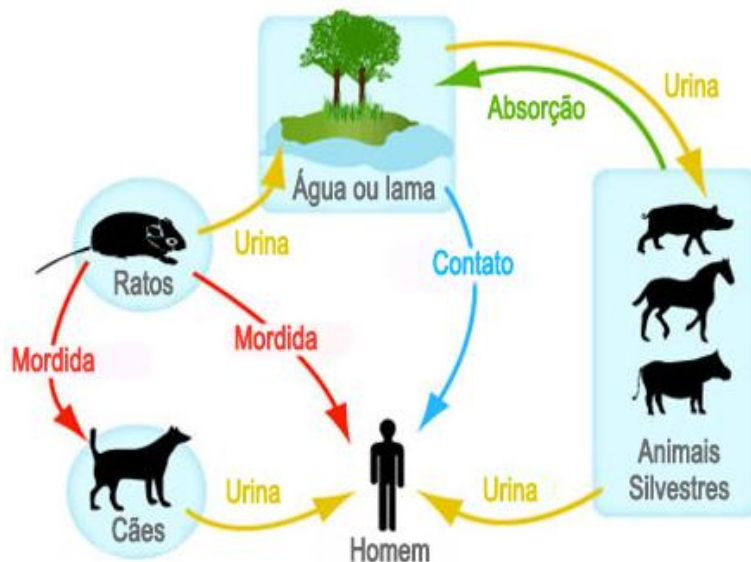


Figura 2.9. Ciclo de vida da leptospirose.
Fonte: Adaptado de L'Internaute (2011).

Conforme o Departamento de Vigilância Epidemiológica (2005), a bactéria *Leptospira interrogans* tem como principais reservatórios animais sinantrópicos, constituídos pelas espécies *Rattus norvegicus* (ratazana ou rato de esgoto), *Rattus rattus* (rato de telhado ou rato preto) e *Mus musculus* (camundongo ou catita). Nos casos dos roedores, ao serem contaminados, não desenvolvem a doença e tornam-se transmissores, portando a bactéria nos rins e eliminando-a ainda com vida no meio ambiente, desta forma, contaminando as superfícies ao entrar em contato, como água, solo e alimento. Contudo, a bactéria também pode estar presente em outros grupos, como caninos, bovinos, caprinos, equinos, ovinos e suínos.

2.3.3. Modo de transmissão

Veronesi (2005) informa que a contaminação ao homem é propiciada pela exposição direta ou indireta à excretas de roedores infectados, como urina e fezes. A transmissão da doença é mais presente em períodos chuvosos, onde rios, lagos, córregos e a própria rede de esgoto transbordam. Essas águas de enchentes invadem as tocas de ratos, fazendo com que

saíam à procura de novos abrigos, período em que ocorre maior incidência do surto. A leptospirose é transmitida ao homem, especialmente por contato com água ou lama, misturando-se à urina de ratos contaminados, sobretudo, se apresentar cortes ou arranhões na pele, mucosas dos olhos, narinas e boca. Outro meio de infecção é através da ingestão de alimentos, medicamentos e água (consumível pelo homem) contaminados.

2.3.4. Período de transmissibilidade

De acordo com o Departamento de Vigilância Epidemiológica (2005), o período de transmissibilidade da leptospirose é gerado a partir da infecção dos animais, cuja eliminação da *Leptospira* ocorre através da urina e permanece durante meses, ou até mesmo anos no meio ambiente. O ser humano pode ser fonte de infecção; entretanto é extremamente rara, podendo acontecer pelo contato com urina, sangue, secreções e tecidos de pessoas infectadas.

2.3.5. Suscetibilidade e imunidade

O Departamento de Vigilância Epidemiológica (2005) afirma que, em geral, todo indivíduo é suscetível à infecção por leptospirose. Inclusive uma mesma pessoa pode contrair a doença mais de uma vez, haja vista que a imunidade adquirida com a contaminação é de sorotipo específico, ou seja, o ser humano fica imune apenas de um sorovar (sorotipo) dos muitos tipos existentes.

2.3.6. Aspectos clínicos e laboratoriais

No subtópico seguinte, serão abordadas as informações técnicas da leptospirose humana, como as manifestações clínicas e os diagnósticos epidemiológico e laboratorial.

2.3.6.1. Manifestações clínicas

Segundo Veronesi (2005), as manifestações clínicas da leptospirose humana podem se manifestar sob duas formas:

- a) Anictérica (Fase de Leptospirosemia): responsável por 90% a 95% dos casos. Os sintomas característicos desse quadro são febres altas, calafrios, cefaléia intensa e mialgia, principalmente, na panturrilha.
- b) Ictérica (Síndrome de Weil): corresponde ao restante dos casos notificados, em torno de 5% a 10% dos casos, sendo considerada a forma grave de leptospirose. Este quadro apresenta os mesmos sintomas descritos anteriormente,

acompanhado de severa disfunção hepática, onde a icterícia é o sinal relevante e associado à disfunção renal, hemorragias pulmonares, alterações cardíacas, o que geralmente, associa-se à taxa de mortalidade elevada.

O autor ainda informa que, no geral, o quadro clínico típico da leptospirose humana é caracterizado por febre alta, mialgia, cefaleia, prostração, congestão conjuntiva, dor na panturrilha, vômito e diarreia. Nos casos mais graves, podem surgir outros sintomas, tais como disfunção hepática, com aparecimento de icterícia, insuficiência renal, alteração respiratória, alteração cardíaca, fenômenos hemorrágicos e meningismo.

2.3.6.2. Diagnóstico epidemiológico

O Departamento de Vigilância Epidemiológica (2005) afirma que leptospirose humana apresenta distribuição mundial. No Brasil, é uma doença epidêmica, principalmente em períodos chuvosos, devido às enchentes acompanhadas à aglomeração populacional de baixa renda em condições inadequadas de serviços básicos de saúde, bem como infraestrutura, saneamento e o alto índice de roedores infectados com a bactéria.

Segundo Ishizuka (2008), o diagnóstico epidemiológico, quando conduzido adequadamente, identifica características determinantes diretas e indiretas de qualquer doença. O diagnóstico epidemiológico baseia-se no conhecimento ecoepidemiológico que permite caracterizar o surto na população estudada.

Considerando que esta doença pode ser diagnosticada sob dois aspectos, epidemiológico e laboratorial, o modelo de RBC proposto nesta dissertação tem como objetivo auxiliar na diagnose de casos relacionados ao diagnóstico epidemiológico, que analisa características espaciais, ambientais, socioeconômicas e, sobretudo, epidemiológicas, relacionadas à leptospirose humana e a área de estudo.

2.3.6.3. Diagnóstico laboratorial

Atualmente, há exames específicos para a diagnose da leptospirose, entre eles o método laboratorial, cuja escolha depende da fase evolutiva do paciente. Conforme o Departamento de Vigilância Epidemiológica (2005), as fases constituem-se de duas formas. Na fase aguda, isto é, durante o período febril, onde as bactérias podem ser visualizadas no sangue através de exame direto e na fase imune, momento no qual as bactérias podem ser encontradas na urina, cultivadas ou inoculadas. Pelos problemas encontrados na realização dos exames citados anteriormente, os métodos sorológicos são eleitos para o diagnóstico da

leptospirose. Os testes mais utilizados são Elisa-IgM, a macroaglutinação e a microaglutinação.

O Departamento de Vigilância Epidemiológica (2005) ainda alega que existem exames inespecíficos, os quais são relevantes para o diagnóstico e acompanhamento clínico da doença, como hemograma, coagulograma, transaminases, bilirrubinas, ureia, creatinina e eletrólitos, gasometria, elementos anormais e sedimentos no exame sumário de urina, raios X de tórax e eletrocardiograma.

2.3.7. Incidência da Leptospirose no Pará

Conforme Avila-Pires (2006, p. 3), “o costume de se atribuir os surtos epidêmicos a enchentes e ratos esconde, na verdade, cadeias ecológicas mais complexas”. O autor aborda a falsa correlação entre os surtos de leptospirose com os períodos chuvosos, onde ocorrem enchentes e alagamentos. Tanto para epidemiologistas, quanto para leigos, a dispersão da bactéria está relacionada ao tempo precipitado. Entretanto, a relação entre os surtos e enchentes sugere ser, no mínimo, controversa, como é demonstrado no capítulo 4, na seção 4.1.2.2, onde se observa que o período de maior incidência da doença nos bairros pesquisados não ocorre em épocas chuvosas.

De acordo com dados tabulados e analisados do Ministério da Saúde (2011), a notificação de casos confirmados de leptospirose no estado do Pará tem caído consideravelmente nos últimos anos. Este fenômeno é notado no gráfico 2.2.

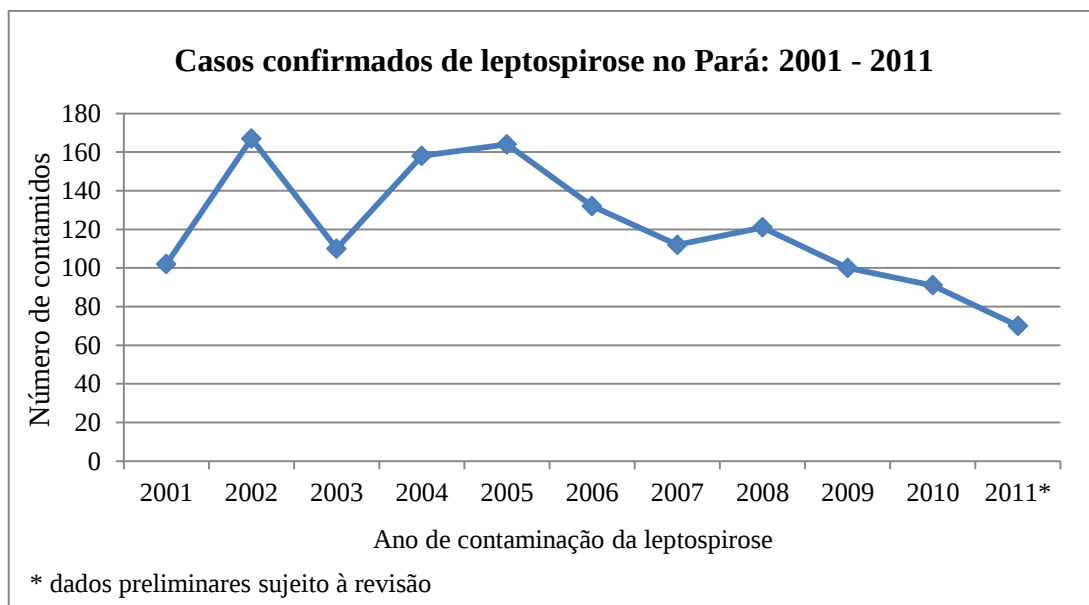


Gráfico 2.2. Casos de leptospirose no Pará, do ano de 2011 a 2011.
Fonte: SINAN/SVS/MS, 2011.

Os bairros do Guamá, Jurunas e Montese, situados na cidade de Belém, estado do Pará, são um dos que mais contribuem para esse índice negativo, como ilustra o gráfico 2.3. O saneamento básico precário encontrado nessas regiões e o baixo nível de escolaridade de seus habitantes são outros fatores alarmantes para tal evento.

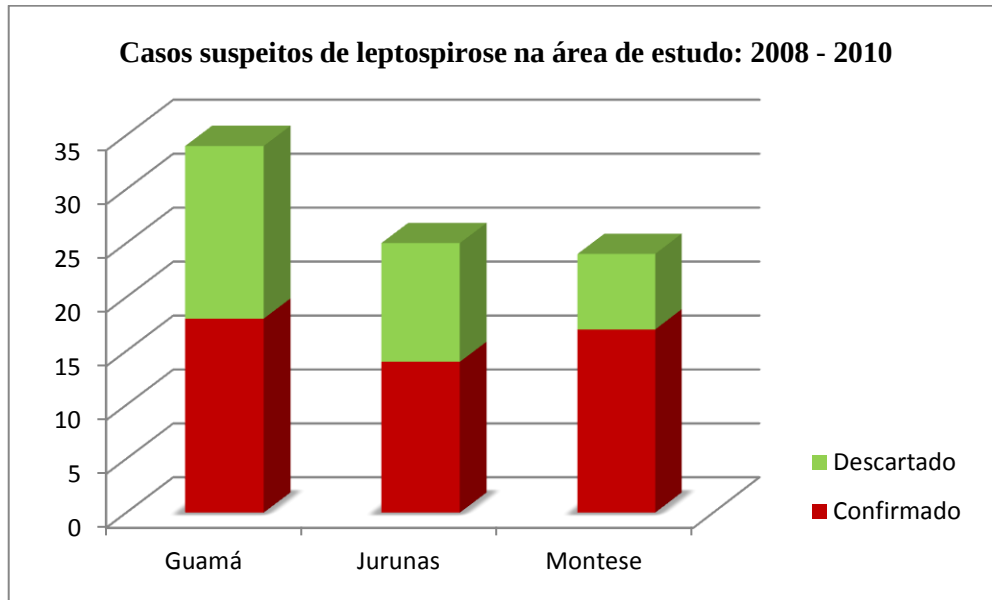


Gráfico 2.3. Casos suspeitos de leptospirose na área de estudo: 2008 - 2010.
Fonte: SINAN/SVS/MS, 2011.

Contudo, essas não são as únicas causas. Os alagamentos também são decorrentes da falta de infraestrutura básica, urbanização desenfreada, ausência de espaços verdes e acúmulo de lixo nas ruas, decorrentes da falta de coleta regular de lixo.



Figura 2.10. Fotos da urbanização desenfreada e ausência de infraestrutura básica.
Fontes: LabGeo/IEC/SVS/MS, 2008.

A figura 2.10a evidencia claramente a urbanização sem controle do bairro do Guamá, um dos motivos da proliferação da doença. Na figura 2.10b é possível observar as deficiências presentes neste bairro, aliados ao processo intenso de urbanização, por pessoas de baixa renda

e pouca instrução, além da ausência de saneamento e infraestrutura básica, onde as populações vivem em condições desumanas, facilitando assim, o alastramento de várias epidemias, como a leptospirose.

A leptospirose apresenta em seu quadro nosológico⁶ vários sintomas, quando não tratados, podem, inclusive, levar a óbito. Apresenta-se, na tabela 2.4, as manifestações clínicas de 47 pacientes contaminados com leptospirose, nos anos de 2008, 2009 e 2010, nos bairros em estudo. Os sintomas mais frequentes são mialgia, febre, cefaleia, prostração e vômito, os quais são comuns em muitas doenças endêmicas, entretanto quando ligados à icterícia, dor na panturrilha, alteração respiratória e outros sinais, geralmente, caracterizam a incidência da leptospirose.

Manifestações clínicas em 47 pacientes com leptospirose: 2008 - 2010						
Sinais e sintomas	Sim	%	Não	%	N/A	%
Mialgia	46	97,9	0	0,0	1	2,1
Febre	46	97,9	1	2,1	0	0,0
Cefaleia	41	87,2	3	6,4	3	6,4
Icterícia	39	83,0	3	6,4	5	10,6
Prostração	35	74,5	8	17,0	4	8,5
Dor na Panturrilha	35	74,5	9	19,1	3	6,4
Vômito	35	74,5	10	21,3	2	4,3
Congestão	24	51,1	18	38,3	5	10,6
Diarreia	24	51,1	20	42,6	3	6,4
Hemorragia Pulmonar	14	29,8	28	59,6	5	10,6
Outras Hemorragias	14	29,8	28	59,6	5	10,6
Alteração Respiratória	13	27,7	28	59,6	6	12,8
Insuficiência Renal	13	27,7	29	61,7	5	10,6
Alteração Cardíaca	7	14,9	31	66,0	9	19,1
Meningismo	2	4,3	36	76,6	9	19,1

Tabela 2.4. Manifestações clínicas em 47 pacientes com leptospirose: 2008 - 2010.
Fonte: SINAN/SVS/MS, 2011.

2.4. Conclusões

Este capítulo abordou a Fundamentação Teórica, que forneceu embasamento para esta dissertação, ao descrever as técnicas de Raciocínio Baseado em Casos e Geoprocessamento, e também, apresentar a bactéria *Leptospira interrogans*, causadora da leptospirose. O capítulo seguinte explorará o Estado da Arte dos Sistemas Baseados em Conhecimento, dando ênfase aos estudos feitos em RBC nos últimos anos.

⁶ Quadro clínico de uma doença.

Capítulo 3: Estado da Arte

Neste capítulo, serão explorados os cinco artigos mais pertinentes relacionados ao tema desta dissertação de mestrado. Os artigos foram selecionados minuciosamente entre diversas fontes de pesquisas secundárias e terciárias, coincidentemente, todos publicados no *Institute of Electrical and Electronics Engineers* (IEEE), para a formulação do Estado da Arte dos Sistemas Baseados em Conhecimento (SBC), com ênfase nos projetos de pesquisas recentes sobre Raciocínio Baseado em Casos e sua aplicação relacionada ao tema desta pesquisa.

O primeiro artigo denominado de “Sistema de Raciocínio Baseado em Casos para Diagnóstico Histopatológico”, cujo título original é “*Case-Based Reasoning System for Histopathology Diagnosis*”, tem como objetivo desenvolver um sistema de RBC para auxiliar os estudantes de Medicina no diagnóstico histopatológico (KATEDEE; SANRACH; THESAWADWONG, 2010).

O segundo artigo titulado de “Sistema de RBC para Diagnóstico de Pacientes”, que tem como título original “*CBR System for Diagnosis of Patients*”, possibilita a diagnose automática para pacientes com leucemia a partir de dados de *microarrays*⁷ (PAZ *et al.*, 2008).

O terceiro artigo nominado como “Sistematização da Assistência em Enfermagem Pediátrica aplicando Raciocínio Baseado em Casos” e o título original é “*Care Systematization in Pediatric Nursing Applying Case-Based Reasoning*”, objetiva desenvolver uma aplicação para auxiliar os profissionais de enfermagem no raciocínio clínico, conservando as suas experiências como uma coleção de casos para pesquisas futuras (MENDES *et al.*, 2009).

