



Universidade Metodista de Piracicaba
Faculdade de Ciências Exatas e da Natureza
Mestrado em Ciência da Computação

**SISTEMA DE APOIO À DECISÃO PARA AUDITORIA INTERNA COM
RACIOCÍNIO BASEADO EM CASOS**

ALEXANDRE CARVALHO DANTAS

ORIENTADOR: PROF. DR. RAFAEL FERREIRA ALVES

Dissertação apresentada ao Mestrado em Ciência da Computação, da Faculdade de Ciências Exatas e da Natureza, da Universidade Metodista de Piracicaba – UNIMEP, como requisito para obtenção do Título de Mestre em Ciência da Computação.

PIRACICABA
2005

SISTEMA DE APOIO À DECISÃO PARA AUDITORIA INTERNA COM RACIOCÍNIO BASEADO EM CASOS

AUTOR: ALEXANDRE CARVALHO DANTAS

ORIENTADOR: PROF. DR. RAFAEL FERREIRA ALVES

Dissertação de Mestrado defendida e aprovada em 13 de dezembro de 2005,
pela Banca Examinadora constituída dos Professores:

Prof. Dr. Rafael Ferreira Alves

UNIMEP

Prof. Dr. Luiz Eduardo Galvão Martins

UNIMEP

Prof. Dr. Manuel de Jesus Mendes

UNICAMP

Dedico esse trabalho à

Minha esposa Daniela pelo apoio, persistência e compreensão.

Ao meu filho Luís Henrique, que perdeu alguns momentos de atenção para que esse trabalho fosse realizado.

Aos meus pais que sempre me incentivaram a enfrentar novos desafios.

A Deus que sempre esteve comigo

AGRADECIMENTOS

Ao professor Rafael Ferreira Alves pela orientação e Incentivo dispensado ao desenvolvimento deste trabalho.

Aos colegas de trabalho pela colaboração e compartilhamento de experiências e conhecimentos.

A todas as pessoas que contribuíram para a realização desse trabalho, de forma direta ou indireta.

Resumo

A melhoria dos resultados apresentados pelos sistemas de apoio à decisão (SAD) disponíveis tem proporcionado ganhos consistentes de tempo, segurança e recursos. Entretanto, ainda hoje existe uma carência de SAD para áreas específicas como a auditoria interna, pois os sistemas dessa área focam muito mais a automatização de uma tarefa operacional do que o auxílio a decisões, devido à dificuldade em estruturar os conhecimentos envolvidos. Este trabalho tem como objetivo desenvolver um protótipo de um sistema de apoio à decisão utilizando o raciocínio baseado em casos (RBC) que possa realizar testes na área de auditoria interna de uma instituição financeira e através dos resultados obtidos, possibilitar a análise comparativa das vantagens ou desvantagens de sua adoção. A escolha do RBC como técnica principal no SAD, é de certa forma natural, pois intuitivamente, o ser humano se vale de experiências anteriores para a tomada de decisões. Essa escolha busca, portanto, reduzir a grande dificuldade de qualquer sistema especialista para auditoria, que é a definição de regras de conhecimento. Essa contribuição, somada a uma metodologia de seleção das melhores técnicas de cálculo de similaridade e recuperação de casos possibilitam a obtenção dos melhores resultados e o desenvolvimento de sistemas que sejam ao mesmo tempo avançados e viáveis com uma implantação gradativa e segura.

PALAVRAS-CHAVE: Inteligência Artificial, Sistemas Especialistas, Sistemas de Apoio à Decisão, Raciocínio Baseado em Casos.

A DECISION SUPPORT SYSTEM TO INTERNAL AUDIT WITH CASE BASED REASONING

Abstract

The improvement in the results presented by the available Decision Support Systems (DSS) has promoted consistent gains of time, security and resources. However, nowadays still exists a lack of DSS for specific areas like internal audit, because the systems are much more focused in automate an operational task than in supporting decisions, due to the difficulty in organizing the involved knowledge for these areas. The goal of this work is to develop a prototype of a DSS using Case Based Reasoning (CBR) that is able to accomplish tests in the internal audit area of a financial institution and based on the reached results, supports the comparative analysis of the advantages and disadvantages of implementing it. Choosing the CBR as the main technique of the DSS is natural, in a certain way, because human beings tend to use their previous experiences to support their decision-making process. So, this choice intends to reduce a huge difficulty of any specialist system for audit, that is the knowledge rules definition. This contribution, combined to a methodology for selection of the best techniques for similarity calculation and cases recovery, make possible to achieve the best results, developing systems that are advanced and viable at the same time and implanting it in a safe and gradual way.

KEYWORDS: Artificial Intelligence, Expert Systems, Decision Support Systems, Case Based Reasoning.

SUMÁRIO

LISTA DE FIGURAS	X
LISTA DE QUADROS	XI
LISTA DE ABREVIATURAS	XII
1 <u>Introdução</u>	1
1.1 <u>Contextualização da Pesquisa</u>	1
1.2 <u>Objetivos</u>	2
1.3 <u>Justificativa da Pesquisa</u>	3
2 <u>Revisão Bibliográfica</u>	6
2.1 <u>Informação e Decisão</u>	6
2.2 <u>Sistemas de Apoio à Decisão – SAD</u>	8
2.2.1 <u>Sistemas e Sistemas de Informação</u>	8
2.2.2 <u>Breve Histórico dos Sistemas de Apoio à Decisão</u>	9
2.2.3 <u>Características de Um Sistema de Apoio à Decisão</u>	10
2.2.4 <u>Recursos de Um Sistema de Apoio à Decisão</u>	11
2.2.5 <u>Etapas de Elaboração de Um Sistema de Apoio à Decisão</u>	12
2.2.6 <u>Uma Comparação Entre SAD, SIE e SIG</u>	13
2.3 <u>Inteligência Artificial - IA</u>	14
2.3.1 <u>Inteligência</u>	15
2.3.2 <u>Breve Histórico da Inteligência Artificial</u>	16
2.3.3 <u>Linhas de Pesquisa da Área de Inteligência Artificial</u>	17
2.4 <u>Sistemas Especialistas - SE</u>	17
2.4.1 <u>Breve Histórico de Sistemas Especialistas</u>	18
2.4.2 <u>Características de Um Sistema Especialista</u>	18
2.4.3 <u>Tarefas Típicas de Sistemas Especialistas</u>	20
2.4.4 <u>Limitações/Problemas dos Sistemas Especialistas</u>	21
2.4.5 <u>Os Componentes de Um Sistema Especialista</u>	22
2.4.6 <u>O Desenvolvimento de Um Sistema Especialista</u>	23
2.5 <u>Raciocínio Baseado em Casos - RBC</u>	24
2.5.1 <u>Breve Histórico do Raciocínio Baseado em Casos</u>	26
2.5.2 <u>Vantagens e Limitações do Raciocínio Baseado em Casos</u>	27
2.5.3 <u>Funcionamento do Raciocínio Baseado em Casos</u>	30
2.5.4 <u>O Desenvolvimento de Um Sistema RBC</u>	32
2.5.5 <u>Ferramentas para RBC</u>	45
2.6 <u>Auditoria</u>	46
2.6.1 <u>Breve Histórico da Auditoria</u>	47
2.6.2 <u>A Auditoria no Brasil</u>	47
2.6.3 <u>Classificação dos Trabalhos de Auditoria</u>	47
2.6.4 <u>Auditoria Interna</u>	48
2.6.5 <u>Auditoria Operacional</u>	49
2.7 <u>A Importância de Um SE Utilizando RBC para a Auditoria</u>	55
3 <u>Metodologia de Pesquisa</u>	57
3.1 <u>Procedimentos Científicos</u>	57
3.2 <u>Classificações da Pesquisa</u>	57
3.3 <u>Métodos Científicos</u>	59
3.4 <u>Análise e Definição do Modelo de Similaridade Adotado</u>	59
3.5 <u>Análise e Definição da Técnica de Recuperação de Casos Adotada</u>	60
3.6 <u>Análise e Definição da Ferramenta de Desenvolvimento Adotada</u>	61

3.7	Estudo de Caso	62
3.7.1	Justificativa	62
3.7.2	Planejamento e Desenvolvimento do Estudo de Caso	63
3.7.3	Análise e Conclusão do Estudo de Caso	64
3.8	Análise dos Resultados	65
4	Análise e Definição do Modelo de Similaridade Adotado.....	66
4.1	Revisão Bibliográfica - Similaridade	67
4.2	Escolha e Avaliação das Características das Métricas de Similaridade	67
4.3	Definição das Características Relevantes para o Sistema	73
4.4	Determinação de Testes a Serem Aplicados	74
4.5	A Metodologia Proposta de Definição da Melhor Métrica de Similaridade para RBC	75
4.6	O Aplicativo DMMS-RBC – Definição da Melhor Métrica de Similaridade para RBC	77
4.7	Análise dos Resultados Obtidos	79
4.8	Conclusões e Definição do Método de Similaridade a Ser Adotado	84
5	Análise e Definição do Modelo de Recuperação de Casos.....	86
5.1	Revisão Bibliográfica	87
5.1.1	Introdução	87
5.1.2	Identificação de Características	88
5.1.3	Busca de Casos	88
5.1.4	Seleção de Casos	97
5.1.5	Esforço Ideal de Recuperação de Casos	97
5.2	Definição das Características Relevantes Para o Sistema a Ser Desenvolvido	98
5.3	Avaliação das Características de Cada Modelo	100
5.4	Seleção dos Modelos a Serem Analisados	103
5.5	Determinação de Testes a Serem Aplicados	104
5.6	Definição de Metodologia e Realização dos Testes	105
5.7	Coleta de Dados	106
5.8	Tabulação de Dados	108
5.8.1	Precisão	108
5.8.2	Facilidade de Uso	109
5.8.3	Quantidade de Informações	110
5.8.4	Flexibilidade	111
5.8.5	Desempenho	111
5.8.6	Consolidação dos Resultados	112
5.9	Análise e Conclusão dos Resultados e Seleção do Modelo a Ser Adotado	112
6	Análise e Definição do Software de Apoio Adotado.....	113
6.1	Pré-Seleção Das Ferramentas Disponíveis	113
6.2	Revisão de Literatura das Ferramentas Pré-Selecionadas	115
6.2.1	GAS-Pro 2003®	115
6.2.2	CASPIAN	116
6.2.3	ALICE ReCall®	117
6.3	Definição das Características Relevantes para o Sistema a Ser Desenvolvido	118
6.4	Determinação de Testes e das Metodologias Aplicadas	120
6.5	Coleta e Tabulação dos Dados	120
6.6	Análise e Conclusão dos Resultados	121

<u>7</u>	<u>Desenvolvimento de Um SAD para Auditoria Interna com RBC</u>	122
<u>7.1</u>	<u>Engenharia de Software</u>	123
<u>7.1.1</u>	<u>Engenharia de Sistemas e análise</u>	123
<u>7.1.2</u>	<u>Projeto Físico e Codificação</u>	131
<u>7.1.3</u>	<u>Testes</u>	137
<u>7.1.4</u>	<u>Implantação/Manutenção</u>	139
<u>7.2</u>	<u>Comparação da Situação Anterior e Posterior a Implantação</u>	142
<u>8</u>	<u>Conclusões</u>	144
<u>8.1</u>	<u>Dificuldades Encontradas</u>	145
<u>8.2</u>	<u>Pesquisas Futuras</u>	146
<u>9</u>	<u>Referências Bibliográficas</u>	147
<u>10</u>	<u>Glossário</u>	157
<u>11</u>	<u>Anexos</u>	159
<u>11.1</u>	<u>Resultados Detalhados da Definição da Métrica de Similaridade Adotada</u>	159
<u>11.2</u>	<u>Questionário de Recuperação de Casos</u>	162
<u>11.3</u>	<u>ER para Um SAD em Auditoria Interna – Registro das Observações</u>	163
<u>11.4</u>	<u>ER para um SAD em Auditoria Interna - Necessidades do Sistema</u>	165
<u>11.5</u>	<u>Lista de Eventos do Sistema AUDIT-AJUD</u>	170
<u>11.6</u>	<u>Diagrama de Contexto do Sistema AUDIT-AJUD</u>	171
<u>11.7</u>	<u>Diagramas de Fluxo de Dados (DFD) do Sistema AUDIT-AJUD</u>	172
<u>11.8</u>	<u>Descrição dos Processos do Sistema AUDIT-AJUD</u>	174
<u>11.9</u>	<u>Modelo Entidade-Relacionamento (MER) do Sistema AUDIT-AJUD</u>	177
<u>11.10</u>	<u>Dicionário de Dados do Sistema AUDIT-AJUD</u>	179
<u>11.11</u>	<u>Resultados dos Testes Comparativos</u>	176

Lista de Figuras

Figura 1: A tomada de decisão e a ajuda ao decisor (FREITAS, 1993)	7
Figura 2: Fases do processo decisório (STAIR, 1998)	11
Figura 3: Nível de tomada de decisões (STAIR, 1998)	11
Figura 4: Componentes de um sistema especialista (STAIR, 1998)	23
Figura 5: Processo de desenvolvimento de um SE (LAUDON & LAUDON 1999a)	24
Figura 6: O ciclo do RBC (AAMODT & PLAZA, 1994)	31
Figura 7: Modelo de funcionamento do RBC (LAUDON & LAUDON, 1999)	32
Figura 8: Impacto do agente RBC no processo de decisão (ANGEHRN & DUTTA, 1998)	34
Figura 9: O ciclo expandido do RBC (6 Rs) (ROTH-BERGHOFFER & IGLEZAKIS, 2001)	35
Figura 10: Princípios da adaptação (JEDLITSCHKA & ALTHOFF, 2001)	42
Figura 11: Estágios de uma auditoria operacional (JUND, 2002)	51
Figura 12: Estágios de uma auditoria operacional do Banco do Brasil S.A.	52
Figura 13: Tela principal do aplicativo DMMS-RBC, versão 1.0.1	78
Figura 14: Similaridade das métricas obtida pela metodologia DMMS-RBC	82
Figura 15: Similaridade dos percentuais obtida pela metodologia e a indicada pelo especialista no cálculo n°. 10	84
Figura 16: Modelo MAC/FAC (KLUWER & FORBUS, 1991)	90
Figura 17: Recuperação de casos orientada a índices com uso de índices já existentes	93
Figura 18: Recuperação de casos orientada a índices com a geração de índices	93
Figura 19: Exemplos de casos com feições compartilhadas	94
Figura 20: Casos organizados com base na rede de feições compartilhadas ...	95
Figura 21: Recuperação baseada em árvores de discriminação	97
Figura 22: O ciclo de vida clássico (PRESSMAN, 1995)	123
Figura 23: Tela de preenchimento de dados para busca de casos no AUDIT-AJUD	133
Figura 24: Formulário para realização de pesquisas no AUDIT-AJUD	135
Figura 25: Exemplo de tela de definição dos atributos do sistema AUDIT-AJUD	135
Figura 26: Diagrama de Contexto do AUDIT-AJUD gerado pelo CASE Studio®	171
Figura 27: Diagramas de Fluxo de Dados do Cadastro gerados pelo CASE Studio®	172
Figura 28: Diagrama de Fluxo de Dados da Rotina de Recuperação de Casos gerado pelo CASE Studio®	173
Figura 29: Modelo Entidade Relacionamento do AUDIT-AJUD gerado pelo CASE Studio®	178

Lista de Quadros

<u>Quadro 1: Características de Dado, Informação e Conhecimento</u>	6
<u>Quadro 2: Comparação entre SIE, SIG e SAD</u>	13
<u>Quadro 3: Exemplo da medida de similaridade local para modelo de impressora</u>	40
<u>Quadro 4: Exemplo de cálculo de similaridade</u>	40
<u>Quadro 5: Fatores relacionadas a escolha da técnica de pesquisa</u>	62
<u>Quadro 6: Resumo dos parâmetros adotados em cada cálculo</u>	79
<u>Quadro 7: Resumo dos resultados obtidos em cada cálculo realizado</u>	81
<u>Quadro 8: Comparativo das características predominantes de cada técnica</u> ...	102
<u>Quadro 9: Critérios de pontuação para as características de Precisão e Desempenho</u>	105
<u>Quadro 10: Critérios de pontuação para Facilidade de uso, Quantidade de informações e Flexibilidade</u>	106
<u>Quadro 11: Simulação de resultados</u>	106
<u>Quadro 12: Resultados de Precisão - situação A</u>	108
<u>Quadro 13: Resultados de Precisão - situação B</u>	109
<u>Quadro 14: Resultado de Facilidade de Uso</u>	110
<u>Quadro 15: Resultado de Quantidade de Informações</u>	110
<u>Quadro 16: Resultados de Desempenho</u>	111
<u>Quadro 17: Consolidação dos resultados</u>	112
<u>Quadro 18: Resumo de aspectos fundamentais para pré-seleção da ferramenta</u>	114
<u>Quadro 19: Matriz AHP para seleção da ferramenta de apoio</u>	120
<u>Quadro 20: Cálculo de Similaridade para um Atributo</u>	136
<u>Quadro 21: Demonstrativo de Cálculo da Similaridade Global</u>	136
<u>Quadro 22: Resultados dos Testes Comparativos</u>	142

Lista de Abreviaturas

ACM - Association for Computing Machinery
AHP - Analytic Hierarchy Process
AUDIBRA - Instituto dos Auditores Internos do Brasil
BC - Banco de Casos
BD - Banco de Dados
CBR - Case Based Reasoning
DFD - Diagrama de Fluxo de Dados
DMMS-RBC - Definição da Melhor Métrica de Similaridade para RBC
DMMR-RBC - Definição da Melhor Métrica de Recuperação de casos para RBC
DSS - Decision Support System
ER - Engenharia de Requisitos
EUA - Estados Unidos da América
EWCBR – European Workshop of CBR
FAC - Few are Chosen
GAS - Gerador Automático de Sistemas
IA - Inteligência Artificial
ICCBR – International Conference of CBR
IEEE - Institute of Electrical and Electronics Engineers
MAC - Many Are Called
MER - Modelo Entidade Relacionamento
OLAP - On-line Analytical Processing
RBC - Raciocínio Baseado em Casos
SAD - Sistema de Apoio à Decisão
SBC - Sistema Baseado em Conhecimento
SE - Sistema Especialista
SIE - Sistema de Informações para Executivos
SIG - Sistema de Informações Gerenciais
SGBD - Sistema Gerenciador de Banco de Dados

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO DA PESQUISA

A tomada de decisões é uma atitude comum para qualquer ser vivo. Com maior ou menor relevância, toma-se decisões praticamente do momento em que se acorda até o fim do dia. O que vem mudando nos dias atuais é o nível de complexidade e a velocidade das decisões a serem tomadas. Os volumes de dados e informações vêm crescendo de forma exponencial e isso, em vez de facilitar a tomada de decisões, muitas vezes torna-a mais complexa e incerta.

A evolução dos recursos de informática, especificamente com relação aos sistemas de apoio à decisão, tem possibilitado a intensificação de seu uso nas mais diversas áreas de trabalho, fazendo com que esses sistemas sejam desenvolvidos para praticamente todo o tipo de atividade, das mais rotineiras até às mais complexas. A busca por melhores resultados e pela rapidez e segurança na tomada de decisões tem feito com que, todas as organizações que utilizam dados e informações para auxiliar esse processo, procurem utilizar os recursos tecnológicos mais modernos.

A área de auditoria, apesar da intensiva utilização dos recursos tecnológicos para apoio a seus processos, ainda não pode ser considerada como usuária tradicional de sistemas de apoio à decisão por questões culturais e técnicas. Isso porque existem aspectos dificultadores para utilização desses sistemas nos trabalhos de auditoria como a resistência dos especialistas em explicitar seus conhecimentos tácitos e a enorme complexidade dessas rotinas, o que praticamente inviabiliza a formação de uma base ampla de regras.

Dessa maneira, a proposta apresentada nesse trabalho busca uma nova solução, que minimize essas duas dificuldades e, gere resultados melhores do que os atuais.

Esse trabalho apresenta uma estrutura que se inicia com uma introdução, onde são feitas a contextualização da pesquisa e a apresentação da estrutura do

trabalho. Nesse mesmo capítulo definem-se também os objetivos e apresenta-se a justificativa da pesquisa. Em seguida há a revisão bibliográfica realizada sobre os assuntos abordados na construção da proposta. A revisão está estruturada de forma a apresentar uma breve descrição dos Sistemas de Apoio a Decisão (SAD), seguida dos conceitos macros de Inteligência Artificial (IA), passando pelo tópico de Sistemas Especialistas (SE) e chegando finalmente na metodologia de Raciocínio Baseado em Casos (RBC), que é o ponto principal desse estudo. É apresentada também, uma visão geral e sucinta sobre a área de auditoria, para que o leitor não familiarizado com essa atividade possa compreender os termos e procedimentos adotados durante as discussões, agrupadas no capítulo 2. No capítulo 3, são apresentadas as metodologias de pesquisa utilizadas, descrevendo os procedimentos, classificações e métodos adotados. Nos capítulos 4 a 7 são detalhados todos os estudos realizados e os resultados obtidos, inclusive com os aspectos técnicos do protótipo desenvolvido. Encerrando, no capítulo 8, estão as conclusões finais e perspectivas de pesquisas futuras, seguidas das Referências Bibliográfica, um pequeno glossário e anexos nos capítulos finais.

1.2 OBJETIVOS

O objetivo dessa dissertação é apresentar uma proposta de um sistema de apoio à decisão para aplicação em auditoria interna, utilizando o raciocínio baseado em casos.

Como o trabalho é desenvolvido com foco nos aspectos tecnológicos, pretende-se enfatizar os detalhes técnicos dessa proposta, estruturando esses conceitos com base nos fundamentos da IA, mais especificamente no ramo de SE para apoio a tomada de decisões. A opção pela técnica de Raciocínio Baseado em Casos (RBC) também é explorada ao máximo uma vez que está nessa opção a maior contribuição à área de pesquisa pelo caráter de inovação da proposta. Os aspectos operacionais do sistema e do trabalho de auditoria são apresentados para compreensão dos leitores, sem nenhuma intenção de discussões ou conclusões sobre seus aspectos técnicos e operacionais. Para isso, algumas etapas do projeto devem ser especificadas:

- a) Realizar uma revisão bibliográfica abrangente e atualizada sobre os principais temas relacionados a esse trabalho:
- Sistemas de Apoio à Decisão (SAD);
 - Sistemas Especialistas (SE);
 - Raciocínio Baseado em Casos (RBC);
 - Auditoria Interna.
- b) Analisar e definir o modelo de cálculo de similaridade mais adequado para ser adotado no sistema proposto;
- c) Analisar e definir a técnica de recuperação de casos mais eficiente e adequada aos procedimentos que serão adotados no sistema proposto;
- d) Selecionar uma ferramenta de desenvolvimento para o sistema proposto;
- e) Desenvolver uma proposta de um SAD utilizando a técnica de RBC adequado à etapa de pré-auditoria interna de uma instituição financeira;
- f) Desenvolver um protótipo que permita avaliar os resultados obtidos na tomada de decisões de auditoria interna baseadas em casos armazenados pela empresa.
- g) Avaliar os resultados obtidos com a utilização do sistema proposto em comparação com a situação anterior do trabalho de auditoria.

1.3 JUSTIFICATIVA DA PESQUISA

O crescimento dos sistemas especialistas de apoio à decisão verificado nos últimos anos é devido aos bons resultados apresentados pelos sistemas atuais. Existem centenas de exemplos de grandes empresas e centros de pesquisa que vêm obtendo ganhos de tempo, segurança e recursos com a utilização dessas ferramentas, de forma consistente e constante. Isso, sem dúvida, é um avanço, mas ainda hoje pode-se afirmar que o desenvolvimento de sistemas especialistas de apoio à decisão ainda tem muito a percorrer, estando mais próximo do início do que do final de seu desenvolvimento.

Essa realidade não é diferente para a área de estudo desse trabalho. Existe atualmente uma carência de SE de apoio à decisão para o trabalho de auditoria interna. Normalmente encontram-se apenas alguns sistemas específicos que apresentam o foco muito mais na automatização de uma tarefa operacional do que no auxílio a decisões. Essa situação ocorre por inúmeros motivos, mas com certeza, a complexidade de estruturação do serviço de um auditor interno, aliada ao alto investimento necessário para o desenvolvimento de um SE de apoio à decisão devido à necessidade de formulação de regras (geralmente instáveis) têm sido os principais entraves para o desenvolvimento e a disseminação de seu uso.

Como esses problemas não afetam apenas a área de auditoria interna, a computação como um todo (e nesse caso, especificamente a área de SE), tem evoluído constantemente para solucionar ou minimizar as dificuldades existentes. Considerando o pouco tempo de pesquisas na área de IA e SE, pode-se considerar que a última década proporcionou um grande avanço tecnológico a essa área. Novas metodologias e técnicas são apresentadas freqüentemente, buscando oferecer facilidades aos desenvolvedores e aos usuários de SE.

Ao mesmo tempo, a área de auditoria tem se mostrado cada vez mais imprescindível na administração de grandes empresas no mercado atual. O trabalho de um auditor atualmente é infinitamente mais complexo do que a tradicional imagem de “checador de números” que os auditores mais antigos carregaram. Diariamente, auditores se deparam com inúmeras situações que exigem uma tomada de decisão que pode representar lucros ou prejuízos monstruosos para as empresas que estão sendo auditadas. Os valores envolvidos em trabalhos de auditoria interna no mundo atual já vêm sendo medido em bilhões de dólares e um campo tão relevante como esse não pode se dar ao luxo de tomar decisões estratégicas e operacionais sem o apoio de recursos tecnológicos existentes. Atualmente, devido à complexidade e velocidade do mercado, um erro isolado é capaz de levar uma empresa à falência. Casos clássicos como o Banco Barrings e os mais recentes como as empresas Enron, WorldCom e Xerox, demonstram que todos os valores

aplicados no auxílio à tomada de decisões nos trabalhos de auditoria devem ser vistos como investimentos e não mais como despesas obrigatórias por lei.

A escolha do RBC como técnica principal de trabalho do SAD proposto, busca reduzir ao máximo a maior dificuldade para o sucesso de qualquer SE para auditoria, que é a definição de regras de conhecimento. Essa contribuição por si só, pode ser o diferencial entre os resultados obtidos até o momento e o sucesso que se espera.

Traçado tal cenário, acredita-se estar demonstrado a relevância que a proposta desse trabalho representa tanto para a área de computação como para a área de auditoria. Ao selecionar as melhores técnicas de similaridade e de recuperação, e dentre as ferramentas de auxílio ao desenvolvimento de sistemas disponíveis, a que apresenta as melhores características para ser utilizada no desenvolvimento de um modelo de SAD para auditoria interna dentro de uma instituição financeira, será possível para essas organizações desenvolver sistemas apropriados a sua realidade, que sejam ao mesmo tempo avançados e viáveis, com uma implantação gradativa e segura.

Sabe-se que o prazo e o escopo de um trabalho como esse não é suficiente para que todo o ciclo de vida de um SE, utilizando RBC seja planejado, executado e avaliado. Entretanto, com a apresentação do projeto por inteiro e o desenvolvimento de um protótipo para utilização, pretende-se evidenciar aspectos positivos de todas as etapas do ciclo.

2 REVISÃO BIBLIOGRÁFICA

2.1 INFORMAÇÃO E DECISÃO

O ato de decidir é inerente ao ser vivo. Mesmo no momento em que uma pessoa optou por não tomar uma decisão, percebe-se que uma decisão acaba de ser tomada. Dessa forma, deve-se buscar sempre o melhor embasamento para tomada de decisões, pois em alguns casos, uma escolha equivocada pode representar prejuízos financeiros ou até riscos de vida.

Diferentemente dos tempos pioneiros da informática, a grande dificuldade na tomada de decisão atualmente não está mais na localização e organização de dados ou de informações, mas sim na transformação dessas informações em conhecimentos úteis para auxiliar a tomada de decisões.

FREITAS (1997) comenta que Davis & Olson definem informação como “um dado que foi processado de uma forma significativa para o receptor e seu valor é real ou percebido no momento, ou em ações prospectivas nas decisões”. Já DAVENPORT & PRUSAK (1999) apresentam uma tabela com as diferenças principais entre dados, informação e conhecimento, demonstrando que o conhecimento é o nível mais elevado na escala e, portanto, deve ser buscado sempre para auxiliar no processo de tomada de decisões.

Quadro 1: Características de Dado, Informação e Conhecimento

<i>Dados</i>	<i>Informação</i>	<i>Conhecimento</i>
Simple observações sobre o estado do mundo	Dados dotados de relevância e propósito	Informação valiosa da mente humana
Facilmente estruturado	Requer análise	Inclui reflexão e síntese.
Facilmente obtido por máquinas	Exige consenso em relação ao significado	De difícil estruturação e transferência
Freqüentemente quantificado	Exige necessariamente a mediação humana	De difícil captura em máquinas
Facilmente transferível		Freqüentemente tácito

Fonte: DAVENPORT & PRUSAK, (1999)

Para TESKEY (1987) dados são simples observações de fatos, informações são conjuntos de dados com uma estruturação e conhecimento é um método de utilização ou modificação da informação. Já a Teoria Interrogativa para Informação e Conhecimento proposta por QUIGLEY & DEBONS (1999), sugere que Dados, Informação e Conhecimento podem ser classificados de acordo com o tipo de pergunta que cada um responde. Segundo a teoria, Dados seriam meros símbolos e números. A Informação seria todo texto que responda as seguintes perguntas: quando, onde, quem e o quê e o Conhecimento seria o texto que responde as perguntas mais complexas: como e porquê.

Percebe-se que apesar de todo avanço tecnológico, a escolha do conhecimento relevante para uma escolha adequada sempre será papel do ser humano. O processo de tomada de decisão, portanto, deve ser o mais estruturado possível. Até os grandes líderes que tomam decisões rápidas e seguras, buscam cada vez mais, basear-se em informações para tomada de decisão.

Dentro de uma organização, grande parte dos esforços estão relacionados ao apoio ao processo de tomada de decisão, que pode ser esquematizado da seguinte maneira (FREITAS, 1993):

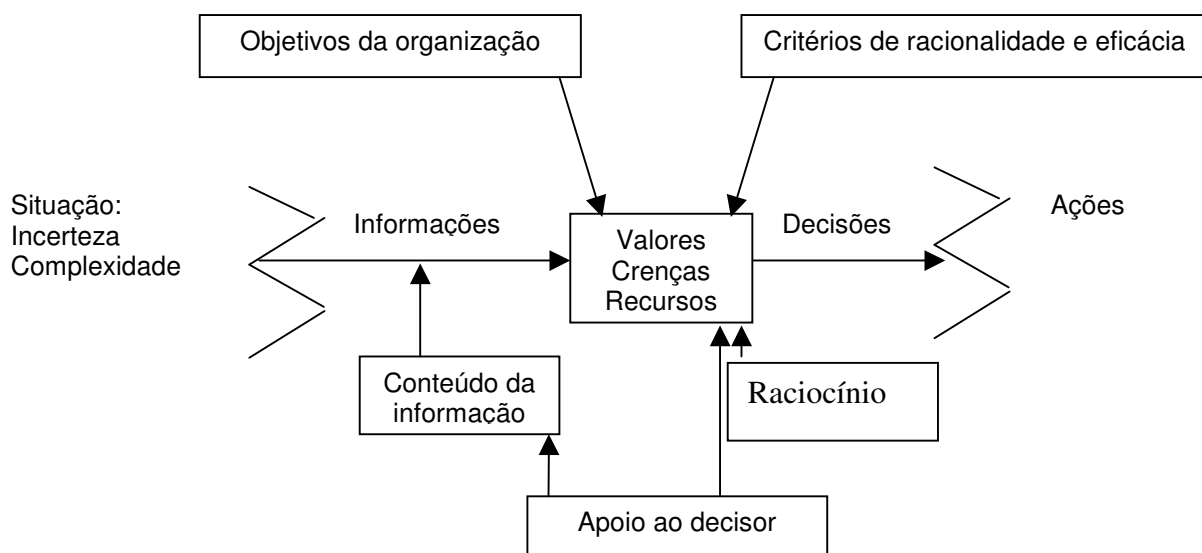


Figura 1: A tomada de decisão e a ajuda ao decisor (FREITAS, 1993)

Percebe-se através do fluxo apresentado, que o processo de tomada de decisão é complexo, não somente pela dificuldade intrínseca da própria

decisão. Todos os outros fatores que influenciam a tomada de decisão como racionalidade, objetivos da organização, valores, crenças e recursos, fazem com que muitas vezes uma decisão, aparentemente simples, se transforme em uma complexa escolha com milhares de resultados possíveis.

2.2 SISTEMAS DE APOIO À DECISÃO – SAD

Antes da definição de sistemas de apoio à decisão, deve-se conhecer o significado dos termos sistemas e sistemas de informação.

2.2.1 SISTEMAS E SISTEMAS DE INFORMAÇÃO

Conforme definido por CASSARRO (2001), um conceito universal para qualquer sistema é o seguinte: “Sistema é um conjunto de partes e componentes, logicamente estruturados, com a finalidade de atender a um dado objetivo”.

Um sistema de informação é um conjunto de componentes inter-relacionados, desenvolvidos para coletar, processar, armazenar e distribuir informação para facilitar a coordenação, o controle, à análise, a visualização e o processo decisório (LAUDON & LAUDON, 1999a). Ele consiste em três entidades que se ajustam: pessoas, organizações e tecnologias.

Um sistema de apoio ou suporte à decisão pode ser definido de diversas maneiras. Resumidamente pode-se definir um SAD como uma ferramenta que procura auxiliar o decisor/gerente na sua tomada de decisão (FREITAS, 1997), mas outras definições ampliam esse conceito. Uma delas define SAD como um conjunto organizado de pessoas, procedimentos, software, banco de dados e dispositivos utilizados para dar suporte à tomada de decisões específicas de um problema (STAIR, 1998). Já para BINDER (1994) um SAD é um sistema computacional interativo que ajuda o tomador de decisões, através da utilização de modelos e dados, a solucionar determinados problemas, aperfeiçoando a qualidade das decisões.

Existem duas classes de SAD: baseados em dados e baseados em modelos. Os primeiros trabalham estatisticamente com grandes bases de dados. Já os

baseados em modelos são mais indicados para situações que apresentem grande número de variáveis independentes (ALTER, 1996). Normalmente, são utilizados pelos níveis mais altos de gerência, mas cada vez mais, vêm sendo utilizados por todos os níveis dentro de uma organização. Eles devem levar em consideração que pessoas tomam decisões e não as máquinas. Portanto, a principal característica do SAD deve ser a estruturação voltada ao auxílio à tomada de decisão em todos os sentidos (STAIR, 1998).

2.2.2 BREVE HISTÓRICO DOS SISTEMAS DE APOIO À DECISÃO

Segundo POWER (2002), um dos primeiros registros históricos sobre SAD foi de 1971, quando Michael S. Scott Morton analisou como os computadores poderiam ajudar ou apoiar os gestores em tomadas de decisão. Nesse mesmo ano, T.P. Gerrity Jr. publicou um artigo com foco em SAD. Nos anos seguintes, alguns pesquisadores publicaram artigos esporádicos sobre o assunto, enquanto algumas universidades criavam protótipos em conjunto com algumas empresas de sistemas sem, entretanto, nenhuma grande mudança no cenário de aplicabilidade comercial dos SAD.

Somente quando Ralph Sprague e Eric Carlson publicaram o livro *Building Effective Decision Support Systems* em 1982, uma coletânea prática e compreensível de como as organizações podem e devem construir um SAD foi disponibilizada. Mas ainda assim, o uso dos SAD não decolava, principalmente pela priorização dos sistemas transacionais e pela pequena disponibilização da informática junto aos usuários finais naquela época.

No início de 1990, a descentralização proporcionada pela evolução dos micro-computadores viabiliza o armazenamento de grandes bases de dados (*data warehousing*) e o processamento analítico on-line (OLAP - On-line Analytical Processing), incentivando a utilização dos SAD. A evolução de modelos matemáticos e dos recursos de software e hardware também tem popularizado sua utilização (POWER, 2002).

2.2.3 CARACTERÍSTICAS DE UM SISTEMA DE APOIO À DECISÃO

Não é objetivo de um SAD substituir decisões do ser humano. O objetivo do SAD é ajudar o decisor na manipulação de grandes volumes de dados. Quase sempre os SAD trabalham na solução de problemas mais complexos e menos estruturados; tentam combinar modelos ou técnicas analíticas com as funções tradicionais do processamento de dados; devem ser interativos, fáceis de usar e possuir interfaces amigáveis; incorporam outras atividades, além do auxílio à tomada de decisão, como meios mais eficientes de distribuição dos resultados (SCHREINER JÚNIOR, 2001).

Normalmente os SAD baseiam-se em grandes bases de dados das empresas – *data warehouse* ou *data marts* – e aproveitam-se do enorme volume de dados para aumentar sua eficiência. Obviamente os SAD não são padronizados e buscam aproveitar a evolução da tecnologia para criar novos e inovadores recursos para facilitar o trabalho dos responsáveis pelo processo de tomada de decisões. Apesar dessa diversidade, existem algumas características que devem ser comuns a todos os SAD, independente do seu tamanho, domínio ou linguagem. Conforme verificado em STAIR (1998), as principais características de um SAD são:

- Capacidade de manipulação de grandes bases de dados;
- Obtenção e processamento de dados de fontes diferentes;
- Proporcionar flexibilidade de relatórios e de apresentação;
- Possuir orientação textual e gráfica;
- Executar análises e comparações complexas e sofisticadas utilizando pacotes de softwares avançados;
- Oferecer suporte às abordagens de otimização e satisfação;
- Executar análises de simulações e baseada em metas.

Usualmente um SAD é composto por três componentes básicos (SAGE, 1991):

- a) **Sistema Gerenciador de Banco de Dados** – É capaz de mostrar aos usuários os tipos de dados que estão disponíveis e como acessá-los.