O quarto artigo intitulado de “Metodologia de Raciocínio Baseado Casos para o Diagnóstico de Gripe Suína”, no qual o título original é “*Case-Based Reasoning methodology for Diagnosis of Swine Flu*”, tem como fim desenvolver um sistema para auxiliar na diagnose da gripe suína (CHAKRABORTY *et al.*, 2011).

O quinto e último artigo chamado de “Diagnósticos de Esclerose Múltipla: Abordagem em Raciocínio Baseado em Casos”, tendo como título original “*Multiple Sclerosis Diagnoses: Case-Based Reasoning Approach*”, tem por finalidade desenvolver uma arquitetura para o sistema de diagnóstico da doença de Esclerose Múltipla (KURBALIJA *et al.*, 2007).

⁷ Conforme Paz *et al.* (2008), *microarray* ou *microarranjo* é uma técnica da Biologia Molecular, que permite medir os níveis de expressão de milhares de genes em um experimento.

Inicialmente, será realizado um resumo dos artigos pesquisados. Em seguida, haverá uma apreciação das suas características em contraste com o tema da dissertação para que os pontos fundamentais, que apoiam o objetivo deste trabalho, sejam abordados.

3.1. Sistema de RBC para Diagnóstico Histopatológico

Segundo Katedee, Sanrach e Thesawadwong (2010), desde a chegada e popularização dos computadores na sociedade contemporânea, o uso da tecnologia tem sido conciliado para o diagnóstico de doenças. Em paralelo, o RBC está experimentando um rápido crescimento e desenvolvimento na área da saúde. A Histopatologia⁸ é uma parte da patologia desafiadora, que exige certo tempo para o ensino e a aprendizagem. Tradicionalmente, são usados microscópios nos cursos de Histopatologia. Recentemente, foram desenvolvidas técnicas vídeo-microscópicas para aperfeiçoar a visualização de imagens e melhorar a experiência de aprendizagem dos alunos. Contudo, estas técnicas possuíam algumas deficiências. Neste contexto, foi proposto um sistema de RBC para o ensino de Histologia Patológica.

O RBC tem sido considerado como uma boa opção para recuperar casos de projetos existentes. Levando em consideração este cenário, os autores construíram um sistema de RBC, através do Protégé⁹, com o *plug-in* do myCBR¹⁰, fornecendo a funcionalidade de raciocínio, como por exemplo, a medida de similaridade.

Por ser uma ferramenta eficiente para trabalhar com ontologias, o Protégé é adequado para a gestão do conhecimento. Através desta ferramenta, os problemas podem ser classificados em classes e subclasses. Por exemplo, o tumor de mama pode ser dividido em diferentes classes, dependendo dos recursos e características. Uma das principais vantagens dos sistemas de RBC é permitir que os atributos da classe pai sejam herdados pelos da classe filho.

Para recuperar casos no RBC, a indexação dos casos é uma tarefa fundamental e precisa ser projetada cuidadosamente. O sistema de RBC, desenvolvido com o Protégé e o *plug-in* do myCBR, tem como tarefas principais projetar a ontologia (classe) e implementar medidas de similaridade para todos os atributos.

⁸ Para Katedee, Sanrach e Thesawadwong (2010), a Histopatologia ou Histologia Patológica é o estudo de como uma determinada doença afeta um conjunto de células.

⁹ De acordo com Katedee, Sanrach e Thesawadwong (2010), Protégé é uma ferramenta considerada eficiente, disponível gratuitamente na Internet. Esta plataforma de código aberto é um programa adequado para a construção de um modelo de ontologia baseada no conhecimento, que fornece bom suporte para a construção de modelos de conhecimento e inserção de dados.

¹⁰ O myCBR é um ambiente de desenvolvimento de RBC em Java, desenvolvido pelo Centro de Pesquisa Alemão de Inteligência Aplicada (myCBR: <http://mycbr-project.net>).

A pesquisa resultou no desenvolvimento de um sistema de RBC para o diagnóstico histopatológico. Para simular os processos por técnica de elementos finitos, os estudantes de Medicina também podem recuperar os casos de sucessos como referência, com o objetivo de se prepararem para as condições de simulação exigidas pela área.

3.1.1. Análise do artigo

Este artigo faz uma análise do uso da tecnologia para auxiliar na análise de doenças. Cabe destacar o crescimento da utilização dos recursos tecnológicos como forma de suporte ao diagnóstico médico. Levando em consideração este atual momento, o RBC tem se manifestado como uma alternativa satisfatória na resolução de casos voltados para a área da saúde.

O RBC, em sinergia com o advento e a popularização dos computadores na sociedade moderna, tem sido avaliado como uma opção para a resolução de problemas das mais variadas temáticas, como pode ser notado no caso de uso deste artigo. O desenvolvimento do sistema de RBC, na linguagem de programação Java, com o Protégé e o *plug-in* do myCBR é relacionado ao fato dessas ferramentas permitirem trabalhar com ontologias, o que as tornam apropriadas para administrar grandes volumes de dados.

O myCBR oferece uma interface que permite a usuários inexperientes criarem sistemas de RBC com facilidade. Também tem flexibilidade para admitir que usuários experientes possam desenvolver aplicações mais complexas. No entanto, esta ferramenta, quando comparada a outros *shells*¹¹, não tem a mesma robustez, como é o caso do CBR-Works¹², utilizado nesta pesquisa. Em contrapartida, o CBR-Works já foi um dos *shells* mais usados e poderosos existentes no mercado, mas entrou em desuso, pois a Empolis, empresa responsável por desenvolver o software, descontinuou o projeto.

3.2. Sistema de RBC para Diagnóstico de Pacientes

Segundo Paz *et al.* (2008), *microarray* tornou-se uma ferramenta essencial na pesquisa genômica, tornando-se possível investigar a expressão gênica global em todos os aspectos relacionados a doença humana. Para os autores, a técnica de *microarray* é baseada em um Banco de Dados de fragmentos de gene chamado Etiquetas de Sequências Expressas (originalmente, conhecido como *Expressed Sequence Tags* – ESTs), que são usados para

¹¹ De maneira genérica, os *shells* são ferramentas que fazem a intermediação entre o usuário e a máquina.

¹² O CBR-Works é um ambiente de desenvolvimento de RBC, desenvolvido pela TecInno e Empolis (Empolis: <http://www.empolis.com>).

identificar as transcrições do gene, sendo de fundamental importância na descoberta de genes e determinação da sequência genética.

O processo de estudo de um *microarray* é chamado de análise de expressão e consiste em uma série de fases: coleta de dados, pré-tratamento de dados, análise estatística e interpretação biológica. Essas fases são basicamente divididas em: normalização e filtragem; agrupamento e classificação. As fases podem ser automatizadas e incluídas em um sistema de RBC. O primeiro passo é fundamental para alcançar tanto uma normalização de dados e realizar uma filtragem inicial para reduzir a dimensão do conjunto de dados que serão trabalhados. É importante aplicar uma técnica de pré-processamento que facilite e automatize a tomada de decisões sobre as variáveis que serão vitais para o processo de classificação. Além disso, a escolha de uma técnica de agrupamento permite agrupar os dados de acordo com as variáveis que determinam o comportamento do grupo. Depois de organizar em grupos, é possível extrair o conhecimento e classificar os pacientes que apresentam maiores similaridades.

Atualmente, a técnica de *microarray* está sendo habitualmente utilizada para o diagnóstico de câncer, como leucemias. A leucemia é uma doença que tem um potencial significativo para a cura se detectado precocemente. Esta doença é uma forma de câncer no sangue, provenientes de um mau funcionamento da medula óssea que tende a produzir células anormais e de diminuir a produção de células normais. A pesquisa propõe o desenvolvimento de um sistema de RBC, que fornece uma técnica de classificação baseada em experiências anteriores, a partir da detecção de padrões cancerígenos dos pacientes em dados de *microarrays*.

O objetivo da pesquisa é desenvolver um sistema de RBC que permita a identificação de pacientes com vários tipos de câncer. O modelo visa melhorar a classificação do câncer com base em dados em *microarrays*. O sistema proposto neste artigo apresenta uma nova síntese que traz vários subcampos da Inteligência Artificial (técnicas de filtragem, *clustering*¹³, Redes Neurais Artificiais¹⁴ e extração de conhecimento¹⁵). As vantagens de aplicar novos métodos para análise de dados são para filtrar e agrupar os dados. O primeiro método combina várias técnicas de filtragem para reduzir drasticamente a dimensão dos

¹³ *Clustering* é uma tarefa descritiva de Mineração de Dados, que pode ser definida como o processo de organização de entidades em *clusters* pelo seu grau de similaridade. Isto é, são grupos cujos elementos são similares de alguma forma, com base em semelhanças percebidas entre as entidades (NASCIMENTO, 2005).

¹⁴ As Redes Neurais Artificiais (RNAs) são modelos matemáticos e computacionais semelhantes às estruturas neurais e biológicas, cuja capacidade computacional é adquirida por aprendizado e generalização (REZENDE, 2005).

¹⁵ A extração de conhecimento é uma técnica de Redes Neurais Artificiais, conhecida pelo bom desempenho que costuma obter quando utilizada em várias aplicações (REZENDE, 2005).

dados. O segundo permite agrupá-los, levando em consideração a similaridade existente entre eles.

No estudo de caso proposto neste artigo, realizou-se uma filtragem das principais variáveis, recuperando as mais importantes para o domínio em questão. A fase de reutilização adapta as soluções para realizar o *clustering*. Uma vez que este agrupamento é realizado, o próximo passo é extrair o conhecimento. Na fase de revisão, houve uma avaliação por especialistas para aprovar a solução proposta. Finalmente, a fase de retenção permite ao sistema aprender com as experiências obtidas nas três fases anteriores.

Esta pesquisa apresentou um sistema de RBC que permite o diagnóstico automaticamente para pacientes com câncer, por meio do uso de dados de *microarrays*. O modelo combina técnicas para a redução da dimensão do conjunto de dados original e um novo método de agrupamento (*clustering*) para a classificação de pacientes. O sistema funciona de forma semelhante como os profissionais operam em laboratório, mas é capaz de trabalhar com grandes quantidades de dados e tomar decisões automaticamente, reduzindo significativamente o tempo necessário para fazer uma previsão e a taxa de erro humano.

3.2.1. Análise do artigo

Este artigo faz uma análise do uso de *microarrays* como uma ferramenta para a pesquisa genômica, fator importante na descoberta de genes e determinação da sequência genética. A utilização do *microarray*, associado às técnicas de Inteligência Artificial, como pré-processamento de dados e *clustering* (agrupamento), que facilite a escolha das variáveis que serão empregadas, servem para reduzir a dimensão dos dados e podem contribuir positivamente no desempenho de um sistema de RBC.

É válido tratar os dados de maneira adequada para possibilitar que o sistema obtenha resultados satisfatórios. Isto é, aplicar técnicas de filtragem para recuperar somente os dados relevantes ao problema. Somado a isso, o *clustering* permite agrupar os dados conforme as variáveis que determinam o comportamento do grupo. Então, é possível abstrair o conhecimento e classificar os casos de acordo com a similaridade.

No caso de uso deste artigo, aplicaram-se alguns métodos de Inteligência Artificial, como: técnicas de filtragem, *clustering* e extração de conhecimento. A vantagem de aplicar tais métodos para analisar os dados é para reduzir e agrupar os dados que serão trabalhados, além de filtrar as principais variáveis para o domínio em questão, possibilitando ao sistema resultados mais precisos e confiáveis.

3.3. Sistematização da Assistência em Enfermagem Pediátrica aplicando RBC

De acordo com Mendes *et al.* (2009), as etapas teóricas e científicas realizadas na enfermagem, desde o início até o fim do tratamento do paciente, são chamadas de Sistematização da Assistência de Enfermagem (SAE). Recentemente, aumentou o interesse na implantação e utilização da SAE no Brasil, pois seu uso facilita a gestão e recuperação de informações relevantes para tomada de decisões em diagnósticos futuros, visando às necessidades de cada paciente, indicando as intervenções mais apropriadas ou ações que devem ser realizadas por enfermeiros, contribuindo significativamente para aprimorar a qualidade no atendimento. No entanto, a SAE tem algumas dificuldades de implementação, pois é visto como um processo lento, complexo e burocrático.

Esses fatores apontam a necessidade de melhorar o sistema manual utilizado. Contudo, os autores afirmam que a área da enfermagem é de natureza humana e não há padrões bem definidos. Neste cenário, os sistemas convencionais e de técnicas computacionais não são eficientes para apoiar a tomada de decisão dos enfermeiros, porque os problemas são compreendidos e solucionados com base no conhecimento e em experiências passadas. Esta condição é ideal para aplicar a técnica de RBC.

O objetivo desta pesquisa foi desenvolver um sistema de RBC para auxiliar os profissionais de enfermagem na tomada de decisões, favorecendo a geração de conhecimento, levando à aprendizagem de novos diagnósticos, além de contribuir para a implementação da SAE. O protótipo foi desenvolvido na linguagem de programação Java e usou o *framework*¹⁶ jCOLIBRI¹⁷, que implementa a técnica de RBC.

3.3.1. Análise do artigo

Este artigo faz uma análise da Sistematização da Assistência de Enfermagem, que permite aplicar, do início ao fim de um tratamento, os conhecimentos técnico-científicos de enfermagem. Com a facilidade de gerir e recuperar informações que são relevantes para o processo de tomada de decisões em diagnosticar doenças e indicar as intervenções mais adequadas para cada paciente, aumentou o interesse de implementar este sistema no Brasil, com o intuito de aprimorar a qualidade no atendimento médico.

¹⁶ “*Framework* é um conjunto de objetos *extensível* para funções relacionadas” (LARMAN, 2007, p. 622).

¹⁷ O jCOLIBRI é um ambiente de desenvolvimento de RBC em JAVA, desenvolvido pelo Grupo de Aplicações de Inteligência Artificial da Universidade Complutense de Madri (jCOLIBRI: <http://gaia.fdi.ucm.es/research/colibri/jcolibri>).

No caso de uso deste artigo, para implementação utilizou-se o jCOLIBRI, devido ser um dos *shells* com a maior quantidade de material disponível na Internet para pesquisa, além integrar suas bibliotecas com a linguagem Java. Para a modelagem foram seguidos os mesmos passos realizados por Gresse von Wangenheim e von Wangenheim (2003) para construir um aplicativo em RBC.

3.4. Metodologia de RBC para o Diagnóstico de Gripe Suína

Conforme Chakraborty *et al.* (2011), o RBC apresenta uma tecnologia essencial na construção de sistemas inteligentes voltados para o diagnóstico médico, que pode ajudar significativamente na melhoria da tomada de decisão.

Os autores propõe o desenvolvimento de um sistema, denominado de *Swine Flu Diagnostic Assistant* (SFDA)¹⁸, para o diagnóstico de gripe suína. Levando em consideração o funcionamento do RBC, a base de casos é composta por vários casos médicos, incluindo os sintomas comuns encontrados nos pacientes que sofrem de gripe suína e suas prescrições. O conhecimento inicial sobre a gripe suína é baseado em uma coleção de casos deste agravo. Este trabalho tem o propósito unicamente acadêmico. Portanto, a experimentação prática e testes são sugeridos como trabalhos futuros.

O sistema recebe um conjunto de sintomas de um novo paciente, comparando-o com os casos anteriores da base, usando um algoritmo de similaridade, e retorna os casos mais similares. Os casos semelhantes recuperados incluem o tratamento sugerido que o paciente foi submetido, contribuindo ao médico tomar a melhor decisão para cada caso consultado.

O RBC é aplicado para permitir ao especialista de plantão tomar a decisão adequada ao diagnosticar se um paciente tem gripe suína ou não, dependendo dos atributos do caso de cada paciente. Os autores consideraram alguns parâmetros de entrada em base experimental, mas em uma situação real, os parâmetros que definem um caso de gripe suína são maiores e mais complexos. Uma solução médica automatizada proporciona economia de tempo, significando um atendimento mais rápido, consequentemente aumentando a eficiência do desempenho no atendimento médico.

O sistema desenvolvido contém quatro módulos principais: a interface com o usuário, o módulo de RBC, o módulo computacional e o módulo administrativo. No estudo de caso deste artigo, o usuário é um médico (ou um especialista responsável pelo diagnóstico da gripe

¹⁸ Em português, Assistente de Diagnóstico de Gripe Suína. Segundo Chakraborty *et al.* (2011), é um sistema de RBC, desenvolvido na linguagem de programação Java, com o intuito de detectar casos de gripe suína para fins acadêmicos.

suína) e sua interação com o ambiente do sistema é através de menus, tabelas e gráficos, que lhe permitem simular a condição do paciente. O modelo computacional é baseado em regras e a similaridade é baseada em inferências para fornecer decisões de diagnóstico. O módulo de RBC realiza a recuperação dos casos com base na similaridade dos casos reais de gripe suína. O módulo administrativo oferece ao administrador a capacidade de adicionar, remover e modificar os casos da base, quando necessário.

A representação dos casos é feita com casos reais de gripe suína retirado do site *National Health Service* (NHS). Um esforço tem sido feito para encontrar uma correlação entre os casos iniciais e os parâmetros de entrada¹⁹. Os pesos são atribuídos a cada atributo em suposições baseadas em conclusões por profissionais experientes da NHS. É usado o algoritmo de recuperação da similaridade pelo Vizinho mais Próximo para calcular a Pontuação de Similaridade (*Similarity Score* – SS) entre os casos da base e o problema consultado, como ilustrado na fórmula 3.1:

$$SS (Query Case, Compared Case) = \sum_{i=1}^n \frac{(W_i * |I_iQC|)}{\sum_{i=1}^n W_i}$$

Fórmula 0.2. Fórmula de similaridade pelo Vizinho mais Próximo para calcular a SS.
Fonte: Chakraborty *et al.* (2011).

Este trabalho foi um protótipo simples desenvolvido, com fins teóricos, para detectar casos de gripe suína. Este modelo pode ser estendido para o diagnóstico de várias categorias de gripe e febre. Neste contexto, diversos “Assistentes de Diagnóstico” podem ser usados em ambientes distribuídos com atualizações em tempo real, alimentados por *Extensible Markup Language* (XML)²⁰, em intervalos regulares de tempo. Isso reforçaria o uso e utilidade dos “Assistentes de Diagnóstico” em larga escala no mundo, que beneficiaria a comunidade médica em geral.

3.4.1. Análise do artigo

Este artigo faz uma análise de sistemas inteligentes desenvolvidos para o diagnóstico médico, que contribuem na tomada de decisão. Cabe destacar que o RBC apresenta-se como uma técnica eficaz para a construção de Sistemas de Apoio à Decisão. Levando isto em

¹⁹ Os parâmetros de entrada são os atributos dados aos casos para obter o melhor desempenho do sistema. A atribuição correta dos parâmetros de entrada incide na recuperação dos casos mais similares ao problema em questão.

²⁰ “A XML é uma tecnologia para criar linguagens de marcação que descrevem dados de praticamente qualquer tipo de uma forma estruturada” (DEITEL *et al.*, 2003, p. 152).

consideração, um dos focos em aplicar o RBC é possibilitar ao especialista tomar a decisão adequada ao diagnosticar um paciente, independente da doença.

No caso de uso deste artigo, desenvolvido na linguagem de programação Java, o sistema foi dividido em quatro módulos: a interface com o usuário, o módulo de RBC, o módulo computacional e o módulo administrativo. É interessante a divisão do sistema em quatro partes, sendo cada uma independente da outra, pois não força que um mesmo usuário seja responsável pelos diagnósticos e pela administração. O módulo de RBC realiza a recuperação dos casos, com base na similaridade dos casos reais da doença, por meio de um dos algoritmos de recuperação mais utilizados nos sistemas de RBC: a similaridade pelo Vizinho mais Próximo.

O modelo, desenvolvido no presente artigo, pode ser usado para fazer o diagnóstico de outras doenças, além da proposta inicialmente: a gripe suína. Desta forma, um sistema médico, que gere análises confiáveis de forma automatizada, proporciona um atendimento mais eficiente, aumentando o desempenho no atendimento nos centros de saúde.

3.5. Diagnósticos de Esclerose Múltipla: Abordagem em Raciocínio Baseado em Casos

De acordo com Kurbalija *et al.* (2007), o RBC tornou-se uma técnica de sucesso para Sistemas Baseados em Conhecimento em diferentes domínios. A Medicina é um domínio adequado para aplicação de RBC, visto que o conhecimento dos especialistas consiste na mistura do aprendizado acadêmico (objetivo) e da experiência prática (subjetivo). Estes dois tipos de conhecimento podem ser claramente separados e representados de forma apropriada. O conhecimento objetivo pode ser representado nas formas de regras ou funções, enquanto o conhecimento subjetivo é contido nos casos. Nos últimos anos, o RBC tem sido bastante aplicado para o diagnóstico médico.

Uma das principais vantagens desta técnica é a possibilidade de aplicá-la a quase qualquer domínio. Essa abordagem é apropriada para domínios menos estudados, onde as regras e as conexões entre os parâmetros não são conhecidos. No entanto, os autores consideram que a etapa de adaptação dos casos pode ser um problema sério na Medicina. Algumas técnicas de adaptação, típicas para utilização médica, são: o foco na recuperação²¹ e a generalização de casos²².

²¹ Segundo Schmidt e Gierl (2000), o foco na recuperação é uma ideia para evitar problemas na etapa de adaptação dos casos. O propósito desta técnica é construir sistemas apenas para recuperação. Estes são programas que somente recuperam casos similares e os apresentam, como informação, para o usuário.

²² Conforme Schmidt e Gierl (2000), a generalização de casos é a especificidade extrema de casos isolados. Uma ideia diferente é generalizar a partir de casos isolados em protótipos abstratos ou classes.

De certa forma, o diagnóstico de doenças raras ou negligenciadas, como a Esclerose Múltipla (EM), é um objeto de estudo interessante para ser investigado. A EM é uma doença crônica, de causa desconhecida, que afeta o sistema nervoso (cérebro, medula espinhal e nervos periféricos), em forma de lesões multifocais²³ (distribuídas), provocando danos de visão, força muscular, sensibilidade, coordenação, linguagem e controle da bexiga. Também pode afetar as funções cognitivas.

Depois de inserir os casos no sistema, o próximo desafio é extrair os dados de diagnósticos prévios de EM. Este problema é relativamente complexo, pois os dados devem ser obtidos a partir do histórico médico dos pacientes dos últimos 10 anos. Cada caso tem 72 características diferentes, que representam as observações mais importantes no processo de diagnóstico de EM.

No geral, a técnica de RBC parece ser apropriada para o uso médico, como em Sistemas Baseados em Conhecimento. A principal intenção do Sistema de Apoio à Decisão proposto foi ajudar especialistas na tomada de decisões complexas. Além disso, a contribuição importante deste sistema é tentar reduzir a lacuna entre médicos experientes e iniciantes, por meio da manipulação do conhecimento existente, a fim de tomar uma decisão final. Outra contribuição significativa é ajudar aos médicos a reduzir o tempo de diagnóstico ou o número de exames clínicos necessários para diagnósticos confiáveis.

3.5.1. Análise do artigo

Este artigo faz uma análise do RBC aplicado para o diagnóstico de doenças raras, como a Esclerose Múltipla, que é um objeto de estudo interessante a ser pesquisado. Sendo assim, no caso de uso deste artigo, os autores desenvolveram um sistema de RBC, na linguagem de programação Java, para diagnosticar casos desta doença. Um dos maiores desafios neste domínio é captar os dados de diagnósticos prévios de EM, visto que cada caso tem 72 características diferentes, que representam as informações relevantes no processo de decisão no diagnóstico de EM.

O sistema de RBC, apresentado neste artigo, foi desenvolvido na linguagem de programação Java, mas sem nenhum *shell* ou *framework*. Atualmente, não há necessidade de desenvolver um sistema de RBC sem o apoio de uma ferramenta, como CBR-Works, myCBR ou jCOLIBRI, pois elas auxiliam e agilizam a construção do sistema.

²³ De acordo com Barros, Rissi e Rech (2008, p. 1), “lesões multifocais são lesões múltiplas distribuídas pelo órgão e separadas entre si por tecido não afetado”.

3.6. Conclusões

O modelo de RBC proposto nesta dissertação assemelhasse aos modelos apresentados em alguns artigos no que diz respeito ao auxílio no diagnóstico médico. Em contrapartida é original em alguns aspectos, bem como, ao aplicar a técnica de RBC para análise epidemiológica da leptospirose e, também, usar técnicas de Geoprocessamento para expressar visualmente as características que determinaram a ocorrência do surto na área de estudo.

Este capítulo abordou o Estado da Arte, levando em consideração as pesquisas mais recentes sobre Raciocínio Baseado em Casos que estão relacionadas ao tema de estudo desta dissertação. O capítulo seguinte faz uma análise do surto de leptospirose humana, ao aplicar as técnicas de RBC e Geoprocessamento.

Capítulo 4: Formulação do Modelo de Análise Epidemiológica

Neste capítulo, apresenta-se uma análise do surto de leptospirose humana, a partir do uso das técnicas de RBC e Geoprocessamento. Inicialmente, é feito o desenvolvimento do sistema, desde a atribuição dos pesos das variáveis até as etapas de construção do modelo, denominado de SisLeptos, levando em consideração as informações coletadas dos pacientes, como antecedentes epidemiológicos e dados clínicos. Em seguida, utilizou-se o Geoprocessamento para identificar e classificar a incidência da leptospirose no bairro do Guamá, nos anos de 2008, 2009 e 2010, por meio dos dados georreferenciados desta área.

4.1. Implementação do sistema de RBC

O presente trabalho foi desenvolvido para a plataforma Web, na linguagem de programação *Active Server Pages* (ASP), cuja escolha da tecnologia foi proporcionada, entre outros motivos, pela versatilidade, visto que possibilita a construção de diversos tipos de aplicações para Internet, além da ferramenta CBR-Works ser compatível somente com o servidor *Internet Information Services* (IIS) da Microsoft®, o mesmo usado pelo ASP. O RBC foi modelado e implementado com o auxílio do CBR-Works, responsável por construir as regras de produção e prover soluções de acordo com o grau de similaridade entre os casos contidos na base.

Desta forma, propôs-se a formulação de um modelo de análise epidemiológica, usando RBC e Geoprocessamento. Para compor o modelo, desenvolveu-se um sistema, denominado de SisLeptos, com o intuito de analisar a incidência da leptospirose humana em 3 bairros da cidade de Belém, estado do Pará, que aplicou técnicas de Mineração de Dados para verificar a relevância dos índices dos casos, e encontrar padrões entre os casos notificados nos bairros do Guamá, Jurunas e Montese, nos anos de 2008, 2009 e 2010, bem como as possíveis fontes de infecção e os principais sintomas dos pacientes infectados com a bactéria *Leptospira interrogans*.

A Mineração de Dados foi importante para verificar a relevância dos índices e atribuir pesos de acordo com a realidade do problema, para posteriormente facilitarem a recuperação de situações úteis ao objetivo do sistema.

4.1.1. Atribuição dos pesos das variáveis

Nesta seção são descritas as atividades de pré-processamento e os padrões identificados, com o objetivo de atribuir pesos às variáveis do sistema para obter acuidade

diagnóstica da leptospirose. Esta etapa é feita com a utilização do software TreeMap 4.1²⁴, que utiliza técnicas de visualização de informação, com o intuito de descobrir padrões em grandes volumes de dados.

4.1.1.1. Atividades de pré-processamento

Os dados relativos à incidência da leptospirose humana foram obtidos a partir da disponibilização dos mesmos pelo Sistema de Informação de Agravos de Notificação (SINAN, 2011) e cedidos pela SESPA, baseados nos dados presentes na ficha cadastral de leptospirose (vide Anexo A), que posteriormente foram depurados e georreferenciados. Na sequência, estes dados foram manipulados para realizar um pré-processamento da base, o qual é dividido em três etapas: seleção, limpeza e transformação dos dados, a fim de representar corretamente os casos a serem processados pelo sistema de RBC.

	H	I	J	K	L	M	N	O	P	Q	R
1	ANTECEDENTES EPIDEMIOLÓGICOS										
2	ÁGUA OU LAMA	CRIAÇÃO DE ANIMAIS	CAIXA D'ÁGUA	FOSSA	LOCAL COM ROEDORES	PLANTIO	RIO	CONTATO C/ ROEDORES	ALIMENTOS	TERRENO BALDIO	LIXO OU ENTULHO
3	Sim	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Sim
4	Não informado	Não informado	Não informado	Não informado	Sim	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado
5	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado
6	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado
7	Sim	Ignorado	Ignorado	Ignorado	Sim	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado
8	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado
9	Sim	Não	Não	Não	Sim	Não	Não	Sim	Não	Não	Sim
10	Sim	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não
11	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado
12	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Não informado	Ignorado	Ignorado	Ignorado
13	Sim	Não	Não	Não	Sim	Não	Não	Sim	Não	Não	Não
14	Não	Sim	Não	Não	Sim	Não	Sim	Sim	Não	Não	Sim
15	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado
16	Sim	Sim	Sim	Sim	Sim	Não	Não	Sim	Não	Sim	Sim
17	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não
18	Não	Não	Não	Sim	Sim	Não	Não	Não	Não	Não	Sim
19	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado
20	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado
21	Não	Não	Não	Não	Sim	Não	Sim	Sim	Não	Não	Sim
22	Sim	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não
23	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado
24	Sim	Não	Não	Não	Não	Não	Sim	Sim	Não	Não	Sim
25	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado	Não informado

Figura 4.11. Casos com os antecedentes epidemiológicos da leptospirose.

Inicialmente, identificam-se os atributos com maior relevância dentre as colunas da tabela, conforme ilustra a figura 4.1, para obter os dados mais importantes do banco, como “ÁGUA OU LAMA”, “LOCAL COM ROEDORES”, “CONTATO C/ ROEDORES”, “TERRENO BALDIO” e “LIXO OU ENTULHO”. Em seguida, descartaram-se os campos “CRIAÇÃO DE ANIMAIS”, “CAIXA D'ÁGUA”, “FOSSA”, etc., devido o número expressivo de dados preenchidos como “Ignorado” ou “Não Informado”, visto que este tipo

²⁴ O TreeMap é um software de Mineração de Dados, desenvolvido durante a década de 1990, na Universidade de Maryland pelo professor Ben Shneiderman (TreeMap: <http://www.cs.umd.edu/hcil/treemap>).

de informação causaria inconsistência na base de casos do RBC, o que geraria dúvida ao sistema no momento de analisar um novo problema, resultando em diagnósticos imprecisos.

Por fim, realiza-se a padronização dos dados que não estavam em um mesmo formato, deixando-os com valores semelhantes; por exemplo, todos os moradores do bairro do Guamá passaram a ser representados com a String “Guamá”, ao invés de “GUAMA” ou “GUAMÁ”, o mesmo aconteceu aos residentes dos bairros do Jurunas e Montese.

MesDaSuspeita	Idade	Sexo	Bairro	AguaOuLama	LocalComRoedores	ContatoComRoedores	TerrenoBaldio	LixoOuEntulho	Febre	Mialgia	Cefaleia	Prostracao	Congestao	DorNaPanturrilha	Vomito
3	71	1	Guamá	1	1	1	2	1	1	1	1	1	2	2	2
4	17	2	Guamá	9	1	9	9	9	1	1	1	2	2	1	1
5	27	1	Guamá	9	9	9	9	9	1	1	1	1	1	1	1
6	32	2	Guamá	9	9	9	9	9	1	1	1	9	9	9	9
7	0	1	Guamá	1	1	1	9	9	1	1	1	1	2	1	1
8	15	2	Guamá	9	9	9	9	9	1	1	1	9	9	9	9
9	24	1	Guamá	1	1	1	2	1	1	1	1	1	1	2	2
10	10	1	Guamá	1	1	2	2	2	1	2	1	2	2	2	2
11	48	1	Guamá	9	9	9	9	9	1	1	1	1	1	1	1
12	11	1	Guamá	9	9	9	9	9	1	1	1	2	2	2	2
13	0	1	Guamá	1	1	1	2	2	1	1	1	1	1	1	1
14	17	1	Guamá	2	1	1	2	1	2	1	2	2	1	2	2
15	9	1	Guamá	9	9	9	9	9	1	1	1	9	9	9	9
16	14	1	Guamá	1	1	1	1	1	1	1	1	1	9	9	9
17	70	1	Guamá	2	1	2	2	2	1	1	1	1	1	2	1
18	70	1	Guamá	2	1	2	2	1	1	1	1	1	1	1	1
19	39	2	Guamá	9	9	9	9	9	1	1	1	1	1	9	9
20	14	1	Guamá	9	9	9	9	9	1	1	1	9	9	9	9
21	23	1	Guamá	2	1	1	2	1	1	1	1	1	1	1	1
22	11	2	Guamá	1	2	2	2	2	1	1	1	1	2	2	2
23	7	2	Guamá	9	9	9	9	9	1	1	2	1	2	1	1
24	23	1	Guamá	1	2	1	2	1	1	1	1	1	2	2	2
25	27	2	Jurunas	9	9	9	9	9	9	9	9	9	9	9	9

Figura 4.12. Casos filtrado e padronizado utilizado no Treemap.

As informações correspondentes aos antecedentes epidemiológicos e dados clínicos (sinais e sintomas), como exibidos na figura 4.2, referentes à “Sim”, “Não” são representadas por “1” e “2”, respectivamente. Os campos preenchidos como “Ignorado” ou “Não informado” é representado por “9”.

4.1.1.2. Padrões identificados

É usada a técnica de Mineração de Dados (do inglês, *Data Mining*) para obter informações dos padrões identificados nos casos e, em seguida, transformá-los em novos conhecimentos. Desta forma, os conhecimentos adquiridos são interpretados e servem para atribuir pesos as variáveis do sistema, com a finalidade de aumentar a confiabilidade e a qualidade das soluções providas pelo SisLeptos.

Observa-se, na figura 4.3, que os meses com maior número de notificações de leptospirose, entre os bairros da região metropolitana de Belém, correspondem a fevereiro (com 45 casos confirmados e 28 descartados), março (com 35 casos confirmados e 39

descartados) e janeiro (com 30 casos confirmados e 34 descartados), realçados pela cor verde, amarela e azul, respectivamente. Cada mês é dividido pelos casos confirmados e descartados.