- b) **Sistema Gerenciador de Modelos** – Sua função é transformar os dados em informações que serão úteis para a tomada de decisão.
- c) **Sistema Gerenciador de Diálogo** – Oferece aos usuários interfaces amigáveis, permitindo a obtenção de todos os recursos do SAD.

2.2.4 RECURSOS DE UM SISTEMA DE APOIO À DECISÃO

Normalmente um SAD oferece:

- Suporte nas fases da solução de problemas (figura 2);

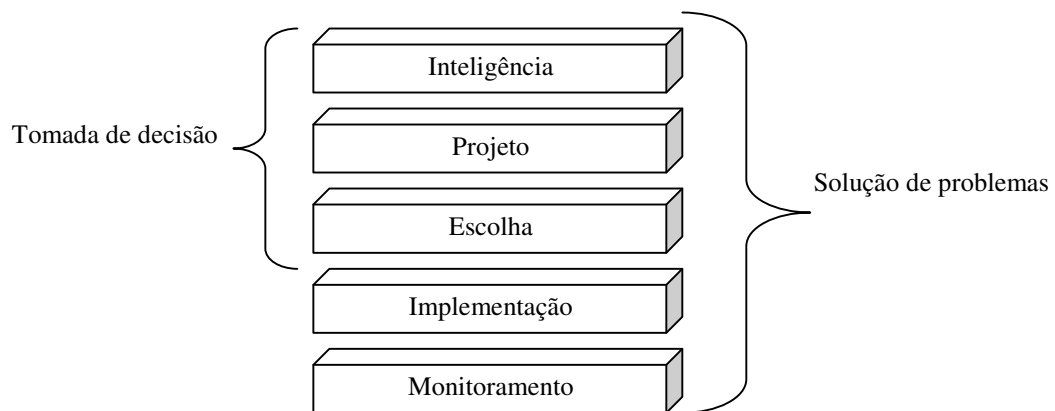


Figura 2: Fases do processo decisório (STAIR, 1998)

- Suporte para decisões que devem ser tomadas com pouca frequência e também para aquelas repetitivas ou muito frequentes;
- Suporte para tomada de decisões de problemas não estruturados e também para aqueles altamente estruturados;
- Suporte para diversos níveis de tomada de decisão (figura 3):

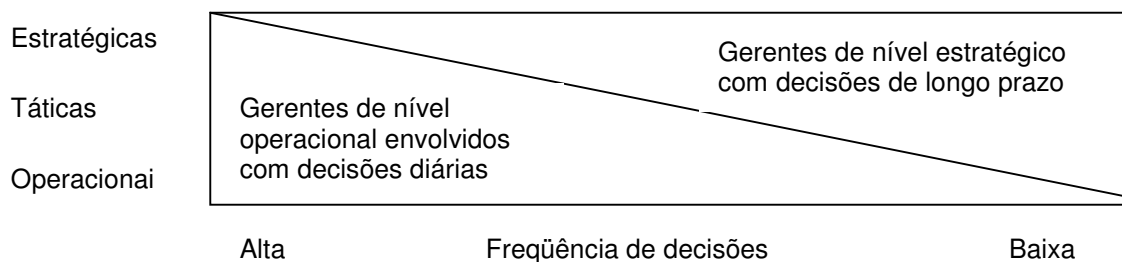


Figura 3: Nível de tomada de decisões (STAIR, 1998)

2.2.5 ETAPAS DE ELABORAÇÃO DE UM SISTEMA DE APOIO À DECISÃO

A elaboração de um SAD não é igual ao processo de um sistema tradicional, mas inclui as mesmas quatro etapas: Projeto, Desenvolvimento, Implementação e Avaliação. Destacam-se a seguir, algumas particularidades dessas etapas na elaboração de um SAD.

2.2.5.1 Projeto

É o processo de identificação dos pontos-chaves para a tomada de decisão através da especificação e análise de requisitos. Os projetos de um SAD são classificados em quatro categorias:

- O modelo racional que foca a seleção das alternativas mais eficientes;
- O processo organizacional que enfatiza o compartilhamento de várias unidades da organização;
- O modelo que reflete a racionalidade limitada para achar uma solução;
- Outros modelos.

2.2.5.2 Desenvolvimento

A própria natureza de um SAD requer uma técnica diferente de desenvolvimento quando comparada com um sistema transacional tradicional, principalmente pela ausência de uma teoria abrangente sobre o processo decisório (SPRAGUE JÚNIOR & WATSON, 1991). O desenvolvimento de um SAD possui algumas características:

- Deve ter feedback rápido dos usuários;
- Deve permitir mudanças rápidas e fáceis;
- Deve promover uma grande interação entre os usuários finais, os projetistas, a gerência (ou os patrocinadores do programa) e a tecnologia envolvida.

Devido a essas características, o desenvolvimento de um SAD é normalmente realizado com base na metodologia espiral.

2.2.5.3 Implementação

Diferentemente dos sistemas transacionais, a utilização de um SAD não pode ser imposta. Ela deve ser vista como importante pelo usuário e devido a isso, a etapa da implementação é primordial. Características como individualidade dos usuários, cultura organizacional, estilo de gerência, questões demográficas e fatores externos, deve ser minuciosamente analisadas e planejadas sob o risco de todo o investimento ser desperdiçado.

2.2.5.4 Avaliação

É concentrada na análise de custos e benefícios através da comparação de cenários em dois momentos. O problema é que em decisões, os indicadores não são padronizados, não se repetem e nunca são influenciados apenas por um fator. É comum uma decisão correta, baseada no SAD, apresentar muitos “padrinhos”, enquanto todas as falhas são atribuídas somente ao novo sistema.

2.2.6 UMA COMPARAÇÃO ENTRE SAD, SIE E SIG

Apesar de já serem utilizados há anos, os SAD ainda são confundidos com outros sistemas não transacionais. Em revistas, jornais e sites na internet, os termos SAD, Sistemas de Informação para Executivos (SIE) e Sistemas de Informações Gerenciais (SIG) são usados indiscriminadamente nas mais diversas situações. Isso pode ser evitado, observando-se características específicas que diferenciam uma categoria de sistema das outras:

Quadro 2: Comparação entre SIE, SIG e SAD

Dimensão	SIE	SIG	SAD
Foco	Acesso ao status	Processamento de informações	Análise, apoio à decisão.
Usuários típicos	Executivos seniores	Executivos de nível médio e baixo, as vezes sênior	Analistas, profissionais, gerentes.
Objetivo	Conveniência	Eficiência	Eficácia

Dimensão	SIE	SIG	SAD
Aplicação	Análise de ambientes, avaliação de desempenho, identificação de problemas e oportunidades.	Controle de produção, projeção de vendas, análise financeira e gerenciamento de recursos humanos.	Diversas áreas onde são tomadas decisões gerenciais
Banco de dados	Especial	Da empresa	Especial
Apoio à decisão	Indireto, principalmente para problemas de alto nível e não estruturados.	Direto ou indireto, principalmente problemas estruturados, utilizando modelos.	Apoio a atividades decisórias semi ou não estruturadas, <i>ad hoc</i> .
Tipo de Informação	Novos itens, informações externas e operações internas	Relatórios sobre operações internas, fluxo estruturado.	Informações de apoio para situações específicas
Uso principal	Acompanhamento e controle	Controle	Planejam., organização e controle.
Tratamento das informações	Filtra e resume informações, acompanha dados e informações cruciais.	As informações são fornecidas a um grupo diversificado de usuários que as manipula	Informações fornecidas pelo sistema transacional são usadas como entrada para o SAD

Fonte: Adaptado de SPRAGUE JR. & WATSON (1991)

2.3 INTELIGÊNCIA ARTIFICIAL - IA

A Inteligência Artificial baseia-se na idéia de que é viável modelar (e codificar esse modelo em computadores) o funcionamento da mente humana. Trata-se um conjunto de atividades relacionadas, que busca capturar algum aspecto da inteligência dos seres humanos e de seu modo de ser (LAUDON & LAUDON 1999). Entretanto, antes de analisar a IA, é importante definir inteligência para que não se criem falsas expectativas.

2.3.1 INTELIGÊNCIA

Etimologicamente, a palavra inteligência vem do latim *inter* (entre) e *legere* (escolher). Portanto, inteligência pode ser definida como a capacidade de escolha entre uma coisa e outra. Entretanto, mesmo sendo estudada a milhares de anos, ainda não existe uma definição consensual para inteligência.

Há uma grande diferença entre possuir conhecimento e ser inteligente. Conforme se pode ver em BARONE *et al.* (2003), nem todos que possuem grande conhecimento podem ser considerados inteligentes. O que torna a pessoa inteligente é a habilidade de adquirir novos conhecimentos, o chamado aprender a aprender constantemente. Esse raciocínio é muito importante na análise de Inteligência Artificial e Sistemas Especialistas que será seguida.

Para MCCARTHY & HAYES (1969), “Uma máquina é inteligente se ela é capaz de solucionar uma classe de problemas que requerem inteligência para serem solucionados por seres humanos”. Já na visão de GONZÁLES (1990), o único diferencial da IA com relação à computação convencional é o fato dela trabalhar com regras e instruções que reproduzem conhecimentos humanos.

RUSSELL & NORVIG (1995) organizaram as diversas definições existentes em quatro categorias:

- Sistemas que pensam como os seres humanos;
- Sistemas que agem como seres humanos;
- Sistemas que pensam racionalmente;
- Sistemas que agem racionalmente.

Segundo STAIR (1998), o objetivo da IA não é substituir totalmente o ser humano na tomada de decisões, mas replicá-la em problemas bem definidos, auxiliando as organizações no alcance de suas metas.

No trabalho de DOYLE *et al.* (1996) os esforços da IA estão divididos em quatro visões:

- Construção de máquinas inteligentes;

- Formalização de conhecimento e mecanismos de raciocínio;
- Desenvolvimento de modelos computacionais para entender a psicologia e o ambiente;
- Tornar o trabalho com computadores tão simples e automático como o trabalho com especialistas.

- Existem duas linhas de pesquisa para construção de sistemas baseados em IA: a conexionista e a simbólica. A conexionista visa à modelagem através da simulação do cérebro humano. A linha simbólica segue o modelo lógico tradicional, e será a linha seguida nesta pesquisa.

2.3.2 BREVE HISTÓRICO DA INTELIGÊNCIA ARTIFICIAL

Alguns fatos podem ser considerados os “verdadeiros atos pioneiros da IA”, como: as definições de regras de produção (McCulloch and Pitts *Neuron Model*) em 1943, a questão proposta por TURING em 1950: “*As máquinas podem pensar?*” Isto deve começar com as definições dos termos “máquinas” e “pensar” (TURING 1950), a criação do Algoritmo de Markov para controlar a execução de regras em 1954 e os trabalhos de Minsky e Edmonds nos primeiros sistemas baseados em IA da década de 50. Isso demonstra que essas idéias já “borbulhavam” na área científica desde aquela época, apesar de não se tratar de uma linha de pesquisa (RUSSELL & NORVIG, 1995)

Mas foi em 1956, durante uma conferência no *Dartmouth College*, que John McCarthy, descreveu a capacidade dos computadores em imitar as funções do cérebro humano, como Inteligência Artificial. Naquela época os cientistas acreditavam que já na década de 60 existiriam computadores com “inteligência” igual a dos seres humanos. Logo em seguida, foi criada a linguagem de IA LISP, seguida por várias outras inovações. Entretanto, a previsão dos cientistas com relação aos avanços da IA não se concretizou nem parcialmente.

2.3.3 LINHAS DE PESQUISA DA ÁREA DE INTELIGÊNCIA ARTIFICIAL

Conforme visto anteriormente a IA tem várias disciplinas. Dentre as diversas linhas existentes, serão focadas as linhas de pesquisas mais estruturadas da Inteligência Artificial, que são segundo LAUDON & LAUDON (1999a):

- **Linguagem Natural** – reconhecer e gerar a linguagem humana (escrita/falada)
- **Sistemas Perceptivos** – simular os sentidos humanos
- **Robótica** – executar tarefas normalmente feitos por seres humanos
- **Redes Neurais** – simular a estrutura física dos cérebros
- **Softwares Inteligentes** – desenvolver softwares com técnicas de IA
- **Sistemas Especialistas** – simular conhecimentos e a tomada de decisões humanas

A proposta de um SAD para auditoria interna é baseada na tecnologia de sistemas especialistas, pois ela apresenta características mais adequadas. Assim sendo, a pesquisa estará restrita a essa área de estudo da IA.

2.4 SISTEMAS ESPECIALISTAS - SE

Sistemas Especialistas são aplicações de software relativamente recentes. Assim como muitos conceitos da área de informática, os SE são várias vezes definidos de maneira errônea ou conflitante com outras áreas de pesquisa.

Popularmente SE são conhecidos por sistemas modelados com base no comportamento de especialistas humanos apresentando dois componentes:

- **Base de conhecimento** - conjunto de regras que representam a maneira que um perito agiria;
- **Mecanismo de interface** - o mecanismo que controla a execução do processo de raciocínio sustentado pela base de conhecimento.

Há diversas definições de SE que com outras palavras, reafirmam a anterior. Segundo JACKSON (1999), SE são programas de computadores que representam e reagem com base no conhecimento de algum especialista, com o objetivo de resolver algum problema ou sugerir ações. Já Laudon afirma: SE

são softwares que buscam capturar conhecimento em domínios limitados do saber e experiência humanos, aplicando esse conhecimento na solução de problemas (LAUDON & LAUDON, 1999a).

2.4.1 BREVE HISTÓRICO DE SISTEMAS ESPECIALISTAS

A primeira idéia de um SE foi apresentada por Post em 1936, ainda que identificado como Sistemas de Produção. Ela permaneceu esquecida até a década de 70 quando foi redescoberta como uma ferramenta de modelagem da psicologia humana.

Os primeiros sistemas classificados especificamente como SE foram os sistemas DENDRAL e MYCIN de 1971 e 1976, respectivamente. Desde então, muitos SE foram sendo desenvolvidos nos mais diversos domínios, apresentando resultados práticos e econômicos dos mais diversos. Atualmente, os SE são desenvolvidos em sua grande maioria, com o auxílio de *shells* de SE, que são ferramentas de auxílio ao desenvolvimento. Elas facilitam muito o desenvolvimento do módulo de memória de trabalho e regras, permitindo que a equipe de desenvolvimento se concentre na codificação do conhecimento dos especialistas, facilitando muito a disseminação dos SE (BITTENCOURT, 1998).

2.4.2 CARACTERÍSTICAS DE UM SISTEMA ESPECIALISTA

As características que diferenciam os SE dos outros sistemas de informações existentes podem ser resumidas em três (FALSARELLA & CHAVES, 2003):

- Trata-se de um sistema que armazena o conhecimento e as experiências de especialistas em bases de conhecimento;
- Utiliza mecanismos de inferência integrados às bases de conhecimento para resolver – ou auxiliar a resolver – problemas;
- Possibilita a inclusão de novos conhecimentos nas bases de conhecimento, sem eliminar os conhecimentos já armazenados.

LAUDON & LAUDON (1999) destacam outra característica importante que distingue os SE de sistemas tradicionais: a quantidade e a magnitude das

regras. Segundo eles, SE podem, facilmente, ter de 200 a 10.000 regras, muito mais do que os sistemas tradicionais, que podem ter de 50 a 100 instruções SE-ENTÃO. Além disso, as regras em um SE, tendem a ser interconectadas e reunidas em um grau muito maior.

Outra confusão muito grande está relacionada com as diversas linhas de pesquisa da IA. Como para muitos a IA é uma grande incógnita, com a falsa impressão de futurismo, todos os tipos de sistemas baseados em técnicas de IA podem ser erroneamente chamados de SE. Os SE podem ser distinguidos de outros aplicativos de IA principalmente pelas seguintes características:

- O SE lida com problemas subjetivos de grande complexidade que requerem uma considerável expertise humana. Muitos programas de IA são verdadeiros veículos de pesquisa, que focam problemas matemáticos abstratos ou versões simplificadas de problemas reais com o objetivo de atingir *insights* ou refinar as técnicas. Os SE, por outro lado, resolvem problemas científicos ou de interesse comercial;
- O SE pode exibir alto desempenho em termos de velocidade e confiança como ferramenta usual. Muitos programas baseados em IA não rodam muito rapidamente e podem conter falhas;
- O SE deve ser capaz de explicar e justificar as soluções e recomendações na ordem e forma adotada, da maneira que o usuário seja convencido de que elas estão corretas. Programas de pesquisa são tipicamente rodados apenas pelos seus criadores ou outros pesquisadores. Um SE é usado por um grande número de usuários, e deve ser de fácil utilização.

Muitas vezes o termo Sistema Baseado em Conhecimento (SBC) é usado como sinônimo de SE. Na verdade, apesar da grande semelhança de conceitos o primeiro apresenta um escopo mais amplo. Um SBC é aquele cujo desempenho é baseado na aplicação de regras para uma representação simbólica do conhecimento, em vez de empregá-la com base em métodos algorítmico e estatístico. Deste modo, um programa capaz de conversar sobre o tempo pode ser um SBC, mesmo que este programa não expresse uma

“expertise” em meteorologia, mas um SE da área deve ser capaz de fornecer previsões do tempo. A maior diferença entre os dois concentra-se na qualidade da informação armazenada e se ele é um conhecimento de especialistas.

2.4.3 TAREFAS TÍPICAS DE SISTEMAS ESPECIALISTAS

Pelas características de um SE pode-se afirmar que não há limitações na utilização desse tipo de instrumental desde que existam recursos (humanos e tecnológicos) disponíveis para desenvolvê-lo. O que ocorre na realidade é que, devido ao pouco tempo (em termos históricos) de desenvolvimento dessa linha de pesquisa, os casos mais comuns de utilização ainda são restritos. De qualquer maneira, já se encontra SE nas mais diversas áreas, desempenhando as mais diversas tarefas, sendo muito utilizados para (JACKSON, 1999):

- A interpretação de dados (como os sinais de um sonar);
- Diagnósticos de falhas/mau funcionamento (como falhas em equipamentos e doenças em seres humanos);
- Análise estrutural de objetos complexos (como a composição química de produtos);
- Configuração de objetos complexos (como sistemas computacionais);
- Planejamento de seqüência de ações (como o executado em robôs).

A utilização de SE especificamente no contexto de apoio a decisões não é nova, nem pode ser considerada uma escolha de risco. Já há algum tempo essa prática vêm sendo estudada com pode ser analisado nos trabalho de AL-ANI *et al.* (1987), HOPLIN (1987) e WYSK (1990). Dessa forma, a proposta que será apresentada nesse trabalho se encaixa nos SE de interpretação de dados com foco no apoio a tomada de decisões, uma vez que ele buscará agilizar a análise de muitos casos arquivados de auditoria, facilitando a pesquisa e permitindo o trabalho e a tomada de decisões com base em um volume de dados muito grandes, o que praticamente é inviável de ser feito manualmente.

2.4.4 LIMITAÇÕES/PROBLEMAS DOS SISTEMAS ESPECIALISTAS

Os problemas/limitações dos SE mudaram com o aperfeiçoamento dessa área de pesquisa. As resistências dos que viam os SE como um modismo, foram diminuindo com a utilização cada vez mais freqüente e consolidada das ferramentas. Conforme descrito por HARDAWAY & WILL (1990) as barreiras para difusão dos SE têm ficado cada vez mais semelhantes às barreiras enfrentadas por todos os tipos de sistemas de informação.

Apesar de ter se tornado um tipo de sistema aplicável em praticamente todas as áreas de pesquisas ou negócios, os SE apresentam algumas características que limitam sua utilização e podem até mesmo, inviabilizar a sua adoção em determinadas áreas.

Pela enorme amplitude que o conhecimento gera em termos de conjunto de regras em um pequeno processo, muitas vezes um SE apresenta erros grosseiros em tarefas aparentemente simples para o ser humano. Isso ocorre devido à incapacidade dos SE em fazer analogias e raciocinar a partir de conhecimentos primários (uma vez que eles só conhecem as regras que lhes foram programadas). Portanto, SE devem ser usados apenas em situações nas quais o domínio de conhecimento é restrito envolvendo não mais do que alguns poucos milhares de regras.

Outros problemas surgem no momento do desenvolvimento. Os SE tendem a ter um custo muito elevado, principalmente nos primeiros anos de vida (até atingir o amadurecimento) e com isso, o seu *break-even* tende a ser muito longo. Conforme afirma KELLER (1991), “Em geral a razão para o alto custo de dos sistemas baseados em conhecimento não é o software de inferência. O custo maior concentra-se na aquisição do conhecimento especializado para construir a base de conhecimento”. Essa questão é muito mais relevante do que aparenta e deve ser friamente analisada no planejamento do desenvolvimento do SE.

Por fim, deve-se avaliar todo SE como um sistema suscetível a erros. WILLIANS (1990) aponta diversas questões relevantes com relação às falhas

de SE como: a importância de uma validação extremamente detalhada dos SE, os tipos de erros cometidos pelos SE e a quem cabe a responsabilidade por essas falhas. Mais do que responder essas questões deve-se usar essa referência como um alerta para a importância de considerar a possibilidade de falhas em qualquer SE desenvolvido.

Por tudo isso se deve ter prudência antes de se lançar na construção de um SE. Em todos os casos, recomenda-se que o seguinte *check-list* (baseado no trabalho de VIDAL *et al.* (2000)) seja respondido antes de qualquer ação:

- A tarefa que será abordada pelo SE é prioritária na organização?
- A tarefa proposta pode realmente ser executada por esses SE?
- Pode-se usar um compilador? Qual?
- Pode-se usar um sistema existente ou mesmo juntá-lo ao módulo SE?
- Foram definidos os critérios de desempenho para avaliação dos resultados do SE?
- A rentabilidade do SE será suficiente para justificar seu custo?
- Existe pessoa preparada para desenvolver o SE? E para operá-lo?
- Equipamentos existentes permitirão o desenvolvimento e o uso do SE?
- Como será feita a manutenção do SE e da base de conhecimento?

2.4.5 OS COMPONENTES DE UM SISTEMA ESPECIALISTA

Um SE apresenta, em geral uma arquitetura com três módulos: uma base de regras, uma memória de trabalho e um motor de inferência. A base de regras e a memória de trabalho formam a chamada Base de Conhecimento do SE, onde está representando o conhecimento sobre o domínio. O motor de inferência é o mecanismo de controle do sistema que avalia e aplica as regras de acordo com as informações da memória de trabalho (BITTENCOURT, 1996). Já STAIR (1998) apresenta um modelo composto com cinco elementos, dispostos conforme o esquema a seguir:

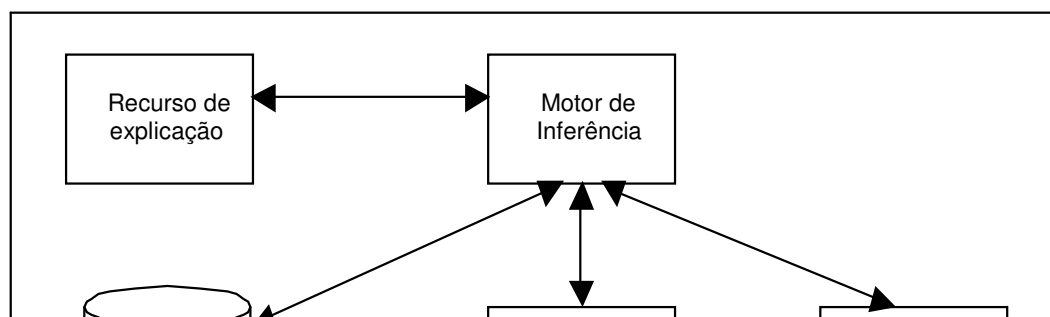


Figura 4: Componentes de um sistema especialista (STAIR, 1998)

Nota-se nesse esquema que apesar de mais detalhado, os componentes básicos (base de regras, memória de trabalho e um motor de inferência) estão presentes através do Banco de Conhecimentos, Aquisição de Banco de Conhecimentos e Motor de Inferência respectivamente; são apresentados complementarmente também os componentes Recurso de Explicação e Interface com o Usuário, que completariam o SE para STAIR (1998).

2.4.6 O DESENVOLVIMENTO DE UM SISTEMA ESPECIALISTA

O desenvolvimento de um SE difere da construção de sistemas convencionais, uma vez que dependem muito das versões experimentais e das freqüentes e intensas interações com o usuário (LAUDON & LAUDON 1999a), consumido muito tempo nas versões preliminares, testes e aperfeiçoamento.

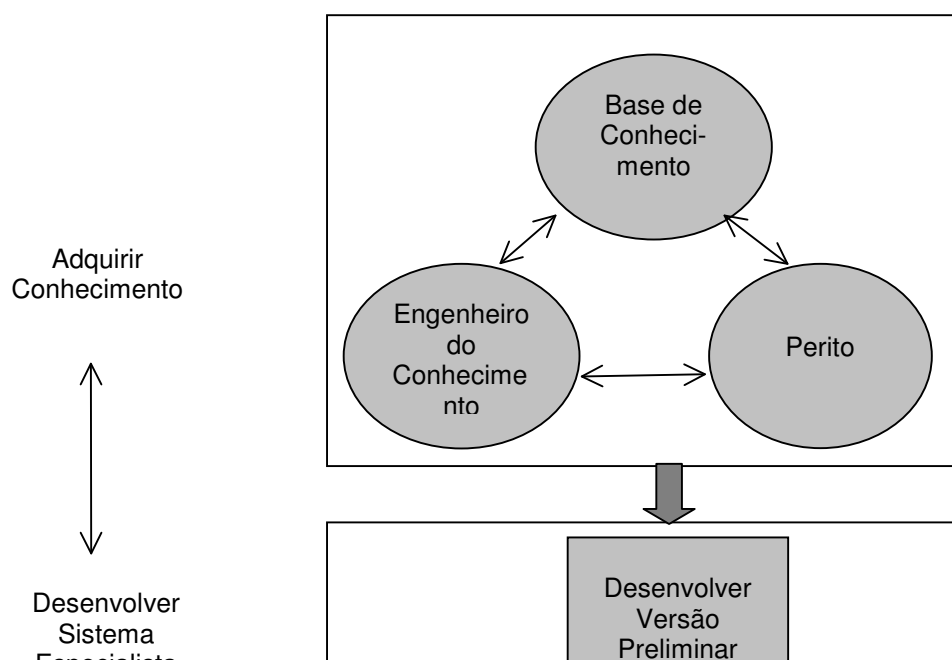


Figura 5: Processo de desenvolvimento de um SE (LAUDON & LAUDON 1999a)

Nessa visão o processo de desenvolvimento de um SE é precedido pela etapa de aquisição do conhecimento. A captação dos conhecimentos dos peritos, realizados pelos engenheiros do conhecimento, servem tanto para a formação da base de conhecimento como para um melhor embasamento para o desenvolvimento do SE pelos responsáveis.

Outra questão relevante no desenvolvimento de um SE é a escolha pelo modo de desenvolvimento: desenvolvimento próprio, desenvolvimento com o apoio de uma *shell* ou a aquisição de uma ferramenta. Como já foi discutido, existe um grande risco do desenvolvimento prolongar-se, (tornando a questão custo extremamente relevante), o que indicaria a aquisição de um software especialista já desenvolvido. Por outro lado, o desenvolvimento próprio possibilita um maior controle sobre as características e os componentes do sistema resultante. Portanto, como a maioria das decisões em software, não há uma decisão única. O que não deve ser esquecido, é que toda escolha deve ser precedida de uma análise muito bem feita.

2.5 RACIOCÍNIO BASEADO EM CASOS - RBC

Os SE têm como premissa à codificação de conhecimento de especialistas, através da construção de regras, mas atualmente, devido ao amadurecimento da existência e da utilização dos sistemas de informações transacionais tradicionais nas empresas, não se deve desprezar o enorme conteúdo de dados (que podem ser transformados em informações e conhecimento)

armazenados em seus bancos de dados. Além disso, a dificuldade em se transformar certas experiências em regras para a utilização em SE, tem feito com que a importância dos sistemas de RBC cresça muito.

Como uma área de pesquisa relativamente nova, o RBC apresenta diversas definições semelhantes, com pequenas diferenças entre um autor e outro. Dentre os principais podemos citar:

- “São sistemas que procuram casos armazenados com características semelhantes ao problema que deve ser resolvido. É utilizada uma técnica que busca resolver novos problemas através da adaptação de soluções utilizadas em problemas anteriores” (KOLODNER, 1993);
- “É uma técnica de Inteligência Artificial (IA) usada para representação de conhecimento e inferência” (BARONE *et al.*, 2003);
- “RBC é uma abordagem para a solução de problemas e para o aprendizado com base em experiência passada. É a solução de novos problemas por meio da utilização de casos anteriores já conhecidos” (WANGENHEIM & WANGENHEIM, 2003).

As descrições de experiências passadas por especialistas humanos, representadas como casos, são armazenadas em um banco de dados (BD) para recuperação posterior, chamado banco de casos (BC). Deve-se armazenar tanto os casos bem sucedidos como os fracassos, havendo apenas um indicativo do tipo de experiência para que o SE possa aproveitar esse exemplo.

“O RBC pode ser mais bem classificado como uma metodologia do que como uma técnica” (SOVAT & CARVALHO, 2001). Independente dessa questão, o RBC é classificado como um tipo de SE, com características específicas. Diferentemente dos SE baseados em normas, o conhecimento disponível para os sistemas de RBC é continuamente expandido e refinado pelos usuários.

Apesar dessa diferença, os dois métodos (baseado em normas e em casos) devem ser vistos muito mais como complementares do que como antagônicos uma vez que a maioria das áreas de aplicação tem os dois tipos de

situações/problemas; e o desenvolvimento de SE híbridos é a tendência natural em qualquer projeto mais abrangente (JACKSON, 1999).

2.5.1 BREVE HISTÓRICO DO RACIOCÍNIO BASEADO EM CASOS

Pode-se considerar que o ponto de partida do RBC na IA foi dado por Schank e Abelson em seus trabalhos sobre Memória Dinâmica e no modelo cognitivo de uma função central de lembrança de situações passadas conforme relata WANGENHEIM & WANGENHEIM (2003). Eles propuseram que o conhecimento humano fique armazenado como roteiros que permitem inferências sobre relacionamentos causais entre ações.

O primeiro sistema baseado em casos foi o CYRUS desenvolvido por Janet Kolodner na Universidade de Yale que se aplicava a soluções diplomáticas baseado em casos descritos pelo ex-secretário de estado dos EUA Cyrus Vance – daí o seu nome - (WANGENHEIM & WANGENHEIM, 2003). Ele serviu de base para outros sistemas RBC como o MEDIATOR – Simpson 1986, o PERSUADER – Sycara 1988, CHEF – Hammond 1989, JULIA – Hinrichs 1992 e o CASEY – Koton 1989 (*apud* AAMODT & PLAZA, 1994).

Na Europa as pesquisas sobre RBC iniciaram-se um pouco mais tarde do que nos EUA. Um dos primeiros resultados apresentados foi o sistema MOLTKE em 1989, seguido pelo sistema de diagnóstico médico de Plaza e Mantaras em 1990 e o sistema CREEK de Aamodt em 1991 (AAMODT & PLAZA, 1994). Em 1992 iniciou-se o primeiro projeto de pesquisa cooperativa internacional em RBC chamado INRECA que desenvolveu uma metodologia de suporte às atividades de desenvolvimento em RBC (BERGMANN & ALTHOFF, 1998). Essa metodologia de apoio à decisão apresentava características híbridas (RBC e raciocínio indutivo) e deu origem a vários sistemas industriais como o CASSIOPEE (sistemas de manutenção de turbinas para aviões Boeing), GDF (controle de medidores de gás) e sistemas de *help desk* para unidades de terapia intensiva (Rússia) e diagnóstico de robôs (França), dentre outros (WANGENHEIM & WANGENHEIM, 2003)

Atualmente são encontradas aplicações de RBC nas mais variadas funções como planejamento, projetos de engenharia, assistência jurídica, áreas de manutenção, detecção de fraudes e atividades de ensino. Além disso, existem conferências e workshops anuais regulares sobre o assunto (como a *ICCBR – Internacional Conference of CBR* e o *EWCBR – European Workshop of CBR*). Isso, entretanto não significa que o assunto já esteja amadurecido em termos de pesquisa. Autores como Plaza, Aamodt, Leake e Schank afirmam que o RBC ainda tem muitos pontos a serem pesquisados e desenvolvidos.

2.5.2 VANTAGENS E LIMITAÇÕES DO RACIOCÍNIO BASEADO EM CASOS

O RBC apresenta vantagens muito significativas quando comparado a outras técnicas de representação do conhecimento. O principal ponto a favor do RBC é que ele procura soluções, relembrando situações similares anteriores, o que é muito parecido com a forma que as pessoas tentam solucionar um problema (WANGENHEIM & WANGENHEIM, 2003). Outra vantagem evidente é a possibilidade de atualização constante, uma vez que o incremento de casos em sua base de casos não exige um trabalho tão refinado como a atualização de regras, geralmente feita por um engenheiro de conhecimento. Dentre as principais vantagens apontadas temos segundo BARONE *et al.* (2003):

- A implantação do sistema e a extração do conhecimento são realizadas de forma mais fácil;
- Reutilização do conhecimento existente na base de dados;
- Disponibilização de uma Base de Conhecimento;
- Aprendizagem automática através de novos casos;
- Possibilidade de trabalho em domínios de difícil formalização;
- Justificativas consistentes podendo incluir alertas sobre prováveis conseqüências da abordagem sugerida.

Outras vantagens sumarizadas por KOLODNER (1993) e JACKSON (1999):

- Regras em uma base de regras são modelos, enquanto casos em uma base de casos são constantes. Por ter variáveis, as regras não descrevem soluções diretamente;
- Para se trabalhar com regras é necessário encontrar resultados exatos e idênticos. Na recuperação de casos pode-se trabalhar por similaridade, facilitando a utilização do modelo em situações mais amplas;
- Enquanto as regras são aplicadas em todo o ciclo de tomada de decisão (até a solução total do problema ou resposta negativa pelo sistema), o uso de casos pode utilizar analogias e soluções parciais, o que em muitas vezes é melhor do que não se encontrar nenhuma resposta.

Além disso, o RBC tem capacidade de incrementar a qualidade da solução de problemas conforme sua experiência aumenta, mesmo em casos onde o conhecimento está mal definido ou incompleto (LEAKE, 1996).

Uma das maiores vantagens do RBC atualmente é sua diversidade de aplicabilidade. A técnica do RBC já foi testada e utilizada com sucesso em praticamente todas as áreas (indústria, comércio, serviços e pesquisas) sendo que os resultados normalmente tem sido muito favoráveis. Essa aplicabilidade garante que as vantagens apontadas já tenham sido aferidas como no caso do RBC utilizado para estimativa de custos de um projeto web (MENDES *et al.*, 2002) que onde a opção pelo RBC foi devido as seguintes vantagens da técnica:

- Estudos recentes, comparando o RBC com outras técnicas, indicaram resultados melhores para o objetivo proposto;
- Um segundo estudo, confirmou esses resultados;
- RBC é um método intuitivo e há evidências de que os especialistas aplicam o raciocínio analógico ao fazerem estimativas;
- RBC é mais simples e flexível, quando comparado com modelos de algoritmos;
- RBC pode ser usado em dados quantitativos e qualitativos, refletindo de forma mais próxima os dados encontrados na vida real.

Comparações recentes de diversas técnicas também indicam resultados muito positivos para a solução baseada em RBC. O trabalho de SHEPPERD & KADODA (2001) comparando quatro das técnicas mais utilizadas (regressão, baseado em regras, baseado em casos e redes neurais) indica aspectos muito positivos para o RBC.

Pela junção de todas essas características descritas anteriormente, o RBC já é apontado como uma das técnicas mais adequadas para apoio à tomada de decisões. Segundo TSATSOULIS *et al.* (1997), “RBC e a Teoria da Decisão podem ser complementares. A Teoria da Decisão ajuda o RBC a tratar com as incertezas do domínio do problema; RBC ajuda a Teoria da Decisão tratar problemas complicados com muitas variáveis”.

Como limitações do RBC pode-se salientar a sua pouca efetividade em aplicativos onde as regras (e não as experiências) são os primeiros referenciais para solução de problemas e onde os casos são únicos, apresentando pouquíssima reutilização (ALLEN, 1994).

Atualmente, existem pesquisas que propõe a utilização de sistemas híbridos, utilizando-se técnicas de SE baseado em casos em conjunto com as técnicas baseadas em regras. Apesar do aumento da complexidade de desenvolvimento, pode-se verificar em (BEN-BASSAT *et al.*, 1998), (CERCONE *et al.*, 1999), (CHI & KIANG, 1999) e (SHUGUANG *et al.*, 2000) que a performance dos sistemas híbridos é consideravelmente melhor, sendo que o nível mínimo de acerto é atingido em um tempo bem inferior ao uso das duas outras técnicas isoladamente. Outras propostas são relacionadas ao uso de RBC em conjunto com outras técnicas (lógica fuzzy e redes neurais, por exemplo) para que as limitações do RBC sejam minimizadas e o sistema final apresente as melhores características de cada técnica. Uma das propostas nessa linha é a da USP de São Carlos, o Ambiente de Desenvolvimento RaBeCa, que pode ser analisado nos artigos de SOVAT & CARVALHO (2001) e SOVAT *et al.* (2001). Outro estudo publicado tem origem italiana e pode ser conhecido no trabalho de BANDINI & MANZONI (2001).