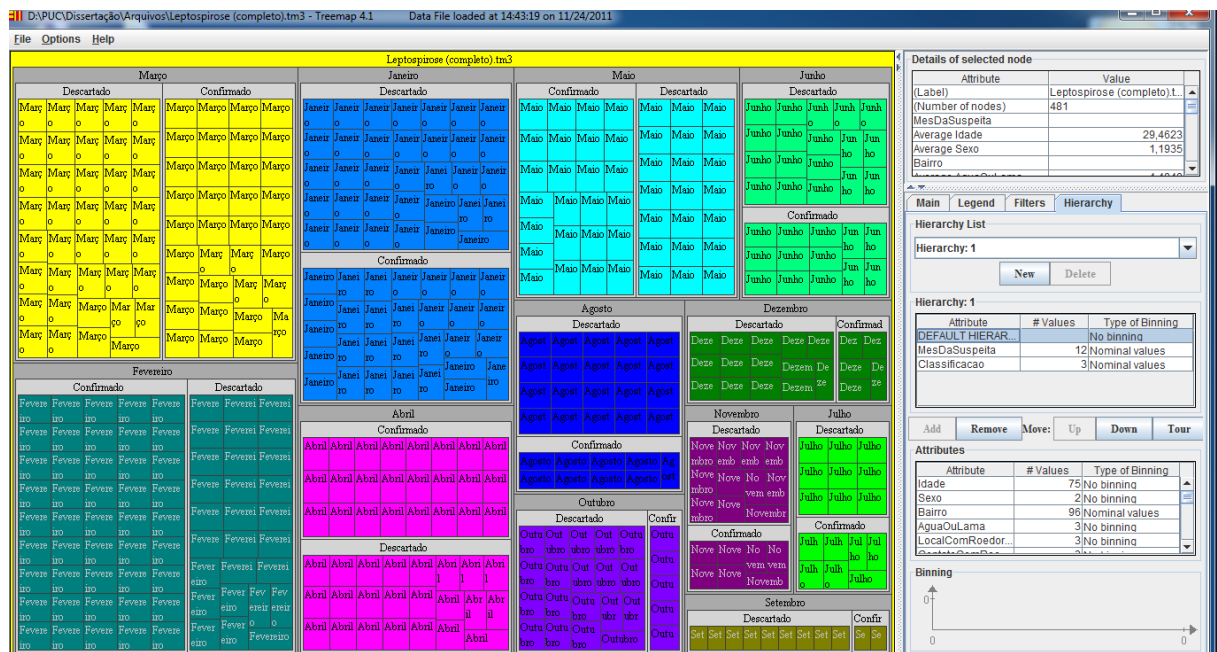


Figura 4.13. Casos suspeitos de leptospirose agrupados por meses.

Contudo, na área estudada, os meses com o número mais elevado de casos notificados e confirmados de leptospirose são março (com 9 casos positivos), abril (com 5 casos positivos) e maio (com 5 casos positivos), como é notado na figura 4.4, destacados pela cor amarela, rosa e azul celeste, respectivamente.

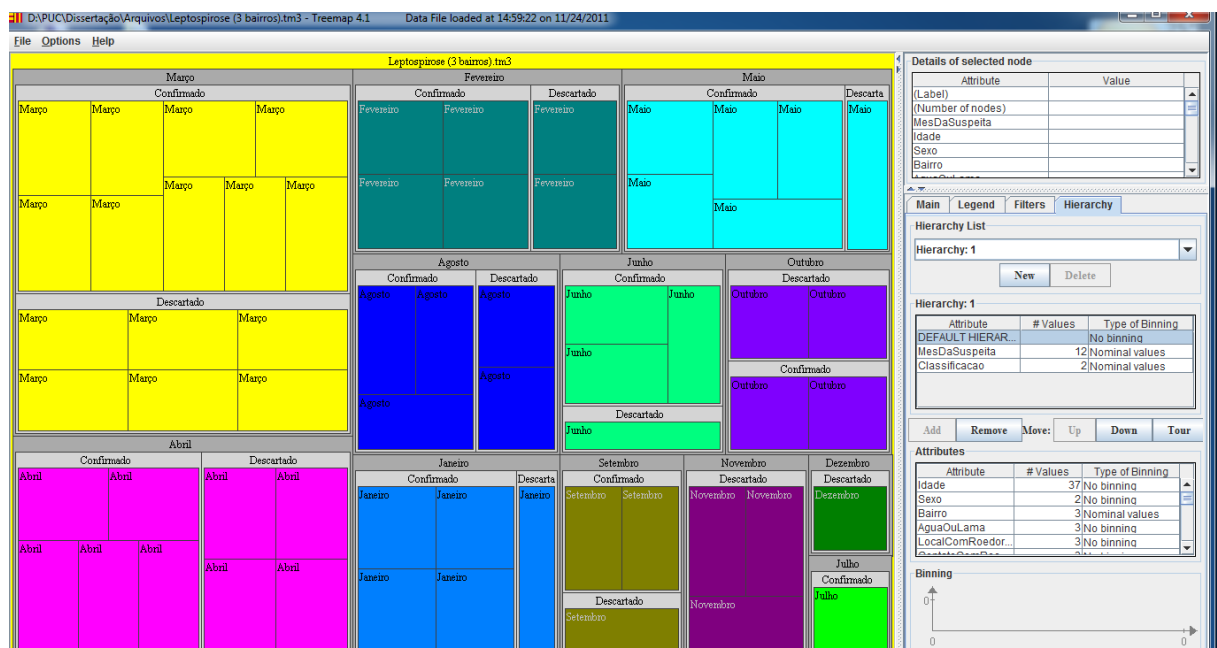


Figura 4.14. Casos suspeitos de leptospirose, na área de estudo, agrupados por meses.

Neste contexto, é possível atribuir pesos de acordo com o resultado analisado, considerando março como o mês mais propício para a infecção da leptospirose, sendo que no período chuvoso serão atribuídos pesos iguais, pois essa época possui características semelhantes, bem como alagamentos e enchentes.

Contudo, segundo Avila-Pires (2006), existe falsa correlação da ocorrência do surto com os períodos de fortes chuvas e alagamentos. Em princípio, o maior número de casos diagnosticados nesta época pode ser ligado às chuvas, que, em Belém, ocorrem com maior frequência nos meses de janeiro e fevereiro, apesar das chuvas serem comuns em todo o ano.

Em seguida, são analisados os casos notificados do surto, agrupando-os por idade, sexo e classificação (confirmado e descartado). Neste cenário, observa-se, na figura 4.5, que os indivíduos do sexo masculino, ilustrados com a cor azul, entre 18 a 65 anos de idade, pertencentes ao grupo economicamente ativo, representam 55,3% dos casos notificados, como demonstra a tabela 4.1. Somado a população feminina, os dados representam 63,8% dos pacientes com leptospirose. Com isso, atribuem-se pesos conforme as análises obtidas, considerando que a idade com maior índice está relacionada à população assalariada; consequentemente é atribuído peso maior a este grupo.

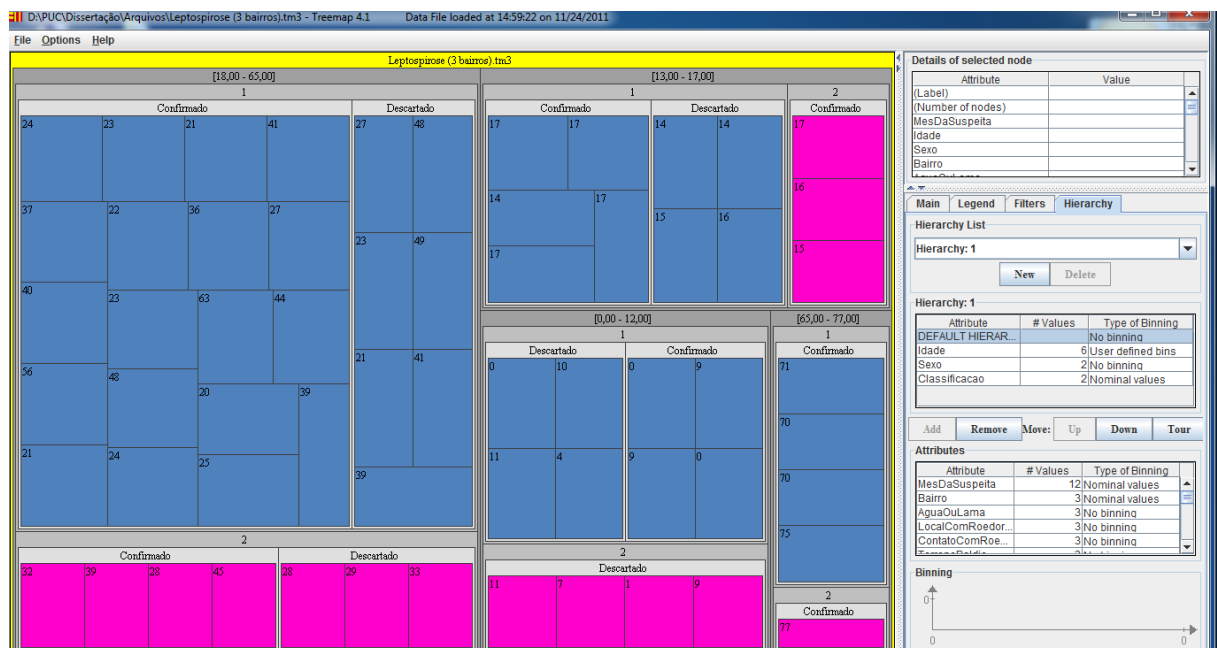


Figura 4.15. Faixa etária e gênero, na área de estudo, com maior incidência da leptospirose.

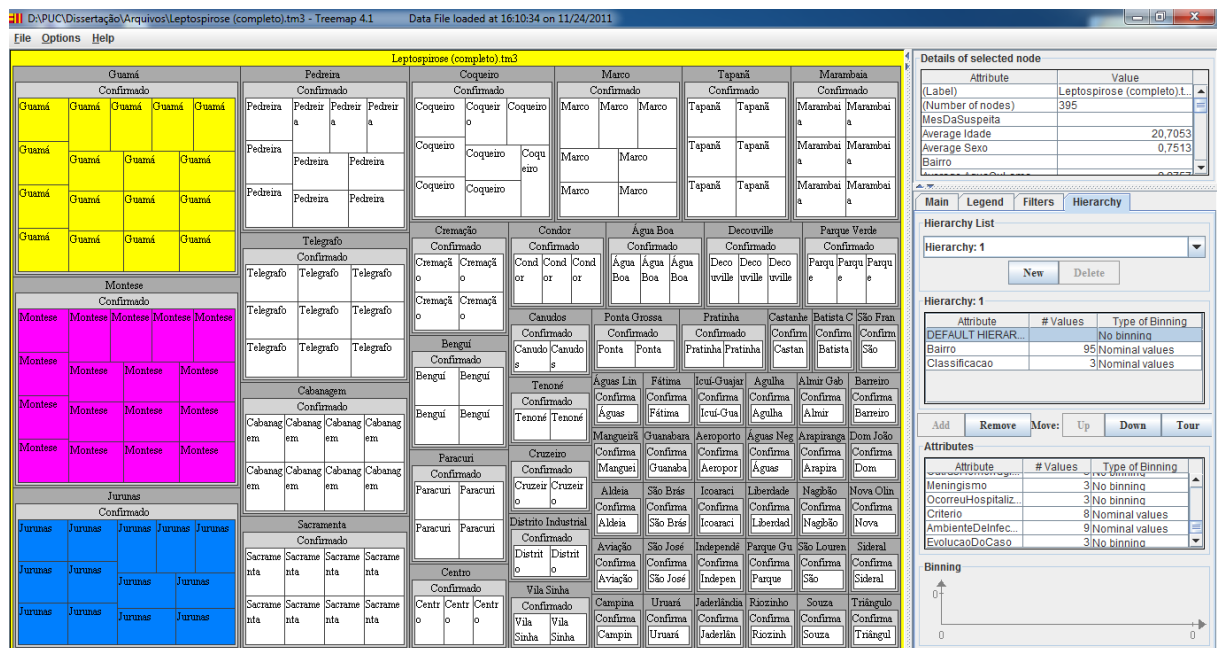
Casos confirmados de leptospirose na área de estudo: 2008 - 2010

Faixa Etária	Masculino	%	Feminino	%	Total	%
0 a 12 anos	4	8,5	0	0,0	4	8,5
13 a 17 anos	5	10,6	3	6,4	8	17,0
18 a 65 anos	26	55,3	4	8,5	30	63,8
Acima de 65 anos	4	8,5	1	2,1	5	10,6
Total	39	83,0	8	17,0	47	100,0

Tabela 4.5. Casos confirmados de leptospirose na área de estudo: 2008 - 2010.

Fonte: SINAN/SVS/MS, 2011.

Observa-se na figura 4.6 que os bairros com maior índice de casos confirmados de leptospirose são Guamá, Montese e Jurunas, representados por amarelo, rosa e azul, respectivamente. Pode-se relacionar o número expressivo de contaminações da doença, por estas áreas estarem vulneráveis a surtos endêmicos, decorrentes da aglomeração populacional de baixa renda em condições precárias de serviços básicos de saúde e educação.

**Figura 4.16.** Casos confirmados de leptospirose agrupados por bairro.

4.1.2. Desenvolvimento do SisLeptos

O desenvolvimento do SisLeptos utiliza experiências passadas de casos suspeitos de leptospirose humana, confirmados e descartados, para auxiliar na diagnose da doença. Para tal, aplica-se o RBC que é conhecido como uma técnica de representação do conhecimento, cuja principal característica é a validação probabilística dos resultados.

Após as análises dos casos contidos no Banco de Dados para que fossem filtrados e selecionados os campos mais relevantes ao objetivo do sistema, assim, eliminando os atributos desnecessários para a criação da base de casos do RBC. A seguir, são abordadas as etapas de desenvolvimento do SisLeptos.

4.1.2.1. Representação dos casos

Representação dos casos é a forma de contextualizar as informações contidas na base, através dos atributos relevantes, onde esses atributos são responsáveis por ensinar casos que possam ser utilizados para solucionar problemas futuros. Esta etapa é importante para analisar a estrutura do sistema, como: informações básicas dos pacientes com suspeita de leptospirose, antecedentes epidemiológicos, sintomas e a classificação dos casos, diagnosticados como confirmados ou descartados. Finalmente, produz-se um resultado probabilístico com a possibilidade do caso de entrada estar ou não com leptospirose, conforme ilustra a figura 4.7.

Caso N
Informações Básicas: <i>ID, mês da suspeita, sexo e idade</i>
Antecedentes Epidemiológicos: <i>Água ou lama de enchentes, local com sinais de roedores, contato diretamente com roedores, lixo e terreno baldio</i>
Sintomas: <i>Febre, mialgia, cefaleia, prostração, congestão conjuntiva, dor na panturrilha, vômito, diarreia, icterícia, insuficiência renal, alteração respiratória, alteração cardíaca, hemorragia pulmonar, outras hemorragias e meningismo</i>
Diagnóstico: <i>Classificação</i>
Resultado: <i>Probabilidade de estar ou não com leptospirose</i>

Figura 4.17. Representação dos casos do sistema SisLeptos.

Os componentes (atributos) selecionados representam o conhecimento adquirido do especialista. A aquisição do conhecimento servirá para indicar soluções satisfatórias aos problemas do SisLeptos. Os casos de leptospirose são armazenados em uma base de casos para que sejam recuperados futuramente.

Assim, os componentes adotados representam o conhecimento contido no SisLeptos e auxiliam na diagnose dos casos de leptospirose humana. Os componentes são responsáveis por determinar a interpretação do contexto onde estão inseridos, por meio da contextualização das informações obtidas pelo sistema. A escolha dos componentes para representar os casos

serve, num modo geral, para realizar a seleção dos casos armazenados na base e, através do cálculo de similaridade, analisar os casos similares que serão utilizados para solucionar o problema em questão.

O banco de casos do SisLeptos foi modelado a partir das análises obtidas com as técnicas de Mineração de Dados do BD de leptospirose, como pode ser observado na figura 4.8. É utilizado o Sistema de Gerenciamento de Banco de Dados (SGBD) Microsoft® Access, devido o CBR-Works adicionar automaticamente na sua base os registros contidos em uma tabela do SGBD citado, evitando, desta forma, que os casos sejam inseridos manualmente.

ID	MesDaSuspeita	Idade	Genero	AguaOuLama	LocalComRoedores	ContatoComRoedores	TerrenoBaldio	LixoOuEntulho	Febre	Mi
1	Fevereiro	71	M	Guamá	Sim	Sim	Não	Não	Sim	Sim
2	Março	17	F	Guamá	Não informado	Sim	Não informado	Não informado	Não informado	Sim
3	Abril	27	M	Guamá	Não informado	Não informado	Não informado	Não informado	Não informado	Sim
4	Março	32	F	Guamá	Não informado	Não informado	Não informado	Não informado	Não informado	Sim
5	Abril	0	M	Guamá	Sim	Sim	Ignorado	Ignorado	Ignorado	Sim
6	Março	15	F	Guamá	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Sim
7	Abril	24	M	Guamá	Sim	Sim	Sim	Não	Sim	Sim
8	Março	10	M	Guamá	Sim	Sim	Não	Não	Não	Sim
9	Maio	48	M	Guamá	Não informado	Não informado	Não informado	Não informado	Não informado	Sim
10	Junho	11	M	Guamá	Ignorado	Ignorado	Não informado	Ignorado	Ignorado	Sim
11	Maio	0	M	Guamá	Sim	Sim	Sim	Não	Não	Sim
12	Maio	17	M	Guamá	Não	Sim	Sim	Não	Sim	Não
13	Abril	9	M	Guamá	Não informado	Não informado	Não informado	Não informado	Não informado	Sim
14	Agosto	14	M	Guamá	Sim	Sim	Sim	Sim	Sim	Sim
15	Setembro	70	M	Guamá	Não	Sim	Não	Não	Não	Sim
16	Setembro	70	M	Guamá	Não	Sim	Não	Não	Sim	Sim
17	Novembro	39	F	Guamá	Não informado	Não informado	Não informado	Não informado	Não informado	Sim
18	Outubro	14	M	Guamá	Não informado	Não informado	Não informado	Não informado	Não informado	Sim
19	Agosto	23	M	Guamá	Não	Sim	Sim	Não	Sim	Sim
20	Novembro	11	F	Guamá	Sim	Não	Não	Não	Não	Sim
21	Setembro	7	F	Guamá	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Sim
22	Outubro	23	M	Guamá	Sim	Sim	Não	Sim	Sim	Sim
23	Janeiro	28	F	Jurunas	Ignorado	Ignorado	Ignorado	Ignorado	Ignorado	Sim
24	Fevereiro	21	M	Jurunas	Sim	Sim	Sim	Sim	Sim	Sim
25	Janeiro	28	F	Jurunas	Sim	Sim	Sim	Não	Sim	Sim

Figura 4.18. Banco de casos do sistema SisLeptos.