2.5.3 FUNCIONAMENTO DO RACIOCÍNIO BASEADO EM CASOS

Num primeiro momento, pode-se ter a impressão que o funcionamento de um sistema RBC baseia-se, resumidamente, na execução de consultas em grandes bancos de dados. O funcionamento de um sistema RBC, entretanto, é muito mais complexo, pois diferentemente de um banco de dados simples, ele realiza buscas de casos semelhantes, mas não idênticos ao problema proposto. Essa característica é muito importante, pois na maioria das vezes, os casos que correspondem exatamente a uma nova situação podem simplesmente não existir, e, portanto, a utilização de um SGBD não funcionaria adequadamente nessa situação (LEAKE, 1996).

Outra idéia que pode gerar equívoco é de que a técnica de RBC é semelhante à técnica baseada em redes neurais, uma vez que ela também trabalha com casos. Entretanto, essa característica isoladamente não pode indicar uma similaridade total entre as duas técnicas, uma vez que na rede neural, diferentemente do RBC:

- São utilizadas análises estatísticas baseadas em grandes volumes de casos;
- Casos muito importantes ou significativos não devem receber uma relevância maior, a não ser por técnicas artificiais e suspeitas (ZELEZNIKOW, 1995).

O ciclo de funcionamento de um sistema RBC mais aceito foi apresentado por AAMODT & PLAZA (1994) e mostra, após a solicitação de pesquisa feita pelo usuário, os quatro processos principais chamados de 4 Rs: Recuperação, Reutilização, Revisão e Retenção.

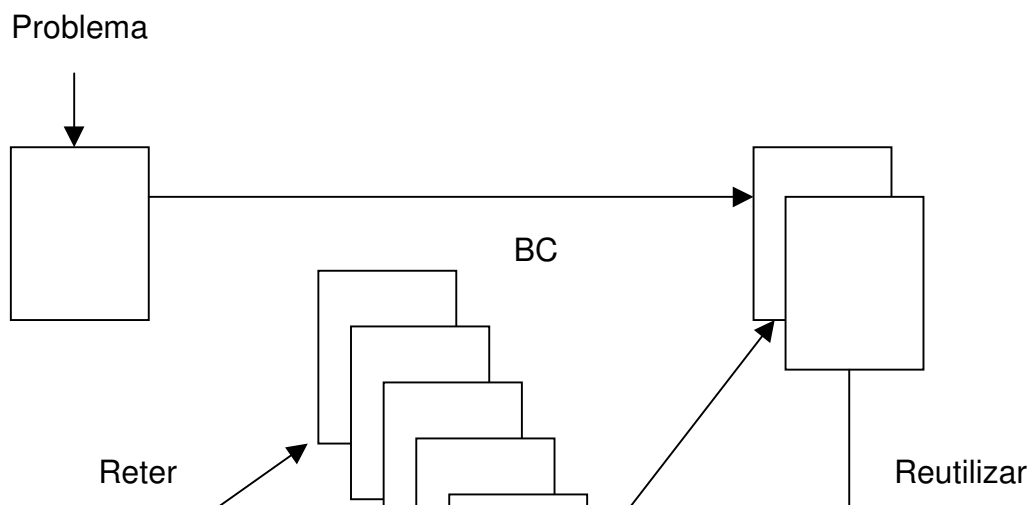


Figura 6: O ciclo do RBC (AAMODT & PLAZA, 1994)

A decisão de reter soluções quase sempre precisará da participação do usuário, o que a princípio pode parecer uma limitação do RBC, mas na verdade é um incentivo à participação do usuário no processo de apoio à tomada de decisão, como afirma WATSON (1997). Para LAUDON & LAUDON (1999b), o ciclo pode ser representado de outra maneira, onde os processos não ficam tão claramente evidenciados, mas o detalhamento de cada passo é mais claro.

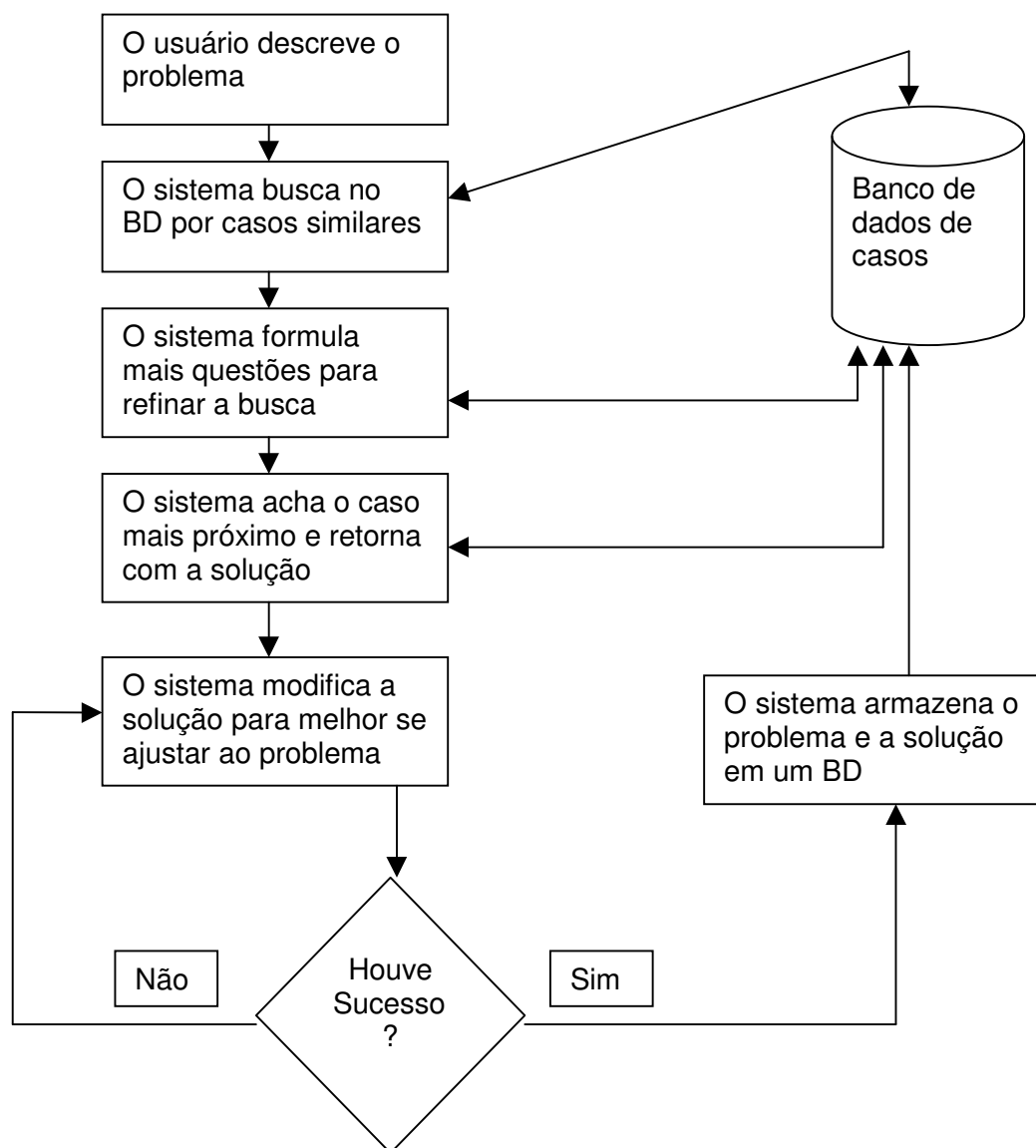


Figura 7: Modelo de funcionamento do RBC (LAUDON & LAUDON, 1999)

2.5.4 O DESENVOLVIMENTO DE UM SISTEMA RBC

Por melhor construído que seja um sistema RBC, o seu desempenho estará diretamente relacionado com sua “experiência” (base de casos). Quanto maior e melhor for a seleção de casos que comporá a base de casos, mais precisos e completos tendem a ser os resultados apresentados.

O desenvolvimento de um sistema RBC pode ser resumido pelo cumprimento das seguintes etapas conforme LORENZI (1998) *apud* BARONE *et al.* (2003):

- Seleção das bases de informação que comporão a base de casos;
- Definição dos atributos relevantes dessa informação;
- Definição dos índices para recuperação dos casos;
- Definição dos métodos de recuperação mais adequados;
- Definição das formas de adaptação dos casos recuperados para se adequarem aos problemas propostos;
- Definição dos processos de aprendizagem (avaliação por parte do sistema se a solução deve ser armazenada ou não).

A etapa de desenvolvimento é fundamental para o sucesso ou fracasso de um sistema baseado em RBC. Cada uma das seis etapas listadas anteriormente é suficientemente relevante para determinar o resultado de todo o desenvolvimento. Apesar do RBC apresentar a tendência de melhorar com o tempo e uso, uma definição de casos para composição da base inicial ruim, pode fazer com que os usuários se desinteressem pelo sistema nas primeiras utilizações, impedindo que o sistema mostre sua verdadeira capacidade e utilidade.

Deve-se identificar, no início do desenvolvimento, em qual categoria o SE com RBC se enquadra para adequar o desenvolvimento à essas características. Conforme DUTTA *et al.* (1997) um sistema RBC pode ser classificado como:

- **Convencional** – busca automatizar a solução de problemas;
- **Auxílio à decisão** – ajuda na tomada de decisões fornecendo informações;
- **Estimulante** – busca estimular o aprendizado através da resolução dos conflitos com base em experiências concretas, observação reflexiva, contextualização abstrata e experimentação ativa.

A proposta desse estudo será o desenvolvimento de um RBC de auxílio à decisão, categoria que vem se destacando no desenvolvimento de sistemas RBC.

ANGEHRN & DUTTA (1998) ressaltam o impacto do RBC no processo de tomada de decisões, evidenciando o grande papel da técnica no aprendizado

que deve ocorrer durante o processo. Pode-se avaliar esse impacto através do esquema retratado na figura a seguir.

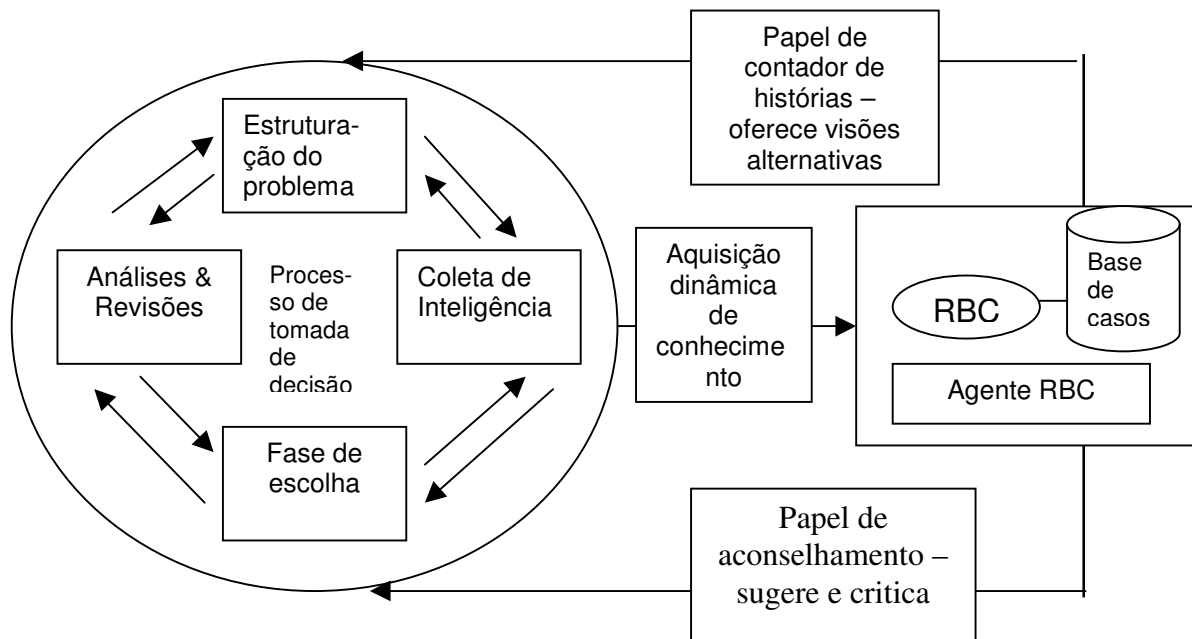


Figura 8: Impacto do agente RBC no processo de decisão (ANGEHRN & DUTTA, 1998)

Com relação às técnicas de desenvolvimento, pode-se trabalhar com os mesmos recursos utilizados no desenvolvimento de outros tipos de SE, com pequenas adaptações às características peculiares do RBC.

Com o incremento do uso de SE com RBC, técnicas anteriormente utilizadas somente no desenvolvimento de sistemas tradicionais têm sido testadas com sucesso para o RBC, como, por exemplo, a modelagem formal de argumentos analisada por BENCH-CAPON (1999) e a técnica de validação *Case Library Subset Test Technique* apresentada por GONZALEZ *et al.* (1998).

Outra questão extremamente relevante para o desenvolvimento de um SE com RBC é o adequado planejamento da manutenção posterior da base de casos (IGLEZAKIS *et al.* 2001). Esse procedimento, muitas vezes pouco lembrado no desenvolvimento de um RBC pode fazer com que um sistema inicialmente eficiente apresente queda na eficiência em pouco tempo, fazendo com que o sistema fique desacreditado e possa ser abandonado pelos usuários. Visando

destacar a relevância da manutenção em sistemas RBC, ROTH-BERGHOFFER & IGLEZAKIS (2001) propuseram um ciclo expandido para os sistemas RBC, introduzindo 2 novos Rs (Reavaliação e Reparação) aos 4 já tradicionais (Recuperação, Reutilização, Revisão e Retenção), conforme a figura a seguir:

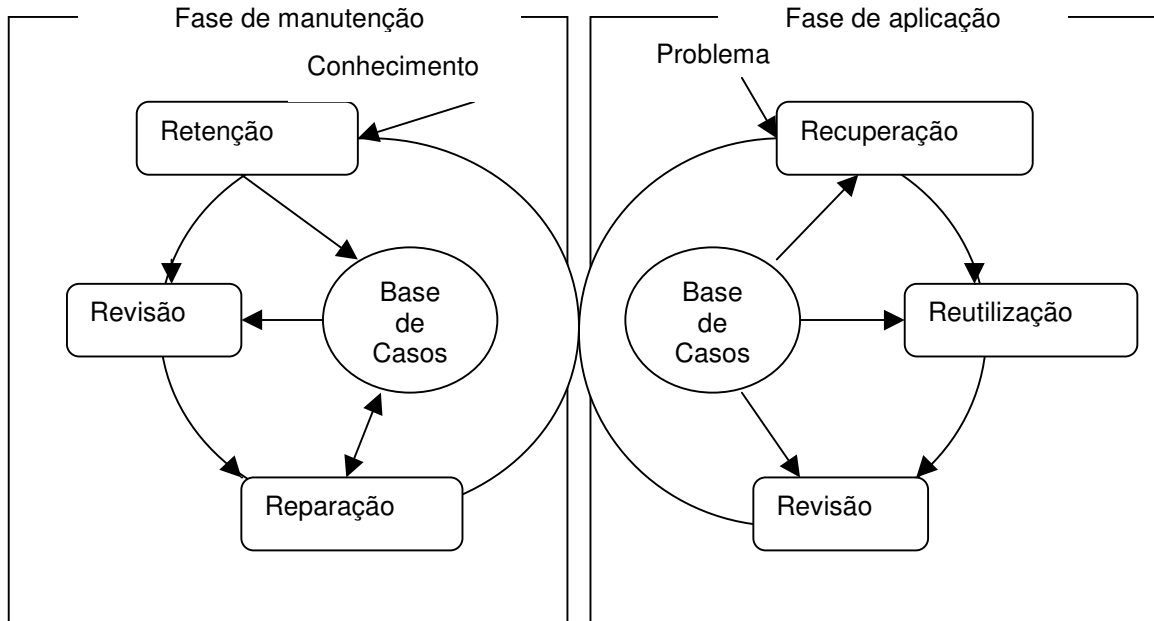


Figura 9: O ciclo expandido do RBC (6 Rs) (ROTH-BERGHOFFER & IGLEZAKIS, 2001)

2.5.4.1 Casos

Definir corretamente o que será um caso em um sistema RBC pode ser o diferencial entre o sucesso e o fracasso de todo o investimento. Caso pode ser definido como uma peça de conhecimento contextualizado representando uma experiência ou um episódio concreto. Um caso contém a lição passada, que é o conteúdo do caso e o contexto em que a lição pode ser usada (WANGENHEIM & WANGENHEIM, 2003). Pode-se descrevê-lo também como situações-problema e suas respectivas soluções, detalhando as características que possam vir a ser úteis, tanto para a localização do caso, quanto para o aproveitamento da experiência pela pessoa que o está pesquisando. Outra definição encontrada afirma que caso é um pedaço contextualizado de conhecimento, representando uma experiência (KOLODNER, 1993). Muitos associam um conjunto de casos a uma biblioteca (JACKSON, 1999), mas comparar casos a livros não é adequado uma vez que:

- Um caso é semelhante a um livro por conter conhecimento, entretanto, o conhecimento existente em um caso já está “preparado” para a solução de um problema bem específico, como uma regra pronta para ser aplicada, enquanto em um livro ele geralmente está mais abrangente e precisa ser ordenado para a utilização;
- Um caso pode apresentar uma solução em um contexto bem particular, descrevendo o resultado do ambiente envolvido com a adoção da solução adotada. Um livro também pode conter essas informações, mas elas não estarão em um formato que o sistema possa compreender;
- A profundidade de um caso é geralmente menor do que a de um livro.

Não se deve “padronizar” excessivamente os casos (obrigando o preenchimento de todas as informações para todos os casos, por exemplo), uma vez que eles apresentam características diferentes; por outro lado, se não houver a definição de alguns pontos básicos em cada caso, não será possível trabalhar com um sistema RBC. O que é comum a todos os casos é que eles devem representar uma situação ou uma experiência obtida em um contexto. O caso registra um conhecimento em nível operacional (KOLODNER, 1993). Existem diversas idéias de quais seriam os componentes básicos de um caso, sendo que cada uma apresenta vantagens e desvantagens, de acordo com o uso proposto. Pode-se, entretanto partir de três elementos básicos que obrigatoriamente farão parte de qualquer caso a ser utilizado em um sistema RBC (WANGENHEIM & WANGENHEIM, 2003):

- Descrição do problema (ou do fato)
 - Metas a serem alcançadas na resolução do problema;
 - Restrições nestas metas;
 - Características da situação e relações entre suas partes.
- Descrição da solução:
 - A solução propriamente dita;
 - Passos de raciocínio para resolver o problema;
 - Justificações para decisões que foram adotadas na resolução do problema;

- Soluções aceitáveis que não foram escolhidas (soluções alternativas) ou soluções não aceitáveis que foram excluídas;
 - Expectativas sobre resultados.
- Conclusão:
 - O resultado propriamente dito;
 - Se o resultado preencheu ou não as expectativas;
 - Se o resultado foi sucesso ou falha;
 - Explicação do não preenchimento da expectativa e/ou falha;
 - Estratégia de reparo;
 - O que poderia ter sido feito para evitar um problema;
 - Apontar para uma próxima possível solução (resultado da aplicação do reparo).

2.5.4.2 Base de Casos

Base de casos (BC) é o conjunto que engloba a base de conhecimento de um sistema RBC (BARONE, 2003). Como todo sistema baseado em IA, a representação do conhecimento tem uma enorme relevância no processo de desenvolvimento. No caso dos sistemas RBC, há dois aspectos principais a serem considerados: os modelos de memória e a organização dos casos.

Nos tipos de memória, pode-se trabalhar com os seguintes modelos existentes:

- Árvores e grafos;
- Redes semânticas;
- Frames;
- Memória dinâmica e;
- Categoria e exemplares.

Na organização dos casos, leva-se em conta a natureza dos dados e as etapas de recuperação, reutilização e retenção do sistema RBC. Pode-se armazenar casos numa estrutura plana, numa estrutura hierárquica ou ainda em estruturas intermediárias (planas que apontem para um nível de índices) (BARONE, 2003). Uma falsa impressão desmistificada pelo estudo de KADOLA *et al.* (2000) é a de que um sistema RBC terá eficiência proporcional ao tamanho da

BC. Deve-se avaliar a quantidade de casos que proporciona uma melhor relação de eficiência, uma vez que uma super alimentação da BC, além da queda de desempenho, resultará em custo elevado de alimentação e manutenção, sem necessariamente garantir melhores resultados.

2.5.4.3 Indexação

A indexação em uma base de casos é fundamental para que um sistema RBC tenha habilidade e eficiência na identificação de casos adequados para a solução de problemas. Segundo WANGENHEIM & WANGENHEIM (2003) pode-se definir Índices como: combinações de seus atributos mais importantes, que permitem distingui-lo de outros e identificar casos úteis para uma dada descrição de problema. Um bom índice trabalha com características relevantes dos casos para que o processo de pesquisa e recuperação seja simples e o mais automatizado possível. A indexação em um sistema de RBC é diferente de um SGBD, uma vez que um SGBD cria índices com poucos campos e um sistema RBC pode trabalhar com dezenas ou até centenas de campos na indexação.

De acordo com o tipo de uso do SE com RBC a ser desenvolvido, pode-se melhorar consideravelmente o desempenho do sistema com a utilização de algumas técnicas/práticas de indexação como a eliminação de termos muito comuns em consultas textuais e o uso de expressões indicativas que orientam o usuário no momento da consulta ao sistema, conforme DAGOSTINI *et al.* (1999) e BARRETO *et al.* (2001)

Técnicas de indexação podem ser manuais e automáticas. Os procedimentos de indexação automáticos são baseados em duas propostas de indexação. Na primeira os casos são indexados com base em características e dimensões comuns a todos os casos e que tendem a serem úteis para localização. Na segunda, cada caso é analisado individualmente e suas características mais relevantes são selecionadas para composição do índice, independente do outros casos existentes na base (KOLODNER, 1993).

2.5.4.4 Recuperação de Casos

O objetivo da recuperação de casos é encontrar um caso ou um pequeno conjunto de casos na base de casos que contenha uma solução útil para o problema ou situação atual (WANGENHEIM & WANGENHEIM, 2003).

AAMODT & PLAZA (1994) identificam três fases do processo de recuperação de casos que são:

- Identificação de características;
- Cálculos de equiparação/similaridade;
- Seleção de casos.

Como a identificação de características deve ser trabalhada no desenvolvimento do sistema, através de um processo de engenharia de requisitos muito bem realizado, restam as outras duas tarefas como extremamente relevantes para os resultados que serão gerados pelo sistema: o cálculo de similaridade e a seleção dos casos. Em ambos os casos, podem-se adotar diferentes critérios, sendo que cada uma das opções existentes apresenta vantagens e desvantagens dependendo das características do sistema e do domínio de trabalho. Algumas das técnicas mais utilizadas para cálculo de similaridade em RBC são *Nearest neighbor*, *Nearest neighbor ponderado*, Distância Euclidiana, Distância Euclidiana ponderada, Distância de Manhattan. Para seleção de casos podem ser citadas as técnicas de Recuperação Seqüencial, Recuperação de Dois Níveis, Recuperação com Árvores k-d, Redes de Recuperação de Casos, dentre outras.

Um exemplo da etapa de recuperação de casos pode ser dado com uma situação fictícia de um sistema RBC para manutenção em impressoras. Para a recuperação de casos, é necessária a realização de um cálculo de similaridade. Em sistemas RBC, a similaridade entre dois atributos, é usada para determinar a proximidade entre uma pesquisa e um caso armazenado e com base nessa proximidade, retornar os casos mais adequados à pesquisa realizada. Pode-se calcular essa similaridade com base no resultado global ou local. No primeiro caso, os resultados de todos os atributos do caso afetam o resultado e no segundo esse cálculo é feito individualmente para cada atributo.

Esse cálculo indica, através de um número real (entre 0 e 1), o nível de igualdade entre o caso avaliado e o problema a ser solucionado. A seguir é apresentado um modelo de cálculo de similaridade baseado em uma tabela de importância e semelhança. Para entender o seu funcionamento, será utilizada uma situação fictícia para localização de casos de defeitos em impressoras:

Quadro 3:Exemplo da medida de similaridade local para modelo de impressora

Valor da similaridade	Robotron 100	Robotron 200	Robotron Matrix 600	Robotron Matrix 800
Robotron 100	1,0	0,8	0,2	0,2
Robotron 200	0,8	1,0	0,2	0,2
Robotron Matrix 600	0,2	0,2	1,0	0,9
Robotron Matrix 800	0,2	0,2	0,8	1,0

Fonte: WANGENHEIM & WANGENHEIM (2003)

Observa-se que é buscado através dos níveis de similaridade indicados, “informar” ao sistema que as impressoras Robotron 100 e Robotron 200 são mais “parecidas” entre si do que as Robotron Matrix 600 e 800, que pertencem à outra série. Isso é importante para que o sistema possa calcular onde estão os casos mais semelhantes quando o problema apresentado não encontrar uma situação idêntica existente na base de casos.

Quadro 4:Exemplo de cálculo de similaridade

Atributos	Importância (Peso)	Situação atual	Caso 1 da BC	Valor da Similaridade
------------------	---------------------------	-----------------------	---------------------	------------------------------

Atributos	Importância (Peso)	Situação atual	Caso 1 da BC	Valor da Similaridade
Descrição	Alta (5)	Impressora não imprime em cores	Impressora não funciona	0
Modelo	Baixa (1)	Imprex 200	Imprex Plus 600	0,6 (similar)
Luz de estado do papel	Média (2)	Apagada	Apagada	1,0 (igual)
Luz do estado de tinta colorida	Média (2)	Acesa	Apagada	0,0
Luz do estado de tinta preta	Média (2)	Apagada	Apagada	1,0 (igual)
Estado do interruptor	Média (2)	Não conhecido	Ligado	0

Valor da similaridade global: $(5*0+1*0,6+2*1,0+2*0)/(5+1+2+2+2+2) = 0,3$

Fonte: WANGENHEIM & WANGENHEIM (2003)

A situação atual apresenta as características do problema a ser solucionado e a coluna Caso 1 da Base de Casos mostra as características de um caso armazenado para que possa ser usado como modelo de solução do problema. Através dos valores de similaridade atribuídos anteriormente (Quadro 3) e dos pesos, é realizado o cálculo da similaridade do problema apresentado em relação ao caso 1. O cálculo realizado para similaridade apresentou um resultado de 0,3. A partir daí, o próximo passo dependerá da técnica de seleção de casos adotada. No exemplo, será adotada a recuperação seqüencial, onde a medida de similaridade é calculada para todos os casos existentes na base de casos. Feito isso, basta que o sistema selecione os n casos mais similares ao problema em questão para apresentar como resultado.

Tanto a técnica de recuperação de casos quanto a métrica de cálculo de similaridade possuem diversas formas de utilização. Como as opções por uma dessas formas são dois dos objetivos desta pesquisa, nos capítulos 4 e 5 esses assuntos serão aprofundados para possibilitar análises fundamentadas em cada um dos casos.

2.5.4.5 Reutilização de Casos

Tendo sido localizado um caso (ou alguns casos) apropriado para seleção do problema, a próxima etapa consiste em adaptar esse caso selecionado para a formulação da proposta de solução. O caso localizado pode ser utilizado em seu formato original ou, se for preciso sofrer adaptações necessárias (AAMODT & PLAZA, 1994). Essas adaptações podem ser manuais ou automáticas. Para adaptação, existem questões centrais que devem ser observadas (WANGENHEIM & WANGENHEIM, 2003):

- Quais são os aspectos de uma situação (descrita por um caso) que devem ser adaptados;
- Quais modificações são razoáveis para adaptar o caso;
- Quais são os métodos de adaptação aplicáveis para esses aspectos;
- Como controlar o processo de adaptação para saber que as modificações estão sendo realizadas no rumo certo.

2.5.4.5.1 Estratégias de Adaptação

Existem inúmeras técnicas de adaptação, cada uma com pequenas diferenças em relação às outras. O processo de adaptação de um caso em um sistema RBC pode ser esquematizado genericamente conforme o modelo apresentado por JEDLITSCHKA & ALTHOFF (2001), (figura 10).

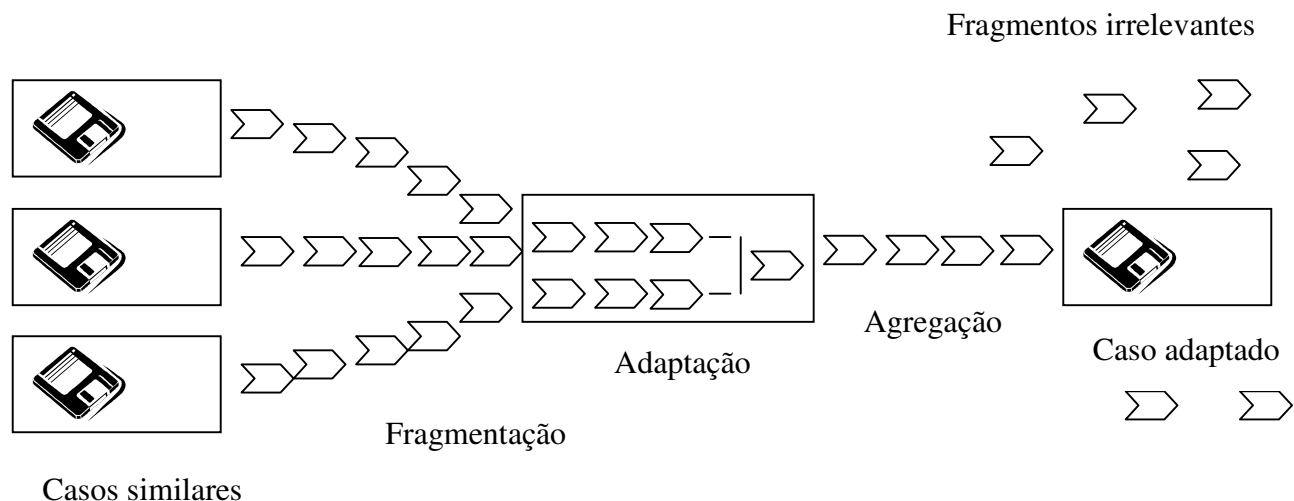


Figura 10: Princípios da adaptação (JEDLITSCHKA & ALTHOFF, 2001)

Como as adaptações podem ser de diversos tipos, existem desde estratégias mais simples até algumas extremamente complexas. É importante definir quais são as situações que podem ser utilizadas, para preparar o sistema com as técnicas adequadas. Na adaptação automática, existem técnicas a serem seguidas, que enfocam duas questões: a) que parte do caso recuperado pode ser aproveitada na questão sob análise; e b) as diferenças entre o caso passado e o atual. Uma visão geral das técnicas de adaptação para RBC mais conhecidas e utilizadas pode ser representada conforme o resumo apresentado por WILKE & BERGMANN (1998):

- Adaptação nula
- Adaptação transformacional
 - o Adaptação substitucional
 - o Adaptação estrutural
- Adaptação gerativa/derivacional
 - o Inicialização
 - o One-shot replay
 - o Replay entrelaçado
- Adaptação composicional
- Adaptação hierárquica

Esse resumo apresenta as diversas técnicas organizadas: da técnica mais simples e limitada - Adaptação nula – para as mais complexas como a Adaptação gerativa, a composicional e a hierárquica. Deve-se analisar, de forma bem detalhada a melhor técnica a ser adotada, pois nem sempre a técnica mais complexa é a melhor opção, dependendo da realidade e das necessidades dos usuários do sistema RBC.

2.5.4.6 Revisão de Casos

A revisão de casos é a oportunidade de aprendizado a partir de uma falha, que ocorre quando a solução gerada na fase de reuso não se mostrou adequada (AAMODT & PLAZA, 1994). Ela é constituída de duas tarefas (WANGENHEIM & WANGENHEIM, 2003):

- Avaliação da solução gerada, definindo-se pelo aproveitamento ou não do caso;
- Caso a opção seja pelo não aproveitamento, deve-se utilizar o conhecimento específico sobre o domínio, fornecido pelo usuário para corrigir a solução proposta pelo caso.

2.5.4.7 Retenção de Novos Casos

É o processo de incorporação, ao conhecimento já existente, daquilo que é útil de um novo episódio de solução de um problema (WANGENHEIM & WANGENHEIM, 2003). Dessa forma, a base de casos estará constantemente e continuamente em atualização.

Para realizar a retenção de novos casos, algumas questões devem ser observadas e definidas para que o procedimento ocorra de forma satisfatória, segundo AAMODT & PLAZA (1994):

- **Extração de dados** – na retenção deve-se avaliar que tipos de dados devem ser retidos. Pode incorporar casos inéditos, casos que alteram um conhecimento ultrapassado ou com informações contraditórias. A relevância está na definição de características relevantes que devem ser usadas para avaliação da possibilidade de extração de dados.
- **Indexação** – Esse problema é foco, pois a retenção de um novo caso exige a identificação das informações que serão usadas para indexação.
- **Integração** – é a etapa final da retenção. Inclui a adição de novos casos ou a exclusão de casos obsoletos, indicando a aprendizagem ocorrida.

Existem basicamente três opções de sistema RBC com relação ao recurso de retenção de casos (WANGENHEIM & WANGENHEIM, 2003):

- **Sem retenção de casos** – Alguns poucos sistemas de RBC ainda não utilizam o recurso de retenção de novos casos, ficando restritos aos casos modelados e alimentados na base de casos durante o processo de desenvolvimento do sistema;
- **Retenção de solução de problemas** – Essa é a forma mais comum de retenção de novos casos em sistemas de RBC. Cada novo caso de

sucesso pode ser acrescentado à base de casos (de forma automática ou após uma filtragem) fazendo com que o uso torne o sistema cada vez mais completo. Existem sistemas que além dos casos de sucesso, guardam também os insucessos para que os erros sejam evitados;

- **Retenção de documentos** – Nesse caso, o processo de incorporação de novos casos é feito independentemente do sistema, através da incorporação de documentos ou novos casos disponibilizados em pacotes adquiridos ou formatados de forma independente.

A opção pela retenção de solução de problemas é, com certeza, a escolha ideal na maioria dos casos, pois isso torna o sistema mais eficiente com o uso e o passar do tempo (KOLODNER & MARK 1992), e será a opção adotada no SAD para auditoria dessa pesquisa, em conjunto com a técnica da retenção de documentos.

2.5.5 FERRAMENTAS PARA RBC

Apesar de tratar-se de uma área de estudo relativamente recente, os SE com RBC já apresentam uma lista extensa de ferramentas de desenvolvimento. Será apresentada a seguir uma pequena amostra.

2.5.5.1 Ferramentas de Desenvolvimento RBC - Área Acadêmica

- Case Advisor 4 - (Information Service Agents Lab at Simon Fraser University)
- CASPIAN - (Depart. of Computer Science at Aberystwyth, University of Wales) *
- CASUEL Parser - (Research Group "Artificial Intelligence / Knowledge Based Systems" of Prof. Richter at University of Kaiserslautern)
- CBR Design Explorer - (AI Applications Institute at University of Edinburgh) *
- CBR Framework for Bioprocess - (Bioprocesses Group at Biotechnology and Food)
- CBR Tools - object oriented software library in JAVA

- CBR-Works product family - CBR shell (Empolis Knowledge Management)
 - Intellix Demo Applications - Intellix Tools
- * = Ferramentas gratuitas

2.5.5.2 Ferramentas de Desenvolvimento RBC - Área Comercial ou Industrial

- Case Advisor 4 - (ISA Lab at Simon Fraser University)
 - CBR Design Explorer - (AI Applications Institute at University of Edinburgh) *
 - CBR Framework for Bioprocess - (Bioprocesses Group Biotechnology and Food)
 - CBR-Works - (Empolis Knowledge Management GmbH)
 - CPR/CPR Web Server - (Haley Enterprise Inc.)
 - Easy Reasoner - (Haley Enterprise Inc.)
 - Intellix Demo Applications – (Intellix Tools)
 - KATE Software Suite - (AcknoSoft)
 - Knowledge Builder 4.1 - (ServiceSoft)
 - Knowledge-Pak Architect - (ServiceWare Inc.)
- * = Ferramentas gratuitas

Apesar dessa grande variedade de ferramentas para desenvolvimento com base em RBC, não temos ainda uma ferramenta completa, que possa ser utilizada em qualquer situação com perfeição. Análises técnicas dessas ferramentas, como as realizadas por WATSON (1995) orientam o usuário a escolher a ferramenta que apresentar as características mais adequadas a sua realidade de desenvolvimento. Essa seleção criteriosa para a adoção de uma ferramenta de auxílio para o desenvolvimento do SAD para auditoria é um dos objetivos dessa pesquisa e será realizada no capítulo 6.

2.6 AUDITORIA

A palavra auditoria tem sua origem etimológica, do latim *audire*, que significa ouvir. Inicialmente o termo era usado apenas para trabalhos de revisão dos registros contábeis, mas hoje auditoria tem um sentido mais amplo. O

ESCRITÓRIO DO AUDITOR-GERAL DO CANADÁ (1995) conceitua a auditoria como “a ação independente de um terceiro sobre uma relação *accountability*, objetivando expressar uma opinião ou emitir comentários e sugestões sobre como essa relação está sendo obedecida”.

2.6.1 BREVE HISTÓRICO DA AUDITORIA

A auditoria não tem registro do nome de seu criador ou da primeira pessoa a trabalhar como auditor. Acredita-se que a auditoria tenha surgido na Itália em torno de 1580, mesma época em que a profissão de contador foi reconhecida. O papel do auditor então, seria conhecido como guarda livros e sua função era analisar e revisar 100% dos lançamentos contábeis em busca de erros ou inconsistências. Com o tempo, o auditor ampliou o seu foco, passando a avaliar as demonstrações financeiras para os investidores não administradores.

2.6.2 A AUDITORIA NO BRASIL

Existem pouquíssimos registros sobre a origem da auditoria no Brasil, mas é certo que seja inglesa. O primeiro parecer de auditoria ocorreu em 1903 (KANITZ, 1979 *apud* JUND 2002), relativo ao exame de livros da São Paulo Tramway, Light and Power Company. Desde 1909 a Arthur Andersen S/C já tinha se instalado no Rio de Janeiro, sendo seguida pela Price Waterhouse em 1915, mas esses trabalhos estavam restritos a auditoria das filiais de empresas multinacionais a pedido das matrizes. A auditoria só viria a ser formalmente organizada no país em 1957 e fortalecida em 1972 com a congregação e auto-disciplinação dos profissionais, através do Instituto dos Auditores Independentes do Brasil.

2.6.3 CLASSIFICAÇÃO DOS TRABALHOS DE AUDITORIA

Existem diversas classificações para os trabalhos de auditoria. De acordo com o foco ou o modelo, o trabalho de auditoria pode receber nomes diversos como: Auditoria de Sistemas; Auditoria a Distância; Auditoria Especial; Auditoria de Gestão; Auditoria Contábil; Auditoria Fiscal; Auditoria de Contas;

Auditoria Orçamental; Auditoria Articulada; Auditoria de Irregularidades e Auditoria Eletrônica.

Esse trabalho estará baseado na classificação apresentada por ARAÚJO (2001) quanto à forma de realização e quanto ao objetivo dos trabalhos.

- **Classificação quanto à forma de realização** - Os trabalhos de auditoria podem ser realizados por dois tipos de equipes de auditores:
 - o **Auditoria interna:** é a auditoria realizada por profissionais vinculados à entidade auditada. Funciona como órgão de assessoramento. Será detalhada a seguir;
 - o **Auditoria externa:** é a auditoria realizada por profissionais qualificados, que não são empregados da administração auditada, com o objetivo precípuo de emitir uma opinião independente sobre a adequação ou não das demonstrações contábeis.
 - o OBS: Em algumas situações específicas alguns trabalhos de auditoria podem ser executados por equipes mistas (auditores internos e auditores externos).
- **Classificação quanto ao objetivo** - Os trabalhos de auditoria podem ter focos distintos o que altera a forma e o domínio do trabalho:
 - o **Auditoria contábil:** avalia as demonstrações contábeis;
 - o **Auditoria Operacional:** avalia o desempenho e a eficácia das operações, os sistemas de informações e de organização, e os métodos de administração, a propriedade e o cumprimento das políticas administrativas; e a adequação e a oportunidade das decisões estratégicas;
 - o **Auditoria Integrada:** é a realização da auditoria contábil tradicional juntamente com a auditoria operacional.

2.6.4 AUDITORIA INTERNA

Para a Associação dos Auditores Internos (<http://www.iaa.com>), Auditoria interna é uma atividade de consultoria, independente e objetiva, que tem como finalidade agregar valor e aprimorar as operações de uma organização, ajudando à empresa atingir seus objetivos por meio de avaliações sistemáticas

e estruturadas que contribuem para a melhoria dos processos de gerenciamento de risco, de controle e de governança.

A auditoria interna é uma atividade recente que tem se difundido principalmente em empresas públicas ou de economia mista. O papel do auditor interno, entretanto, está sendo cada vez mais valorizado, exigindo deste profissional, competência em diversas áreas de negócios. Apesar dessa valorização, na maioria dos casos, a implantação da auditoria interna em uma empresa ocorre, não por opção administrativa, mas sim por exigência legal. No Brasil a existência da auditoria interna é regulamentada pelo Decreto no 3.591, de 06/09/2000 (BRASIL, 2000), para as entidades da Administração pública Federal Indireta.

2.6.5 AUDITORIA OPERACIONAL

A Auditoria Operacional ou *Performance Audit* pode ser definida resumidamente como avaliação e análise do desempenho de uma organização.

2.6.5.1 Breve Histórico da Auditoria Operacional

Com as mudanças no mundo de negócios, a auditoria precisou oferecer o tipo de informação desejada pelos seus patrocinadores. A auditoria operacional surgiu dessa evolução, focada na avaliação da economia, eficiência e eficácia, mediante a revisão dos processos administrativos. Por questões legais, a auditoria operacional surgiu na área governamental, no início da década de 70 nos EUA e a partir de 1985 (oficialmente em 1988) no Brasil. No setor privado, ela tem evoluído conforme os investidores exigem informações mais apuradas para terem mais segurança com relação aos valores aplicados na empresa. Os recentes casos de fraude em grandes empresas (*Banco Barrings, Enron, Xerox*) demonstram como as informações contábeis tradicionais não tem sido suficientes para transmitir tranquilidade e segurança aos investidores.

2.6.5.2 Objetivos

Segundo JUND (2002), o objetivo geral da auditoria operacional é assessorar a administração no desempenho efetivo de suas funções e responsabilidades, avaliando se a organização, departamento, atividades, sistemas, funções, operações e programas auditados estão atingindo os objetivos organizacionais e gerenciais com eficiência, eficácia e economia, na obtenção e utilização dos recursos (financeiros, materiais, humanos e tecnológicos), bem como em observância às leis e regulamentos aplicáveis e com a devida segurança.

2.6.5.3 Metodologia de Trabalho da Auditoria Operacional

A definição de uma metodologia para auditoria operacional não é simples. Ela deve ser ao mesmo tempo abrangente (para que “englobe” qualquer tipo de processo) e detalhada (para que nenhum ponto relevante fique fora do trabalho). Uma proposta de fluxo do trabalho de auditoria operacional apresentada por JUND (2002) pode ser representada graficamente da seguinte maneira:

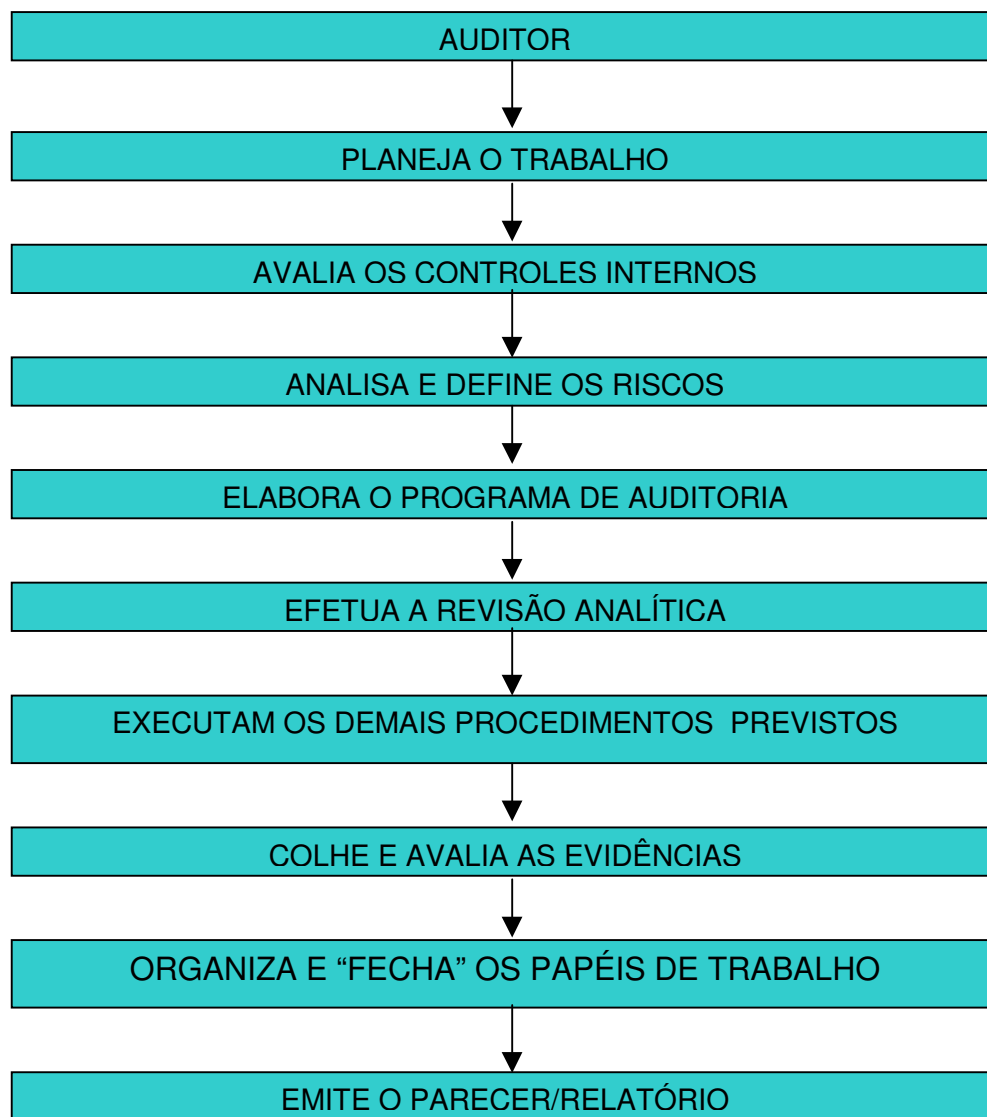


Figura 11: Estágios de uma auditoria operacional (JUND, 2002)

Nesse fluxo percebe-se que o trabalho apresenta uma seqüência unidirecional, onde o auditor deve realizar o trabalho do começo ao fim, independente dos resultados encontrados em cada fase. Para evitar essa situação, já se trabalha em um processo mais avançado, onde algumas etapas possuem procedimentos metodológicos mais definidos e apresentam opções de retorno para revisões, apoiando-se principalmente nos avanços tecnológicos recentes, como os sistemas de informação extremamente detalhados e a disseminação desses recursos entre os auditores. Como exemplo, o fluxo do processo de

auditoria operacional utilizado atualmente pela Auditoria Interna do Banco do Brasil S.A. pode ser representado da seguinte maneira:

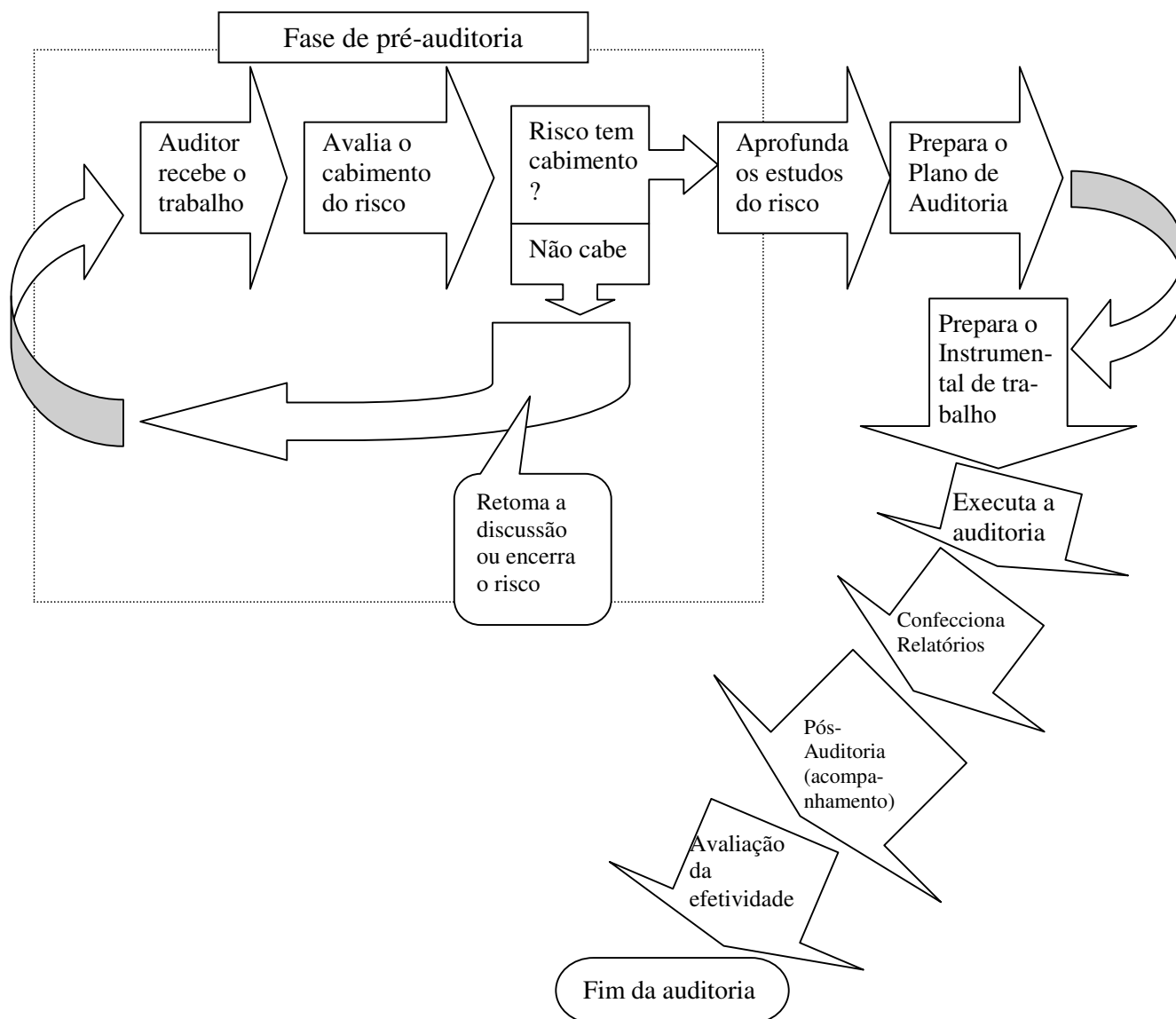


Figura 12: Estágios de uma auditoria operacional do Banco do Brasil S.A.

A proposta de SAD para auditoria interna que será apresentada neste trabalho abrangerá os procedimentos de pesquisas da pré-auditoria. Apesar das fortes evidências que os resultados alcançados com o pleno funcionamento do sistema resultarão em benefícios diretos em todas as etapas da auditoria, como planejamento da auditoria, execução e pós-auditoria, essas etapas não serão focadas nessa pesquisa.

2.6.5.3.1 Pré Auditoria

Pré-Auditoria pode ser vista como a busca pela maior quantidade possível de informações antes de iniciar os seus trabalhos para:

- Evitar erros de estratégia;
- Aproveitar trabalhos semelhantes já realizados;
- Buscar a maior quantidade de informações existentes sobre o assunto;
- Preparar-se tecnicamente para a tarefa a ser realizada.

A Pré-Auditoria precede a elaboração e aplicação do que popularmente é conhecido como o trabalho de um auditor – a aplicação de testes. A elaboração das conclusões de auditoria, a comunicação dos resultados e a realização da pós-auditoria também são uma consequência das decisões tomadas durante a pré-auditoria.

Etapas

Existem diferentes divisões de etapas do processo de pré-auditoria. Na proposta de ARAÚJO (2001), por exemplo, a pré-auditoria está sendo chamada de Planejamento e engloba as etapas de Análise Geral, Análise Específica e Revisão e Avaliação dos Controles Internos. Já na metodologia adotada pela Auditoria Interna do Banco do Brasil S.A., as etapas da pré-auditoria são três:

- **Levantamento e Análise de Dados** - Forma ordenada e sistematizada de coleta e avaliação de dados com o objetivo de produzir informações seguras ao auditor na identificação dos principais riscos e controles necessários, a definição do escopo do trabalho e a elaboração do Plano de Auditoria;
- **Detalhamento dos Riscos e Definição do Escopo** - Para facilitar o entendimento do processo a ser auditado, o Risco Macro pode ser desdobrado em Riscos Detalhados através da Pré-Auditoria. Os Riscos Detalhados geralmente representam uma probabilidade concreta de ocorrer perdas ou custo excessivo;
- **Informações Adicionais** - Resumo de todas as informações relevantes obtidas na etapa de pré-auditoria.

A Pré-Auditoria pode apresentar como resultado um relatório parcial acompanhado de sugestão de adiamento ou encerramento dos trabalhos, caso os riscos identificados já sejam de pleno conhecimento do gestor, desde que o auditor tenha convicção de que a continuidade dos trabalhos não agregaria valor significativo. Isso deve ser visto como um ganho da auditoria, uma vez que houve economia de recursos da empresa que seriam despendidos em um trabalho que pouco valor agregaria; ou porque os riscos já estavam minimizados ou porque os ajustes necessários já foram tomados na fase de pré-auditoria.

Dificuldades na Realização da Pré-Auditoria

A pré-auditoria ainda é extremamente complexa devido ao grande volume de dados e a dificuldade em localizar as informações desejadas. Imagine uma empresa, que realiza por ano mais de dois mil trabalhos completos de auditoria, totalizando, portanto, aproximadamente 10 mil relatórios de auditoria em 5 anos. Se essa unidade estiver adequadamente estruturada e informatizada, poderá obter rapidamente informações administrativas de cada uma dessas atividades como custo, prazo, responsáveis, etc., mas dificilmente localizará com rapidez as auditorias com as características A, B ou C (semelhantes a um determinado caso) quando necessário. Isso pouparia tempo em atividades como levantamento de dados organizacionais, confecção de planejamento, preparação de instrumental, etc. Por exemplo, em uma fraude de cartões de crédito ocorrido em uma agência bancária do interior de Goiás pode-se ter acesso rápido a todos os relatórios de auditoria emitidos nos últimos anos sobre fraudes ocorridas em cartões de crédito (ou produtos relacionados) em dependências com as mesmas características e analisarmos os procedimentos que os auditores adotaram para elucidar os fatos?

Essas características tornam o trabalho de auditoria um campo propício ao desenvolvimento de novos SAD. Qualquer avanço com relação à localização de características semelhantes e de experiências anteriores será muito útil para a tomada de decisão quanto a continuidade ou não e a forma de organização de rotinas de trabalho que poderão economizar recursos significativos para as

empresas, uma vez que a estrutura de auditoria interna e o seu custo de manutenção é sempre muito alto devido a especialização de seu pessoal, além de minimizar os riscos na decisão tomada.

2.7 A IMPORTÂNCIA DE UM SE UTILIZANDO RBC PARA A AUDITORIA

A opção pela utilização de um SAD em auditoria já é uma realidade para as grandes empresas da área. Segundo YANG & VASARHELYI (1994) são usados principalmente para detecção de fraudes e transações suspeitas. Nesse artigo pode-se analisar uma extensa lista de SE utilizados em auditoria, com indicação de seu nome, área de domínio, além da referência para localização de maiores detalhes, caso desejado. Pode-se também verificar as duas categorias que os SE podem ser classificados quanto a sua finalidade:

- Usados para suportar o próprio processo de auditoria nas atividades de pré-auditoria, planejamento, emissão de relatório, etc.;
- Usados como ferramenta de execução de auditoria, (SE para detecção de contas suspeitas na contabilidade, fraudes em cartões de crédito, etc)

A proposta apresentada neste trabalho se encaixa na primeira categoria, uma vez que focará o processo de pré-auditoria e seus resultados.

Além da certeza de que a utilização dos SE, em trabalhos de auditoria, já é uma realidade, é importante frisar que os resultados alcançados têm sido positivos. Não há dúvida, portanto, de que um SE bem estruturado pode facilitar muito o trabalho de um auditor, assim como ocorre em todo tipo de trabalho complexo e com regras pouco estruturadas como é a auditoria.

Essa segunda característica, entretanto, restringe um pouco a aplicação de algumas técnicas de SE conforme estudado anteriormente. Não é conveniente investir-se muitas horas de especialistas e recursos financeiros em uma tabela de regras, por exemplo, que só atenderá a 2 ou 3 % dos casos analisados ou que estará desatualizada em poucos meses de uso.

Após a análise das características de SAD, do RBC e de todo o processo de auditoria interna com seus avanços e limitações atuais, fica claro que já

existem tecnologias na área que possibilitam um enorme avanço no trabalho de auditoria a um custo proporcionalmente baixo. A maioria, entretanto, utiliza a tecnologia de SE baseados em regras, com raras exceções (como por exemplo, o estudo apresentado por DIZ *et al.* (2002)). Como já foi discutido anteriormente, isso tem se mostrado um grande dificultador para a ampliação da abrangência desses sistemas nos trabalhos de auditoria. Por esse motivo, será apresentada uma proposta para o desenvolvimento de um SAD para o trabalho de auditoria interna, utilizando a técnica de RBC.

3 METODOLOGIA DE PESQUISA

As metodologias de pesquisas utilizadas nesse trabalho visam viabilizar o processo de desenvolvimento da pesquisa cientificamente. Além disso, essas metodologias devem sustentar os resultados alcançados, uma vez que procedimentos e instrumentais inovadores estão sendo avaliados e muitas situações que ainda não foram analisadas na prática devem ser estudadas e mensuradas.

3.1 PROCEDIMENTOS CIENTÍFICOS

A pesquisa é, segundo DEMO (1991), “um questionamento sistemático crítico e criativo, mais a intervenção competente na realidade, ou o diálogo crítico permanente com a realidade em sentido teórico e prático”. Entretanto, todo trabalho de pesquisa formal deve seguir metodologias adequadas. Para GIL (1999), a pesquisa tem um caráter pragmático, é um “processo formal e sistemático de desenvolvimento do método científico. O objetivo fundamental da pesquisa é descobrir respostas para problemas mediante o emprego de procedimentos científicos”.

3.2 CLASSIFICAÇÕES DA PESQUISA

Analisado como um todo, esse projeto apresenta algumas características que são determinantes na seleção das técnicas de pesquisa aqui utilizadas. Diversos autores apresentam diferentes classificações de pesquisa, mas alguns tipos já podem ser considerados como aceitos na área de pesquisa científica. Baseado no trabalho de SILVA & MENEZES (2001), que em alguns momentos utiliza classificações apresentadas por GIL (1999) alguns aspectos foram adotados nessa pesquisa:

- **Quanto a sua natureza** - esse projeto pode ser classificado como uma Pesquisa Aplicada, uma vez que os conhecimentos gerados visam aplicações práticas, envolvendo interesses e situações reais.
- **Quanto à forma de abordagem do problema** - a pesquisa apresenta uma parte Qualitativa, já que diversas avaliações são baseadas em

critérios subjetivos, sem tradução em números. Entretanto, é buscado sempre que possível, a realização de avaliações Quantitativas, mesmo que em alguns momentos seja necessário um exercício de conversão de valores qualitativos em quantitativos.

- **Quanto aos seus objetivos** - o projeto utiliza em parte a pesquisa Exploratória, utilizando um amplo levantamento bibliográfico, análise de pesquisas similares, entrevistas com pessoas vinculadas aos assuntos e um estudo de caso. É também utilizada a pesquisa Explicativa no momento em que se busca explicar o porquê das coisas, trabalhando com as formas de Pesquisa Experimental e Pesquisa Expost-facto.

Do ponto de vista dos procedimentos técnicos, são utilizados:

- **Pesquisa Bibliográfica** - através de análises e fundamentos elaborados a partir de material já publicado, constituído principalmente de livros clássicos e atuais, artigos de periódicos, com foco especial nos artigos disponibilizados pela *ACM - Association for Computing Machinery* e pela *IEEE - Institute of Electrical and Electronics Engineers*, além de materiais disponibilizados em *sites* considerados referências em suas respectivas áreas, na Internet;
- **Pesquisa Experimental** - determinado o objeto de estudo, são selecionadas variáveis discriminantes, as formas de controle e de observação dos efeitos que a variável produz no objeto, para seleção dos itens de resultados mais favoráveis;
- **Levantamento** - no momento em que a pesquisa envolve o levantamento de dados diretamente com as pessoas envolvidas com o trabalho de auditoria interna, principalmente no processo de engenharia de requisitos e na composição da base de casos;
- **Estudo de caso** - através da realização de um estudo profundo e exaustivo do sistema proposto (através de um protótipo) de maneira que se permita a sua ampla e detalhada análise. O estudo de caso é fundamental para confirmar ou desmistificar questões levantadas na pesquisa bibliográfica e no levantamento, além de fundamentar as conclusões finais do trabalho;

- **Pesquisa Expost-Facto** - pela realização de análises após o desenvolvimento de um protótipo, utilizando a ferramenta de apoio ao desenvolvimento de sistemas selecionada durante o projeto, buscando comparar os resultados obtidos com os resultados anteriores;
- **Pesquisa Participante** - por gerar uma ampla e constante interação entre o pesquisador e os membros de uma equipe de auditoria interna. A exploração dessa técnica pretende evitar interpretações inadequadas entre os conceitos (principalmente de auditoria) e a prática, além de envolver os usuários finais desde o desenvolvimento do projeto.

3.3 MÉTODOS CIENTÍFICOS

Segundo GIL (1999), a investigação científica depende de um “conjunto de procedimentos intelectuais e técnicos” para que seus objetivos sejam atingidos: os métodos científicos. Para SILVA & MENEZES (2001), método científico é o conjunto de processos ou operações mentais que se devem empregar na investigação.

As bases lógicas necessárias para uma investigação podem ser: dedutiva, indutiva, hipotético-dedutiva, dialética e fenomenológico (GIL, 1999). Nesse trabalho as bases lógicas que pautam os métodos adotados são:

- **Método dialético** - por ser de interpretação dinâmica e principalmente por considerar que fatos não podem ser encarados fora de um contexto social, político e econômico. É adequado também para pesquisas qualitativas (GIL, 1999);
- **Método dedutivo** - por apresentar uma cadeia de raciocínio em ordem descendente, do geral para o particular, até chegar à conclusão.

3.4 ANÁLISE E DEFINIÇÃO DO MODELO DE SIMILARIDADE ADOTADO

Em sistemas RBC, a seleção de um modelo de similaridade entre dois objetos, é usado para determinar a proximidade entre uma pergunta e um caso armazenado e com base nessa proximidade, retornar os casos mais adequados a pesquisa realizada. Existem diversas técnicas como *Nearest neighbor*, *Nearest neighbor* ponderado, Distância Euclidiana, Distância Euclidiana ponderada, Distância de Manhattan, dentre outros. para realização

da análise e definição do modelo mais indicado, adotou-se um processo sistematizado, que compreendem várias etapas que podem ser detalhadas da seguinte forma:

- Escolha dos modelos a serem analisados;
- Revisão de literatura;
- Avaliação das características de cada modelo;
- Definição das características relevantes para o sistema a ser desenvolvido;
- Determinação de testes a serem aplicados;
- Definição de metodologia e realização dos testes;
- Coleta de dados;
- Tabulação de dados;
- Análise e discussão dos resultados;
- Conclusão da análise dos resultados;
- Seleção do modelo a ser adotado.

3.5 ANÁLISE E DEFINIÇÃO DA TÉCNICA DE RECUPERAÇÃO DE CASOS ADOTADA

Em sistemas RBC, a recuperação de casos pode ser realizada de várias formas. Existem diversas técnicas como a Recuperação Seqüencial, Recuperação de Dois Níveis, Recuperação com Árvores, Redes de Recuperação de Casos, dentre outras. Para realizar a análise e definição da técnica mais indicada, diversas questões como a forma de representação de casos, a estruturação da base de casos e o modelo de similaridade têm que ser analisados. Para isso adotou-se nesse trabalho um processo sistematizado, que compreende diversas etapas que podem ser descritas da seguinte forma:

- Escolha das técnicas a serem analisadas;
- Revisão de literatura;
- Avaliação das características de cada técnica;
- Definição das características relevantes para o sistema a ser desenvolvido;
- Determinação de testes a serem aplicados;

- Definição de metodologia e realização dos testes;
- Coleta de dados;
- Tabulação de dados;
- Análise e discussão dos resultados;
- Conclusão da análise dos resultados;
- Seleção da técnica a ser adotada.

3.6 ANÁLISE E DEFINIÇÃO DA FERRAMENTA DE DESENVOLVIMENTO ADOTADA

Por existir diversas ferramentas de apoio ao desenvolvimento de sistemas foi necessário definir uma metodologia para seleção da ferramenta mais indicada para essa proposta. Assim, adotou-se nesse trabalho uma seqüência de pesquisas, análises e testes para certificar que a ferramenta selecionada apresente a melhor combinação de recursos e desempenho para as características do sistema desenvolvido. As etapas são:

- Pré-seleção, dentre as inúmeras ferramentas disponíveis, de três que apresentem em suas características explícitas, a melhor combinação de recursos e disponibilidade;
- Revisão de literatura sobre essas ferramentas pré-selecionadas;
- Definição das características relevantes para o sistema a ser desenvolvido;
- Avaliação das características de cada ferramenta;
- Determinação de testes a serem aplicados;
- Definição de metodologia e realização dos testes;
- Coleta de dados;
- Tabulação de dados;
- Análise e discussão dos resultados;
- Conclusão da análise dos resultados;
- Seleção da ferramenta a ser adotada.

3.7 ESTUDO DE CASO

3.7.1 JUSTIFICATIVA

A importância do desenvolvimento de um processo formal de investigação da proposta desse projeto está na necessidade de que o estudo das técnicas RBC e da sua aplicabilidade no auxílio à tomada de decisões para auditoria interna seja profundo e exaustivo, de maneira que se chegue a um amplo, detalhado e conclusivo resultado.

Basicamente, existem três tipos de técnicas de investigação:

- **Experiências formais** – pesquisas específicas em pequenos focos, com grande controle das variáveis que podem afetar seus resultados;
- **Estudos de caso** – focado em um caso específico, com o objetivo de obter informações que possam ser generalizadas para todos os casos semelhantes. O controle sobre as variáveis já não é tão grande;
- **Pesquisa/visão abrangente** – estudo retrospectivo de uma situação, procurando relações entre os resultados. Baseia-se em grandes levantamentos onde o controle sobre as variáveis fica prejudicado.

Nesse projeto são utilizadas técnicas de investigação de estudo de caso, baseadas na proposta de FENTON & PFLEEGER (1997). Essa técnica de investigação foi escolhida pela sua adequabilidade ao cenário previsto, conforme podemos verificar na tabela abaixo:

Quadro 5: Fatores relacionadas a escolha da técnica de pesquisa

Fator	Experiências formais	Estudo de Caso
Nível de controle	Alto	Baixo
Dificuldade de controle	Baixa	Alta
Facilidade de replicação	Alta	Baixa
Custo de replicação	Baixo	Alto

Fonte: FENTON & PFLEEGER (1997)

Essas características discriminantes justificam a escolha do estudo de caso uma vez que o controle dos trabalhos de auditoria pode ser considerado complexo quando se necessita de muitas replicações, devido à diversidade de características existentes em cada trabalho. Além disso, o custo da ampliação do número de experiências tornar-se um dificultador devido à remuneração dos funcionários especializados envolvidos e o prazo relativamente longo de realização/conclusão dos processos.

3.7.2 PLANEJAMENTO E DESENVOLVIMENTO DO ESTUDO DE CASO

Todo estudo de caso deve ter as seguintes etapas: concepção, definição, preparação, execução, análise, disseminação e tomada de decisão/conclusão (FENTON & PFLEEGER, 1997), portanto, todas essas etapas foram realizadas, seguindo as técnicas mais adequadas, visando o atingimento de resultados confiáveis. As amostras utilizadas nos testes, por exemplo, seguiram técnicas probabilísticas, mantendo a segurança contra a falta de confiabilidade que uma seleção não intencional pode representar. Nos casos onde a observação foi a técnica adotada, buscou-se sempre a adoção da observação sistemática por permitir um planejamento, e realizou-se em condições controladas (como registro dos resultados, com testes de confirmação) para responder aos propósitos preestabelecidos. Quando se utilizou entrevista para obtenção de informações junto a usuários, elas foram realizadas através de entrevistas estruturadas, com perguntas abertas e fechadas.

Ainda segundo FENTON & PFLEEGER (1997), o estudo de caso pode ser organizado de três maneiras:

- **Baseado em um projeto “irmão”** – seleciona-se um projeto similar ao que será testado, propiciando uma comparação entre os resultados obtidos com as diferentes técnicas;
- **Baseado em uma linha “mestra”** – define-se um conjunto de valores (baseados em dados de vários projetos) como um “projeto padrão”, para servir como referência;

- **Baseado em uma seleção aleatória** – divisão do projeto em partes, onde uma parte adota a nova técnica estuda e a outra não. Não se trata de uma experiência formal, uma vez que o projeto não foi selecionado através de uma amostra aleatória.

Para esse estudo de caso foi adotada a organização baseada em um projeto “irmão”, visando uma comparação ente os resultados obtidos.

Com relação ao desenvolvimento do protótipo propriamente dito, o trabalho foi feito dentro de métodos consagrados de engenharia de requisitos, gerenciamento de desenvolvimento de projetos e validação e verificação de resultados. Pequenas adaptações foram necessárias devido às características específicas do RBC, mas evitou-se ao máximo a falta de método no trabalho. Para coleta de dados necessária na alimentação do BC do protótipo, foram utilizados casos reais de auditoria, apenas com alterações dos valores e nomes envolvidos, devido à questão do sigilo bancário. As simulações também envolveram situações reais do trabalho de um auditor interno, buscando assim, obter resultados que realmente representem a aplicabilidade (ou não) de cada característica ou função do protótipo desenvolvido.

3.7.3 ANÁLISE E CONCLUSÃO DO ESTUDO DE CASO

Os resultados obtidos foram tabulados e apresentados de forma a facilitar a compreensão dos resultados. Foi tomado o cuidado com dados discrepantes, que possam vir a distorcer o resultado total do estudo. Somente após todas essas etapas é que foi possível analisar as propostas e discutir os resultados.

O desenvolvimento desse protótipo baseado na realidade dos trabalhos de auditoria apresenta como facilitadores a experiência e o trabalho atual do autor, que permitiu uma ampla e imediata avaliação das propostas, além de ampliar a quantidade de opiniões de especialistas no assunto foco do sistema (auditoria interna), pelo livre acesso do autor a esse ambiente de trabalho.

O protótipo desenvolvido é utilizado como apoio a tomada de decisões na fase de pré-auditoria. Portanto, ele deve conter uma interface amigável para

possibilitar a utilização por parte de todos os auditores, com conhecimento de informática ou não, e deve trabalhar com dados existente na base de dados atual do sistema transaccional existente. Outra característica importante é que ele tenha flexibilidade para possibilitar a pesquisa de trabalhos de auditoria com poucas ou com até 10 informações, sem apresentar resultados com mais de cinco anos, devido a desatualização que eles apresentariam.

3.8 ANÁLISE DOS RESULTADOS

Na análise dos resultados buscou-se uma conclusão abrangente da viabilidade ou não da adoção do RBC em SE utilizados no apoio a decisões para o trabalho de auditoria interna. Foi preciso considerar todas as variáveis envolvidas, uma vez que o investimento de qualquer empresa para o desenvolvimento de um SE não pode ser tomado sem um nível de confiabilidade muito elevado.

Além das conclusões inicialmente desejadas, inúmeras outras questões relacionadas à pesquisa surgiram durante os trabalhos e não poderiam ser ignoradas. Dessa forma, esses pontos foram registrados como aspectos importantes para futuras pesquisas.

4 ANÁLISE E DEFINIÇÃO DO MODELO DE SIMILARIDADE ADOTADO

Uma das decisões mais importantes e ao mesmo tempo complexas do planejamento e desenvolvimento de uma aplicação RBC está na escolha do modelo de determinação das similaridades global e local (LEE, 1998). Provavelmente, devido a essa relevância e também pela diversidade de situações e características abrangidas pelos sistemas RBC, existem inúmeros modelos de cálculo de similaridade. Como a escolha de um desses modelos é fundamental para o futuro sistema e sua troca após o início do desenvolvimento é extremamente complexa, a decisão deve ser tomada pelos responsáveis com todo embasamento e segurança possíveis.

Apesar da importância e complexidade da escolha, não foi encontrado em pesquisas realizadas durante esse trabalho, um método com critérios objetivos de escolha amplamente aceito pela comunidade de desenvolvimento em RBC. Muitas vezes a escolha do modelo de similaridade é feita com base em: experiências anteriores e facilidade de uso e/ou conhecimento do responsável.

Nessa pesquisa, um dos objetivos é escolher através de análise técnica, o melhor método de cálculo de similaridade para um SAD para auditoria interna utilizando o RBC. Para isso, foi adotado um processo sistematizado detalhado a seguir:

- Revisão bibliográfica;
- Escolha dos modelos a serem analisados;
- Avaliação das características de cada modelo;
- Definição das características relevantes para o sistema;
- Determinação de testes a serem aplicados;
- Definição de metodologia e realização dos testes;
- Coleta e tabulação de dados;
- Análise e discussão dos resultados;
- Conclusão e seleção do modelo a ser adotado.

4.1 REVISÃO BIBLIOGRÁFICA - SIMILARIDADE

Similaridade pode ser definida como um conceito intuitivo que descreve a percepção de um observador humano da semelhança existente entre dois itens (TURNER, 1989). É utilizado, portanto para indicar o quão semelhantes ou “parecidos” dois registros (que podem ser compostos por campos em um banco de dados ou imagens e sons, por exemplo) são entre si. Para que a similaridade seja mensurada, é necessária a adoção de uma métrica de similaridade, que é a apuração através de uma função do grau de similaridade entre dois casos (LEE, 1998). A medida de similaridade seria, portanto o resultado de um julgamento de semelhança entre casos, através de um modelo matemático (WANGENHEIM & WANGENHEIM, 2003). A apuração dessa medida de similaridade é uma das fases mais importantes e cruciais para a eficiência do RBC como um todo, segundo KOSLOSKY (1999).

Os métodos de recuperação para apuração de similaridade podem ser divididos em classificatórios, eliminatórios e numéricos. Nos métodos classificatórios os casos são divididos em categorias, facilitando sua recuperação ao restringir as buscas àquela categoria. Nos métodos eliminatórios são adotadas restrições, que fazem com que a cada passo, o número de casos analisados diminua consideravelmente. Os modelos para determinação de similaridade baseada no método numérico medem o nível de similaridade entre o item analisado e os itens armazenados no BC, atribuindo um resultado numérico com base em modelos de cálculos matemáticos que apuram a proximidade ou distância entre dois casos. Além dos diversos modelos de cálculo, existem também vários princípios que normalmente definidos em conjunto com o modelo de cálculo escolhido, fazendo com que os resultados se alterem de acordo com o princípio adotado.