4.1.2.2. Indexação dos casos

A indexação dos casos é considerada uma das etapas mais importantes para o desenvolvimento do SisLeptos, pois a mesma apresenta caminhos e atalhos para a recuperação dos casos. Levando em consideração que os índices de um caso são tão significantes quanto os índices dos livros de uma biblioteca, a indexação é fundamental para que os casos sejam recuperados corretamente, sendo comparados através da similaridade do caso consultado com os casos contidos na base. Neste cenário, a figura 4.9 representa as características indexadas e não-indexadas do sistema.

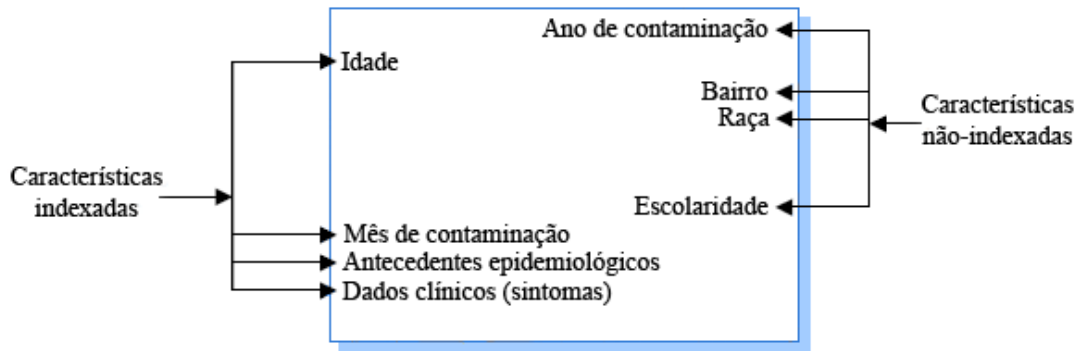


Figura 4.19. Representação das características indexadas e não-indexadas do sistema.

Apesar do sucesso de alguns métodos automáticos de indexação, como por exemplo, os métodos de *checklist* e por comparação, optou-se por usar o modo manual, por representar com maior precisão as características indexadas responsáveis em recuperar os casos da base.

Para cumprir esta etapa de maneira eficiente, fez-se necessário a assistência de um especialista ou meios didáticos, como livros e artigos médicos, e de técnicas de Mineração de Dados (abordados neste capítulo, no tópico 4.1.1.2) para determinar a escolha e os valores adequados dos índices de cada caso. Desta forma, atribuíram-se pesos, de acordo com as características analisadas dos fatores que determinam a leptospirose humana, onde os índices mais importantes receberam pesos maiores que os demais, como demonstra a figura 4.10. Observa-se na indexação que quanto mais precisa e adequada for a atribuição dos índices, maior é similaridade recuperada entre os casos para solucionar um novo problema.

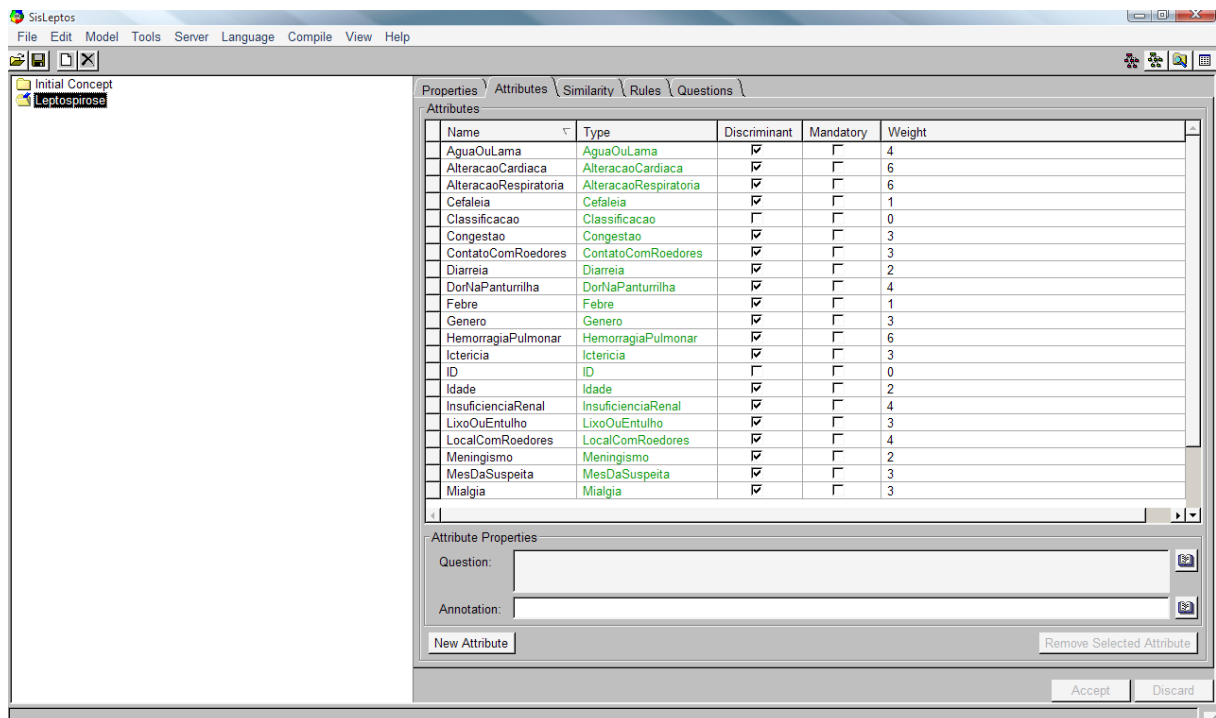


Figura 4.20. Indexação dos casos do sistema.

4.1.2.3. Recuperação dos casos

A recuperação dos casos é uma etapa crucial para atingir o objetivo do sistema, ou seja, recuperar problemas similares ao caso consultado e, posteriormente, definir a probabilidade dos mesmos serem confirmados ou descartados. A indexação adequada dos casos influencia diretamente nesta etapa, haja vista que os índices mais relevantes, isto é, aqueles que influenciam diretamente na solução do problema, servirão como base para a recuperação de casos úteis ao sistema.

Os casos são comparados e classificados pelos métodos *Matching* e *Ranking*, responsáveis por analisar os atributos conforme o grau de similaridade entre os casos existentes na base e o caso de entrada. Em seguida, são ordenados de acordo com o grau de maior semelhança em relação ao caso inicial. Desta forma, para que os casos sejam recuperados com eficácia, o SisLeptos propõe auxiliar apenas na diagnose de casos epidemiológicos, como detalhados no capítulo 2, tópico 2.3.6.

4.1.2.4. Adaptação dos casos

A adaptação tem como objetivo transformar um caso, quando necessário, para solucionar o problema em questão. O tipo de adaptação e o algoritmo usado pelo SisLeptos foi a adaptação estrutural e o algoritmo de adaptação nula. O primeiro adapta regras diretamente na solução recuperada, enquanto o segundo aplica a solução do caso mais similar recuperado para satisfazer uma nova situação. A partir da aplicação dos elementos citados anteriormente, o SisLeptos analisa a similaridade entre os casos contidos na base e o problema de entrada para auxiliar em um novo diagnóstico.

4.2. Geoprocessamento aplicado à Análise da Leptospirose

Esta dissertação também tem como objetivo analisar e demonstrar a distribuição espaço-temporal, da incidência da leptospirose humana, no bairro do Guamá, na cidade de Belém, capital do estado do Pará, nos anos de 2008, 2009 e 2010, no intuito de fornecer subsídios à adoção de medidas preventivas para o controle da incidência da doença, face às variáveis que a determinam.

As bases cartográficas contidas nesta dissertação referentes aos limites dos bairros, bacias hidrográficas e altimetria para delimitar a área de estudo, foram produzidas em parceria com Instituto Evandro Chagas, órgão de pesquisa científica do Ministério da Saúde, que desenvolve estudos relacionados aos processos epidemiológicos que ocorrem na Amazônia

Legal, sendo que as imagens foram captadas por satélites na escala de 1:50.000. Posteriormente, usam-se os dados processados pelo Sistema de Informação Geográfica *ArcView* versão 9.3 para realizar análises das variáveis ambientais e socioeconômicas do bairro do Guamá, devido a sua capacidade de expressar visualmente as informações georreferenciadas produzidas por diversas plataformas, como por exemplo imagens vetoriais geradas a partir da utilização de Sistemas de Posicionamento Global (GPS), bem como diferentes formatos gráficos.

Após as análises dos padrões de infecção pela bactéria *Leptospira interrogans*, foi identificado estatisticamente que a maior incidência da doença estava localizada no bairro do Guamá. No sentido de validar esta análise, usou-se a técnica de Geoprocessamento de dados ecoepidemiológicos para analisar o potencial de risco de transmissão da doença nesta área. Para tal, foi executado o georreferenciamento de todos os casos suspeitos do ano de 2008, 2009 e 2010, bem como de acidentes geográficos presentes na área de estudo. O bairro do Guamá localiza-se na parte sul da cidade de Belém, estado do Pará, o qual sofre forte influência do rio Guamá, como ilustra a figura 4.11.

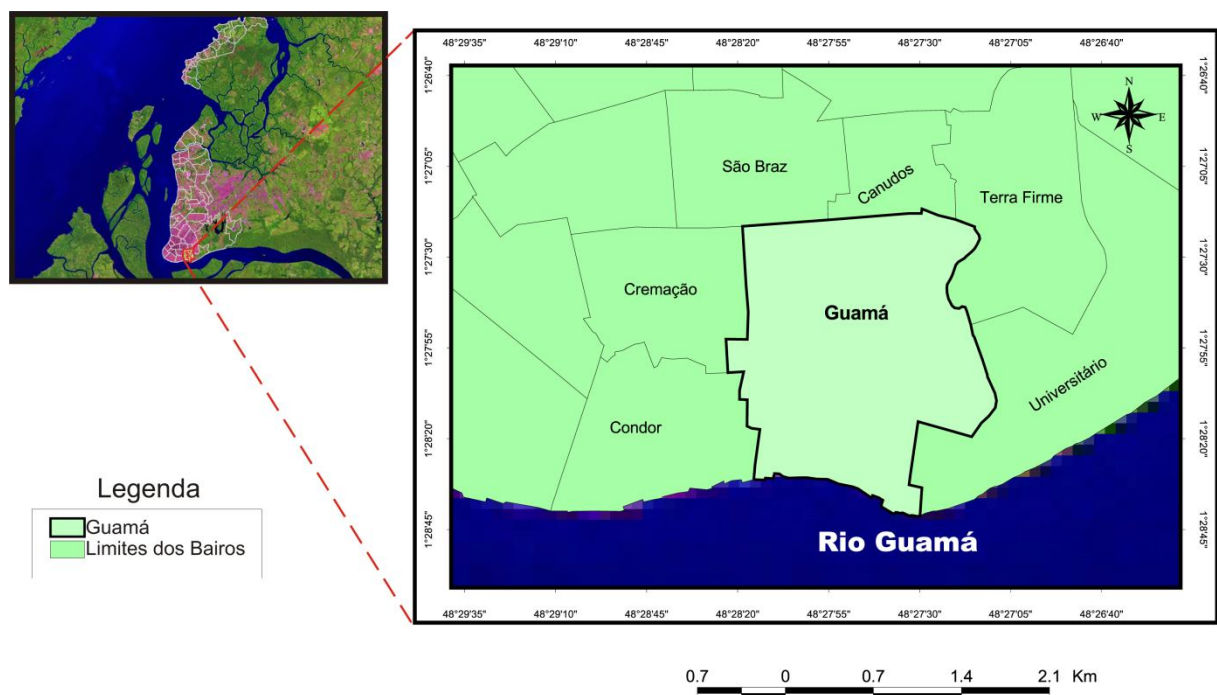


Figura 4.21. Mapa da área de estudo, o bairro do Guamá.
Fonte: LabGeo/IEC/SVS/MS, 2011.

Segundo Silva (2000), a imagem vetorial consiste na tentativa de um objeto representar e definir precisamente todas as posições, cumprimentos e dimensões geográficas. Desta forma, para que diferentes níveis de altimetria pudessem ser analisados foram

desenvolvidas bases cartográficas vetoriais (figura 4.12), que se caracterizam por reunir digitalmente elementos cartográficos, como localidades, limites, vegetação, altimetrias, entre outros, que ao serem georreferenciadas possibilitaram o inter-relacionamento destes bancos de dados.

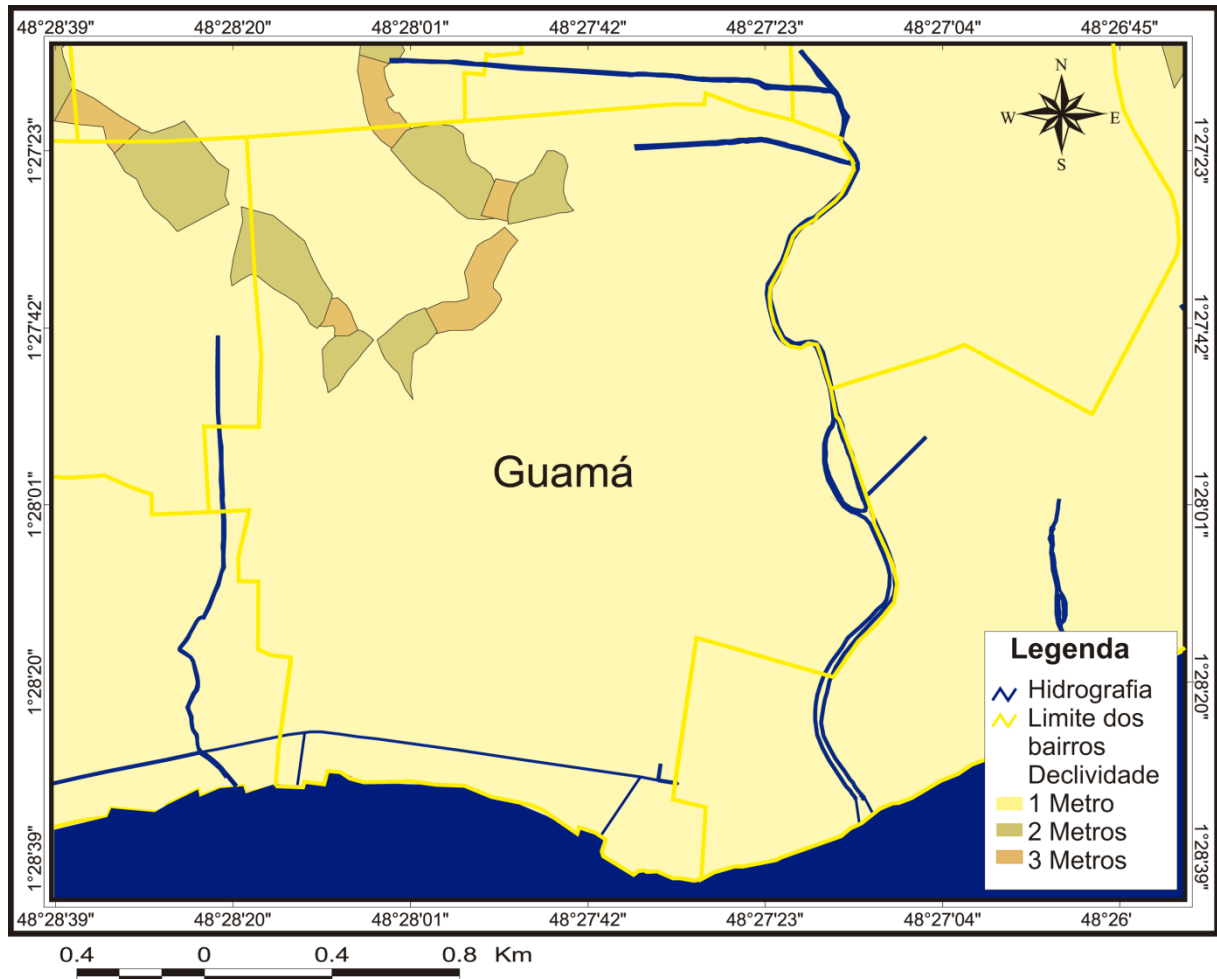


Figura 4.22. Mapa da altimetria da área de estudo, o bairro do Guamá.
Fonte: LabGeo/IEC/SVS/MS, 2011.

Os dados georreferenciados ao serem gerados possibilitaram que fosse avaliada a importância das características ambientais da área no processo epidemiológico da doença estudada. Dentre as variáveis ambientais que foram georreferenciadas, tem-se a altimetria e a drenagem do solo. Outra variável observada foi o padrão de distribuição das populações humanas, onde é possível observar na figura 4.13, gerada por satélite, as invasões proporcionadas pela urbanização desenfreada. Desta forma, observa-se que a distribuição espacial dos casos confirmados de pacientes com leptospirose está relacionada à área de influência do igarapé do Tucunduba.

É importante observar o padrão de ocupação humana, devido a maior parte das pessoas contaminadas habitarem áreas ambientalmente frágeis, que geralmente são

desconhecidas pelo poder público e são pessoas com menor poder aquisitivo, desprovidas de serviços básicos de saúde, onde a partir de trabalhos de campos foram encontradas grandes concentrações de lixo e ratos, que são vetores da doença.

Do ponto de vista de drenagem, pôde-se observar que as pessoas infectadas habitam as áreas de influência do igarapé Tucunduba e seus tributários²⁵, influenciados pelo rio Guamá, o que caracteriza como área de risco para essas populações.

O bairro do Guamá se caracteriza por possuir duas cotas altimétricas bem definidas que podem ser observadas na imagem 4.13, onde a “Área 1” corresponde ao terreno com cotas altimétricas²⁶ de 2 a 3 metros e a “Área 2” caracteriza-se por possuir um terreno mais baixo, com altimetria de até 1 metro.