4.2 ESCOLHA E AVALIAÇÃO DAS CARACTERÍSTICAS DAS MÉTRICAS DE SIMILARIDADE

Para que a similaridade seja mensurada, é necessária a adoção de uma métrica de similaridade, que é a apuração através de uma função, do grau de similaridade entre valores cujas naturezas são determinantes e relevantes para os casos (LEE, 1998). De acordo com o resultado obtido, é possível distinguir

se um caso apresenta similaridade com o outro, mesmo que não se tenham muitos casos para servirem de comparação. A similaridade entre os casos é definida através de uma função que apresenta resultados entre 0 (casos sem similaridade) e 1 (casos idênticos). Pode-se apurar a similaridade de forma local (calculando-se a similaridade existente entre cada uma das informações ou campos dos casos) ou de forma global (resultado único da similaridade entre os casos). Outra tarefa que pode ser necessária é a atribuição de pesos às características, que afetará diretamente os resultados das métricas ponderadas (MICHIE *et al.*, 1994). Segundo AAMODT & PLAZA (1994), a avaliação da similaridade pode ser sintática (onde os atributos são comparados pela semelhança de seus conteúdos) ou semântica (onde a avaliação é mais profunda, abrangendo o significado de cada um deles). Um exemplo simples pode ser o caso de verificação da similaridade entre duas palavras – concerto e conserto. A comparação sintática (principalmente nos casos onde a comparação é feita pela taxa de *substring* comum que será apresentada a seguir) resultaria em um índice de similaridade elevado (87,5%) enquanto uma apuração semântica levaria em conta o sentido da palavra no texto em que ela estiver inserida, o que resultaria provavelmente em um percentual de similaridade muito pequeno ou até inexistente, uma vez que o significado de concerto pode ser uma apresentação de uma orquestra enquanto a palavra conserto poderia estar relacionada à correção de algum problema. Apesar de tão diferentes, tanto a identificação da similaridade sintática quanto da semântica, podem ser realizadas de forma automatizada (DEKANG, 1998) sendo, entretanto mais difícil se automatizar a identificação da similaridade semântica.

Outra distinção ocorre entre as medidas cardinais e ordinais (JÚLIO, 2005). As medidas cardinais apresentam resultados numéricos expressando, por exemplo, uma medida de distância. As medidas ordinais adotam a escala ordinal que permite apenas a aplicação das relações: maior que (>), menor que (<) ou igual a (=) sobre seus valores. As classificações, escores, rankings, são exemplos de escalas ordinais, mesmo que sejam expressas através de números e geralmente precisam de tabelas referenciais para serem

automatizadas em sistemas RBC. Entre as métricas de apuração de similaridade de casos cardinais, as mais conhecidas e utilizadas são os algoritmos de vizinhança (WATSON & MARIR, 1994). A seguir serão apresentadas algumas dessas métricas com uma breve descrição de suas características e da forma de cálculo.

- **Coeficiente de casamento simples ou contagem de características (*Features Count*)** – apura-se o número de fatores (ou campos) coincidentes entre a consulta e o caso armazenado. Os campos correspondentes e os divergentes são avaliados com a mesma representatividade, não havendo diferenciação de peso entre eles. Uma das principais limitações desse modelo é que qualquer diferença existente entre dois registros (mesmo que seja uma letra) faz com que o resultado da similaridade local deixe de ser 100% e passe a ser 0%. A representação matemática mais usual é a seguinte:

$$\text{Similaridade entre (N,C)} = \frac{\sum \text{Similares}}{\sum \text{Casos}}$$

- **Coeficiente de casamento ponderado** – apura-se o número de fatores (ou campos) coincidentes entre a consulta e o caso armazenado, atribuindo-se maior valor (peso) à correspondência ou à divergência. Apresenta a mesma limitação descrita no Coeficiente de Casamento Simples. A representação matemática mais usual é a seguinte:

$$\text{Similaridade entre (N,C)} = \frac{\sum \text{Similares} * \text{Peso}}{\sum \text{Casos} * \text{Peso}}$$

- **Vizinho mais próximo** – Atualmente é uma das métricas mais estudadas e utilizadas no RBC, também chamado *nearest neighbor* (GREENSPAN *et al.*, 2001) Trabalha com a idéia de que os casos podem ser vistos como pontos no espaço, localizados de acordo com um ou mais índices e onde o caso mais similar é o mais próximo. Exige a definição de um valor que identifique essa posição no espaço e após essa definição, a apuração dos resultados passa a ser muito simples e intuitivo. O cálculo é realizado através da

apuração da distância entre os dois pontos (das representações gráficas do caso armazenado e do avaliado).

$$\text{Similaridade entre (N,C)} = \sum_{i=1}^n \phi(N_i, C_i)$$

- **Vizinho mais próximo ponderado** – Trabalha com o mesmo conceito do *nearest neighbor*, só que também considera a importância de cada informação no cálculo, ponderando os índices.

$$\text{Similaridade entre (N,C)} = \sum_{i=1}^n \phi(N_i, C_i) * w_i$$

- **Distância Euclidiana** – é outra metodologia que trabalha com a distância geométrica real entre dois pontos em um espaço qualquer (WANGENHEIM & WANGENHEIM, 2003). É muito utilizada e trata da distância em um espaço multidimensional (JÚLIO, 2005).

$$\text{Similaridade entre (N,C)} = \sqrt{\sum_{i=1}^n (n_i - c_i)^2}$$

- **Distância de Manhattan (Métrica do quarteirão)** – é uma medida que pondera todas as diferenças de forma idêntica. É conhecido como métrica do quarteirão (*city block*), pois em um modelo urbano com ruas perpendiculares, corresponde ao menor caminho. Diminui os efeitos dos *outliers*. A representação matemática é apresentada a seguir.

$$\text{Similaridade entre (N,C)} = \sum_{i=1}^n |n_i - c_i|$$

- **Distância Chebychev** – é uma medida recomendada para casos onde é necessário trabalhar com um grau máximo de diferença entre dois objetos, em qualquer dimensão (JÚLIO, 2005)

$$\text{Similaridade entre (N,C)} = \text{Máximo } |N_i - C_i|$$

No cálculo de similaridade devem ser analisados e definidos também alguns princípios que serão adotados. Essa combinação de metodologias de apuração de similaridade e diversos princípios ampliam enormemente a quantidade de combinações disponíveis para testes, fazendo com que a adoção de uma metodologia de apoio à escolha do melhor método de similaridade apresente milhares de cenários a serem analisados através da realização de simulações e testes. Alguns dos princípios normalmente considerados são:

- **Medidas de similaridade assimétrica** – Para valores assimétricos dois casos positivos são simétricos, enquanto dois casos negativos não o são necessariamente. Como exemplo, imagine um campo lógico que informe se o caso teve origem no Banco Central. Caso os dois registros estejam marcados, pode-se afirmar sem dúvida que os campos são similares. Já no caso deles não estarem marcados deve-se analisar se a similaridade existe, pois cada caso pode ter tido uma origem diferente, apesar de nenhum dos dois vir do Banco Central. Por isso, pode-se definir se os campos negativos (onde os campos lógicos não estão marcados) verificados na consulta e na base são considerados similares ou não e caso positivo, qual o percentual de similaridade será atribuído a eles.
- **Consideração de dados incompletos** – Quando alguns dos fatores podem estar incompletos, é necessário que se defina uma estratégia de trabalho para que se possa trabalhar com essas incompletezias, sem que isso prejudique os resultados alcançados. Pode-se trabalhar apenas com fatores completos tanto no caso armazenado como no caso consulta, mas isso pode reduzir em muito a precisão da pesquisa. Pode-se trabalhar com todos os fatores completos do caso consulta, atribuindo a não similaridade quando os casos da base estiverem incompletos. Ou então, podem-se considerar os fatores faltantes como similares, o que aumentará de uma maneira geral os valores de similaridades apontados.
- **Cálculo de similaridade entre campos texto** – Normalmente, na informática, um campo texto só é considerado igual a outro campo texto se todos os caracteres da string estiverem iguais. No RBC isso não é interessante, pois muitas vezes deve-se considerar palavras ou pedaços

de palavras iguais como similares, pelo menos parcialmente. Por isso, podem-se adotar princípios distintos para o cálculo de similaridade entre campos texto onde o resultado pode ser obtido de diversas formas:

- o **Correspondência exata** – A similaridade é calculada (100%) só para campos texto que apresente o mesmo conteúdo. Qualquer diferença entre os campos indica ausência de similaridade (0%).
- o **Correção ortográfica** – A similaridade é calculada com o resultado de caracteres coincidentes entre os dois campos (letra e posição).
- o **Contagem de palavras** - A similaridade é calculada com o resultado de palavras inteiras coincidentes entre os dois campos.
- o **Taxa de maior *substring* comum** - A similaridade é calculada com o resultado de partes de palavras coincidentes entre os campos.
- **Cálculo de similaridade entre campos numéricos** – O cálculo de similaridade entre dois campos numéricos é feito normalmente através de um cálculo da diferença entre eles. Esse cálculo, entretanto, pode ser realizado de diversas formas. Serão apresentadas as mais comuns (todas podem apresentar resultados entre 0 e 100%):
 - o **Função escada** – Só resulta em 100% se os itens forem idênticos ou a diferença entre eles estiver dentro de um limiar definido pelo usuário como aceitável. Qualquer outra situação resulta em 0%.
 - o **Função linear** – Calcula a similaridade considerando a distância entre dois valores e os intervalos máximos, que são definidos pelo usuário ou apurados entre o maior e o menor registro existente.
 - o **Função assintótica** – Calcula a similaridade considerando a distância entre dois valores e um gradiente definido pelo usuário. Esse gradiente pode fazer com que uma pequena variação nos dados cause uma mudança muito significativa na similaridade final.
 - o **Função escalar** - Calcula a similaridade através de uma função logarítmica. É interessante quando se deseja que a similaridade calculada entre dois valores pequenos seja menor que a calculada para dois números muito grandes, mesmo que a diferença nominal seja a mesma.

- **Consideração de dados não-binários** – A verificação da similaridade em muitos casos trabalha com valores diferentes do sim ou não (ou ativado/desativado). Nessas situações pode-se adotar um modelo simples de cálculo, que é feito através da divisão do número de fatores correspondentes pelo número total de fatores.

4.3 DEFINIÇÃO DAS CARACTERÍSTICAS RELEVANTES PARA O SISTEMA

Quando se analisa um SAD para uso em áreas de trabalho complexas e pouco estruturadas como a auditoria interna, a dificuldade já começa na própria seleção das características mais relevantes do sistema. Sabe-se que algumas generalizações podem ser feitas, mas dificilmente elas podem ser classificadas como fatores determinantes e excludentes, pois sempre podem surgir situações onde aquele fator indique um resultado oposto às conclusões iniciais.

No caso de SAD para auditoria interna utilizando o RBC, pelas suas características, pode-se imaginar (e até mesmo exemplificar) “n” casos onde cada uma das características que influenciam no cálculo da similaridade podem ser extremamente relevantes e “n” casos onde essa mesma característica possa ser considerada pouco determinante para o bom funcionamento. Por exemplo, para uma empresa que trabalha predominantemente com dados numéricos, muito mais relevantes do que o método de cálculo adotado serão os **parâmetros de cálculo de similaridade entre campos numéricos (função escada, linear, assintótica e escalar)**. Já no caso de auditorias com registros em campos texto livre, a escolha do método a ser adotado passa a ser extremamente relevante, já que o cálculo de similaridade do casamento simples, por exemplo, dificilmente atingiria os níveis de precisão esperados.

Essa situação mostra que há a necessidade de se adotar uma definição de relevância das características mais avançada, que possibilite uma definição de forma dinâmica, de acordo com cada realidade.

Para isso adotou-se um critério que busca trabalhar com as características mais relevantes a cada caso. Essa proposta, apesar de muito mais complexa, possibilitou uma precisão muito maior, uma vez que, em cada caso, todos os

princípios como medidas de similaridade assimétrica, consideração de dados incompletos, cálculo de similaridade entre campos texto e numéricos e consideração de dados não-binários são analisados de acordo com o interesse do analista. Além disso, é possível comparar os resultados de vários modelos de cálculo de similaridade de forma rápida, oferecendo segurança na tomada de decisão da escolha do melhor modelo, possibilitando uma apuração pragmática, sem definições baseadas em resultados passados.

4.4 DETERMINAÇÃO DE TESTES A SEREM APLICADOS.

Devido à complexidade e diversidade de situações já citadas, torna-se necessário à adoção de testes com critérios muito abrangentes. Como primeira condição, surge a necessidade de que os testes avaliem uma quantidade considerável de modelos de cálculos, uma vez que, o número de modelos existentes é muito grande. Deve-se viabilizar testes com um número entre três e sete modelos, sendo o ideal é que se avaliem cinco modelos.

Outra questão relevante é a adoção de princípios como medidas de similaridade assimétrica, consideração de dados incompletos, cálculo de similaridade entre campos texto e numéricos e consideração de dados não-binários. Mesmo com a multiplicidade de combinações de princípios possíveis, deve-se viabilizar testes que considere todas as combinações possíveis, não se descartando a adoção de nenhum deles.

Finalmente, como o RBC trabalha com a idéia de aproveitar as informações de casos já registrados, é importante que o modelo de cálculo de similaridade escolhido reproduza, da forma mais próxima do possível, os resultados que os especialistas indicam como sendo os casos mais similares. Para isso, devemos comparar os resultados obtidos nos testes aplicados ao BC, aos resultados indicados por especialistas para esse mesmo BC. Essa indicação dos especialistas reflete a similaridade real entre os casos analisados, indicada com base na experiência profissional dos mesmos, como se eles “calculassem” mentalmente o nível de similaridade que cada caso armazenado apresenta com àquele sob comparação. Esse conceito dos SE, permite que a metodologia utilize esses parâmetros como referência para a comparação das

metodologias, apontando, com altíssimo nível de precisão e segurança, a métrica de similaridade para RBC mais adequada para cada caso.

4.5 A METODOLOGIA PROPOSTA DE DEFINIÇÃO DA MELHOR MÉTRICA DE SIMILARIDADE PARA RBC

Conforme já discutido, a melhor solução para definição da métrica de similaridade é o desenvolvimento de uma nova metodologia. Isso, apesar de complexo, permite segurança e qualidade no trabalho, visto que elimina a maioria dos problemas detalhados anteriormente como dificuldade de análise de todas as combinações possíveis e de obtenção da opinião de especialistas. Além disso, há a possibilidade de que essa metodologia sirva de apoio para outras situações onde essa decisão tenha que ser tomada, além dessa pesquisa.

Para criação da metodologia chamada de Definição da Melhor Métrica de Similaridade para RBC (DMMS-RBC), partiu-se de conceitos simples e já amplamente utilizados no RBC. O critério adotado baseia-se no conceito de busca do método de cálculo de similaridade que apresenta resultados mais similares ao resultado ideal, previamente definido pelo(s) especialista(s). Ela se aplica a quaisquer métricas de apuração de similaridade local e global dos métodos numéricos sintáticos ou semânticos, dos mais simples aos mais complexos. Assim busca-se atender ao maior número de usuários e de necessidades.

São realizadas diversas etapas que através da adoção da simulação de resultados podem fornecer um resultado objetivo para auxiliar a decisão de definição do método de cálculo de similaridade para cada caso. As etapas são:

- I. Definição dos princípios que serão adotados nos cálculos que serão realizados para escolha do melhor método;
- II. Seleção ou definição de um conjunto de casos aleatórios para testes. Apesar de não haver um número mínimo de casos, é evidente que uma base mais extensa proporcionará resultados mais consistentes, apesar do aumento da dificuldade de realização de cálculos manuais;

- III. Seleção ou definição de um caso X para ser usado como base da consulta do cálculo de similaridade;
- IV. Definição pelos especialistas da ordem de similaridade de todo o conjunto de casos ou em parte dele em relação ao caso X definido para consulta-teste;
- V. Definição pelos especialistas dos percentuais de similaridade de cada caso em relação ao caso X definido para consulta-teste;
- VI. Apuração da ordem de similaridade de cada um dos casos armazenados na base de casos em relação ao caso X, definido para consulta-teste com a metodologia A. Esse procedimento deve ser realizado para todas as metodologias desejadas, armazenando-se os resultados de cada caso em cada metodologia para posterior comparação.
- VII. Apuração do percentual de similaridade de cada um dos casos armazenados na base de casos em relação ao caso X, definido para teste com a metodologia A. Esse procedimento deve ser repetido para todas as metodologias desejadas, armazenando-se os resultados de cada caso em cada metodologia para posterior comparação.
- VIII. Finalizadas as apurações, deve-se comparar os resultados de ordem e de similaridade obtidos através de cada metodologia ao resultado ideal apontado pelo(s) especialista(s). Essa etapa é realizada em duas fases. Inicialmente, com todos os resultados apurados, realiza-se mais um cálculo de similaridade, comparando-se dessa vez a classificação apontada como a provável pelo especialista e as classificações verificadas em cada uma das metodologias. Verifica-se nessa fase qual é o método que mais se aproximou do resultado ideal com relação a classificação dos casos; A segunda fase é semelhante a descrita anteriormente, alterando apenas o foco da comparação que passa a ser feita entre o percentual de similaridade mais provável apontado pelos especialistas e os percentuais calculados em cada uma das metodologias testadas. Da mesma forma, verifica-se no final qual é o método que mais se aproximou do resultado ideal com relação ao percentual de similaridade dos casos.
- IX. Os dois resultados (dos percentuais e das classificações reais frente à sugerida pelos especialistas) podem ser coincidentes ou não. Como eles

são apenas referencias para a decisão do analista, a metodologia não define qual dos resultados terá maior ou menor relevância e conseqüentemente, qual dos “vencedores” deverá ser adotado.

Essa metodologia, apesar de simples, apresenta inúmeras vantagens quando comparada a qualquer outra forma de escolha de métricas para cálculo de similaridade, sendo a eliminação da subjetividade a principal delas.

4.6 O APLICATIVO DMMS-RBC – DEFINIÇÃO DA MELHOR MÉTRICA DE SIMILARIDADE PARA RBC

Apesar da simplicidade lógica da metodologia proposta, sua aplicabilidade manual não se mostrou viável pelos seguintes motivos:

- Como a metodologia trabalha com comparações de casos, é interessante que o maior número de casos possível sejam definidos/registrados no BC para melhorar a qualidade da análise realizada. Esse aumento de registros, entretanto é inversamente proporcional à facilidade de apuração manual dos resultados, o que ampliaria o tempo necessário para realização dos cálculos;
- Mesmo não sendo muito complexos, os cálculos necessários são numerosos e detalhados, podendo causar erros de difícil detecção;
- A grande quantidade de princípios existentes no cálculo de similaridade fazem com que o número de combinações de configurações possíveis para simulações/testes atinja quantidades que praticamente inviabilizem os trabalhos sem o auxílio de um sistema baseado em computador;
- Da mesma forma que o volume de casos, o número de metodologias analisadas deve ser o maior possível, aumentando a complexidade.

Portanto, para agilizar os testes e a análise das inúmeras combinações possíveis na metodologia, foi desenvolvido um aplicativo DMMS-RBC – Definição do Melhor Método de Similaridade para Raciocínio Baseado em Casos - em linguagem Visual Basic® 6.0, utilizando a ferramenta de desenvolvimento de sistemas GAS 2003 Enterprise®. O sistema viabiliza a realização relativamente rápida e simples de testes completos com inúmeras

combinações de modelos de cálculo e princípios possíveis sendo flexível, permitindo a introdução de novos modelos e princípios sem muita dificuldade, além da revisão de pesos e paradigmas adotados. Essa flexibilidade, somada aos recursos e facilitadores existentes no aplicativo, permitem que qualquer pessoa, com um mínimo conhecimento sobre RBC, possa definir o melhor método de cálculo de similaridade e justificar essa escolha.



Figura 13:Tela principal do aplicativo DMMS-RBC, versão 1.0.1

Todos os campos que devem ser utilizados nos casos podem ser ativados ou desativados, ter seu nome informado, trabalhar com peso, ser validado ou não, dentre outros parâmetros. Os princípios Medidas de Similaridade Assimétrica, Consideração de Dados Incompletos, Metodologia de Cálculo de similaridade entre campos texto, numéricos e datas, etc. foram implementados como recursos que podem ser ativados ou desativados pelo usuário.

Nessa primeira versão, cinco métricas estão disponibilizadas para comparações no sistema DMMS-RBC. Tratam-se do Coeficiente de Casamento Simples e Ponderado, Vizinho Mais Próximo Simples e Ponderado e Distância Euclidiana. Cada métrica pode ser ativada/desativada de acordo com a necessidade do usuário. A operacionalização do sistema é simples e permite a realização de centenas de testes/análises diferentes em questão de minutos, pela facilidade de alterações nos parâmetros e pela possibilidade de recálculo com manutenção de históricos para comparações de resultados. Além disso, diversos recursos como construção de consultas, gráficos e relatórios pelo próprio usuário fazem com que o aplicativo seja totalmente

- Legenda:**
- A = Considera similar campos completos tanto na consulta quanto na base
 - B = Considera similar campos completos na consulta e incompletos na base
 - C = Considera similar campos incompletos na consulta e completos na base
 - D = Considera similar campos incompletos na consulta e incompletos na base
 - E = Considerar similar quando ambos os campos lógicos estiverem desativados
 - F = Calculo de campo texto = Correspondência exata
 - G = Calculo de campo texto = Correção ortográfica
 - H = Calculo de campo texto = Contagem de palavras
 - I = Cálculo de campo texto = Taxa da maior *substring* comum*
 - J = Calculo de campo n°. e data = Função escada
 - K = Calculo de campo n°. e data = Função linear
 - L = Calculo de campo n°. e data = Função assintótica
 - M = Cálculo de campo n°. e data = Função escalar*
 - N = Campo texto n°. 3
 - O = Campo texto n°. 7
 - P = Campo lógico n°. 2
 - Q = Campo numérico n°. 4
 - R = Campo tabela interna n°. 1
 - S = Campo data n°. 1
 - T = Campo data n°. 2
- * Parâmetros não adotados por não se aplicarem aos trabalhos de auditoria

Os resultados dos cálculos apresentados no quadro a seguir demonstram um aspecto importante. A metodologia DMMS-RBC deve ser usada preferencialmente para comparar a classificação obtida pelas métricas em relação à fornecida pelo especialista. Isso porque, segundo opinião dos próprios especialistas é muito complexo indicar um percentual de similaridade para cada caso apenas no “sentimento” de similaridade (ou seja, o conhecimento do especialista). Já a ordenação é uma coisa mais intuitiva, apesar de também se tornar complexa quando o número de casos supera uma ou duas dezenas. Por esse motivo, os resultados relacionados à classificação (R1 e C1) foram mais relevantes na escolha da melhor métrica do que os relacionados ao percentual (R2 e C2).

Quadro 7: Resumo dos resultados obtidos em cada cálculo realizado

Nº. Cálculo	Metodologia																			
	Casamento Simples				Casamento Simples Ponderado				Vizinho Mais Próximo				Vizinho Mais Próximo Ponderado				Distância Euclidiana			
	R1	C1	R2	C2	R1	C1	R2	C2	R1	C1	R2	C2	R1	C1	R2	C2	R1	C1	R2	C2
01	67,0730	5	73,1600	1	77,2922	4	66,0000	3	94,0954	2	59,6800	5	96,5909	1	70,5600	2	89,0628	3	59,8000	4
02	67,0730	5	73,1600	1	77,2922	4	66,0000	3	94,8996	2	62,3200	4	96,6686	1	70,0800	2	87,6812	3	60,5200	5
03	67,0730	5	73,1600	1	77,2922	3	66,0000	2	70,7988	4	60,3600	4	80,9049	2	62,1200	3	91,3206	1	59,8800	5
04	67,0730	5	73,1600	1	77,2922	4	66,0000	2	93,1671	2	59,7600	5	96,3090	1	62,9200	3	88,8041	3	61,7600	4
05	67,0730	5	73,1600	1	77,2922	4	66,0000	2	93,5003	2	58,4400	5	96,3634	1	62,4800	3	88,6194	3	61,6400	4
06	63,4576	5	71,1200	3	73,6768	4	71,3600	2	90,9778	2	67,2400	5	94,6520	1	73,7200	1	90,6959	3	67,3600	4
07	63,4576	5	71,1200	3	73,6768	4	71,3600	2	93,1048	2	69,2000	4	95,9695	1	74,8400	1	89,8220	3	65,0800	5
08	68,6883	5	69,4800	2	77,3691	4	66,0800	3	94,9439	2	62,3200	4	96,5917	1	70,0000	1	87,6388	3	60,6000	5
09	61,3037	5	72,6400	1	72,1845	4	56,4400	2	90,0852	2	38,3200	5	92,7351	1	49,6400	3	88,3294	3	43,8000	4
10	71,1663	5	66,0000	1	85,2607	3	58,0400	2	94,3745	1	31,9200	5	88,0580	2	39,6000	4	83,5789	4	41,0800	3

Legenda:

R1 = Resultado em percentual da similaridade da classificação obtida pela metodologia em comparação com a classificação indicada pelo especialista.

C1 = Classificação obtida pela metodologia na comparação da classificação obtida pela metodologia em comparação com a classificação indicada pelo especialista.

R2 = Resultado em percentual da similaridade do percentual obtido pela metodologia em comparação com o percentual indicado pelo especialista.

C2 = Classificação obtida pela metodologia na comparação do percentual obtido pela metodologia em comparação com o percentual indicado pelo especialista.

Para o cálculo de similaridade realizado entre casos de auditorias, a importância dos campos texto indica a adoção do princípio de contagem de palavras e com função linear para números e datas. Essa configuração de parâmetros indicou a metodologia do Vizinho Mais Próximo Ponderado como a mais apropriada tanto na comparação feita entre os percentuais de similaridade calculados x apontados pelo especialista, quanto na classificação (ranking) calculada x apontada pelo especialista. Na primeira comparação o percentual obtido pela métrica do vizinho mais próximo ponderado foi 73,72% (enquanto todas as outras quatro métricas apresentaram resultados inferiores a 72%) e na segunda o resultado foi 94,69592 (enquanto todas as outras quatro métricas apresentaram resultados inferiores a 91%), conforme gráfico a seguir.

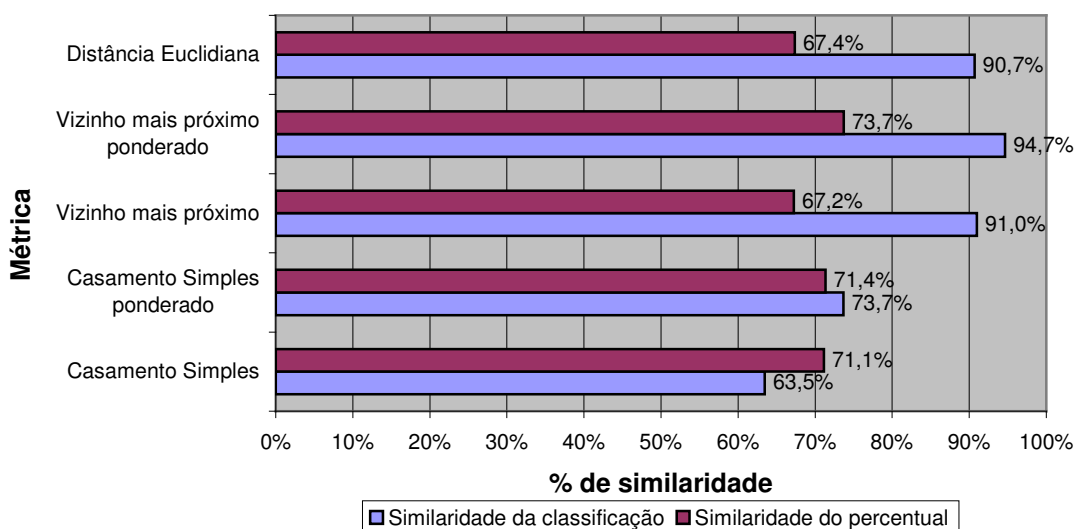


Figura 14: Similaridade das métricas obtida pela metodologia DMMS-RBC

OBS: O detalhamento dos resultados do cálculo que reproduziu a parametrização ideal para o trabalho de auditoria está no anexo 11.1.

Esses resultados não se alteram significativamente mesmo quando os princípios de Similaridade entre dados incompletos são alterados de *Não considerar similar os casos onde o campo da consulta está completo e o da base incompleto* para *Considerar similar os casos onde o campo da consulta*

está completo e o da base incompleto, já que nos dados de auditoria utilizados praticamente todos os campos sempre estão preenchidos. Se considerarmos a comparação entre as classificações apontadas pelo especialista e os calculados pelo aplicativo, a técnica que apresentou os melhores resultados continua sendo a do Vizinho Mais Próximo Ponderado com um resultado final de 96,59175%, enquanto na comparação dos percentuais o resultado da melhor métrica (também a do Vizinho Mais Próximo Ponderado) foi de 70,00%.

Nos cálculos de nº. 01 a 05, 09 e 10, os resultados apresentados na comparação dos percentuais calculados e apontados indicam a métrica do Casamento Simples como a mais adequada, mas esse resultado (já esperado) só ocorreu, pela dificuldade de apontamento pelo especialista do percentual de similaridade de cada caso. Além disso, no cálculo nº. 09 o princípio de Consideração de dados incompletos foi alterado para uma situação diferente da realidade da auditoria e no cálculo nº. 10, vários campos relevantes foram desativados. Essas situações, mesmo não refletindo as características da auditoria, foram simuladas para demonstrar que em situações muito parecidas, a adoção de uma métrica pode ser indicada e não aquela que sempre vinha sendo utilizada, comprovando a importância da adoção da DMMS-RBC.

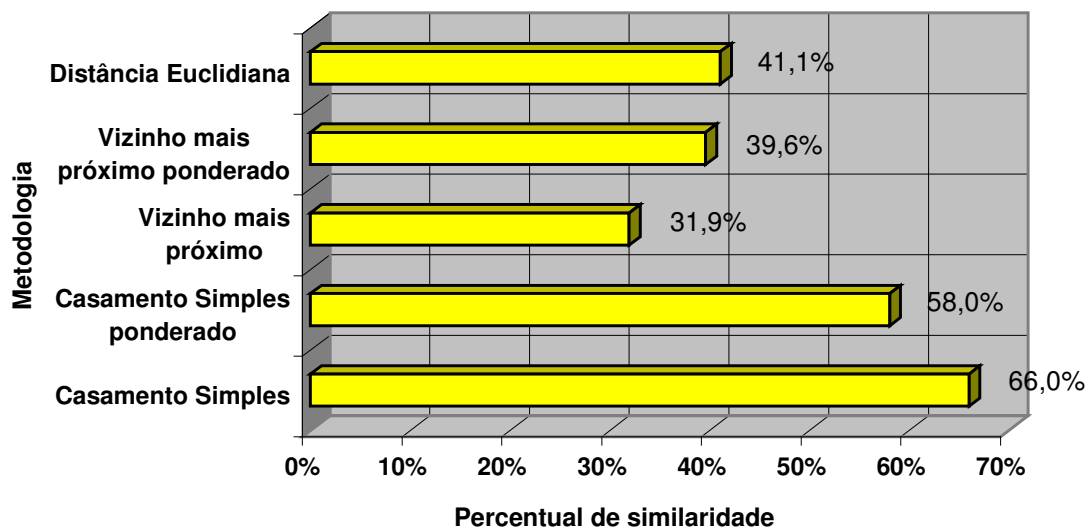


Figura 15: Similaridade dos percentuais obtida pela metodologia e a indicada pelo especialista no cálculo nº. 10

4.8 CONCLUSÕES E DEFINIÇÃO DO MÉTODO DE SIMILARIDADE A SER ADOTADO

A adoção da metodologia de DMMS-RBC permitiu uma escolha embasada e com um grande número de simulações e testes. As métricas apontadas não são necessariamente aquelas mais simples ou as mais adotadas, o que já demonstra que a metodologia ajuda a impedir uma decisão equivocada. Isso somado a facilidade de utilização (com o sistema) indicam que a adoção da metodologia é de grande valia para o processo de seleção do método.

Sendo assim, sabe-se com segurança que a adoção do modelo de cálculo de similaridade indicados pela métrica do Vizinho Mais Próximo Ponderado é adequada para as características dos casos de auditoria que compõem a base de casos do SAD para auditoria interna com RBC que o trabalho propõe.

Pode-se dar continuidade às pesquisas através da realização de testes maiores, com maior número de casos e principalmente, com a utilização em outros tipos de métricas, com características distintas. Além disso, uma nova versão do aplicativo DMMS-RBC pode possibilitar a adoção de novos recursos

como tabelas de contingências, símbolos ordenados e não ordenados, além da utilização de um número maior de métricas. Isso será extremamente interessante, devido a grande quantidade de métricas existentes e a diversidade de características entre elas. Exemplos de outras métricas que poderão ser usadas: Jaccard, Sorensen-Dirce, Anderberg, Ochiai I e II, Rogers e Tanimoto, Russel e Rao (MEYER, 2002), Chebychev, Distância Percentual, Distância Forte (FERNANDES, 2003), Distância de Hamming, Euclidiana Ponderada (WANGENHEIM & WANGENHEIM, 2003), ADG Action-Distance Based (TONIDANDEL E RILLO, 2003) além de várias extensões destes.

A disponibilização do aplicativo DMMS-RBC na internet para análise, possibilitará que, tanto a metodologia como o aplicativo sejam aperfeiçoados, permitindo a continuidade no desenvolvimento da técnica.

5 ANÁLISE E DEFINIÇÃO DO MODELO DE RECUPERAÇÃO DE CASOS

Outra decisão extremamente relevante no planejamento e desenvolvimento de um sistema RBC é escolher o melhor método de recuperação de casos. A análise técnica é fundamental, para que essa decisão no caso de um SAD para auditoria interna utilizando o RBC seja adequada. O objetivo da recuperação é localizar dentre os inúmeros casos armazenados, aqueles mais similares, iniciando-se com a descrição do problema e terminando com a escolha (ou recuperação) do melhor caso (REIS, 1997).

Pode-se subdividir esse processo nas seguintes fases:

- **Identificação das características:** registro no sistema das características do caso a ser pesquisado;
- **Busca dos casos mais similares:** levantamento com base nas características informadas no passo anterior dos casos mais similares;
- **Refinamento da busca:** processo mais elaborado de seleção que permite um aprofundamento dos cálculos de similaridade, aplicado apenas nos casos selecionados na etapa de busca. Essa fase só ocorre em modelos de recuperação de casos que prevêem esse refinamento, adotados quando o volume de casos armazenados no sistema é grande;
- **Seleção:** nessa fase, após a ordenação por similaridade dos casos de acordo com a métrica ou algum critério de classificação escolhido, o caso mais adequado para comparação como a pesquisa é selecionado.

Assim como no cálculo da similaridade, existem várias formas de se realizar a recuperação de casos em sistema RBC, apesar de todas elas seguirem as fases anteriormente descritas. Essa diversidade de opções está relacionada principalmente a busca dos melhores resultados possíveis em cada tipo de sistema RBC. Não há um método que possa ser considerado o melhor, pois dependendo das características, os resultados da mesma técnica podem ser totalmente diferentes. Por esse motivo, a definição do modelo a ser adotado passa a ser uma questão de extrema importância no trabalho a ser

desenvolvido. Para essa escolha, usou-se um processo sistematizado compreendido de várias etapas que podem ser detalhadas da seguinte forma:

- Revisão bibliográfica;
- Definição das características relevantes para o sistema;
- Avaliação das características de cada modelo;
- Seleção dos modelos a serem analisados;
- Determinação de testes a serem aplicados;
- Definição de metodologia e realização dos testes;
- Coleta de dados e tabulação de dados;
- Análise dos resultados e seleção do modelo a ser adotado.

5.1 REVISÃO BIBLIOGRÁFICA

5.1.1 INTRODUÇÃO

O objetivo da recuperação de casos é encontrar um caso ou um pequeno conjunto de casos na base de casos que contenha uma solução útil para o problema ou situação atual (WANGENHEIM & WANGENHEIM, 2003).

Segundo (KOSLOSKY, 1999), três fatores são fundamentais na recuperação:

- I. **Eficiência:** é a velocidade com que um sistema recupera casos;
- II. **Precisão:** é o grau que os casos recuperados podem ser usados para alcançar as metas propostas;
- III. **Flexibilidade:** é o grau de recuperação de casos para raciocínios inesperados.

Nem sempre se consegue conciliar esses três fatores facilmente, uma vez que as grandes bases de casos, apesar de serem adequadas para que os melhores resultados sejam localizados, são problemáticas quanto ao aspecto de desempenho de recuperação de casos nos sistema RBC e flexibilidade. Cabe ao analista responsável pelo sistema RBC encontrar o ponto de equilíbrio que possibilite a obtenção dos melhores resultados.

5.1.2 IDENTIFICAÇÃO DE CARACTERÍSTICAS

Todas as fases são importantes para o sucesso do processo de recuperação, mas pode-se apontar a fase de identificação das características como a mais complexa (WANGENHEIM & WANGENHEIM, 2003), pois:

- Tem utilização intensa do conhecimento, o que é sabidamente complexo devido às características subjetivas do assunto;
- Exige, em alguns casos, interação intensa e ativa com o usuário;
- Existe diversidade de características entre os casos armazenados na BC e o caso a ser usado como base de consulta;
- Há dificuldade na definição das características determinantes para a localização do caso mais similar;
- As características que podem ser utilizadas para localização do caso mais similar são diversas e complexas;
- Existe a possibilidade de identificação parcial ou incorreta das características por desconhecimento do usuário ou mesmo por definição do sistema com foco na tecnologia e não no usuário (CARDOSO, 1999).
- Existe a necessidade de adoção de levantamento de dados dirigido, onde a resposta a uma pergunta inicial é que determinará os próximos passos da identificação das características.

5.1.3 BUSCA DE CASOS

É considerada menos complexa, pois é uma fase automatizada, não necessitando de intervenção do usuário. Há, entretanto, a necessidade de estudo e planejamento apurado dos procedimentos adotados, pois a simples adoção de uma busca seqüencial comparativa pode não se tornar viável devido ao desempenho resultante, da mesma forma que, uma busca composta por duas ou mais etapas de busca, pode se mostrar por demais complexa. Os tipos de busca mais utilizados e suas principais características são:

5.1.3.1 Recuperação Seqüencial

Realizada através da leitura e análise da similaridade de todos os casos armazenados na base de casos. Após essa apuração total, os casos são

ordenados em forma de ranking e os casos mais similares ficam disponíveis para a fase seguinte de seleção. É o tipo de busca de maior aplicação por ser extremamente simples de ser implementado, além das seguintes vantagens:

- É o tipo de busca comprovadamente mais completo e correto, pois é aplicado em 100% dos casos;
- Elimina a importância, para efeito da recuperação de casos, da indexação da BC, uma vez que todos os casos serão analisados;
- Permite a adição de novos casos à base de forma simples;
- Permite consultas instantâneas, independentes da similaridade;
- Algoritmo de busca pode ser simples.

Mas o tipo de busca seqüencial não apresenta apenas vantagens e até mesmo por isso, não é o único adotado. Seu principal aspecto negativo está no desempenho com grandes BC, por efetuar cálculos para todos os registros, independentemente dele apresentar muita ou nenhuma similaridade com o caso pesquisado. Outras desvantagens do método de busca são:

- Esforço de recuperação máximo e constante, independente da complexidade da consulta e do número de casos a serem recuperados;
- Pode selecionar casos pouco adequados, dependendo da precisão do algoritmo adotado;
- Inadequada com a identificação de características dirigida, já que a vantagem de redução de análises que as diversas fases do processo proporcionariam é perdida ao se analisar todo o BC.

5.1.3.2 Recuperação de Dois Níveis

Procura reduzir ao máximo o número de comparações de casos, com critérios adequadamente definidos, para manter a confiabilidade do processo. Isso é feito através da exclusão de casos cujas características que indique a inexistência de similaridade antes da realização de cálculos. A estruturação e indexação do BC são aspectos fundamentais para o funcionamento adequado da técnica, baseando-se fortemente nos métodos desenvolvidos para banco de dados relacionais. A recuperação de dois níveis é baseada no modelo *Many*

Are Called / Few are Chosen (MAC/FAC) (KLUWER & FORBUS, 1991) – muitos são chamados, poucos são escolhidos – motivado por estudos em ciências cognitivas. Ele trabalha da seguinte maneira:

- **Pré-seleção** – Nessa fase, associada à sigla MAC, muitos casos passam pelo primeiro critério de seleção por ele ser menos complexo. Normalmente, o critério de filtragem utilizado é o da comparação simples, onde algumas informações (geralmente campos) são comparadas e somente os casos onde essas informações são idênticas é que são mantidos. Isso permite uma eliminação de um número considerável de casos que, por não apresentarem quase nenhuma semelhança com a consulta formulada, podem ser descartados de imediato, sem a necessidade de uma avaliação mais apurada.
- **Ordenação dos casos pré-selecionados/seleção mais refinada** – Fase correspondente a sigla FAC, implica na ordenação daqueles casos que aparentemente possuem mais similaridade com o caso consulta e por esse motivo foram mantidos após a fase de pré-seleção. A ordenação é muito mais apurada e complexa, para possibilitar um nível de precisão adequado às necessidades de um sistema RBC. Nessa etapa, a recuperação é seqüenciada, e todos os casos pré-selecionados são analisados, independentemente da similaridade.

A recuperação de dois níveis pode ser apresentada através do modelo:

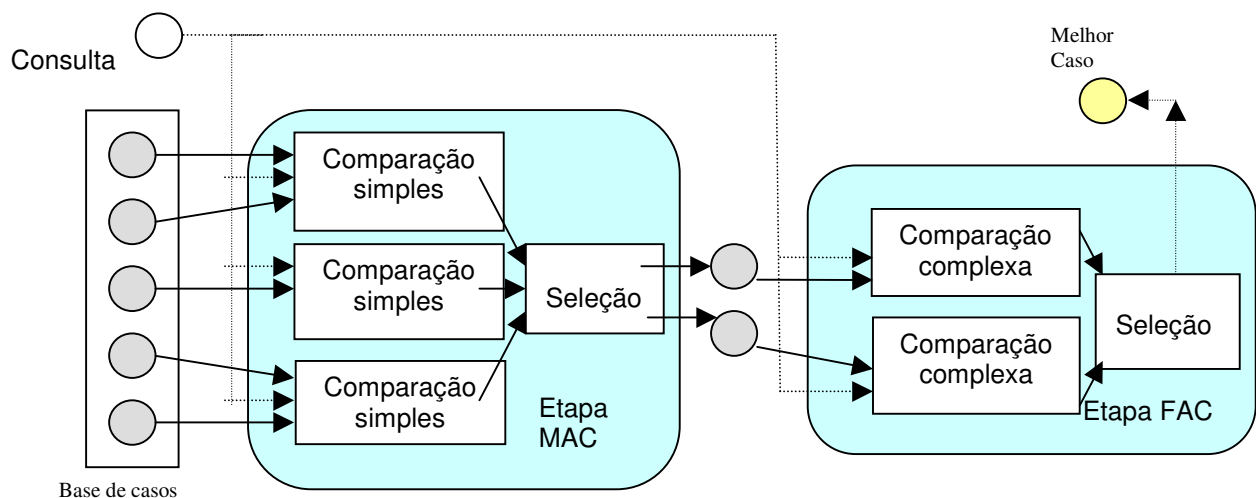


Figura 16: Modelo MAC/FAC (KLUWER & FORBUS, 1991)

Após o detalhamento do método, percebe-se que a recuperação de dois níveis é na verdade uma combinação de duas metodologias distintas, realizadas em conjunto justamente para aproveitamento das vantagens de cada uma. Isso possibilita uma melhora de performance, desde que o número de casos potencialmente similares (aqueles que passam pela pré-seleção) não seja grande e o volume de casos do BC seja grande.

A vantagem da melhoria de performance pode ser pouco interessante, entretanto, se ela vier acompanhada da ocorrência de possíveis erros da recuperação em dois níveis. Esses erros podem ser de dois tipos:

- Casos sem similaridade ou muito pouco similares podem “passar” pela fase de pré-seleção, só sendo eliminados na fase de ordenação dos casos pré-selecionados ou seleção mais refinada;
- Um caso similar pode não ser selecionado na fase de pré-seleção, ficando de fora do processo de recuperação de casos definitivamente.

Apesar de qualquer erro ser sempre indesejável, é evidente que o primeiro tipo é menos prejudicial que o segundo, pois só terá como consequência um aumento no tempo de processamento, sem prejudicar a precisão dos resultados apresentados. Já o segundo problema pode fazer com que, casos importantes não sejam avaliados, prejudicando irremediavelmente os resultados aplicados. Por esse motivo, é sempre mais prudente, utilizar critérios conservadores na fase de pré-seleção, somente eliminando casos que indiscutivelmente não tenham similaridade com a consulta formulada. Com esse intuito e devido à evolução de hardware, é comum hoje em dia, que os sistemas RBC já adotem critérios de apuração de similaridade na fase de pré-seleção, relativamente complexos, como igualdade ou similaridade local parcial, que apesar de reduzirem um pouco o desempenho do processo, minimizam a exposição aos riscos descritos anteriormente.

5.1.3.3 Recuperação Baseada em Índices

Baseia-se na utilização da indexação do BC para agilizar e possibilitar uma recuperação de casos eficiente, precisa e flexível. Parte da premissa que apenas os casos onde os índices sejam coincidentes devem ser analisados.

A recuperação baseada em índices se mostra extremamente eficiente em casos onde o BC armazena um volume de registros muito grande ou estejam armazenados em meios externos. Apresenta duas fases básicas: na primeira uma estrutura de índices adequada é gerada e na segunda fase ocorre à recuperação propriamente dita (WANGENHEIM & WANGENHEIM, 2003).

Para a condução da primeira fase pode-se adotar duas metodologias:

- **Aproveitamento dos índices já existentes no BC** (TANAKA & SUEDA, 1993). A escolha por esse método apresenta como vantagens principais à rapidez do processo e a facilidade de utilização. A desvantagem é que em algumas consultas, o índice existente pode não ter uma estrutura adequada e o resultado não ser adequado;
- **Geração de índices a cada processo de recuperação** (LENZ & BURKHARD, 1995). Esse método é vantajoso justamente por possibilitar a adoção dos índices mais adequados a cada caso, já que o índice é construído a cada processo de recuperação de acordo com as características daquela busca. Essa precisão obviamente gera como desvantagem a lentidão do processo.

Uma representação do ciclo de recuperação de casos orientada a índices com o aproveitamento de índices já existentes seria:

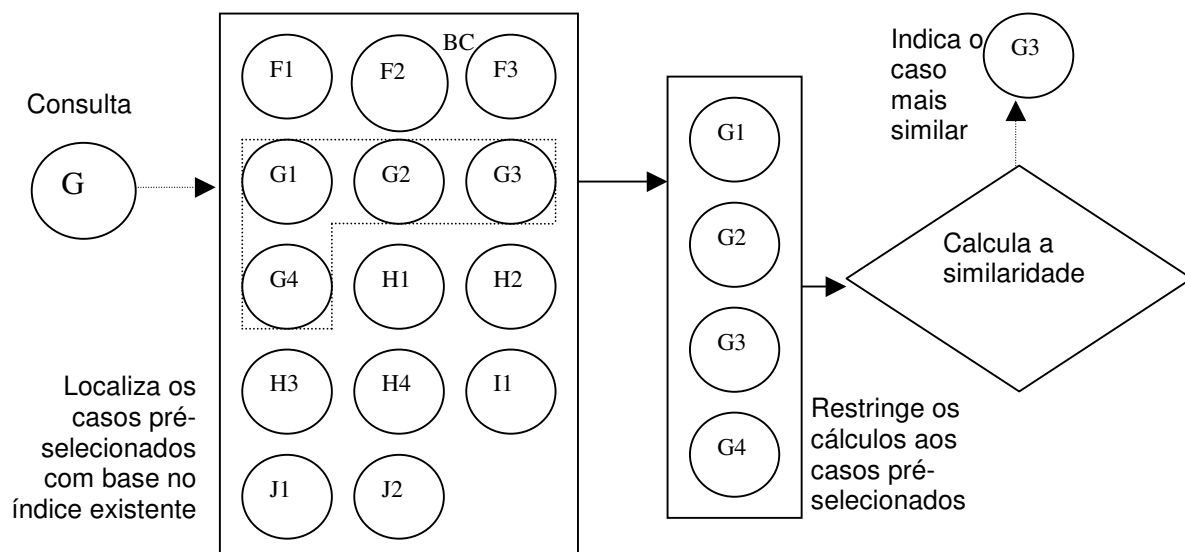


Figura 17: Recuperação de casos orientada a índices com uso de índices já existentes

Já a recuperação de casos com geração de índices a cada processo, é bem semelhante, acrescentando apenas o processo de criação ou reorganização do índice antes da localização dos casos pré-selecionados.

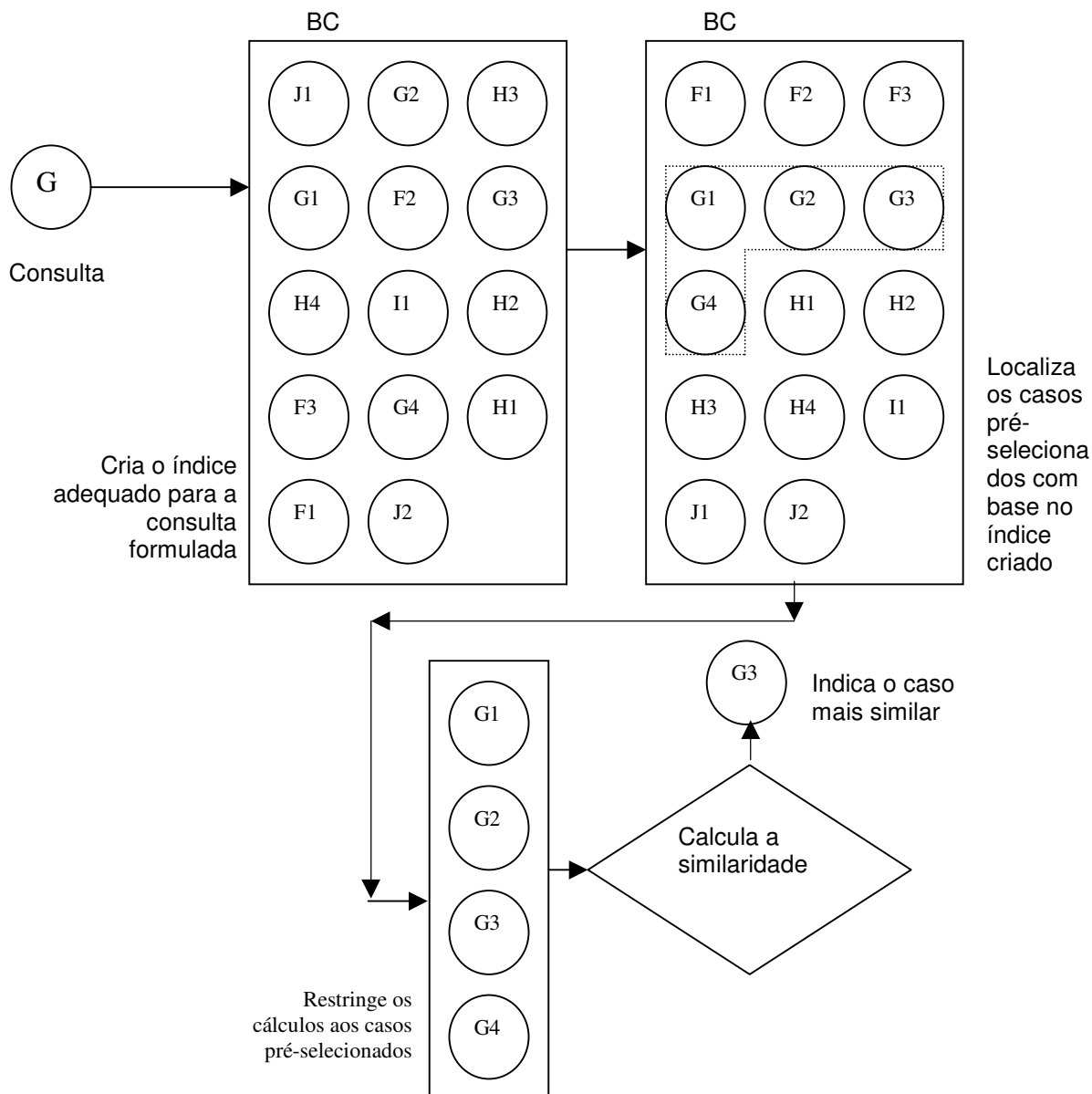


Figura 18: Recuperação de casos orientada a índices com a geração de índices

Quando uma BC é consultada freqüentemente pelos mesmos campos, pode-se adotar tanto o aproveitamento dos índices existentes como a geração de índices específicos. Isso porque na geração de índices específicos, caso os

últimos índices gerados sejam mantidos, essa freqüência verificada nas consultas fará com que o tempo perdido na construção dos índices seja limitada as primeiras consultas realizadas.

A recuperação orientada a índices também apresenta algumas desvantagens em relação a outros métodos de recuperação de casos adotados:

- **Custos elevados** – tanto pela necessidade de processamento constante como pelo aumento de espaço de armazenamento;
- **Esforço de implementação** – a análise e implementação de estrutura de índices, rotinas de consultas flexíveis e regras de consultas e algoritmos mais complexos, aumentam o trabalho de planejamento e implementação de um sistema RBC.

5.1.3.4 Recuperação Baseada em Rede de Feições Compartilhadas

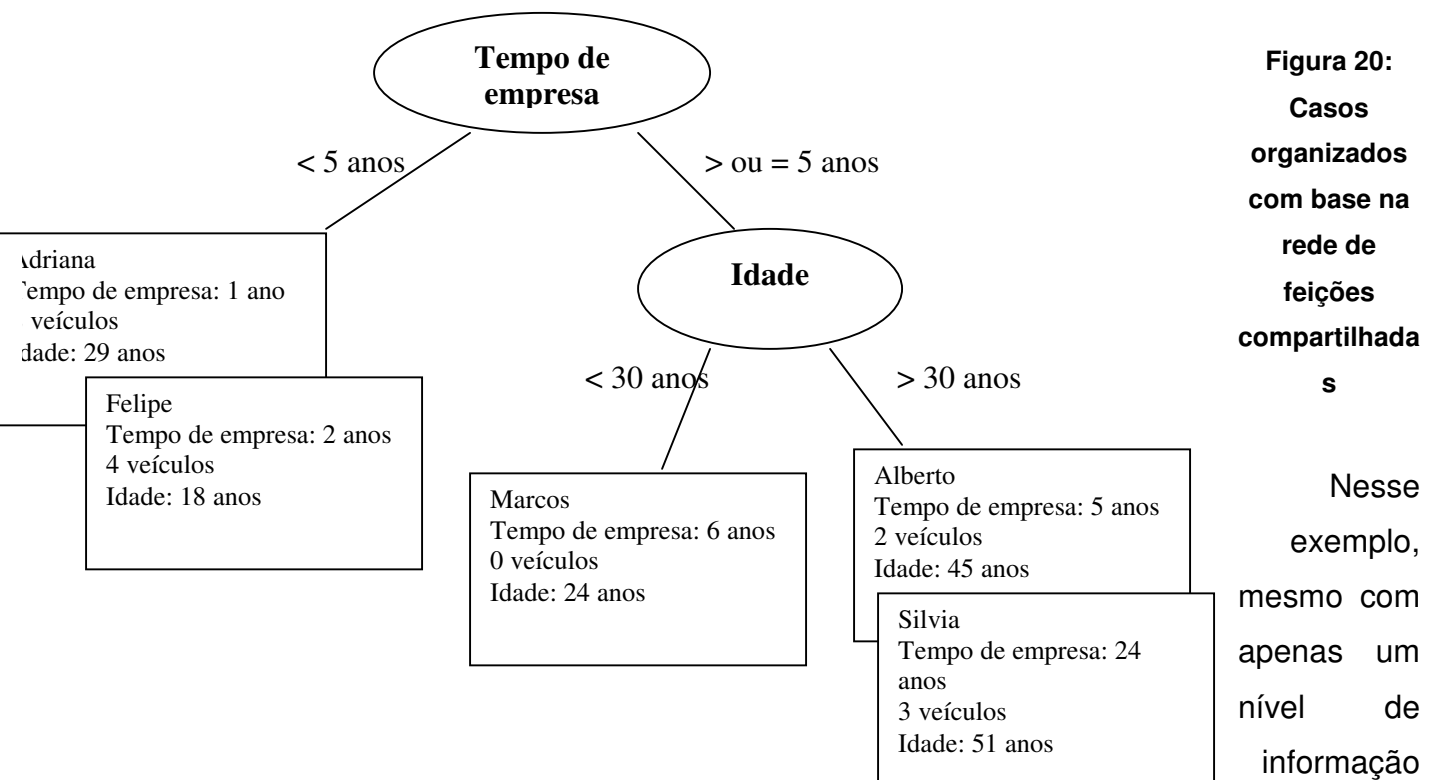
Nessa organização os casos que compartilham as características mais significativas para a recuperação de casos são organizados em grupos, com o objetivo de facilitar e agilizar as buscas. Ela otimiza a busca quando comparada à busca seqüencial, por exemplo, pois com a organização dos casos em grupos, grande parte do BC (que não faz parte desse grupo) não precisará ser comparada com o problema sob pesquisa, por ter baixa ou nenhuma possibilidade de apresentar qualquer semelhança.

Por exemplo:

Alberto Tempo de empresa: 5 anos 2 veículos Idade: 45 anos	Marcos Tempo de empresa: 6 anos 0 veículos Idade: 24 anos	Adriana Tempo de empresa: 1 ano 8 veículos Idade: 29 anos
Felipe Tempo de empresa: 2 anos 4 veículos Idade: 18 anos	Silvia Tempo de empresa: 24 anos 3 veículos	

Figura 19: Exemplos de casos com feições compartilhadas

Ao se trabalhar com a técnica de feições compartilhadas, pode-se organizar os casos anteriormente apresentados da seguinte maneira (se a característica mais relevante para a recuperação dos casos for, por exemplo, o tempo de trabalho na empresa):



discriminante, a utilização da rede de feições compartilhadas já possibilitou uma redução de pelo menos 40% nos registros a serem pesquisados. Por outro lado, a organização com base na rede de feições compartilhadas ocupa grande espaço na memória, aumentando a complexidade de tratamento na atualização da base de casos. Além disso, nem sempre é viável definir atributos significativos para todas as pesquisas possíveis, e caso essa definição não seja bem feita, casos similares podem não ser localizados e isso é um problema muito maior que a queda de performance do sistema.

5.1.3.5 Recuperação Baseada em Árvores de Discriminação

A idéia principal da recuperação com árvores é a de estruturar o espaço de busca com base na densidade observada e utilizar essa estrutura pré-computada para recuperação eficiente (WESS *et al.*, 1993). A lógica é muito semelhante a da rede de feições compartilhadas, distinguindo-se no fato que

as características mais significativas são utilizadas para obter a maior discriminação possível, tentando diminuir ao máximo, o número de casos pré-selecionados e não formando grupos de acordo com as feições compartilhadas. Isso possibilita que, onde o número de características discriminantes seja considerável, a rotina de recuperação localize casos bem específicos, pesquisando um percentual bem pequeno de casos da base.

A dificuldade está justamente na definição das características que sejam ao mesmo tempo discriminantes e compatíveis com os inúmeros tipos de pesquisas que possam ser formalizadas. Se bem definidos a rotina localizará apenas casos extremamente similares, o que é o ideal em qualquer sistema RBC. Entretanto, critérios muito focados podem fazer com que a rotina não localize nenhum caso similar (mesmo que eles existam), causando sérios riscos à imagem de confiabilidade do sistema. Na prática, devem-se definir as características mais discriminantes para a recuperação e colocá-las mais próxima à raiz, ou seja, nas primeiras condições de filtro. No caso, por exemplo, de uma recuperação de casos de defeitos em equipamentos de informática, o tipo de equipamento deve ser a primeira característica a ser analisada (ficando próxima à “raiz” da árvore). Em seguida, pode-se utilizar a descrição do defeito como segunda característica de análise.

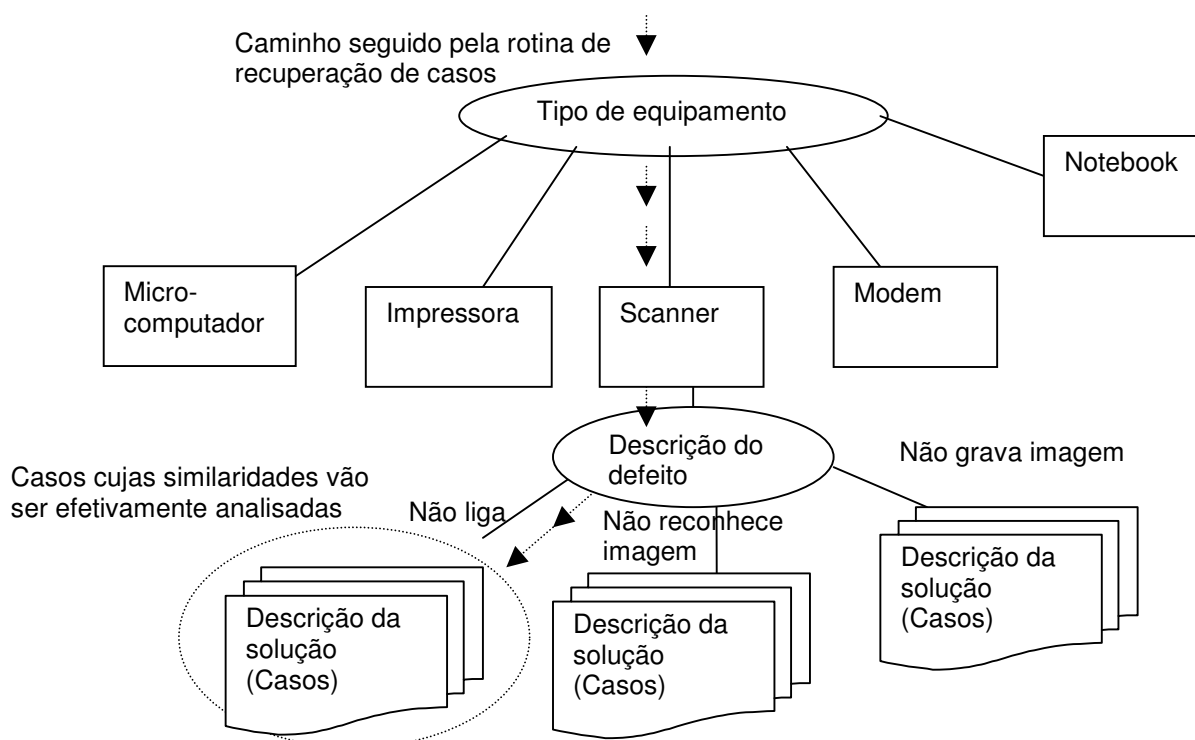


Figura 21: Recuperação baseada em árvores de discriminação

Esse é o conceito básico, já existindo diversas extensões de técnicas baseadas em árvores, que possibilitam a manipulação de situações mais complexas como atributos desconhecidos (MANAGO *et al.*, 1993) e desordenados e recuperação com testes de limites virtuais (WESS *et al.*, 1993).

5.1.4 SELEÇÃO DE CASOS

O sistema pode implementar o método de seleção usando várias formas: perguntando ao usuário, combinação de heurísticas ou aplicação de regras (COSTA, 1999). A maioria dos sistemas RBC utiliza um conjunto de heurísticas ou regras pré-definidas (ou parametrizáveis) para definir quais e quantos dos casos mais similares, devem ser selecionados. As regras mais usuais são:

- Número máximo de casos a serem recuperados;
- Percentual mínimo para que um caso seja recuperado (MANTARAS & PLAZA, 1995);
- Limite de data para recuperação de casos (casos muito antigos não são recuperados, independentemente do nível de similaridade).

A utilização da interatividade com o usuário é um recurso interessante quando se tem como usuários, pessoas experientes ou pelo menos com um nível de conhecimento equivalente para todo o grupo. Caso contrário, é melhor automatizar ao máximo a rotina de recuperação, evitando assim, dificuldades de uso e até entendimentos equivocados que possam fazer com que o sistema seja mal visto frente aos usuários.

5.1.5 ESFORÇO IDEAL DE RECUPERAÇÃO DE CASOS

Na análise para adoção do melhor método para recuperação de casos deve ser considerado o desempenho em função do número de casos previsto para a BC. O ideal seria optar pelo método mais seguro, para que o usuário tenha confiança no seu uso, mas como já foi discutido, isso pode ser inviável para certos sistemas. Por outro lado, deve-se considerar que o processo de

recuperação de casos é apenas uma das fases do ciclo de um sistema RBC e, portanto, não pode ser analisada isoladamente (VELOSO *et al.*, 1996).

5.2 DEFINIÇÃO DAS CARACTERÍSTICAS RELEVANTES PARA O SISTEMA A SER DESENVOLVIDO

A escolha da melhor técnica de recuperação de casos para um SAD para auditoria interna utilizando o RBC é muito complexa. Ele deve ser eficiente e flexível para as mais diversas formas de trabalhos existentes na auditoria. Outra dificuldade é a impossibilidade de limitação dos ambientes de trabalho desse sistema. A maioria das organizações que dispõem de equipes de auditoria interna são grandes o suficiente para pressupor a existência de uma estrutura tecnológica razoável, mas não se pode trabalhar com uma exigência excessiva de softwares e hardwares, sob o risco de inviabilizar ou desestimular a adoção e o uso do sistema proposto pelos usuários finais.

Assim, a palavra de ordem da definição das características relevantes para o sistema de apoio a decisão para auditoria interna utilizando o RBC será adaptabilidade. Evidentemente, essa flexibilidade pode dificultar o trabalho a ser realizado, mas o que se espera com essa estratégia é ampliar ao máximo o universo de possíveis usuários do sistema a ser desenvolvido.

Inicialmente, ao analisar os três fatores fundamentais na fase de recuperação destacados por KOSLOSKY (1999) – Eficiência, Precisão e Flexibilidade – já se podem definir alguns pontos:

- O sistema precisa ter um nível de precisão elevado. A auditoria interna trabalha com assuntos estratégicos para as organizações, envolvendo valores financeiros elevados ou mesmo questões éticas e políticas onde uma falha pode ter conseqüências muito graves;
- A eficiência é sempre desejável e relevante, mas nessa proposta ela pode ser considerada apenas como um fator desejável, uma vez que a forma de uso do SAD proposto prevê sua utilização em momentos específicos do trabalho de auditoria (principalmente no planejamento) o que torna a velocidade da recuperação de casos um aspecto menos

relevante, apesar de sempre almejada. Outro aspecto que torna essa variável secundária é a dimensão de um BC de auditoria que, salvo raras exceções, não deve superar os milhares de registros;

- A Flexibilidade será valorizada quando a rotina de adaptação de casos realmente for incorporada pelos usuários do sistema, sendo por isso uma característica fundamental no sistema a ser desenvolvido.

Na fase de identificação das características o SAD para auditoria interna baseado em RBC apresenta:

- Utilização intensa do conhecimento, o que é sabidamente complexo devido às características subjetivas do assunto;
- Dificuldade de definição das características determinantes para a localização do caso mais similar;

Para minimizar os problemas acima, uma característica fundamental na técnica de recuperação de casos adotada é que ela possibilite:

- A definição de características determinantes (que serão usadas na busca) tanto em pequeno número quanto em grandes quantidades;
- Levantamento de dados parciais (com características de preenchimento não obrigatório);

Assim, pode-se definir que os aspectos relevantes para a escolha da técnica de recuperação de casos para o sistema proposto são (em ordem de relevância):

- I. Precisão dos casos recuperados, evitando principalmente a ocorrência da falta de seleção de casos similares;
- II. Facilidade de uso pelos usuários (que apresentarão os mais diversos níveis de conhecimento do uso do aplicativo), evitando a necessidade de muitas intervenções humanas para a precisa recuperação de casos;
- III. Possibilidade de uso de qualquer quantidade de informações na fase de identificação de características;

- IV. Possibilidade de adaptabilidade na recuperação das mais diversas e diferentes estruturas de casos, já que o sistema terá que ser flexível para o uso em todos os tipos de trabalhos executados na auditoria;
- V. Desde que os aspectos anteriormente citados sejam atendidos, é desejável um desempenho razoável que permita a ampliação do uso do sistema, sem necessidade de grandes investimentos.

5.3 AVALIAÇÃO DAS CARACTERÍSTICAS DE CADA MODELO

Com as cinco características mais importantes definidas, pode-se avaliar através de uma comparação direta, cada uma das técnicas apresentadas. Essa pré-seleção indicou as três metodologias que participaram do teste completo.

Na precisão dos casos recuperados, é evidente que o método que apresenta a maior e incontestável precisão é a recuperação seqüencial, já que todos os casos são analisados, eliminando qualquer possibilidade de falha. As técnicas de recuperação em dois níveis, recuperação baseada em índices, recuperação baseada em rede de feições compartilhadas e recuperação baseada em árvores de discriminação, tem todas com certeza, uma precisão menor pela característica de análise parcial do BC, havendo necessidade de um levantamento mais profundo para que possam indicar os melhores resultados.

Na facilidade de uso pelos usuários, o número de intervenções é o fator determinante para a avaliação de cada técnica. Nesse aspecto, a recuperação em dois níveis aparentemente apresenta uma desvantagem na comparação com todas as outras técnicas pela exigência de duas ou mais intervenções pelos usuários a cada busca. Isso, entretanto pode ser pouco significativo se, por exemplo, a técnica de recuperação em dois níveis possibilitar a busca com a utilização de apenas duas ou três informações em cada nível (o que representaria um questionário de 4 a 6 respostas) enquanto uma recuperação seqüencial, por exemplo, apresentar um formulário com 60 campos.

No caso da quantidade de informações preenchidas pelos usuários, o que pode fazer a diferença para a classificação dessa característica é a obrigatoriedade

de preenchimento dos campos do caso a ser recuperado. As técnicas que dependem do preenchimento obrigatório de determinados campos para a realização de uma recuperação eficiente podem dificultar a utilização do SAD, pois a inexistência ou o desconhecimento daquela informação, pode fazer com que o usuário desista do uso. Nesse sentido, as técnicas de recuperação em dois níveis, recuperação baseada em índices, recuperação baseada em rede de feições compartilhadas e recuperação baseada em árvores de discriminação levam uma pequena desvantagem novamente, pois todas elas precisam sempre ter uma ou mais informações obrigatórias para:

- Passar ao segundo nível (recuperação em dois níveis);
- Efetuar a busca pelo índice (recuperação baseada em índices);
- Localizar o grupamento desejado (recuperação baseada em rede de feições compartilhadas);
- Definir o “caminho” a ser seguido pelos “galhos” da árvore (recuperação baseada em árvores de discriminação).

Já a técnica de recuperação seqüencial, por percorrer toda a base de casos, pode se basear em qualquer campo preenchido pelo usuário para a busca, tornando a interface bem mais amigável. É importante salientar que qualquer uma das técnicas sob análise (inclusive a busca seqüencial) pode ter campos de preenchimento obrigatório de acordo com a característica dos casos armazenados, o que eliminaria a diferença descrita.

As técnicas que reduzem a quantidade de casos a serem analisados através de pré-seleções ou agrupamentos (recuperação em dois níveis, recuperação baseada em índices, recuperação baseada em rede de feições compartilhadas e recuperação baseada em árvores de discriminação), podem ignorar muitas informações similares por terem excluído os casos na pré-seleção.

Para que o sistema possa ser adaptável para a recuperação das mais diversas e diferentes estruturas de casos em todos os tipos de trabalhos de auditoria, qualquer uma das técnicas que se baseiam em definição de índices, regras para grupamento ou exclusão apresentam menor flexibilidade do que o

método de recuperação seqüencial. Isso se deve a necessidade de definição dos campos que serão usados para a criação de índices e regras de exclusão ou agrupamento no momento do desenvolvimento do sistema, restringindo em muito a adaptabilidade desejável do aplicativo. É evidente que existem condições, através de técnicas de programação, de parametrizar o sistema de tal forma que esses índices e regras tornem-se variáveis, mas isso pode tornar o sistema extremamente complexo e lento.

O quinto aspecto é relacionado à necessidade de um desempenho adequado que possibilite a ampliação do uso sem necessidade de grandes investimentos. Esse aspecto indica certamente as técnicas com pré-seleções ou agrupamentos (recuperação em dois níveis, recuperação baseada em índices, recuperação baseada em rede de feições compartilhadas e recuperação baseada em árvores de discriminação) em ampla vantagem frente à recuperação seqüencial. Elas devem apresentar um desempenho mais eficiente conforme o número de casos do BC for crescendo. Essa vantagem, entretanto pode só se tornar significativa em termos de tempo de processamento a partir de um número de registros muito grande (dependendo do sistema e do banco de dados adotado, a partir de dezena de milhares de casos). Como o sistema se propõe a ser útil em qualquer trabalho de auditoria interna, fica difícil prever o número médio de casos que serão armazenados e qual o desempenho resultante de uma ou outra técnica. O mais correto seria a realização de testes comparativos entre a técnica de recuperação seqüencial e uma das técnicas com pré-seleção ou agrupamento, utilizando BC de diversos tamanhos, para que se possa verificar a partir de que volume o desempenho começa a ser afetado pela adoção de uma ou outra técnica. Um resumo das características de cada técnica, mostra o seguinte:

Quadro 8: Comparativo das características predominantes de cada técnica

	Seqüencial	Dois níveis	Índices	Grupamento	Árvores
Precisão	Alta	Razoável	Razoável	Razoável	Razoável
Facilidade de uso p/ usuário –intervenções	Alta	Média	Alta	Alta	Alta

	Seqüen- cial	Dois níveis	Índices	Grupa- mento	Árvores
Facilidade de uso p/ usuário –quant.de inf.	Baixa	Alta	Média	Média	Média
Quant.de informações obrigatórias	Alta	Média	Média	Média	Média
Flexibilidade	Alta	Média	Baixa	Baixa	Baixa
Desempenho	Baixo	Alto	Alto	Alto	Alto

Esses resultados indicam uma divisão acentuada entre as técnicas baseadas em pré-seleção e a técnica de recuperação seqüencial, não sendo suficiente para apontar uma das cinco técnicas como a melhor para o desenvolvimento de um SAD para auditoria interna baseado em RBC.

5.4 SELEÇÃO DOS MODELOS A SEREM ANALISADOS

Uma vez que não ocorre uma concentração de vantagens em uma única técnica de recuperação que possibilite sua escolha, deve-se definir, quais as técnicas mais adequadas para serem testadas exaustivamente. Essa seleção deve basear-se nas características avaliadas anteriormente (precisão, facilidades de uso, quantidade de informações obrigatórias, flexibilidade e desempenho). Pelos resultados classificados na tabela 1, pode-se dividir as técnicas em três níveis de resultados. O primeiro, onde está somente a técnica da recuperação seqüencial, apresenta resultados favoráveis em todos os quesitos, exceto na quantidade de informações obrigatórias (que pode ser minimizado no momento da definição do sistema) e no desempenho. O segundo inclui a recuperação em dois níveis que só é classificada como média em três características (facilidade de uso, quantidade de informações obrigatórias e flexibilidade). Por último, estão as três técnicas restantes que apresentam resultados semelhantes entre si, com análise desfavorável da flexibilidade e mediana em precisão, quantidade de informações e facilidade de uso. O ponto forte dessas técnicas está justamente no desempenho superior, mas o quanto essa vantagem é representativa, só pôde ser verificado através de testes com base de dados de diversos tamanhos.