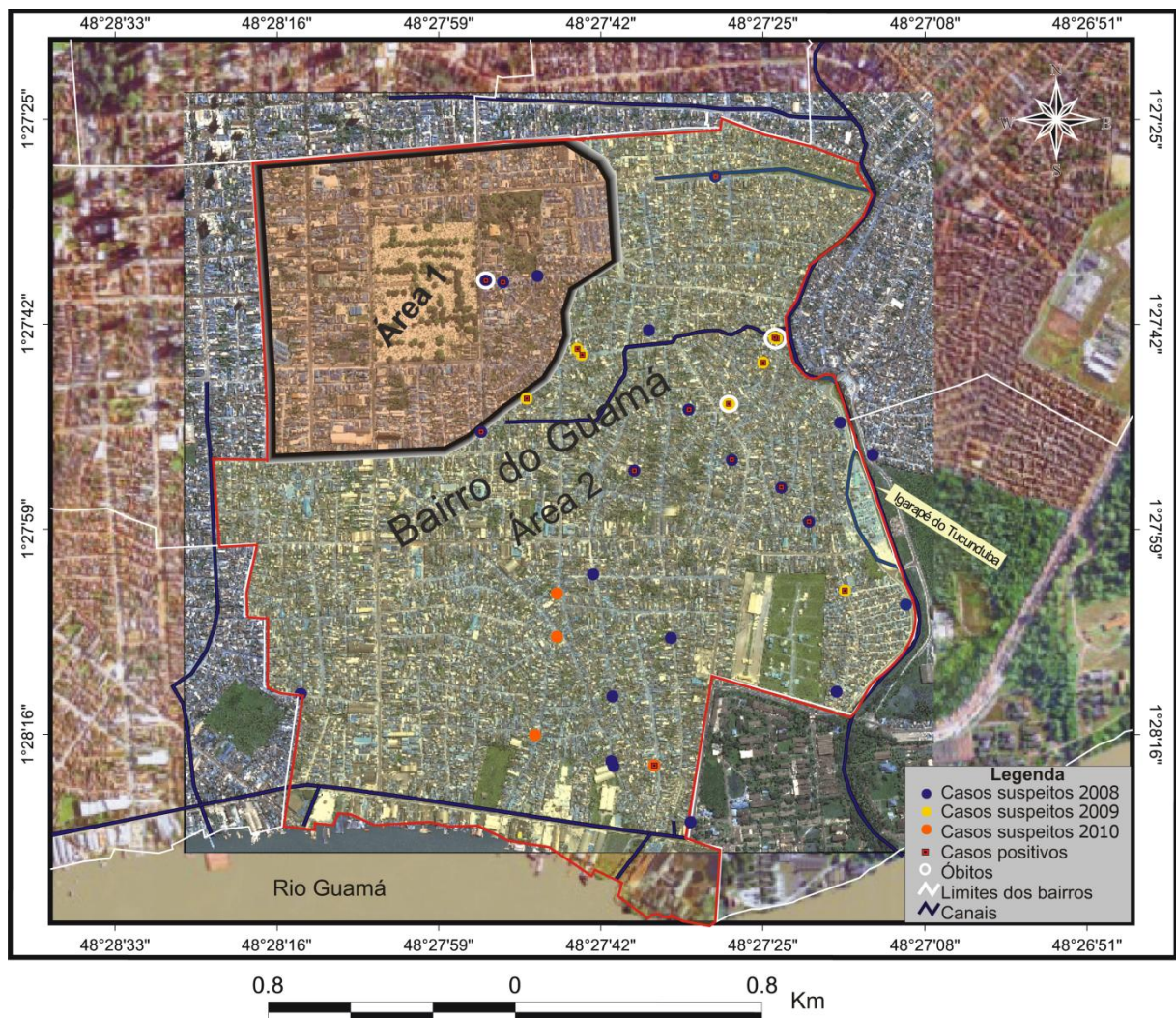


Figura 4.23. Mapa de áreas, pontos e drenagem.
Fonte: LabGeo/IEC/SVS/MS, 2011.

²⁵ Tributários são os córregos, valas ou qualquer outro curso d'água a céu aberto ligado a rios ou igarapés.

²⁶ Cota altimétrica é o nível do terreno de uma determinada região.

4.3. Conclusões

Este capítulo abordou a análise do surto de leptospirose humana. Levando em consideração a aplicação da técnica de RBC, o modelo proposto nesta dissertação e formulado neste capítulo, conhecido como SisLeptos, apresenta resultados satisfatórios (serão abordados no capítulo 5, seção 5.1), a partir da comparação de casos anteriores da doença.

No que diz respeito ao Geoprocessamento, considerando as informações obtidas com a técnica de Geoprocessamento dos dados ambientais e socioeconômicos é possível observar as razões pelas quais o bairro do Guamá apresenta a maior incidência de leptospirose, de acordo com a distribuição espacial dos casos confirmados que seguiu um padrão de área, sendo este fato ter sido, possivelmente, influenciado pela escala.

No próximo e último capítulo, serão apresentadas as Considerações Finais desta dissertação, onde serão abordados os resultados obtidos com a pesquisa, dificuldades encontradas, recomendações e extensões para trabalhos futuros.

Capítulo 5: Considerações Finais

Esta dissertação é original em diversos aspectos, destacando a inovação em integrar o Raciocínio Baseado em Casos (RBC) e Geoprocessamento para identificar e caracterizar casos relacionados à leptospirose humana. A aplicação, em conjunto, das técnicas citadas anteriormente no diagnóstico epidemiológico mostrou-se apropriada e viável, facilitando a identificação de casos da doença. Desta forma, a integração das técnicas possibilitou relacionar os fatores de risco da doença com a sua incidência.

O RBC permite a construção de um protótipo para analisar as fases iniciais, como a representação do conhecimento, a indexação e a recuperação dos casos importantes para estruturar o sistema. Além disso, também armazena, em sua base, casos que obtiveram sucesso e insucesso, com o intuito de aumentar a compreensão do conhecimento do domínio em questão, diminuindo a necessidade de um especialista para avaliar a inserção de um novo caso, devido à quantidade de conhecimentos armazenados. Entretanto, o RBC tem dificuldade em encontrar casos que cobrem todo o domínio, impedindo a recuperação de casos úteis e, consequentemente, soluções satisfatórias ao objetivo do sistema.

Pode-se considerar que os sistemas de RBC são modelos auxiliados por um especialista, sobretudo, para a inclusão de regras de um domínio específico e validação dos novos casos adicionados na base do sistema. Essa técnica pode ser utilizada para diagnóstico, classificação e suporte à decisão em diversas outras áreas do conhecimento.

A ferramenta CBR-Works, desenvolvida na linguagem de programação Java, foi escolhida para implementação por apresentar um ambiente de fácil acesso, ter sido bastante referenciada na literatura da área e facilitar o desenvolvimento do modelo de RBC. Para que o sistema obtenha melhores resultados é importante a inserção de novos casos avaliados por especialistas, visto que quanto mais consolidada a base, mais confiáveis são suas escolhas.

Avaliando os resultados obtidos, observou-se que o SisLeptos se mostrou eficaz quanto ao objetivo proposto, assim como os métodos adotados no decorrer do desenvolvimento desta dissertação, sendo os objetivos alcançados de forma satisfatória, devido a coerência das informações que podem subsidiar a geração de um diagnóstico gerado pelo sistema, conforme foi detalhado no capítulo 5, tópico 5.1. Contudo, levando em consideração que o SisLeptos teve um caráter essencialmente acadêmico, ressalta-se a necessidade de sua validação, com a análise dos dados gerados por um especialista. Usaram-se casos de leptospirose humana de bairros com características ambientais e socioeconômicas

similares das áreas estudadas para testar e avaliar a confiabilidade do modelo formulado nesta dissertação.

De modo geral, um dos pontos fortes alcançados pelo SisLeptos foi a usabilidade, isto é, a facilidade com que os especialistas responsáveis por gerar a diagnose da leptospirose, usando o sistema, têm em realizar determinadas tarefas, além da qualidade dos resultados gerados pelo sistema.

5.1. Resultados Obtidos

Após a formulação do modelo para diagnose da leptospirose humana, usando as técnicas de Raciocínio Baseado em Casos e Geoprocessamento, esta subseção apresenta telas do sistema SisLeptos e os resultados obtidos, relativos ao RBC.

Na tela da figura 5.1 o médico ou responsável pelo diagnóstico da leptospirose entrará com os dados do paciente, como informações básicas, antecedentes epidemiológicos, sinais e sintomas. O SisLeptos também está disponível na língua inglesa e conta com as funcionalidades de consultar um caso suspeito de leptospirose, além das informações sobre os autores, projeto, metodologia, contato e ajuda (vide Apêndice A).

Firefox - SisLeptos - Seja bem vindo

http://localhost/CQLConnect/root/SisLeptos/index.asp

Português English

SisLeptos INÍCIO AUTOR PROJETO METODOLOGIA CONTATO AJUDA

Sistema inteligente para diagnose da leptospirose humana

Página Inicial

INFORMAÇÕES BÁSICAS		ANTECEDENTES EPIDEMIOLÓGICOS	
Mês da suspeita	Não Informado	Água ou Lama	Não Informado
Gênero	Não Informado	Local c/ Roedores	Não Informado
Idade		Contato c/ Roedores	Não Informado
		Lixo ou Entulho	Não Informado
		Terreno Baldio	Não Informado

SINAIS E SINTOMAS			
Febre	Não Informado	Dor na panturrilha	Não Informado
Mialgia	Não Informado	Vômito	Não Informado
Cefaléia	Não Informado	Diarréia	Não Informado
Prostração	Não Informado	Icterícia	Não Informado
Congestão	Não Informado	Insuficiência Renal	Não Informado
		Alteração respiratória	Não Informado
		Alteração cardíaca	Não Informado
		Hemorragia pulmonar	Não Informado
		Outras hemorragias	Não Informado
		Meningismo	Não Informado

CONSULTAR LIMPAR

Figura 5.24. Tela de consulta com informações sobre um caso de leptospirose.

Ao apresentar ao sistema um caso confirmado de leptospirose humana do bairro da Cremação, que apresenta características ambientais e socioeconômicas semelhantes aos bairros estudados, observou-se que o SisLeptos comparou o caso de entrada com os casos

contidos na base, para encontrar características semelhantes a situação atual e, em seguida, aplicou a regra de adaptação para os casos recuperados, com o intuito de sugerir a probabilidade do caso consultado ser confirmado, sendo que nesta circunstância específica foi de 91,9%, conforme ilustra a figura 5.2.

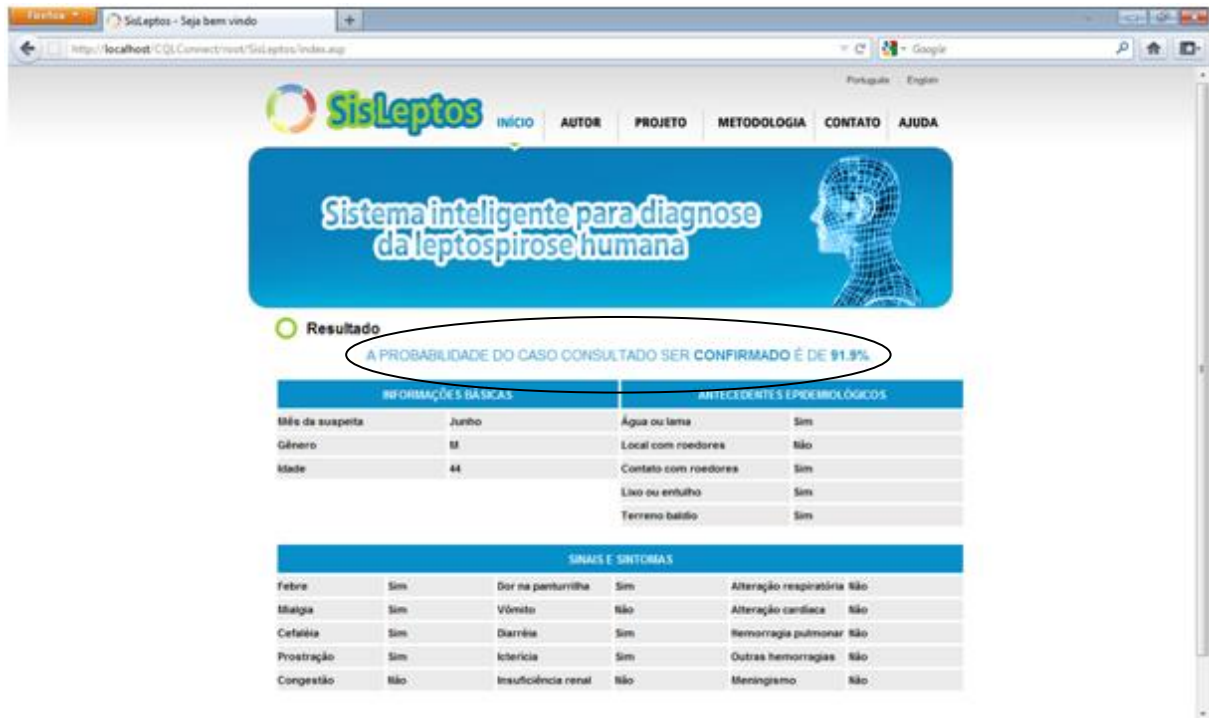


Figura 5.25. Tela com o resultado de um caso confirmado de leptospirose.

Verificou-se que o resultado produzido, representado na figura 5.2, foi decorrente da utilização da técnica de adaptação nula, que é responsável por analisar e classificar a melhor solução entre os casos recuperados para definir a saída adequada ao problema consultado, ou seja, encontrar a situação com maior grau de similaridade em relação caso de entrada.

Considerando um caso descartado de leptospirose do bairro da Cremação, que o SisLeptos analisou a probabilidade do problema consultado ser descartado com 95,5%, conforme mostra a figura 5.3. Este caso foi diagnosticado conforme citado anteriormente, devido o conjunto de sintomas característicos do surto estarem preenchidos como “Não”. Já sintomas como febre, mialgia, cefaleia, vômito e diarreia, apesar informados como “Sim”, são comuns em diversas doenças.

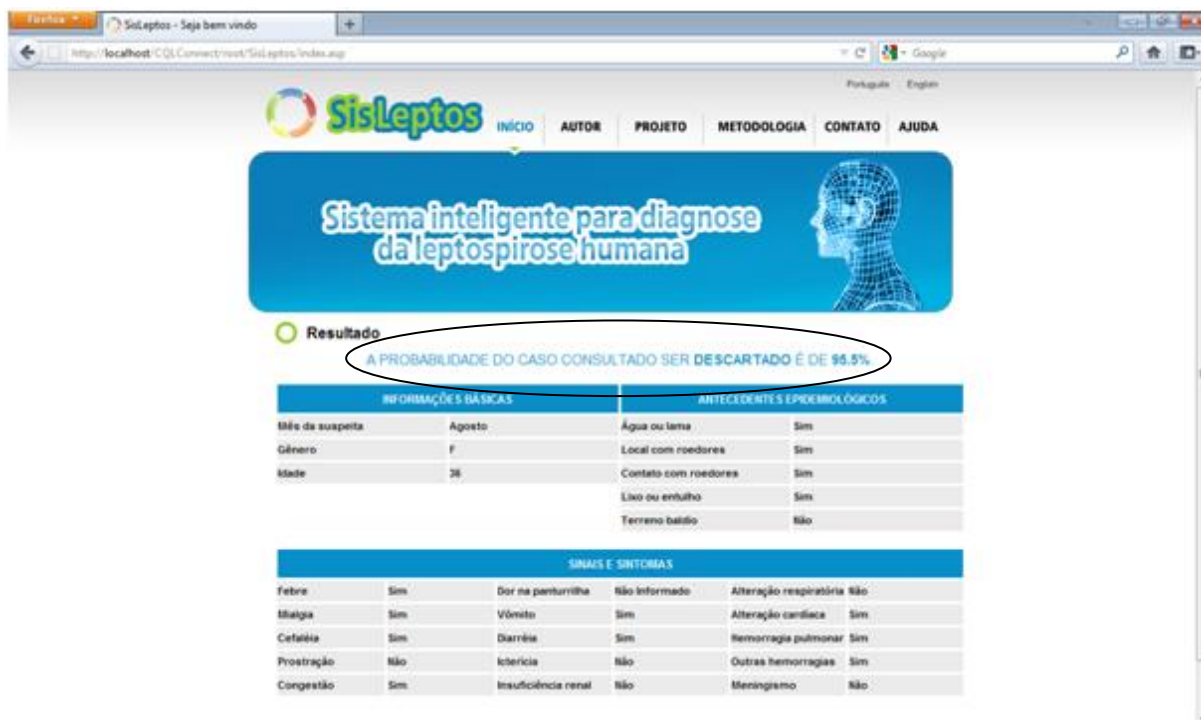


Figura 5.26. Tela com o resultado de um caso descartado de leptospirose.

Até o presente momento os testes realizados com bairros semelhantes ao estudados, como por exemplo, os bairros da Cremação e Pedreira, estão sendo bem sucedidos, com 9 acertos entre 10 casos consultados, atingindo um índice de 90%, entretanto, o SisLeptos ainda não foi testado por nenhum especialista da área, o que poderia garantir maior validação dos resultados fornecidos pelo sistema. Para que o SisLeptos obtenha melhores resultados, ainda é necessário a inclusão de casos já analisados. Portanto, quanto mais sólida a base de casos do sistema, o diagnóstico fornecido será mais preciso e confiável.

Levando em consideração a utilização da técnica de RBC, foi avaliado que a mesma obteve sucesso em caracterizar a leptospirose humana nas áreas de estudo, nos anos de 2008, 2009 e 2010, por analisar questões semelhantes, como características epidemiológicas, socioeconômicas e ambientais. Com isso, a técnica de RBC foi considerada eficiente para diagnosticar casos do surto, sendo possível compreender as relações que influenciaram determinada área ser mais suscetível a infecções que outras.

5.2. Dificuldades Encontradas

O primeiro desafio foi encontrar um ambiente capaz de ser adaptado para a realização das análises com a técnica de RBC. Após uma busca exaustiva, encontrou-se o *shell* CBR-Works, pertencente a Empolis. O CBR-Works foi disponibilizado, devido o caráter acadêmico

do trabalho proposto, possibilitando o desenvolvimento de um sistema inteligente para auxiliar na diagnose da leptospirose humana.

Em relação ao Geoprocessamento, inicialmente, a maior dificuldade encontrada foi na escolha dos processos a serem adotados, devido ao expressivo número de características e variáveis necessárias para que a modelagem representasse, o máximo possível, a realidade do bairro do Guamá.

Observou-se no início do desenvolvimento do SisLeptos, que os índices aplicados resultaram em análises insatisfatórias, visto que foram atribuídos pesos incorretos em relação às características da base. Para melhorar os resultados obtidos pelo sistema, fez-se necessário o suporte de especialistas e o uso de fontes literárias, os quais serviram de apoio para uma melhor atribuição dos pesos aos índices de acordo com a doença, mas não com a realidade do problema, gerando, a princípio, inconsistência ao sistema, carecendo maior aprimoramento das mesmas.

O SisLeptos também encontrou problemas ao cobrir o domínio, devido as limitações da base fornecida pela SESPA, como o não preenchimento de dados relevantes para análise de um caso, relacionados à sua incompletude.

A última dificuldade encontrada foi relacionada à implementação de um algoritmo capaz de viabilizar a adaptação dos casos, considerando que o mesmo utiliza regras complexas de difícil compreensão para serem desenvolvidos em um sistema de RBC, como as técnicas de soluções parametrizadas, abstração e especialização.

5.3. Trabalhos Futuros

A partir desta dissertação é possível derivar um conjunto de temas para futuras pesquisas. Neste sentido, propõe-se como trabalho futuro, o desenvolvimento de um sistema de Raciocínio Baseado em Casos, utilizado uma linguagem de programação orientada a objetos, como por exemplo, Java ou C++, que possibilitem a construção das cinco etapas de desenvolvimento de um modelo RBC. Integrado ao desenvolvimento deste sistema, sugere-se também o uso de Algoritmos Genéticos para auxiliar na etapa de indexação, devido os mesmos serem capaz de atribuir melhores pesos aos índices, de acordo com as características encontradas na base.

Para observar com mais propriedade as relações ambientais que interferem na leptospirose, como por exemplo, a drenagem e a altimetria, indica-se como trabalhos futuros a técnica de tridimensionalização dos dados ambientais, que possibilitarão uma melhor visualização dos dados trabalhados.

Levando em consideração as limitações do ambiente CBR-Works, observou-se que o desenvolvimento de uma rotina gráfica que pudesse ser acoplada ao sistema possibilitaria análises envolvendo imagens e dados (numéricos e textuais) de forma integrada, os quais poderiam gerar análises mais precisas.

Como última consideração, recomenda-se testar e avaliar os resultados do sistema proposto, por meio de métodos estatísticos. Com isso, o sistema ficaria apto para o uso de médicos e especialistas através da Web, visto que, no momento, o SisLeptos está disponível apenas localmente. A disponibilização na Internet possibilitaria o acesso remoto aos dados epidemiológicos, para que profissionais da área da saúde, de órgãos públicos e privados, possam inserir novos casos, os quais seriam recuperados para auxiliar no processo de tomada de decisão para o diagnóstico da leptospirose humana.

Referências Bibliográficas

- ABEL, Mara. **Um Estudo Sobre Raciocínio Baseado em Casos**. 1996. 41f. Monografia (Especialização em Ciência da Computação) – Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1996.
- ALMEIDA FILHO, Naomar de; ROUQUAYROL, Maria Zélia. **Introdução à epidemiologia moderna**. Salvador: Apce Produtos do Conhecimento, 1990. 223 p.
- AVILA-PIRES, Fernando Dias de. Leptospirose e enchentes: uma falsa correlação. **Revista de Patologia Tropical**, Departamento de Medicina Tropical, Instituto Oswaldo Cruz, Rio de Janeiro, v. 35, Ed. 3, p. 199-204, set./dez. 2006.
- BARROS, Claudio; RISSI, Daniel; RECH, Raquel. **Descrição e interpretação macroscópica das lesões**. Encontro Nacional de Diagnóstico Veterinário. Campo Grande: ENDIVET, 2008. 7 p.
- BELLUSCI, Silvia Meirelles. **Epidemiologia**. 4 ed. São Paulo: Senac, 1995. 82 p.
- BORGHETTI, Alex. **Caracterização Geoambiental da Bacia Hidrográfica do Rio das Antas Utilizando Ferramentas de Geoprocessamento**. 2006. 85f. Dissertação (Mestrado em Engenharia Agrícola) – Programa de Pós-graduação em Engenharia Agrícola, Universidade Estadual do Oeste do Paraná, Cascavel, 2006.
- CARVALHO FILHO, L. M. *et al.* **Disseminação e Desenvolvimento do Geoprocessamento: Sistemas Geográficos de Informação para Situações Territoriais Relevantes**. 2000.
- CHAKRABORTY, Baisakhi. *et al.* **Case-Based Reasoning methodology for Diagnosis of Swine Flu**. In: *2011 IEEE GCC Conference and Exhibition (GCC)*, Dubai, Emirados Árabes Unidos, p. 132-135, 2010.
- DALFOVO, Oscar. *et al.* **Raciocínio Baseado em Casos utilizando a Dieta do Tipo Sangüíneo**. Blumenau: Universidade Regional de Blumenau, 1998. 16 p.
- DEITEL, Harvey M. *et al.* **XML: Como Programar**. São Paulo: Bookman, 2003. 972 p.
- DEPARTAMENTO DE VIGILÂNCIA EPIDEMIOLÓGICA. **Guia de Vigilância Epidemiológica**. Secretaria de Vigilância em Saúde. 6. ed. Brasília: Ministério da Saúde, 2005. 815 p.
- FERNANDES, Anita Maria da Rocha. **Inteligência Artificial: Noções Gerais**. 3. imp. Florianópolis: VisualBooks, 2005. 160 p.
- FURTADO, Vasco. **Tecnologia e Gestão da Informação na Segurança Pública**. Rio de Janeiro: Garamond, 2002. 261 p.
- GRESSE VON WANGENHEIM, Christiane. **REMEX: A Case-Based Approach for Reuse of Software Measurement Experienceware**. In: *Proceedings of 3RD INTERNATIONAL CONFERENCE OF CASE-BASED REASONING*, Munich, Alemanha, 1999.

GRESSE VON WANGENHEIM, Christiane; VON WANGENHEIM, Aldo. **Raciocínio Baseado em Casos**. Barueri: Manole, 2003, 300 p.

HEINRICH, Daniel Jonas. **Sistema de Apoio para o Diagnóstico de Defeitos em Equipamentos Eletrônicos Aplicado a Oficinas Eletrônicas Utilizando Raciocínio Baseado em Casos**. 2001. 57f. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Centro de Ciências Exatas e Naturais, Universidade Regional de Blumenau, Blumenau, 2001.

ISHIZUKA, Masaio Mizuno. **Vigilância Epidemiológica em Saúde Animal para o Estado do Espírito Santo**. XXXV Semana Capixaba do Médico Veterinário e III Encontro Regional de Saúde Pública em Medicina Veterinária. Guarapari: CRMVES, 2008. 18 p.

JÚLIO, Márcia Regina Ferro Móss. **Um estudo de métricas de similaridade em sistemas baseados em casos aplicados à Área da Saúde**. 2005. 136f. Dissertação (Mestrado em Computação) – Instituto de Computação, Universidade Estadual de Campinas, Campinas, 2005.

KATEDEE, Suchada; SANRACH, Charan; THESAWADWONG, Thiwapom. **Case-Based Reasoning System for Histopathology Diagnosis**. In: *2010 International Conference on Educational and Information Technology (ICEIT 2010)*, Chongqing, China, p. 496-498, 2010.

KOLODNER, Janet. **Case-Based Reasoning**. San Mateo: Morgan Kaufmann, 1993. 668 p.

KOSLOSKY, Marco Antônio Neiva. **Aprendizagem Baseada em Casos: Um Ambiente para Ensino de Lógica de Programação**. 1999. 113f. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 1999.

KURBALIJA, V. et al. **Multiple Sclerosis Diagnoses: Case-Based Reasoning Approach**. In: *Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)*, Novi Sad, Sérvia, p. 65-72, 2007.

LAGEMANN, Gerson Volney. **RBC para o Problema de Suporte ao Cliente nas Empresas de Prestação de Serviço de Software: O Caso Datasul**. 1998. Dissertação (Mestrado em Engenharia de Produção) – Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 1998.

LARMAN, Craig. **Utilizando UML e Padrões: Uma introdução à análise e ao projeto orientados a objetos e ao desenvolvimento iterativo**. 3 ed. Porto Alegre: Bookman, 2007. 695 p.

LEE, Rosina Weber. **Pesquisa Jurisprudencial Inteligente**. 1998. Tese (Doutorado em Engenharia de Produção) – Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 1998.

MELCHIORS, Cristina. **Raciocínio Baseado em Casos Aplicado ao Gerenciamento de Falhas em Redes de Computadores**. 1999. 151f. Dissertação (Mestrado em Ciência da Computação) – Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1999.

MENDES, Marcio A. *et al.* **Care Systematization in Pediatric Nursing Applying Case-based Reasoning**. In: PICMET 2009. *Portland International Conference on Management of Engineering & Technology*, Oregon, EUA, p. 3011-3017, 2009.

MIRANDA, José Iguelmar. **Fundamentos de Sistemas de Informações Geográficas**. Brasília: EMBRAPA, 2005. 425 p.

MOURA, Ana Clara Mourão. **Geoprocessamento na gestão e planejamento urbano**. 2 ed. Belo Horizonte: Aurora, 2005, 294 p.

NASCIMENTO, Susana. **Fuzzy Clustering via Proportional Membership Model**. Amsterdam: IOS Press, 2005, 200 p.

PAULA, Eduardo Vedor de. **Leptospirose Humana: Uma Análise Climato-geográfica de sua Manifestação no Brasil, Paraná e Curitiba**. In: Anais XII SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, Goiânia, p. 2301-2308, 2005.

PAZ, Juan F. de. *et al.* **CBR System for Diagnosis of Patients**. In: 2008. HIS '08. *Eighth International Conference on Hybrid Intelligent Systems*, Barcelona, Espanha, p. 807-812, 2008.

RAMIREZ, Milton Ramos. **Sistemas Gerenciadores de Banco de Dados para Geoprocessamento**. 1994. 241f. Dissertação (Mestrado em Engenharia de Sistemas e Computação) – Programa de Engenharia de Sistemas e Computação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1994.

REZENDE, Solange Oliveira. **Sistemas Inteligentes: Fundamentos e Aplicações**. 1. imp. Barueri: Manole, 2005, 526 p.

RICH, Elaine; KNIGHT, Kevin; NAIR, Shivashankar B. **Artificial Intelligence**. 3. ed. Nova Delhi: Tata McGraw Hill Education Private Limited, 2010. 588 p.

SCHMIDT, Rainer; GIERL, Lothar. **Case-Based Reasoning for Medical Knowledge-based Systems**. In: MIE/GMDS-2000 *Workshop*, Rostock, Alemanha, p. 1-34, 2000.

SILVA, Jorge Luiz Barbosa da. **Geoprocessamento Aplicado à Identificação de Áreas para Rejeitos e Estimativa de Recurso de Carvão na Região da Mina Leão II**. 2000. 166f. Dissertação (Mestrado em Sensoriamento Remoto) – Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2000.

TEIXEIRA, João de Fernandes. **Inteligência Artificial: uma odisseia da mente**. São Paulo: Paulus, 2009. 64 p.

VERGARA, Walter Hernández. **Simulação cognitiva da tomada de decisão em situações complexas: modelagem do raciocínio humano por Meio de Casos**. 1995. Tese (Doutorado em Engenharia de Produção) – Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 1995.

VERONESI. **Tratado de Infectologia**. 3. ed. São Paulo: Atheneu, 2005. 1271 p.

Webgrafia

CASTOLDI, Augusto Cesar; SANTOS, Marcos de Oliveira dos. (2002). **Raciocínio Baseado em Casos**. Artigo da Universidade Federal de Santa Catarina. Disponível em: <<http://www.inf.ufsc.br/~barreto/trabaluno/IA20022AugMarc.pdf>>. Acesso em: 12 maio 2011.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Estados**, 2011. Disponível em: <<http://www.ibge.gov.br/estadosat>>. Acesso em: 13 jan. 2012.

L'INTERNAUTE (2011). ***Leptospirose: les origines de la contamination***. Disponível em: <<http://www.linternaute.com/sante/magazine/dossier/zoonoses/12.shtml>>. Acesso em: 21 set. 2011.

MARCOS, Soraia Tomich; DIAS, Izabel C. (2005). **As Espécies de Raciocínio**: Dedução, Indução e Abdução. Disponível em <<http://www.unimar.br/pos/semiotica/raciocinio.doc>>. Acesso em: 13 mar. 2011.

MARTINS, Fernando S. V.; CASTIÑEIRAS, Terezinha Marta P. P. (2009). **Leptospirose**. Artigo do Centro de Informação em Saúde para Viajantes. Disponível em: <<http://www.cives.ufrj.br/informacao/leptospirose/lep-iv.html>>. Acesso em: 21 abr. 2011.

SANTANA, Kátia (2008). **Servidores Denunciam Precariedade da Saúde**. Artigo da Procuradoria Federal dos Direitos do Cidadão. Disponível em: <<http://2008.jornaldacidade.net/2008/noticia.php?id=17128>>. Acesso em: 31 jul. 2011.

MINISTÉRIO DA SAÚDE. **Sistema de Informação de Agravos de Notificação**, 2011. Disponível em: <<http://dtr2004.saude.gov.br/sinanweb/tabnet/dh?sinannet/lepto/bases/leptobrnet.def>>. Acesso em: 01 jun. 2011.

SOCIEDADE BRASILEIRA DE INFECTOLOGIA (2009). **Doenças infecciosas e parasitárias matam 44,3 mil pacientes em 1 ano no Brasil**. Disponível em: <http://www.sbinfecto.org.br/default.asp?site_Acao=&paginaId=134&mNoti_Acao=mostraNoticia¬iciaId=5423>. Acesso em: 05 out. 2011.

Referências Complementares

LABGEO. **Fotos da urbanização desenfreada e ausência de infraestrutura básica.** 2008. 2 fotografias, color.

LABGEO. **Banco de Dados de Geoprocessamento.** Belém: Instituto Evandro Chagas, n. 1, 2011. CD-ROM.

SISTEMA DE INFORMAÇÃO DE AGRAVOS DE NOTIFICAÇÃO – SINAN. **Banco de Dados da Leptospirose.** Brasília: SINAN, n. 1, 2011. CD-ROM.

APÊNDICE

APÊNDICE A: TELAS DO SISLEPTOS

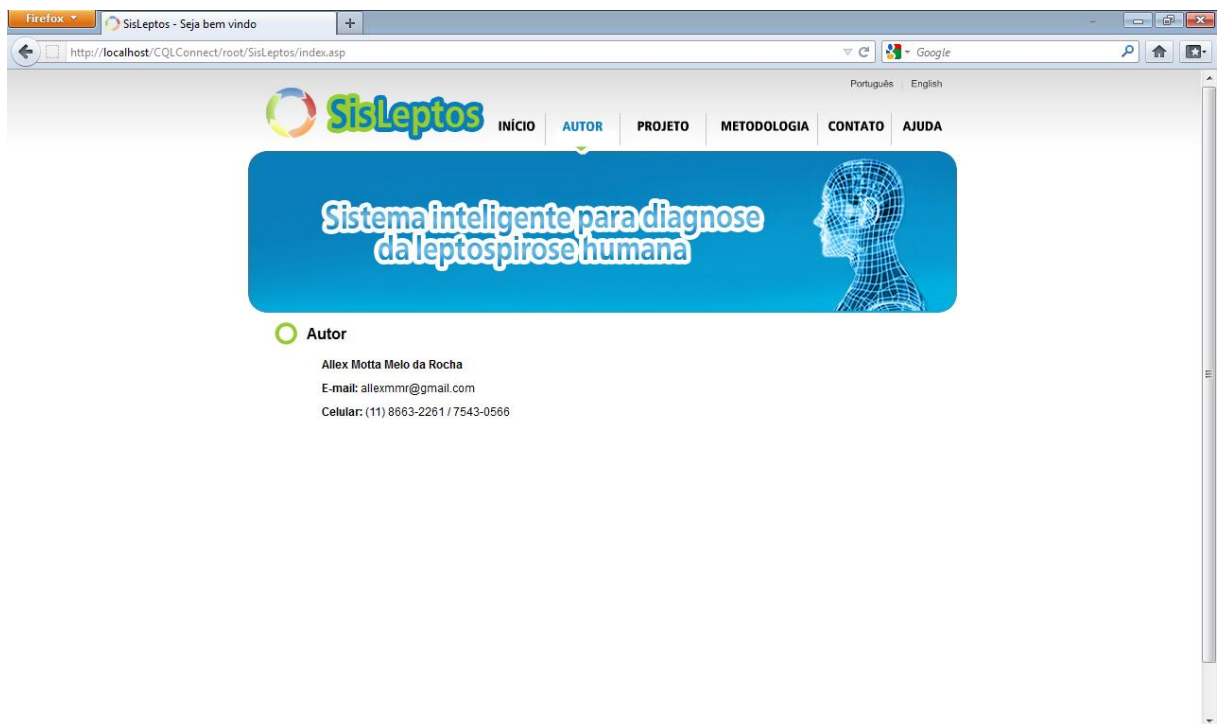


Figura A.1. Tela de autor do SisLeptos, no idioma português.