Portanto, para a continuidade desses estudos, foram selecionadas três técnicas (uma de cada grupamento de resultados): Recuperação seqüencial, Recuperação em dois níveis e Recuperação em grupamento de feições. As três técnicas foram testadas com BC de diversos tamanhos para indicar a técnica mais adequada para o SAD para auditoria interna com RBC.

5.5 DETERMINAÇÃO DE TESTES A SEREM APLICADOS

O primeiro aspecto a ser observado é que os testes realizados deveriam, na medida do possível, apresentar as mesmas características presentes no SAD para auditoria interna com RBC desenvolvido. As principais características já foram detalhadas anteriormente e são as seguintes:

- Possibilidade de parametrização do sistema.
- Possibilidade de utilização por pessoas sem grande conhecimento em informática.
- Possibilidade de utilização para localização de qualquer tipo de caso de auditoria.

Dessa forma, é necessária a criação de uma rotina de testes que contemple:

- Definição e criação de um BC para casos de auditoria;
- Alimentação desse BC;
- Definição e criação de uma rotina de identificação das características que são utilizadas para a busca e recuperação;
- Criação de três rotinas de busca, sendo:
 - o Uma para recuperação seqüencial
 - o Uma para recuperação em dois níveis
 - o Uma para recuperação em grupamento de feições
- Recuperação dos casos mais similares em cada uma das três rotinas
- Análise dos resultados obtidos, considerando:
 - o **Precisão** - Qual o percentual de acerto nos casos indicados por cada uma das três técnicas com a BC contendo: 1.000, 10.000 e 70.000 casos.

- o **Facilidade de uso** – O resultado foi obtido através da realização de uma pesquisa junto a futuros usuários do sistema (auditores);
- o **Quantidade de informações necessárias para cada uma das técnicas de recuperação** - O resultado foi obtido através de uma classificação simples, onde o número de campos necessários em cada uma das três rotinas de recuperação foi comparado;
- o **Flexibilidade** - Após a primeira rodada de testes foram realizadas duas trocas de tipos de caso de auditoria armazenados para verificar qual é o nível de flexibilidade permitido por cada técnica. Além da possibilidade de troca sem necessidade de reprogramação o nível de precisão não deveria variar consideravelmente;
- o **Desempenho** O desempenho foi medido da seguinte maneira: tempo necessário para recuperação de casos em cada uma das técnicas com a BC contendo: 1.000, 10.000 e 70.000 casos.

5.6 DEFINIÇÃO DE METODOLOGIA E REALIZAÇÃO DOS TESTES

Para tabulação dos resultados obtidos, foram adotados os seguintes critérios:

- Nas características Precisão e Desempenho, o resultado numérico foi obtido, com base na tabela a seguir:

Quadro 9: Critérios de pontuação para as características de Precisão e Desempenho

Resultado obtido	Pontos
Melhor das três com resultado mais que 50% superior a média	8
Melhor das três com resultado entre 30% e 49,99% superior a média	7
Melhor das três com resultado entre 15% e 29,99% superior a média	6
Melhor das três com resultado entre 0,1% e 14,99% superior a média	5
Mediana	4
Pior das três com resultado entre 0,1% e 14,99% abaixo da média	3
Pior das três com resultado entre 15% e 29,99% abaixo da média	2
Pior das três com resultado 30% ou mais abaixo da média	1

- No caso das características Facilidade de uso, Quantidade de informações e Flexibilidade, cada uma das técnicas foi classificada como

MELHOR, MEDIANA ou PIOR de acordo com os testes definidos. Os resultados foram associados à tabela de pontuação a seguir:

Quadro 10: Critérios de pontuação para Facilidade de uso, Quantidade de informações e Flexibilidade.

Classificação	Pontos
Melhor	8
Mediana	4
Pior	2

Para exemplificar como a tabulação do resultado final ocorre, segue uma simulação de resultados:

Quadro 11: Simulação de resultados.

Característica analisada	Técnica A	Técnica B	Técnica C
Precisão	Melhor (8)	Mediana (4)	Pior (3)
Facilidade de uso	Mediana (4)	Pior (2)	Melhor (8)
Quant. de inform. necessárias	Pior (2)	Melhor (8)	Mediana (4)
Flexibilidade	Melhor (8)	Mediana (4)	Pior (2)
Desempenho	Pior (3)	Mediana (4)	Melhor (6)
Resultado total	25	22	23

Nesse exemplo, a técnica adotada seria a técnica A, pois o conjunto de resultados apresentados indica essa técnica como a mais apropriada.

5.7 COLETA DE DADOS

Pelos grandes volumes de cálculos e dados envolvidos e pela necessidade de precisão na apuração dos fatores Precisão e Desempenho, optou-se por automatizar o processo de coleta de dados e apuração dos resultados para os testes descritos anteriormente. Para isso foi desenvolvido um aplicativo em Visual Basic ® que permite:

- Alimentação de BC manualmente com verificação de consistência;
- Apresentação de interface amigável, possibilitando (se for necessário) o uso por pessoal não técnico;

- Geração automática de grande volume de casos;
- Alteração de parâmetros de uso (métricas, pesos, etc.) com rapidez e precisão;
- Realização de cálculos com rapidez,
- Realização de vários cálculos para simulação.

O BC foi alimentado com casos que apresentavam informações semelhantes às utilizadas no dia a dia do trabalho de um auditor. Os itens Precisão e Desempenho foram analisados em três momentos (com 1.000, 10.000 e 70.000 casos).

Já os aspectos Facilidade de uso, Quantidade de informações e Flexibilidade apesar de subjetivos também foram avaliados tecnicamente através de uma pesquisa junto a 23 usuários finais (auditores internos do Banco do Brasil S.A.) com experiência média na área de 4,5 anos. Essa pesquisa foi aplicada de forma padronizada a todos os auditores seguindo os seguintes passos:

- Cada participante era convidado a participar da pesquisa, após receber as explicações básicas do trabalho que está sendo desenvolvido e dos objetivos gerais;
- Como todas as entrevistas foram conduzidas pelo mesmo pesquisador, cada participante recebeu exatamente as mesmas explicações e orientações, independentemente do seu nível de conhecimento/experiência em auditoria e/ou informática;
- Utilizando-se o aplicativo desenvolvido para os testes de Desempenho e Precisão, cada auditor foi convidado a realizar uma recuperação de caso em cada técnica, como se estivesse iniciando um novo trabalho de auditoria. O aplicativo apresenta três rotinas de busca, sendo:
 - o Uma para recuperação seqüencial
 - o Uma para recuperação em dois níveis
 - o Uma para recuperação em agrupamento de feições

Em seguida, cada participante classificou cada uma das rotinas como a melhor, a mediana ou a pior com relação aos quesitos Facilidade de uso, Quantidade

de informações e Flexibilidade. É importante salientar que isso era realizado por cada auditor individualmente, sem que houvesse troca de opiniões ou qualquer tipo de comunicação entre os entrevistados. Todos os resultados serão apresentados a seguir.

5.8 TABULAÇÃO DE DADOS

5.8.1 PRECISÃO

Os testes de precisão das três técnicas de recuperação foram feitos de forma automatizada, garantindo que todas as situações testadas passaram pelas mesmas condições (fórmulas matemáticas, equipamentos, volume e características de dados) não havendo riscos de vantagens ou prejuízos indevidos para uma ou outra técnica.

Usando sempre informações ligadas à auditoria interna, realizou-se testes com 100, 1.000 e 70.000 casos armazenados no BC, sempre buscando encontrar os 10 casos mais similares. Os resultados de precisão verificados indicam que, muito mais importante que a quantidade de casos é a escolha dos campos usados para a filtragem ou agrupamento de casos, pois os percentuais de precisão não se alteraram com o aumento da base de casos e sim com a mudança de campos referenciais. No primeiro teste os resultados obtidos foram os seguintes:

Quadro 12: Resultados de Precisão - situação A

Téc. de recuperação	1.000		10.000		70.000		Classific
	Qtd	%	Qtd	%	Qtd.	%	
Seqüencial	10	100	10	100	10	100	1
Em dois níveis	08	80	08	80	08	80	3
Grupamento de feições	09	90	09	90	09	90	2

Em seguida foram alterados dois campos usados inicialmente para filtragem e definição de agrupamentos. Isso fez com que as rotinas usassem outras informações para realizar a primeira filtragem (no caso da recuperação em dois

níveis) ou para realizar o agrupamento baseado em feições. Após essas alterações os resultados apresentaram os seguintes níveis de precisão.

Quadro 13: Resultados de Precisão - situação B

Téc. de recuperação	1.000		10.000		70.000		Classific
	Qtd	%	Qtd	%	Qtd.	%	
Seqüencial	10	100	10	100	10	100	1
Em dois níveis	09	90	09	90	09	90	2
Grupamento de feições	05	50	05	50	05	50	3

Essa significativa queda nos resultados apresentados pela técnica de Grupamento de feições pode ser explicada pela mudança sem critérios dos campos usados para a realização do grupamento. Como a rotina de recuperação de casos foi estruturada para formar grupos com base em alguns campos considerados determinantes no momento da engenharia do software, essa técnica dificulta muito a mudança desse critério com o sistema em uso.

Mais algumas alterações foram realizadas e os resultados sofreram sempre alterações nas técnicas de recuperação em dois níveis e grupamento de feições, ora apontando uma como a segunda classificada ora apontando a outra. Em todos os testes realizados a técnica de recuperação seqüencial se mostrou a mais precisa, ficando claro que, se o sistema RBC precisar de precisão total essa é a técnica mais indicada. As outras duas técnicas avaliadas se alternaram como a mediana e a pior, sendo a definição dos campos de critérios (e a não a técnica) o fator determinante para o nível de precisão obtido, optou-se por classificar ambas as técnicas (em dois níveis e grupamento de feições) como mediana.

5.8.2 FACILIDADE DE USO

A pesquisa realizada junto a 23 auditores internos do Banco do Brasil em São Paulo (SP). Utilizou um questionário com 3 perguntas de múltipla escolha (anexo 11.2) relacionadas ao quesito facilidade de uso. Os resultados foram:

Quadro 14: Resultado de Facilidade de Uso

Téc. de recuperação	Melhor		Mediana		Pior		Ranking	
	Qtd	%	Qtd	%	Qtd.	%	Pts	Clas
Seqüencial	8	34,78	9	39,13	6	26,08	112	2
Em dois níveis	2	8,70	5	21,74	16	69,57	68	3
Grupamento de feições	13	56,52	9	39,13	1	4,35	142	1
Resultado total	23	100	23	100	23	100		

Esse resultado se deve provavelmente a necessidade de preenchimento em duas telas distintas na técnica de recuperação em dois níveis, enquanto as outras duas rotinas podem ser rodadas em apenas um passo (apesar da seqüencial exibir uma quantidade de campos maior).

5.8.3 QUANTIDADE DE INFORMAÇÕES

Para a contagem da quantidade de informações necessárias para cada uma das técnicas foi considerado apenas o número de campos de preenchimento obrigatório para o funcionamento da técnica. Nesse caso a recuperação seqüencial leva muita vantagem, pois não possui exigência de preenchimento para realizar a busca, bastando que um campo esteja preenchido. Já as outras técnicas precisaram que pelo menos os campos usados para filtragem ou seleção do próximo nível de busca sejam preenchidos, fazendo com que eventuais informações incompletas inviabilizem a realização de uma pesquisa. Os resultados consolidados foram:

Quadro 15: Resultado de Quantidade de Informações

Técnica de recuperação	Resultados	
	Quantidade campos obrigatórios	Classificação final
Seqüencial	01	1º
Em dois níveis	02	2º
Grupamento de feições	05	3º

5.8.4 FLEXIBILIDADE

O resultado obtido nesse quesito é bem claro, pois a técnica de recuperação seqüencial possibilita qualquer tipo de mudança (atributos do campo, indexação ou mesmo tipo de caso sob avaliação) sem necessitar de nenhum tipo de adaptação ou mesmo pequenos ajustes na rotina de recuperação do sistema. Dentre as outras duas técnicas, houve uma pequena vantagem para a recuperação em dois níveis, pois as intervenções no código do programa de recuperação que se mostraram necessárias foram mais simples e só afetaram o campo usado na passagem do primeiro para o segundo nível (ou seja, no campo de filtragem). Já a técnica de grupamento de feições por ser mais complexa e envolver mais campos, gerou mais intervenções de programação sendo, portanto a menos flexível nessa análise.

5.8.5 DESEMPENHO

Para os testes de desempenho foram utilizados os seguintes recursos: Notebook Accer de 700 mhz com 128 Mb de memória RAM (mais 192 Mb de memória virtual) e disco rígido de 11 Gb com velocidade de acesso 1800 mb/s com cachê de gravação ativado e Sistema Operacional Windows® 2000 com aplicativo DMMR-RBC, desenvolvido em Visual Basic® 6.0 SP5.

Os tempos obtidos por cada uma das técnicas de recuperação de casos em cada um dos três volumes de registros definidos (sempre com as mesmas condições de uso dos equipamentos) foram os seguintes:

Quadro 16: Resultados de Desempenho

Tempo necessário para rodar a rotina de recuperação de Casos				
	Número de Casos no BC			
Téc. de recuperação	1.000	10.000	70.000	Classif
• Seqüencial	16 seg.	2 min. e 44 seg.	18 min. e 58 seg.	3
• Em dois níveis	1 seg.	40 seg.	2 min e 53 seg.	1
• Grupam.de feições	1 seg.	1 min.	2 min e 59 seg.	2

5.8.6 CONSOLIDAÇÃO DOS RESULTADOS

Após a apuração dos resultados em cada um dos cinco critérios, utilizou-se as regras anteriormente definidas para consolidar os resultados totais.

Quadro 17: Consolidação dos resultados

Característica analisada	Técnica A	Técnica B	Técnica C
Precisão	Melhor (6)	Mediana (4)	Mediana (4)
Facilidade de uso	Mediana (4)	Pior (2)	Melhor (8)
Quant. de inform. necessárias	Melhor (8)	Mediana (4)	Pior (6)
Flexibilidade	Melhor (8)	Mediana (4)	Pior (2)
Desempenho	Pior (1)	Melhor (8)	Mediana(4)
Resultado total	27	22	24

5.9 ANÁLISE E CONCLUSÃO DOS RESULTADOS E SELEÇÃO DO MODELO A SER ADOTADO

A princípio, os resultados tabulados já indicam o método de recuperação seqüencial como o melhor método a ser adotado. Pode haver, entretanto alguma dúvida quanto a escolha uma vez que não houve uma unanimidade nos quesitos analisados. Assim, foi conveniente analisar se os números obtidos poderiam ser realmente uma referencia segura para a seleção do modelo.

Uma vez que os resultados apresentaram grandes diferenças no item desempenho, pode-se questionar se esse quesito não deve ser determinante na escolha da técnica de recuperação. A análise, entretanto foi realizada com volumes muito maiores (10.000 e 70.000 casos) do que a previsão de casos do SAD de auditoria interna que será desenvolvido (máximo de 5.000 casos armazenados). Além disso, os equipamentos atuais estão cada vez mais velozes, fazendo com que a diferença de desempenho seja ainda menor.

Deve-se lembrar que as características Precisão e Flexibilidade foram apontadas como determinantes para a escolha, reforçando a adequabilidade da indicação da técnica de recuperação seqüencial para o sistema proposto.

6 ANÁLISE E DEFINIÇÃO DO SOFTWARE DE APOIO ADOADO

Apesar de possível, a construção de um SAD para auditoria interna com RBC sem nenhum tipo de ferramenta de apoio (específica para RBC ou não) envolve um esforço de programação muito elevado e ainda assim, dificilmente oferece a flexibilidade, agilidade e segurança desejadas.

Entretanto, devido às inúmeras opções existentes e as características específicas do RBC essa escolha precisava ser feita com base em critérios técnicos. Por esse motivo, foram seguidos os seguintes procedimentos:

- Pré-seleção, dentre as inúmeras ferramentas disponíveis, de três que tenham características adequadas, recursos e disponibilidade;
- Revisão de literatura sobre essas ferramentas pré-selecionadas;
- Definição das características relevantes para o sistema;
- Avaliação das características de cada ferramenta;
- Determinação de testes a serem aplicados;
- Definição de metodologia e realização dos testes;
- Coleta de dados e tabulação de dados;
- Análise e discussão dos resultados;
- Conclusão e seleção da ferramenta a ser adotada.

6.1 PRÉ-SELEÇÃO DAS FERRAMENTAS DISPONÍVEIS

Para realização da pré-seleção de ferramentas é preciso avaliar alguns aspectos fundamentais dessa escolha, devido às características inerentes desse trabalho. É fundamental que ela esteja sendo atualizada pelos seus desenvolvedores, para evitar descontinuidades que coloquem toda a pesquisa em risco. Além disso, é fundamental que o software seja disponibilizado para essa pesquisa sem custos. Outro aspecto relevante é que a ferramenta gere um aplicativo final em língua portuguesa, para que a interface com os usuários seja fácil. É fundamental também que a ferramenta seja de uso simplificado, não exigindo conhecimentos profundos de programação (dispensando a necessidade de programação manual) e permitindo que o desenvolvimento seja relativamente simples. Finalmente, o quinto e último aspecto fundamental

é que a ferramenta seja adequada para a geração de aplicativos para qualquer área de utilização, não sendo muito específica (voltada somente para a área médica ou científica, por exemplo).

Como existem ferramentas específicas para RBC e ferramentas genéricas (que geram aplicativos baseados em qualquer técnica), testou-se os dois tipos de maneira idêntica, sendo que todas deveriam atender aos cinco aspectos fundamentais. O quadro a seguir apresenta um resumo dos resultados.

Quadro 18: Resumo de aspectos fundamentais para pré-seleção da ferramenta

Ferramenta	Atualizada	Disponibilizou cópia p/ pesquisa	Aplicativo em português	Programa automatizado	Genérico	Adequada p/ a seleção
• Alice Re-Call ¹	Sim	Sim	Sim	Sim	Sim	Sim
• Case Advisor ¹	Não	Não	N/A	N/A	Sim	Não
• CASPIAN ¹	Sim	Sim	Sim	Parcial	Sim	Sim
• CASUEL ¹	Não	Não	N/A	N/A	N/A	Não
• CBR Design Explorer ¹	Não	Não	N/A	N/A	N/A	Não
• CBR Framework ¹	Sim	Sim	Não	Sim	Não	Não
• CBR Tools ¹	Sim	Sim	Sim	Não	Sim	Não
• CBR-Works ¹	Não	Não	Sim	Sim	Sim	Não
• Delphi ²	Sim	Não	Sim	Sim	Sim	Não
• GAS-Pro ²	Sim	Sim	Sim	Sim	Sim	Sim
• IntellixTools ¹	Sim	Não	Não	Sim	Sim	Não
• Kaidara ¹	Sim	Não	Sim	Sim	Sim	Não

Legenda: (1) = Ferramentas específicas para RBC

(2) = Ferramentas de apoio a desenvolvimento genéricas

N/A = Não foi necessário/possível avaliar

Dessa forma, as três ferramentas pré-selecionadas para a avaliação da adequação foram: Caspian, GAS-Pro 2003 e Alice Re-Call.

6.2 REVISÃO DE LITERATURA DAS FERRAMENTAS PRÉ-SELECIONADAS

6.2.1 **GAS-Pro 2003®**

Ferramenta comercial, disponível no mercado desde 1992, com extensa documentação para auxílio ao desenvolvimento, além de possuir suporte técnico permanente. Serve como ferramenta de apoio ao desenvolvimento de qualquer tipo de aplicação.

6.2.1.1 **Características Técnicas e Recursos**

O GAS-Pro é um software gerador de aplicações para manipulação de banco de dados utilizando a linguagem Visual Basic® que utilizam o banco de dados Jet Engine em um dos três idiomas: português, inglês ou espanhol. A Empresa também disponibiliza uma versão mais completa chamada Enterprise que possibilita a utilização de banco de dados Microsoft SQL Server 7.x® e Oracle 8i® (que não foi disponibilizada para essa pesquisa).

Para essa pesquisa, as características mais relevantes disponíveis são:

- As aplicações geradas contêm todos os recursos e funções necessárias à manipulação do banco de dados, podendo ser prontamente compiladas e utilizadas, sem qualquer esforço de programação.
- Possui modelador próprio que permite ao projetista desenhar a estrutura hierárquica do projeto (árvore) com uma interface bastante amigável.
- Geração de aplicação multi-usuária.
- Através de uma interface amigável, o projetista poderá definir em um ou mais banco de dados do tipo MDB, as tabelas neles contidas, suas estruturas e atributos (campos), bem como índices e as regras de integridade existentes entre essas tabelas (relacionamentos 1-1 e 1- N). Com a definição de mais de um banco de dados a aplicação poderá acessar informações de outros sistemas, utilizando-as em consultas, relatórios e validações bem como incluir registros, criar processamentos e lançamentos em qualquer um dos bancos de dados.

- Pode criar uma senha para que a aplicação final mantenha o banco de dados na forma encriptada impedindo que seus dados possam ser manipulados por outras aplicações/usuários não autorizados.
- O aplicativo final pode ser totalmente personalizado, possibilitando que o usuário gere seu aplicativo de acordo com suas necessidades.
- Possibilita também a programação manual, em qualquer parte do código, caso necessário.

A ferramenta possui além dessas, uma lista de recursos e características extensa, além de vídeos demonstrativos que podem ser vistos nos endereços: [Características do GAS](#), [Novidades do GAS](#) e [Vídeos demonstrativos do GAS](#)

Como limitação da ferramenta, deve-se destacar que ela não tem embutida nenhuma rotina para RBC, o que, apesar de ser tecnicamente viável, exigirá um esforço extra de programação.

6.2.2 CASPIAN

CASPIAN é um sistema capaz de lidar com problemas reais de Raciocínio Baseado em Casos. O CASPIAN usa um arquivo escrito na linguagem CASL para criar uma base de casos na memória, a qual pode ser acessada e modificada para recuperar um caso, sugerir uma solução, etc. Ele foi desenvolvido pela Universidade de Wales – Aberystwyth, mas está disponível para a comunidade de pesquisas em código aberto.

6.2.2.1 Características Técnicas e Recursos

Existem dois manuais documentando os programas e arquivos associados do CASPIAN. Além disso, existem exemplos de aplicativos gerados com ele. A versão distribuída é para ambiente MS-DOS®, mas pode ser usada em uma janela dentro do sistema operacional Windows®.

Na prática, é necessário criar um BC na linguagem CASL usando um editor e em seguida carregá-lo com o CASPIAN. O CASL trata cada caso como um registro de um banco de dados com pequenas diferenças como:

- O caso tem informações em dois grupos: características do caso e características da solução
- Cada caso tem que ter um valor como resultado que pode ser sucesso/fracasso ou um valor entre -5 e 5.
- Cada caso tem um nome associado a ele.

Assim que o CASPIAN fizer as checagens automáticas, o usuário deve entrar com um caso para comparação. Para isso, existem n perguntas pré-programadas que levarão o aplicativo a encontrar o caso similar. Como se trata de um aplicativo para ambiente MS-DOS, não há nenhum recurso de ajuda, listas internas, etc, dificultando o uso para quem não conhece bem o software. O programa realiza automaticamente a indexação para selecionar o(s) caso(s). Nessa recuperação de casos, o CASPIAN pode automaticamente selecionar um caso baseado em resultados de similaridade ou pode listar os casos selecionados para que o usuário faça a seleção manual. Em seguida o usuário pode inserir o caso da consulta no BC ou não de acordo com sua conveniência.

A ferramenta possui três programas exemplo que podem facilitar o entendimento de seu funcionamento. Para usá-los basta baixar os arquivos através dos atalhos abaixo: [CASPIAN](#), Manuais: [CASL](#) e [CASPIAN](#) e [Exemplos](#)

6.2.3 ALICE RECALL®

O ALICE ReCall® é um kit de ferramentas para auxílio no desenvolvimento de SAD passo-a-passo. Ele suporta todos os aspectos do RBC e é intuitivo e de interface amigável. Também é uma ferramenta comercial, disponibilizada pela empresa Isoft – Produtos para *Business Intelligence*, apresentando características de software de grande porte. A documentação disponível para auxílio ao desenvolvimento com essa ferramenta é muito completa.

6.2.3.1 Características Técnicas e Recursos

O ReCall foi projetado para viabilizar um desenvolvimento totalmente customizado dos sistemas. Suas principais características técnicas são:

- Uma linguagem orientada a objetos com taxonomia, descrição de tipos, máscaras e relacionamentos, permitindo a especificação de situações complexas de conhecimento de forma estruturada, permitindo inclusive que os fatos sejam apresentados de forma incompleta e incerta.
- Indexação multi-hierarquizada. O RBC trabalha muito com a indexação e isso pode ser importante para sistemas específicos ou com grandes BC.
- Possibilidade de atribuição de pesos para cada campo, definindo a similaridade local para influenciar no cálculo da similaridade global.
- Mecanismos de adaptação automática autorizada por indicação do usuário ou pelo próprio aplicativo.
- Interface gráfica para o usuário final
- Interatividade com o usuário
- Possibilidade de conexão com outros aplicativos, principalmente com relação à banco de dados.
- Está disponível em duas versões: a) Interface Gráfica – para usuários finais e; b) Biblioteca em C++ para programadores que precisem incluir rotinas RBC em seus aplicativos.
- O módulo de programação está disponível em inglês e francês e é compatível com os sistemas operacionais Windows 95, 98, 2000, Me e NT além do UNIX.

O site da empresa disponibiliza mais documentos com detalhes técnicos, além da possibilidade de solicitação de versão para avaliação: [Documentação complementar](#), [Solicitação de versão para avaliação](#).

6.3 DEFINIÇÃO DAS CARACTERÍSTICAS RELEVANTES PARA O SISTEMA A SER DESENVOLVIDO

Para que o SAD de auditoria interna com RBC pudesse ser desenvolvido conforme as necessidades dessa pesquisa existiam 12 características desejáveis na ferramenta de apoio a ser escolhida, conforme lista a seguir:

Recursos de alta relevância:

- **Interface gráfica amigável e ajustável (definição de aspectos da aplicação final)** - é importante que o SAD tenha uma interface adequada aos usuários (na maioria leigos em informática) e não obrigue o desenvolvedor a adotar telas padrões pré-concebidas uma vez que as organizações financeiras (que são os usuários finais) têm regras de desenvolvimento de aplicativos específicas.
- **Geração de aplicativos para trabalho multi-usuário** - como as organizações que possuem auditoria interna são geralmente empresas de médio ou grande porte é extremamente relevante que o SAD tenha recursos e segurança para o trabalho em rede.
- **Possibilidade de parametrização do aplicativo final pelos usuários** - um dos objetivos dessa pesquisa é oferecer um SAD que possa ser utilizado por equipes de auditoria interna das mais variadas áreas de atuação. Dessa maneira, é muito importante que o aplicativo final possibilite alterações em seus parâmetros e em seus conhecimentos armazenados sem a necessidade de intervenção do programador.
- **Permita a criação de programa independente (nnnnn.exe)** - Para facilitar o uso, evitar alterações nos fontes e evitar a necessidade de distribuição da ferramenta de apoio é fundamental que seja possível compilar o programa escrito gerando um arquivo executável independente para distribuição.
- **Disponibilizar de forma automática rotinas internas de Raciocínio Baseado em Casos** - é desejável que a ferramenta escolhida facilite o desenvolvimento disponibilizando as rotinas RBC. Isso evitará a necessidade de programação e reduzirá o tempo necessário para testes.
- **Possibilitar a programação de rotinas internas de RBC** - Caso a ferramenta de apoio não disponibilize ou disponibilize de forma parcial as rotinas RBC é desejável que pelo menos ela possibilite a programação manual das mesmas.

Recursos desejáveis

- Geração de gráficos pelo usuário final
- Controle automático de senhas

- Facilidade de manutenção/atualização
- Controle de características de campos (banco de campos)
- Controle automático de integridade referencial
- Criação de consultas pelo usuário final

6.4 DETERMINAÇÃO DE TESTES E DAS METODOLOGIAS APLICADAS

Definidos os critérios, definiu-se uma metodologia para os testes. A técnica considerada adequada foi a AHP (*Analytic Hierarchy Process*), pois possibilita uma hierarquização dos critérios, uma visão global do problema e ao mesmo tempo é de utilização relativamente simples (EHRlich, 1996). Essa metodologia consiste em definir os critérios, atribuir pesos para cada um de forma que os mais relevantes influenciem mais o resultado final. Em seguida, atribui-se uma “nota” que é baseada em uma escala subjetiva de 1 (indiferente) a 9 (preferência total) para cada critério de cada ferramenta analisada. A multiplicação de cada nota pelo seu peso correspondente gera um número que ao ser somado indica o resultado obtido pelo item analisado.

6.5 COLETA E TABULAÇÃO DOS DADOS

Com os critérios e metodologia, monta-se a matriz AHP e apura-se o resultado.

Quadro 19: Matriz AHP para seleção da ferramenta de apoio

Critérios	Peso	Ferramentas de apoio					
		ALICE ReCall		Caspian		GAS-Pro 2003	
		Nota	Pontos	Nota	Pontos	Nota	Pontos
Interface amigável/ajustável	10%	5	0,50	4	0,40	8	0,80
Aplicação multi-usuário	7%	6	0,42	1	0,07	9	0,63
Parametrização do aplicativo	15%	5	0,75	1	0,15	7	1,05
Gera arquivo.exe	12%	9	1,08	0	0,00	9	1,08
Geração de rotinas RBC	17%	8	1,36	5	0,85	0	0,00
Programação rotinas RBC	11%	5	0,55	5	0,55	8	0,88
Usuário gera gráficos	4%	7	0,28	0	0,00	7	0,28
Controle de senhas	3%	6	0,18	0	0,00	9	0,27
Facilidade de manutenção	8%	4	0,32	3	0,24	8	0,64
Banco de campos	2%	4	0,08	0	0,00	7	0,14
Integridade referencial	5%	4	0,20	0	0,00	8	0,40
Usuário gera consultas	6%	8	0,48	0	0,00	8	0,48
Somatório	100%		6,20		2,26		6,65

6.6 ANÁLISE E CONCLUSÃO DOS RESULTADOS

A utilização da metodologia AHP dá segurança na escolha e deixa claro que mesmo não sendo uma ferramenta de apoio ao desenvolvimento de aplicativos RBC, a ferramenta escolhida – GAS-Pro 2003 – é a mais adequada, pois possui um conjunto de características que possibilita o desenvolvimento de um SAD para auditoria muito mais adequado a pesquisa. A eventual dificuldade de programação causada pela ausência de rotinas RBC na ferramenta é compensada pelas outras vantagens encontradas na ferramenta, possibilitando a construção de um aplicativo mais completo e flexível.

O resultado tão discrepante da ferramenta CASPIAN pode ser explicado pela própria origem e pelos objetivos pelos quais a ferramenta foi desenvolvida. Como não se trata de um software comercial e sim de apoio a pesquisa, suas características são mais limitadas, sendo mais adequada para uso no ensino e compreensão da técnica do RBC.

Apesar de ter tido resultados inferiores, a ferramenta ALICE ReCall também é bastante adequada para o trabalho com as características levantadas, possibilitando provavelmente, resultados muito bons caso viesse a ser selecionada. Um aspecto que não prejudicaria essa pesquisa, mas certamente seria um dificultador na adoção do Alice ReCall para o desenvolvimento de SAD com RBC pelas organizações interessadas é o seu alto custo (a partir de U\$ 5.000,00 a licença) comparado à ferramenta selecionada – GAS Pro 2003 (a partir de R\$ 1.357,00 a licença) – valores referentes a outubro/2005.

7 DESENVOLVIMENTO DE UM SAD PARA AUDITORIA INTERNA COM RBC

Definidos os modelos de similaridade e recuperação de casos, assim como qual será a ferramenta de apoio utilizada pôde-se iniciar efetivamente o desenvolvimento do aplicativo proposto. Como já foi visto na revisão bibliográfica, para o desenvolvimento de um SE é extremamente relevante avaliar a conveniência dessa decisão através da avaliação de questões relevantes, como os itens do check list proposto por Vidal (VIDAL et. al, 2000).

- A tarefa que será abordada pelo SE é prioritária na organização? **SIM**
- A tarefa proposta pode realmente ser executada por esses SE? **SIM**
- Pode-se usar um compilador? Qual? **SIM, Visual Basic**
- Pode-se usar um sistema existente ou mesmo juntá-lo ao módulo SE? **NÃO, mas haverá o aproveitamento de dados do sistema transacional.**
- Foram definidos os critérios de desempenho para avaliação dos resultados do SE? **SIM**
- A rentabilidade do SE será suficiente para justificar seu custo? **SIM**
- Existe pessoa preparada para desenvolver o SE? E para operá-lo? **SIM**
- Equipamentos existentes permitirão o desenvolvimento e o uso do SE? **SIM**
- Como será feita a manutenção do SE e da base de conhecimento? **Pela equipe de informática e pelos próprios usuários.**

Além desses aspectos, é fundamental estruturar adequadamente o conhecimento especialista, e normalmente esse conhecimento é obtido com um trabalho bem estruturado de engenharia de requisitos junto aos profissionais especialistas da área de atuação do sistema.

As principais ações de estruturação do conhecimento especialista foram a definição dos atributos do caso (itens a serem avaliados em cada questão) e da sua forma de armazenamento no BC (suas correlações e similaridades), realizadas junto a auditores internos e gerentes de auditoria.

O aplicativo AUDIT-AJUD como foi batizado, é basicamente um aplicativo que utiliza as técnicas RBC para ajudar o auditor a verificar a existência de trabalhos similares durante a fase de pré-auditoria/planejamento de um trabalho. Este aplicativo é focado nessa fase do trabalho de auditoria devido a forma que essa pesquisa está sendo conduzida; isso, entretanto não impede que ele venha a ser expandido para outras fases do trabalho de um auditor como os testes de conformidade ou mesmo nas conclusões do trabalho, auxiliando na confecção de relatórios e documentos finais. Suas características também o habilitam a ser usado em auditorias de empresas das mais diversas áreas de atuação, uma vez que foi preparado para aceitar parametrizações muito flexíveis, ajustando-se as diversas formas de trabalho.

7.1 ENGENHARIA DE SOFTWARE

O desenvolvimento do AUDIT-AJUD seguiu os conceitos amplamente conhecidos do ciclo de vida clássico, uma vez que o objetivo final desse desenvolvimento é gerar um aplicativo que possa ser usado nos testes necessários para as verificações e conclusões da pesquisa. As etapas foram:

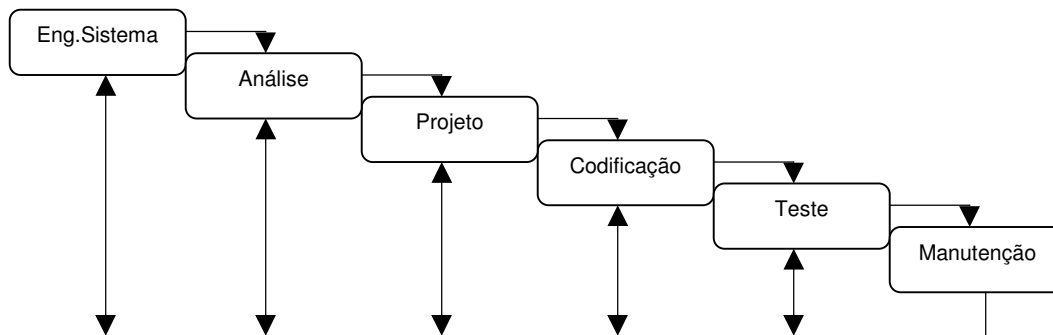


Figura 22: O ciclo de vida clássico (PRESSMAN, 1995)

7.1.1 ENGENHARIA DE SISTEMAS E ANÁLISE

A análise e a engenharia de sistemas envolveram a coleta dos requisitos e uma análise de alto nível do sistema. O trabalho de engenharia de requisitos foi fundamental para o correto desenvolvimento do AUDIT-AJUD e foi realizado conforme a descrição a seguir.

7.1.1.1 Engenharia de requisitos

A Engenharia de Requisitos (ER) é uma sub-área da Engenharia de Software que estuda o processo de definição das funcionalidades e restrições que o cliente do software necessita. É uma área que tem que lidar com conhecimento interdisciplinar envolvendo aspectos de ciências sociais ao exigir o contato e a compreensão dos desejos dos usuários. Para o AUDIT-AJUD a ER foi realizada seguindo as cinco fases do ciclo de ER clássico sendo que a fase de gerenciamento ocorreu simultaneamente durante todo o processo da ER:

Elicitação/Análise de Requisitos

Os objetivos dessas etapas foram delimitar o problema, tornar claro e delimitado as necessidades dos usuários e do próprio sistema. Para isso buscou-se um envolvimento com os usuários para permitir uma análise cuidadosa do domínio da aplicação e do processo que será abrangido.

Descrição do Problema

O problema analisado nessa pesquisa está limitado a uma das etapas do processo de auditoria operacional realizada por equipes de auditoria interna. Essa etapa (chamada pré-auditoria) pode ser definida como a busca realizada pelo auditor interno (antes da execução da auditoria) pela maior quantidade possível de informações para evitar erros de estratégia, aproveitar trabalhos semelhantes já realizados, buscar a maior quantidade de informações existentes sobre o assunto e preparar-se tecnicamente para a tarefa.