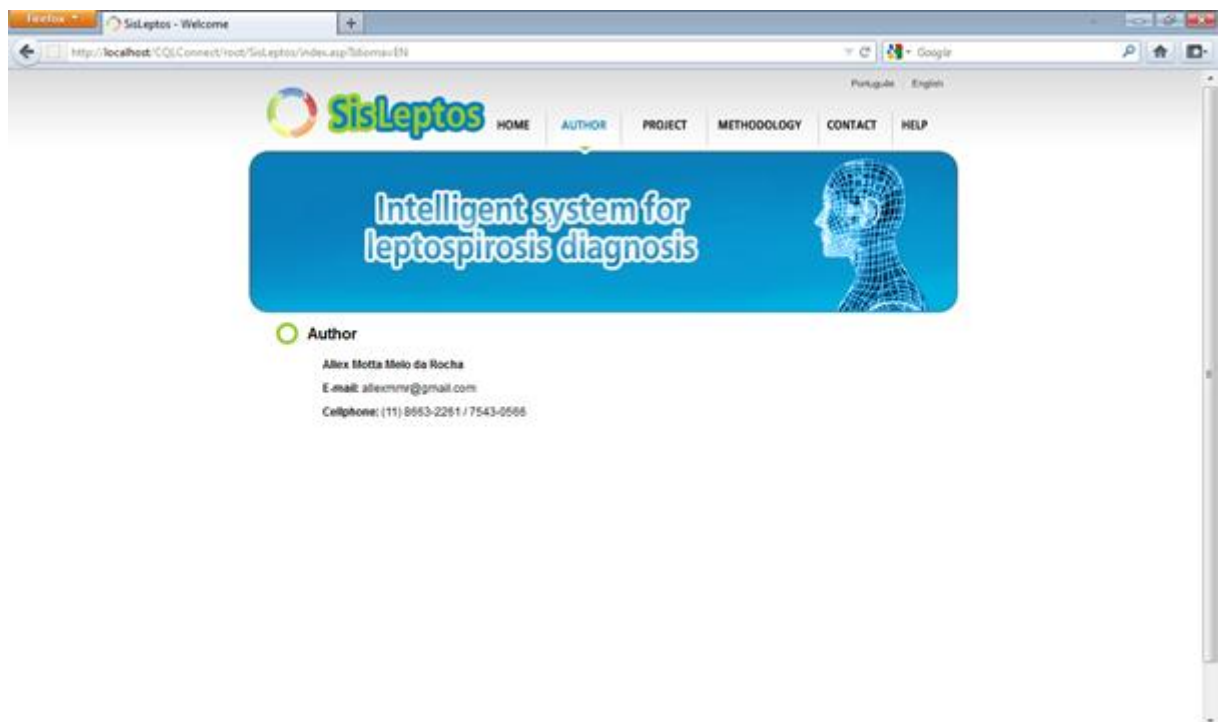


Figura A.2. Tela de autor do SisLeptos, no idioma inglês.



Figura A.3. Tela do projeto do SisLeptos.

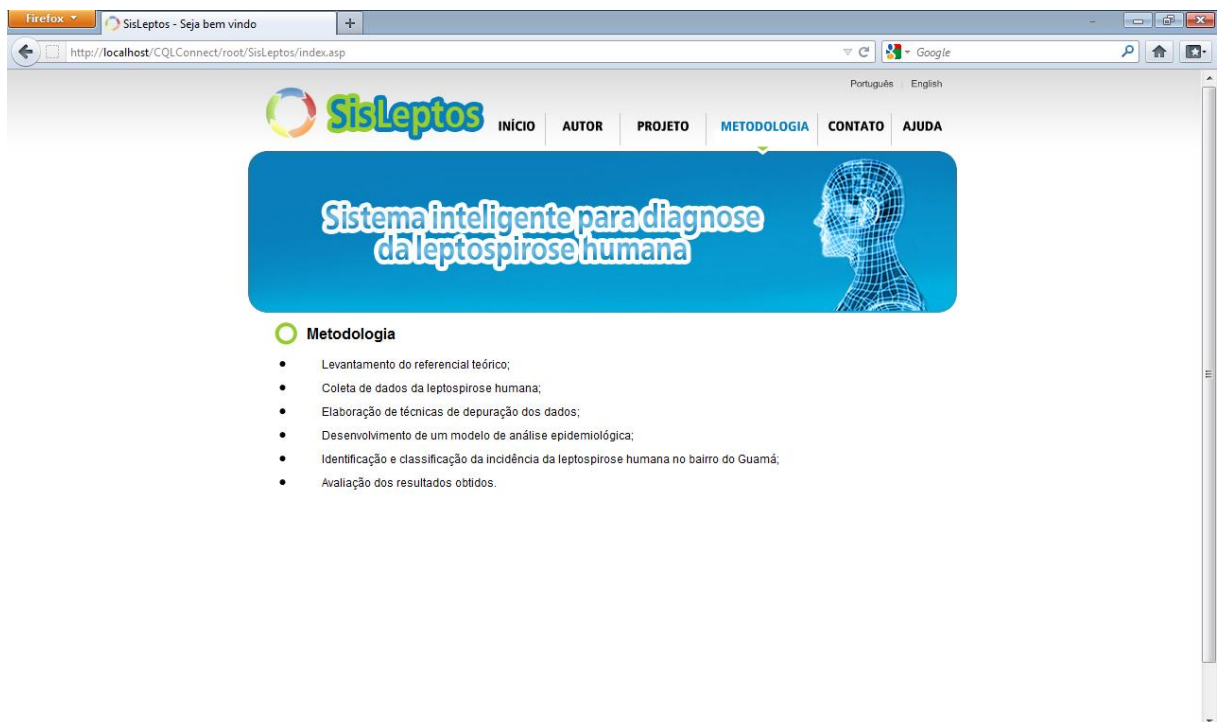


Figura A.4. Tela da metodologia do SisLeptos.

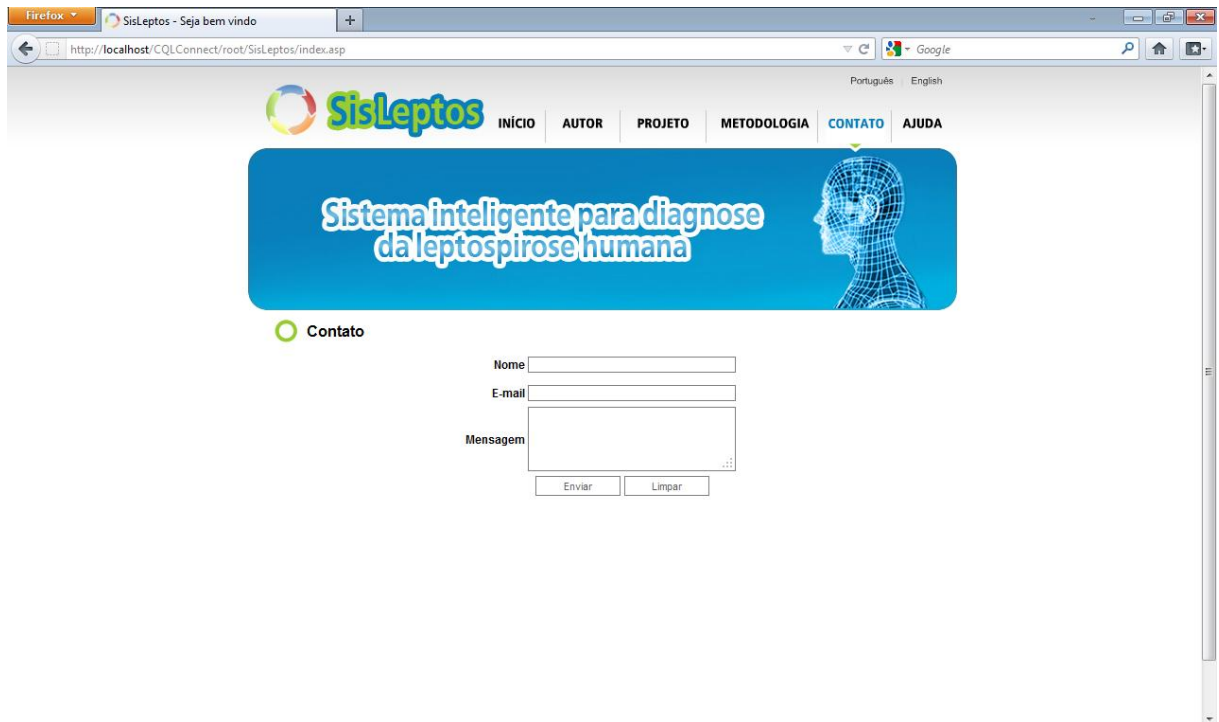


Figura A.5. Tela de contato do SisLeptos.

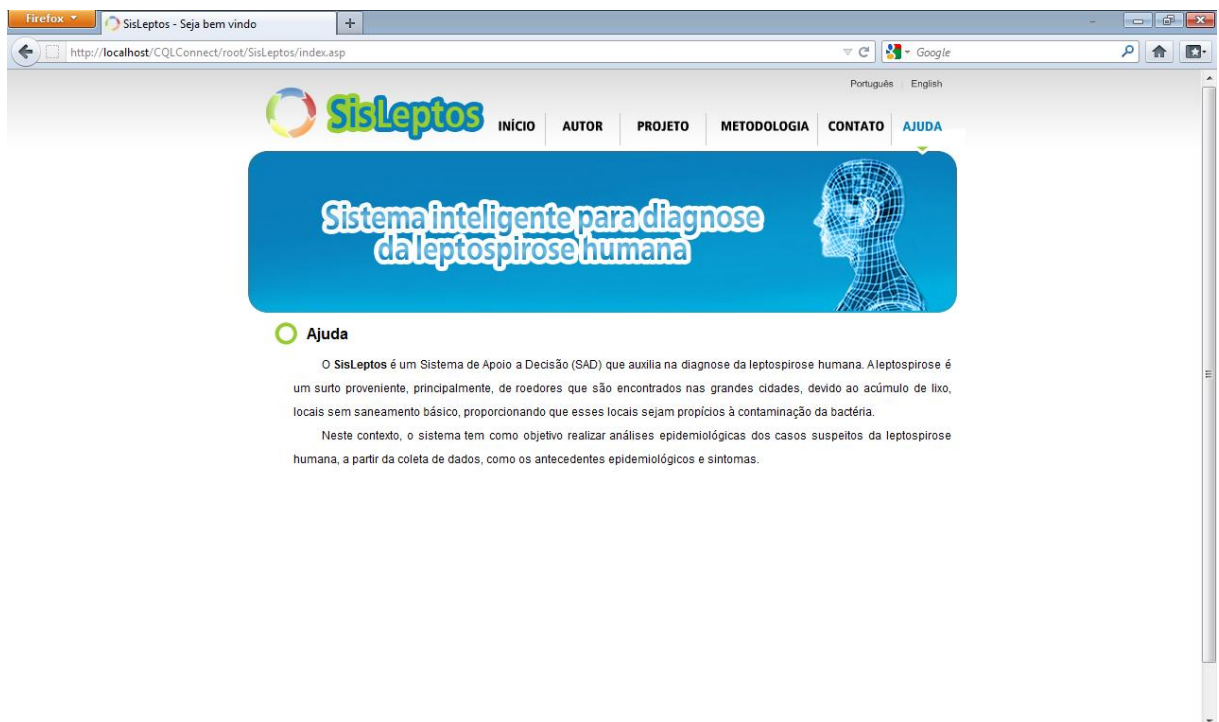


Figura A.6. Tela de ajuda do SisLeptos.

ANEXO

ANEXO A: FICHA CADASTRAL DE LEPTOSPIROSE

Dados do Laboratório				
Sorologia IgM - Elisa 43 Data da Coleta - 1ª amostra 44 Resultado 1ª Amostra 1 - Reagente 2 - Não Reagente 3 - Inconclusivo 4 - Não realizado 45 Data da Coleta - 2ª amostra 46 Resultado 2ª Amostra 1 - Reagente 2 - Não reagente 3 - Inconclusivo 4 - Não realizado				
Microagulinação 47 Data da Coleta - Micro 1ª amostra 48 Micro 1ª Amostra 1º sorovar título 49 Micro 1ª Amostra 2º sorovar título				
50 Resultado MICRO-aglutinação 1ª Amostra 1 - Reagente 2 - Não Reagente 3 - Não realizada 9 - Ignorado				
51 Data da Coleta - Micro 2ª amostra 52 Micro 2ª Amostra 1º sorovar título 53 Micro 2ª Amostra 2º sorovar título				
54 Resultado MICRO-aglutinação 2ª Amostra 1 - Reagente 2 - Não Reagente 3 - Não realizada 9 - Ignorado				
Isolamento 55 Data da Coleta 56 Resultado 1 - Positivo 2 - Negativo 3 - Inconclusivo 4 - Não realizado				
Imunohistoquímica 57 Data da Coleta 58 Resultado 1 - Positivo 2 - Negativo 3 - Inconclusivo 4 - Não realizado				
RT-PCR 59 Data da Coleta 60 Resultado 1 - Positivo 2 - Negativo 3 - Inconclusivo 4 - Não realizado				
Conclusão 61 Classificação Final 1 - Confirmado 2 - Descartado 62 Critério de Confirmação ou Descarte 1 - Clínico-Laboratorial 2 - Clínico-Epidemiológico				
Local Provável da Fonte de Infecção 63 O caso é autóctone do município de residência? 1 - Sim 2 - Não 3 - Indeterminado				
64 UF 65 País				
66 Município 67 Distrito 68 Bairro				
Característica do Local Provável de Infecção 69 Área provável de infecção 1 - Urbana 2 - Rural 3 - Peri-Urbana 9 - Ignorado				
70 Ambiente da Infecção 1 - Domiciliar 2 - Trabalho 3 - Lazer 4 - Outro 9 - Ignorado				
71 Doença Relacionada ao Trabalho 1 - Sim 2 - Não 9 - Ignorado 72 Evolução do Caso 1 - Cura 2 - Óbito por leptospirose 3 - Óbito por outras causas 9 - Ignorado				
73 Data do Óbito 74 Data do Encerramento				
Informações complementares e observações				
Data e Endereço se esteve em Situação de Risco Ocorrida nos 30 dias que Antecederam os Primeiros Sintomas				
Data	UF	Município	Endereço	Localidade
Observações: 				
Investigador Município/Unidade de Saúde Nome Função Sinan NET Assinatura Código da Unid. de Saúde				
LEPTOSPIROSE 15/12/2009 MR COREL Leptospirose SVS 02/02/2007				

LEPTOSPIROSE
terça-feira, 15 de dezembro de 2009 16:13:57

Figura A.7. Primeira página da ficha cadastral com as informações sobre leptospirose.

 REPÚBLICA FEDERATIVA DO BRASIL MINISTÉRIO DA SAÚDE ESTADO DE SÃO PAULO SECRETARIA DE ESTADO DA SAÚDE		SINAN SISTEMA DE INFORMAÇÃO DE AGRAVOS DE NOTIFICAÇÃO FICHA DE INVESTIGA DE LEPTOSPIROSE		Nº
CASO SUSPEITO: Indivíduo com febre, cefaléia e mialgia, que apresente pelo menos um dos seguintes critérios: Critério 1- antecedentes epidemiológicos sugestivos nos 30 dias anteriores à data de início dos sintomas (exposição a situações de risco, vínculo epidemiológico com um caso confirmado por critério laboratorial ou residir/trabalhar em áreas de risco); Critério 2- pelo menos um dos seguintes sinais ou sintomas: sufusão conjuntival, sinais de insuficiência renal aguda, icterícia e/ou aumento de bilirrubinas e fenômeno hemorrágico.				
Dados Gerais	1 Tipo de Notificação	2 - Individual		
	2 Agravado/doença	LEPTOSPIROSE		3 Data da Notificação
	4 UF	5 Município de Notificação	Código (CID10) A 2 7. 9	Código (IBGE)
Notificação Individual	6 Unidade de Saúde (ou outra fonte notificadora)	Código	7 Data dos Primeiros Sintomas	
	8 Nome do Paciente	9 Data de Nascimento		
	10 (ou) Idade	11 Sexo	12 Gestante	13 Raça/Cor
	14 Escolaridade			
	15 Número do Cartão SUS	16 Nome da mãe		
Dados de Residência	17 UF	18 Município de Residência	Código (IBGE)	19 Distrito
	20 Bairro	21 Logradouro (rua, avenida,...)	Código	
	22 Número	23 Complemento (apto., casa, ...)	24 Geo campo 1	
	25 Geo campo 2	26 Ponto de Referência	27 CEP	
	28 (DDD) Telefone	29 Zona	30 País (se residente fora do Brasil)	
	Dados Complementares do Caso			
	31 Data da Investigação	32 Ocupação		
Antecedentes Epidemiológicos	33 Situação de Risco Ocorrida nos 30 dias que Antecederam os Primeiros Sintomas - Contato/ limpeza de:			
	☐ Água ou lama de enchente ☐ Fossa, caixa de gordura ou esgoto ☐ Rio, córrego, lagoa ou represa ☐ Terreno baldio ☐ Criação de animais ☐ Local com sinais de roedores ☐ Roedores diretamente ☐ Lixo/ entulho ☐ Caixa d'água ☐ Plantio/ colheita (lavoura) ☐ Armazenamento de grãos/ alimentos ☐ Outras			
	34 Casos Anteriores de Leptospirose no Local Provável de Infecção nos últimos dois meses			
Dados Clínicos	35 Data de Atendimento	36 Sinais e Sintomas		
	☐ Febre ☐ Congestão conjuntival ☐ Icterícia ☐ Hemorragia pulmonar ☐ Mialgia ☐ Dor na panturrilha ☐ Insuficiência renal ☐ Outras hemorragias ☐ Cefaléia ☐ Vômito ☐ Alterações res piratórias ☐ Meningismo ☐ Prostração ☐ Diarréia ☐ Alterações cardíacas ☐ Outros, quais?			
Atendimento	37 Ocorreu Hospitalização	38 Data da Internação		39 Data de Alta
	40 UF	41 Município do Hospital	Código (IBGE)	
	42 Nome do Hospital	Código		

LEPTOSPIROSE 15/12/2009 MR COREL

Leptospirose SVS 02/02/2007

LEPTOSPIROSE
terça-feira, 15 de dezembro de 2009 16:13:56**Figura A.8.** Segunda página da ficha cadastral com as informações sobre leptospirose.