Por tratar-se de um trabalho pouco estruturado, onde a forma de coleta, por exemplo, dificilmente se repete de um trabalho para o outro, a etapa de pré-auditoria dificilmente pode ser totalmente automatizada. Por esse motivo, o que se pretende nesse estudo é realizar a ER de um processo de pesquisa de casos iguais ou semelhantes que possibilite o reaproveitamento de pesquisas/procedimentos já realizados, diminuindo assim o tempo necessário para a realização de uma pré-auditoria completa e confiável.

Resumidamente pode-se descrever o problema em 5 passos:

1. O auditor recebe um risco (assunto) para ser auditado;
2. O auditor deve classificar esse risco de acordo com suas características para que ele possa ser catalogado;
3. Antes de iniciar as pesquisas de pré-auditoria, o auditor deve pesquisar no SAD a existência de casos iguais ou semelhantes;
4. De acordo com os resultados apresentados, o auditor deve optar por um dos três caminhos:
 - o Abandonar as pesquisas e realizar a pré-auditoria completa manualmente, uma vez que não se localizou caso semelhante;
 - o Realizar novas pesquisas, uma vez que houve a localização de casos que podem ajudar, mas as informações ainda são limitadas (pode haver necessidade de reclassificação do risco – item b);
 - o Encerrar as pesquisas, uma vez que resultados satisfatórios já foram localizados.
5. No caso de casos inéditos, o auditor deve optar pela conveniência do armazenamento desse caso na BC do sistema. Caso o armazenamento seja decidido, na maioria das vezes será necessário o complemento de informações para possibilitar uma perfeita indexação e armazenamento.

Papel dos *Stakeholders* Envolvidos

A definição dos *stakeholders* nesse sistema é relativamente simples uma vez que, apesar do trabalho de auditoria envolver praticamente todos os departamentos e diretorias de uma empresa, esse sistema está restrito tanto em seu uso quanto na alimentação de sua BC aos funcionários da auditoria.

O papel dos auditores pode ser dividido em dois tipos: o simples usuário, que somente poderá executar consultas, sem alimentar o BC e o usuário avançado (ou administrador), que além da utilização básica de consultas, poderá alimentar e expurgar o BC corporativo. Os gerentes de auditoria e os membros do Comitê de Auditoria (Auditor Geral e Gerentes Executivos) também terão permissão para execução de todos os papéis, apesar de na prática eles raramente se ocuparão de tarefas operacionais de manutenção do BC.

Entendimento do Domínio da Aplicação

A aplicação a ser desenvolvida funcionará na Unidade de Auditoria Interna de uma instituição financeira. Trata-se do maior banco de varejo da América Latina, contando atualmente com 80 mil funcionários e uma das maiores bases tecnológicas de processamento de dados do país e do mundo. A unidade de auditoria é totalmente informatizada e conta com aproximadamente 400 funcionários (350 auditores e 50 funcionários ligados à área administrativa). O sistema transacional roda em ambiente web (intranet)

Entendimento do Negócio

O negócio da auditoria interna já foi amplamente detalhado no capítulo 2.6.4.

Entendimento do Problema

Já detalhado no item 7.1.1. Cabe salientar que nesse trabalho o processo de ER está restrito a consulta de casos similares, não abrangendo a etapa de tomada de decisões e de manutenção da BC, por limitações de tempo/espaço.

Entendimento das Necessidades e Restrições das Pessoas Envolvidas

Todos os funcionários da auditoria interna utilizam sistemas de informação automatizados diariamente. Isso, entretanto, não significa que os níveis de conhecimento e de interesse por tecnologia de todos sejam equivalentes. É muito provável que inicialmente, o SAD venha a ser muito mais utilizado por alguns funcionários do que por outros, principalmente pela característica de não obrigatoriedade natural a um sistema não transacional. Isso deve ser minimizado pelo incentivo a utilização que deve ser promovido pelo nível estratégico da auditoria que já demonstrou grande interesse no sucesso da implementação do sistema. Algum tipo de treinamento deverá ser oferecido para que o nível de utilização mínimo seja atingido por todos os usuários, mas acredita-se que, com os resultados obtidos pelos usuários pioneiros, os usuários restantes percebam a vantagem da migração dos procedimentos.

Técnicas Utilizadas para Elicitação/Análise de Requisitos

Para realização da elicitação/análise de requisitos desse sistema foram selecionadas três técnicas tradicionais: observação, questionário, entrevista. A

opção pela observação deve-se ao conhecimento do autor dos procedimentos relacionados ao sistema, já que trata-se de um auditor em exercício da função. Devido a essa particularidade, muitos requisitos podem ser claramente entendidos pelo desenvolvedor com base em sua prática diária, evitando problemas acidentais e essenciais de interpretação. O documento de registro das observações está no anexo 11.3.

A técnica de questionário foi aplicada em uma única rodada, à distância, com o envio para usuários dos mais diversos cargos e níveis de conhecimento/experiência em auditoria e em informática. Esse questionário foi estruturado com orientação aos casos de uso previstos com vistas a facilitar a organização dos requisitos. A utilização do questionário foi importante por duas questões: aumentar o número de usuários envolvidos e permitir a participação de usuários de outras unidades localizadas em diversas regionais de auditoria espalhadas pelo país. O modelo do questionário está no anexo 11.4.

Por último, trabalhou-se com a entrevista para confirmação de questões e elucidação de alguns pontos onde o pairavam algumas dúvidas. Essa entrevista foi conduzida de forma não estruturada (devido à familiaridade do desenvolvedor com o assunto), com base nos resultados obtidos com os questionários anteriores. Foram entrevistados três auditores (um júnior, um pleno e um sênior) e um gerente de auditoria.

Após a aplicação das três técnicas acima, foram identificados preliminarmente 04 principais requisitos de funcionalidades desejadas que passaram a ser identificados da seguinte maneira:

- Consultar a existência de auditorias anteriores;
- Cadastrar casos relevantes no BC;
- Excluir casos obsoletos do BC;
- Emitir relatório estatístico das consultas.

Nessas funcionalidades os seguintes requisitos restritivos foram identificados:

- 1) Consultar a existência de auditorias anteriores;

- o Deve ser restrito a funcionários da unidade de auditoria;
 - o Deve permitir consulta com informação de ao menos três campos;
 - o Deve apresentar todos os casos semelhantes ao proposto;
 - o Deve permitir a impressão de cada caso individualmente ou de todos os casos listados;
 - o Deve ter mensagens de ajuda desativáveis em todos os campos;
 - o Deve permitir a identificação do auditor responsável pelo caso;
 - o Deve exibir um resumo do caso e só detalhá-lo se solicitado.
- 2) Cadastrar casos relevantes no BC;
- o Deve ser restrito a usuários administradores;
 - o Deve permitir o cadastramento de casos dos últimos 20 anos;
 - o Deve exigir o preenchimento de todos os campos obrigatórios;
 - o Deve manter log dos cadastramentos por 5 anos.
- 3) Excluir casos obsoletos do BC;
- o Deve ser restrito a usuários administradores;
 - o Deve excluir os casos com mais de 5 anos automaticamente;
 - o Deve exibir no momento da exclusão uma mensagem de alerta pedindo a confirmação da exclusão;
 - o A exclusão deve ser confirmada pelo auditor responsável pelo cadastramento;
 - o Deve manter log das exclusões por 5 anos.
- 4) Emitir relatório estatístico das consultas;
- o Deve informar o número de consultas realizadas por gerência;
 - o Deve informar os casos mais consultados;
 - o Deve exibir a média de similaridade das consultas;
 - o Deve exibir as consultas por usuário/gerência.

Após essa tabulação, foram realizadas checagens dos fatores como: conflito, ambigüidade, necessidade e viabilidade dos requisitos. Como detectou-se problemas em alguns casos, foi necessário a realização de uma reunião com os usuários envolvidos para resolução desses conflitos. Após a ponderação, alguns requisitos foram rediscutidos e novas soluções foram encontradas. As mudanças ocorreram nos seguintes requisitos:

- **Problema:** O requisito “Deve apresentar todos os casos semelhantes ao proposto” se mostrou impraticável uma vez que a base de casos pode conter milhares de casos e se não houver limite de apresentação à listagem de todos esses casos pode travar o sistema.
- **Solução:** o requisito teve sua redação alterada para: “Deve apresentar os 10 ou 20 casos mais semelhantes ao caso proposto”.
- **Problema:** O requisito “Deve permitir o cadastramento de casos dos últimos 20 anos” se mostrou incompatível com outro requisito: “Deve excluir os casos com mais de 5 anos automaticamente”.
- **Solução:** O requisito “Deve excluir os casos com mais de 5 anos automaticamente” foi excluído.
- **Problema:** O requisito “Deve permitir o cadastramento de casos dos últimos 20 anos” apesar de viável tecnologicamente se mostrou inviável por questões práticas de recuperação de informações.
- **Solução:** O requisito foi excluído.
- **Problema:** O requisito “A exclusão deve ser confirmada pelo auditor responsável pelo cadastramento” é inviável uma vez que o auditor responsável pelo cadastramento pode não estar mais nos quadros da auditoria.
- **Solução:** O requisito foi excluído.
- **Problema:** O requisito “Deve exibir as consultas por usuário/gerência” não deixa claro se o relatório listaria todos os usuários da auditoria e todas as suas consultas, o que seria inviável devido ao volume.
- **Solução:** a redação foi alterada para “Deve exibir as consultas realizadas nos últimos 30 dias por todos os usuários da gerência”

Assim, ficou acordado entre os usuários participantes que a lista de requisitos seria gerada com base na original, alterada com as soluções acima apontadas.

Especificação de Requisitos

Para a especificação de requisitos as técnicas adotadas foram da Análise Essencial (YOURDON, 1992):

- Lista de Eventos

- Diagrama de Contexto
- Diagramas de Fluxo de Dados (DFD's)
- Descrição dos Processos
- Modelo Entidade-Relacionamento (MER) e
- Dicionário de Dados.

Essas técnicas classificadas de semi-formais possibilitam uma representação clara das interações entre os usuários e as funcionalidades do sistema, além do entendimento do fluxo seqüencial e seus requisitos. A ferramenta CASE utilizada foi a CASE Studio 2.21. da Charonware® que possibilita toda a modelagem. Todos os modelos gerados estão disponíveis para consulta nos anexos 11.5 a 11.10.

Validação de Requisitos

Para a validação de requisitos foi seguido o seguinte fluxo:

- a) Planejamento da revisão com o grupo envolvido nas etapas anteriores;
- b) Distribuição da documentação dos requisitos;
- c) Análise individual dos requisitos;
- d) Reunião para equalização;
- e) Análise das propostas/entendimentos;
- f) Correções.

Por se tratar do ambiente de trabalho do desenvolvedor, e o número de participante ter se reduzido por dificuldades de conciliação de datas, foi dispensado o formalismo e a reunião foi coordenada pelo próprio desenvolvedor. Isso, entretanto, não comprometeu os resultados obtidos.

Ficou claro nessa etapa que mesmo com todos os procedimentos anteriores de elicitación e análise, ainda houve entendimentos equivocados de solicitações relativamente simples. Isso ocorreu com a solicitação feita por um gerente de auditoria que ficou registrada como requisito relacionado à emissão de relatórios → “Deve informar os casos mais consultados”. Conforme se verificou posteriormente, o objetivo do usuário com esse requisito era analisar todos os detalhes dos casos mais consultados para avaliar quais procedimentos de

auditoria estão sendo mais demandados enquanto o entendimento do desenvolvedor foi de que bastaria uma informação resumida para levantamento estatístico. Ou seja, a redação adotada não deixou claro o objetivo do requisito e em nenhuma das fases anteriores isso foi revisto/questionado. Outra constatação foi a dificuldade inicial de entendimento por parte de usuários leigos dos modelos de diagramação utilizados (DFD e MER). Foi necessário uma “tradução” dos símbolos para que a maioria dos participantes compreendesse os conceitos ali registrados. Com relação aos testes de requisitos, foram realizados “testes conceituais” para alguns requisitos considerados vitais ao sistema como “Deve exigir o preenchimento de todos os campos considerados obrigatórios” e “A exclusão deve ser restrita a usuários administradores” através da descrição aos usuários do procedimento que será desenvolvido para atender o requisito.

7.1.2 PROJETO FÍSICO E CODIFICAÇÃO

O projeto físico do AUDIT-AJUD foi baseado nos documentos geradas pela análise essencial na fase de especificação de requisitos, os resultados da validação desses requisitos e mais alguns detalhamentos de procedimentos e das características que o aplicativo deveria ter. Além disso, foi feito um projeto prévio de rotinas que seriam geradas na ferramenta adotada como processos pré-definidos, rotinas avulsas e programação extra, para que a modelagem (necessária para codificação) ocorra da forma mais simples possível. Isso também permitiu que algumas dificuldades e até a qualidade do software pudesse ser estimada e avaliada antes do início da codificação.

Para a codificação propriamente dita do AUDIT-AJUD e de todas as rotinas necessárias, como anteriormente definido, foi utilizada a ferramenta de programação GAS 2003®, utilizando o banco de dados MS-Acess®. Foram gerados no total aproximadamente 36.500 linhas de programas fontes na linguagem Visual Basic® que estão disponíveis para consulta juntamente com o manual de utilização do usuário criado para facilitar o uso pelo usuário.

7.1.2.1 Características do AUDIT-AJUD

O aplicativo final possui diversos recursos que facilitam muito o trabalho de recuperação de casos do auditor. Além da Ajuda on-line, ele exibe dicas quando o mouse permanece sobre o campo por mais de 2 segundos, permite montagem de consultas, gráficos e relatórios pelo próprio usuário, controle de acesso por usuário, rotinas de filtragem, pesquisa, corretor ortográfico, etc.

Na utilização prática, conforme pode ser visto na figura 12 do capítulo 2, quando um auditor recebe um trabalho para ser conduzido, existem várias atividades a serem executadas antes do início da auditoria propriamente dita. Ao receber a descrição do trabalho que deve ser realizado, o auditor deve em primeiro lugar, verificar a existência de trabalhos semelhantes ocorridos anteriormente. Essa tarefa, entretanto é extremamente complexa uma vez que os atuais dados disponibilizados pelos sistemas de informação tem foco em informações administrativas (data, custo, valores, etc.) impedindo uma rápida comparação entre o que se quer e o que existe armazenado nos computadores.

Com a adoção do AUDIT-AJUD, o primeiro passo para o auditor será preencher todas as informações na tela de busca de casos, atentando para a necessidade de exatidão em datas e valores. A tela do AUDIT-AJUD parametrizada para essa pesquisa pode ser vista a seguir.

Audit-Ajud - [Caso a ser comparado]

Arquivo Editar Exibir Pesquisas Resultados Janela Ajuda

Parâmetros Banco de Casos Pesquisa Casos Similares Relatório e Consultas Ferramentas de Apoio Encerra o Audit Ajud

Ações sobre o caso de comparação: Incluir Excluir Procurar Filtrar Em grade Gravar Cancelar Fechar

Ações para pesquisa de similaridade: Status atual desse caso: **Ativado** Ativar o caso Realizar Pesquisa

Conjunto de casos: ANÁLISE DE CASOS DE AUDITORIA Código do caso: 1

Nome da demanda	Irregularidades em desconto de che	Início da abrangência	01/01/2003	Custo do trabalho	4.000.00000
Risco auditado	Crédito	Final da abrangência	31/08/2003	Valor em risco	35.000.00000
Funcionário envolvido	João Silvério	Início da auditoria	01/11/2003	Prejuízo	16.000.00000
Unidade solicitante	GRA São Paulo II	Final da auditoria	20/11/2003		
Unidade executante	GRA São Paulo II	Idade do envolvido	01/05/1960		
Dependência auditada	Campinas do Sul				
Situação da demanda	Encerrada				

Trata-se de fraude:

Funcionário envolvido é culpado:

Há participação estratégica:

É auditoria crítica?:

Solicitação do BACEN:

Valor do campo 1 lista interna:

- AGENCIA VAREJO
- AGENCIA CORPORATE
- AGENCIA EMPRESARIAL
- AGENCIA SINGULAR
- GERIE
- DIREÇÃO GERAL

Valor do campo 2 lista interna:

- PRESENCIAL
- À DISTANCIA
- VIA SISTEMA
- ESPECIAL
- DE GESTÃO
- OUTRO

Auditor responsável: FERNANDO ASATO

Resultado obtido: ABERTURA DE INQUÉRITO CO

Navegação: 14 Primeiro Registro da Tabela < Registro Anterior da Tabela > Próximo Registro da Tabela >> Último Registro da Tabela

Figura 23: Tela de preenchimento de dados para busca de casos no AUDIT-AJUD.

Assim como faz um auditor que não dispõe do AUDIT-AJUD, ao receber um novo trabalho o sistema busca em sua memória as experiências de trabalhos semelhantes realizados anteriormente que tiveram sucesso ou não. A grande vantagem do AUDIT-AJUD é que além da precisão e rapidez na recuperação, seu BC deixa de ser pessoal para se tornar corporativo, tornando praticamente ilimitada a capacidade de aperfeiçoamento do sistema.

7.1.2.2 Recursos de Parametrização Disponíveis no AUDIT-AJUD

Na aplicação das métricas de similaridade são descritos alguns princípios que podem ser adotados, cujos detalhes foram descritos em 4.3.

Os princípios para aplicação de métricas de similaridade implementados no AUDIT-AJUD foram:

- Medidas de similaridade assimétrica
- Consideração de dados incompletos
- Cálculo de similaridade entre campos texto

- o Correspondência exata
- o Correção ortográfica
- o Contagem de palavras
- o Taxa de maior *substring* comum
- Cálculo de similaridade entre campos numéricos
 - o Função escada
 - o Função linear
 - o Função assintótica
 - o Função escalar
- Consideração de dados não-binários

A definição de pesos e obrigatoriedade de preenchimento para cada atributo será sempre um parâmetro do sistema final (possibilitando alterações pelo usuário responsável pela manutenção do sistema), assim como os princípios para aplicação de métricas de similaridade implementadas. Isso, apesar de elevar consideravelmente a complexidade da programação do aplicativo, é fundamental para que o sistema se adapte a diferentes realidades e momentos de trabalho que ocorrem às vezes dentro de uma mesma organização, além de permitir aos especialistas de cada organização um ajuste fino e contínuo dos parâmetros de referência definidos por eles.

7.1.2.3 Representação dos Casos

A representação dos casos é feita pela reprodução das informações necessárias para a localização de um caso de auditoria. A figura a seguir contém as informações necessárias para que o usuário do AUDIT-AJUD possa executar uma pesquisa por auditorias similares no sistema (Não há necessidade de preenchimento de todos os campos para a pesquisa).

Registro: 000001 (<i>controle interno do AUDIT-AJUD</i>)	
Nome da demanda de auditoria:	
Data de Início dos trabalhos: / /	
Data de término dos trabalhos: / /	
Data de início da abrangência: / /	
Data do fim da abrangência: / /	
Unidade Solicitante:	Unidade Executante:
Dependência Auditada:	Risco auditado:

Situação da Demanda:	
Funcionário Envolvido:	Data de Nascimento: / /
Trata-se de Fraude? () Sim () Não	
Funcionário Envolvido é culpado? () Sim () Não	
Trata-se de Auditoria Crítica? () Sim () Não	
Há Aspectos Estratégicos? () Sim () Não	
Trata-se de Solicitação do BACEN? () Sim () Não	
Custo Total do Trabalho: R\$	
Valor Total em Risco: R\$	Prejuízo Apurado: R\$
Tipo de Dependência Auditada: () Agência Varejo () Agência Corporate () Agência Empresarial () Agência Singular () Direção Geral	
Tipo de Auditoria Realizada: () Presencial () À Distância () Via Sistema () Especial () De Gestão () Outro	
Auditor Responsável:	
Resultado Obtido:	
Descrição do Problema:	
Descrição da Situação Final:	

Figura 24: Formulário para realização de pesquisas no AUDIT-AJUD

7.1.2.4 Representação dos Casos no AUDIT-AJUD

Para uma representação adequada de casos de auditoria em um sistema RBC, as informações passadas por especialistas foram fundamentais, mas como os próprios especialistas fizeram questão de frisar, não são definitivas e devem ser constantemente recalibradas, como o uso do sistema e conseqüente ampliação do número de casos armazenados na base de casos. Um dos aspectos que pode ilustrar esse aspecto é relacionado ao peso atribuído a cada um dos atributos no aplicativo. A figura a seguir demonstra como esses pesos podem ser cadastrados/alterados facilmente pelos usuários autorizados

e

b

Campos Texto	Campos Lógicos	Campos Numéricos	Campos Lista Interna	Campos Tabela	Campos Data
Informações sobre os campos tipo lógico					
<input checked="" type="checkbox"/>	Campo 1 está ativo	Nome do campo 1	Trata-se de fraude	Peso do campo 1	1,60000
<input checked="" type="checkbox"/>	Campo 1 não preenchido também indica simetria	% de similaridade caso o campo 1 não esteja preenchido			1,00000
<input checked="" type="checkbox"/>	Campo 2 está ativo	Nome do campo 2	Funcionário envolvido é culpado ?	Peso do campo 2	1,50000
<input checked="" type="checkbox"/>	Campo 2 não preenchido também indica simetria	% de similaridade caso o campo 2 não esteja preenchido			1,00000
<input checked="" type="checkbox"/>	Campo 3 está ativo	Nome do campo 3	Há participação estratégica ?	Peso do campo 3	2,00000
<input checked="" type="checkbox"/>	Campo 3 não preenchido também indica simetria	% de similaridade caso o campo 3 não esteja preenchido			1,00000
<input checked="" type="checkbox"/>	Campo 4 está ativo	Nome do campo 4	É auditoria crítica ?	Peso do campo 4	2,50000
<input checked="" type="checkbox"/>	Campo 4 não preenchido também indica simetria	% de similaridade caso o campo 4 não esteja preenchido			1,00000
<input checked="" type="checkbox"/>	Campo 5 está ativo	Nome do campo 5	Solicitação do BACEN	Peso do campo 5	1,00000

7.1.2.5 Métrica de Similaridade do AUDIT-AJUD

O modelo de similaridade definido pelas pesquisas realizadas no capítulo de análise e definição do modelo de recuperação de casos a ser adotado foi o do Vizinho Mais Próximo Ponderado. Um exemplo simples de cálculo de similaridade por essa métrica pode ser conferido a seguir.

Quadro 20: Cálculo de Similaridade para um Atributo

Novo Caso	Valor	Caso da Base	Valor	Distância
Atributo A	0,75	Atributo A	0,93	$ 0,75 - 0,93 = 0,18$
Atributo B	0,17	Atributo B	0,87	$ 0,17 - 0,87 = 0,70$

Nesse exemplo percebemos que o atributo A teve um resultado menor (0,18) do que o atributo B, portanto, nessa comparação o atributo A é mais similar do que o atributo B. É importante salientar que isso indica uma similaridade entre atributos (local) e não entre casos que deve ser calculada de forma global (incluindo todos os atributos de cada caso) e não ponderada (calculada sem a atribuição de pesos). O valor da similaridade pode ser obtido subtraindo 1 da diferença para obtenção da similaridade de cada atributo.

Partindo desse modelo de cálculo, pode-se adotar uma lógica semelhante e complementar para o cálculo da similaridade global entre dois casos. A seguir é apresentado um exemplo onde todos os atributos receberam um peso e no final, os resultados das similaridades locais (entre cada atributo) foram usados para o cálculo da similaridade global (entre cada caso).

Quadro 21: Demonstrativo de Cálculo da Similaridade Global

Atributos	Peso	Caso de comparação	Caso da Base	Similaridade Local	Similaridade Local x Peso
Nome da Auditoria	18	Fraude em poupança	Fraude em c. correntes	0,60714286	10,92857143
Início do trab.	12	01/08/04	15/02/99	0,07456875	0,89482500
Final do trab.	2		15/05/99	0,00000000	0,00000000
Início abrang.	12	01/08/03	15/02/96	0,02846832	0,34161984
Final abrang	12	31/07/03	14/05/99	0,02467243	0,29606915
Origem	10	BACEN	BACEN	1,00000000	10,00000000

Atributos	Peso	Caso de comparação	Caso da Base	Similaridade Local	Similaridade Local x Peso
Riscos envolv.	17	Imagem.	Imagem.	1,00000000	17,00000000
Instruções Relacionadas	4	73.200, 82.220.	73.200, 63.560.	0,50000000	2,00000000
Roteiro Utiliz.	13	Fraudes	Fraudes	1,00000000	13,00000000
Sub-total	100				54,46108542
$S = \frac{\sum Simlocal}{\sum Pesos}$				54,4610/100	0,5446108

7.1.3 TESTES

Conforme definido na metodologia os testes seguiram as seguintes etapas: concepção, definição, preparação, execução, análise, disseminação e tomada de decisão/conclusão (FENTON & PFLEEGER, 1997).

Aspectos relevantes foram observados nas etapas de concepção, definição e preparação dos testes. Para a alimentação do BC do SAD, foram utilizados casos reais de auditoria com alterações dos valores e nomes envolvidos, por questões de sigilo. Os testes também foram realizados com situações reais do trabalho de um auditor, buscando assim, obter resultados que realmente indiquem a aplicabilidade ou não do AUDIT-AJUD nos trabalhos de auditoria.

Outros aspectos demandaram algumas decisões como:

- Qual o tamanho ideal do BC?
- Qual deve ser o tamanho da amostra dos testes para o nível de confiabilidade ser aceitável na técnica baseada em projeto “irmão”?
- Todo tipo de trabalho pode participar dos testes ou existem exceções que devem ser excluídos para evitar distorções nos resultados?

No caso do tamanho ideal do BC, trabalhou-se com 4.000 casos uma vez que na instituição financeira onde os testes ocorreram são realizados em médio 800 trabalhos de auditoria interna por ano e um dos requisitos apontados pelos usuários nos trabalhos de elicitação de requisitos definiu-se que os casos devem

permanecer armazenados por pelo menos 5 anos. Portanto 800 casos x 5 anos = 4.000 casos armazenados.

Para definição da amostras estatísticas necessárias para validar os testes realizados (no total foram 400 recuperações realizadas com o aplicativo) foi necessário definir quais os aspectos precisavam ser avaliados. Como o trabalho de auditoria não é medido apenas pelo tempo gasto, mas também pela sua precisão e qualidade, foram mensurados basicamente esses três aspectos: o tempo de resposta, a precisão dos resultados e a qualidade das informações obtidas. Para atender a essas questões com todo o rigor estatístico necessário, foram seguidos os seguintes passos, conforme descrito em (OLIVEIRA, 2004):

- Montagem de três amostras piloto com 30 testes dentre os 400 realizados com o AUDIT-AJUD para apuração do tempo total de pesquisa de casos similares, precisão dos casos recuperados e qualidade das informações recuperadas respectivamente. Em cada uma das amostras foi apurando a variância desses fatores (tempo médio, precisão inadequada e qualidade inferior ou igual);
- Cálculo do tamanho da amostra necessária para verificação do novo tempo médio de recuperação de informações, da proporção de inadequação no nível de precisão e no grau de qualidade, com nível de confiança de 95%, margem de erro de 3 minutos no caso de tempo de recuperação e 5% na precisão e 10% na qualidade;
- Realização da apuração do tempo médio de recuperação de casos, das proporções de inadequação de precisão e qualidade com o AUDIT-AJUD nas amostras finais;
- Comparações dos:
 - o tempo médio de recuperação de informações manual (informação já existente no sistema transacional) x tempo médio de recuperação de informações com o AUDIT AJUD.
 - o níveis de precisão dos processos manual e automatizado, com base na seguinte classificação: localização ou não localização correta → Adequado; localização incorreta ou não localização em

casos onde deveria haver localização → Inadequado. Dessa maneira se apurou a proporção de adequabilidade em cada uma das amostras.

- Nível de qualidade das informações recuperadas. Para avaliar a qualidade das informações obtidas, um dos auditores que participaram do processo manual de pré-auditoria foi convidado a avaliar na amostra final o nível das informações obtidas pelo processo manual em comparação ao obtido pelo AUDIT-AJUD, classificando o resultado automatizado como Inferior/Igual ou Superior na comparação aos resultados obtidos pelo processo tradicional.

No caso dos tipos de trabalho que fariam parte dos testes, mesmo sabendo-se que existem trabalhos onde a pré-auditoria é extremamente detalhada e outros onde essa etapa é feita em poucas horas, decidiu-se pela aplicação dos testes em todos os casos uma vez que isso reflete a realidade do trabalho de auditoria e os resultados obtidos em casos porventura excluídos poderiam ser determinantes na decisão final pela adoção ou não do AUDIT-AJUD.

7.1.4 IMPLANTAÇÃO/MANUTENÇÃO

O Aplicativo AUDIT-AJUD foi instalado em um equipamento existente no ambiente de trabalho para que os testes fossem realizados da forma mais próxima possível ao dia-a-dia de trabalho. Após a transferência dos 4.000 casos da base de dados transacional para o AUDIT-AJUD, foi iniciado o trabalho de ajustes nessa BC, que se mostrou mais trabalhoso do que se imaginava inicialmente pela complexidade em obter-se informações, principalmente nos casos com mais de 1 ano de registro.

Em seguida foram iniciados os testes com recuperação de casos. Foram feitos 400 processos com os mais diversos tipos de casos relacionados ao trabalho de auditoria. Em seguida os testes estatísticos foram iniciados para verificar a consistência dos resultados obtidos. Os passos realizados foram os seguintes:

I) - Montagem de uma amostra aleatória de 30 casos para apuração da variância através da fórmula $\rightarrow s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$ ou pelo levantamento do proporção do fator avaliado (casos recuperados inadequadamente no caso da Precisão e resultados piores ou iguais no caso da Qualidade). Os cálculos estão detalhados a seguir e os casos que compuseram cada amostra estão listados no anexo 11.12:

- tempo de recuperação: $s^2 = \frac{4753,136}{29} \rightarrow 163,9012$.
- Precisão: Casos com recuperação inadequada $\rightarrow 1/30$ ou 3,33%, Casos com recuperação adequada $\rightarrow 29/30$ ou 96,67%. Variância $\rightarrow 0,9667 \times 0,0333 = 0,03222$
- Qualidade: Casos com qualidade inferior ou igual ao processo manual $\rightarrow 11/30$ ou 36,67%, Casos com qualidade superior ao processo manual $\rightarrow 19/30$ ou 63,33%. Variância $\rightarrow 0,3667 \times 0,6333 = 0,2322$

II) – Apurada a variância, calcula-se a amostra necessária para apuração final através da fórmula $n = \frac{N \cdot z^2 \cdot \sigma^2}{(N-1) \cdot e^2 + z^2 \cdot \sigma^2}$ para estimativa do tempo médio de recuperação e $n = \frac{N \cdot z^2 \cdot p \cdot q}{(N-1) \cdot e^2 + z^2 \cdot p \cdot q}$ para as proporções de precisão e qualidade, onde:

- z = fator z equivalente ao Nível de Confiança desejado (95% $\rightarrow 1,96$)
- σ^2 = Variância dos dados
- p = proporção esperada da ocorrência a ser estimada
- q = proporção esperada da não ocorrência (100% - p)
- e = Margem de erro (expressa em minutos no primeiro caso e em percentual nos dois outros)
- N = tamanho da população (número de recuperações realizadas)

Resolvendo a equação temos:

a) tempo de recuperação: $n = \frac{400.1,96^2.163,9012}{(400-1).3^2 + 1,96^2.163,9012} \rightarrow 59,6727$ que

arredondando passa a **60** itens na amostra.

b) precisão: $n = \frac{400.1,96^2.0,9667.0,0333}{(400-1).0,05^2 + 1,96^2.0,9667.0,0333} \rightarrow 44,1203$ que

arredondando passa a **45** itens na amostra.

c) qualidade: $n = \frac{400.1,96^2.0,6333.0,3667}{(400-1).0,1^2 + 1,96^2.0,6333.0,3667} \rightarrow 73,09411$ que

arredondando passa a **74** itens na amostra.

III) – Dessa forma, os resultados com o uso do AUDIT-AJUD foram apurados (detalhes no anexo 11.12). Pode-se afirmar com 95% de certeza que:

a) O tempo médio de recuperação de casos com o uso do aplicativo é 15 minutos e 13 segundo, podendo variar entre o intervalo de 12,2 e 18,2 minutos (incluindo nesse tempo o processo de análise dos resultados pelo auditor).

b) A precisão do processo de recuperação de casos do aplicativo é de 95,56%, podendo variar no intervalo de 90,56% a 100%.

c) O nível de qualidade apurado com o critério definido nesse estudo para as informações prestadas pelo AUDIT-AJUD é superior ao processo manual em aproximadamente 58,11% das vezes, podendo variar entre 48,11% a 68,11% devido à margem de erro admitida nos cálculos.

IV) – Para comparação dos três quesitos na situação anterior (manual) em relação ao novo processo (automatizado), foram levantados os dados de tempo médio de recuperação e precisão existentes na base de dados importada do sistema transacional. Nos dois casos foi apurada a média de toda a população (censo) uma vez que essas informações já estavam em base de dados, o que facilitou a contabilização. Os números apurados foram os seguintes:

a) O tempo médio de recuperação de casos similares no método manual apurado foi de 570 minutos..

- b) A média de recuperações adequadas manualmente era de apenas 26,72%.
- c) Qualidade: No quesito qualidade não houve necessidade de levantamento de dados anteriores, uma vez que a própria apuração da proporção de qualidade superior já é feita de forma comparativa.

Na eventualidade de continuidade de uso do AUDIT-AJUD, será necessário realizar algumas manutenções em alguns detalhes que os participantes dos testes apontaram durante as recuperações realizadas, com foco principalmente na rotina de ajustes de casos importados para a BC. Essa eventual manutenção poderá utilizar a ferramenta de apoio, o que possibilitará uma enorme facilidade na alteração de estrutura de dados, rotinas e programas, permitindo inclusive a geração de programas adaptadores da estrutura do banco de dados atualmente em uso, evitando a perda de dados já cadastrados.

7.2 COMPARAÇÃO DA SITUAÇÃO ANTERIOR E POSTERIOR A IMPLANTAÇÃO

Após a realização dos testes comparativos os resultados obtidos foram todos tabulados para permitir análises conclusivas.

Quadro 22: Resultados dos Testes Comparativos

Aspecto Analisado	Situação anterior	Situação posterior	Melhoria/ Redução
Tempo médio na recuperação de casos	570 minutos	15,2167 minutos	97,33%
Proporção de Precisão	26,72%	95,56%	257,63%
Qualidade	-	58,11%	58,11%

Tempo de Resposta: A melhoria no tempo de recuperação de casos já era esperada e se confirmou. Cabe salientar que alguns testes ainda demoraram um pouco pela dificuldade de uso do sistema.

Precisão: Os testes de precisão permitiram o entendimento de quais são os principais problemas existentes com relação a esse aspecto: na rotina manual, a maioria das avaliações de inadequação ocorreram por não localização de casos similares, ocasionada principalmente pela dificuldade em realizar uma

busca adequada manualmente. Já no caso da rotina automatizada, os casos de inadequação ocorreram pela localização de casos inadequados (ou seja, não similares) pelo AUDIT-AJUD, o que pode ser explicado tanto pela inexperiência dos usuários como pelo pouco tempo de uso da BC (e a conseqüente falta de ajustes nos conteúdos dos casos armazenados). Esses detalhes permitem ainda a conclusão que esse resultado pode apresentar números ainda melhores com o passar do tempo e o maior uso do AUDIT-Ajud.

Qualidade das Informações: Esse quesito foi o que apresentou o resultado menos inquestionável dos três (apesar de mais da metade dos testes apontar resultados melhores com o AUDIT-AJUD). O aspecto determinante para isso foi a dificuldade de ajustes na BC devido à inexistência de registros dos casos mais antigos para o fornecimento das informações relevantes (principalmente o detalhamento do caso e de sua solução). Apesar de entender que essa dificuldade é relacionada à implantação do AUDIT-AJUD e não com o seu desenvolvimento, o resultado verificado nos testes foi mantido e não prejudica a avaliação da proposta dessa pesquisa. A análise das dessas dificuldades e sugestões para solução estão no próximo capítulo.

8 CONCLUSÕES

Os resultados obtidos precisam ser analisados de forma muito ponderada, pois apesar de todos serem favoráveis, não houve uma padronização nos resultados. Alguns números (tempo de resposta e precisão) indicam que a proposta de adoção da técnica de RBC para a construção de um SAD para auditoria interna é uma decisão acertada, mas os números da avaliação da qualidade das informações podem gerar alguns questionamentos.

A interpretação desses resultados, entretanto, demonstram que essa situação não está diretamente relacionada à escolha da técnica do RBC ou eventuais falhas na definição do aplicativo desenvolvido. Esses números podem ser entendidos quando se avaliam quais são as informações mais relevantes para cada resultado obtido e como essa informação estava antes da implantação do AUDIT-AJUD.

No caso do tempo de resposta e da precisão dos cálculos de similaridade, os campos de informações administrativas (como data da auditoria, riscos auditados, etc.) são determinantes. Essas informações foram facilmente obtidas para construção do BC do AUDIT-AJUD usado no teste, pois já existiam no sistema transacional.

Já as informações mais relevantes para o quesito Qualidade das Informações estariam nos campos Descrição do Problema e Descrição da Situação Final que não existiam nos dados usados como origem para o BC. Como a alimentação manual desses campos é muito trabalhosa, em muitos casos, as informações localizadas estavam resumidas ou até mesmo não existiam (pela impossibilidade de recuperar as informações mais antigas). Isso deixa claro que as dificuldades verificadas nesse quesito podem ser minimizadas facilmente com um trabalho adequado de implantação e ajustes no BC. Além disso, a conscientização dos usuários é fundamental, pois os campos relacionados ao problema só podem ser corretamente preenchidos se os

usuários tiverem paciência e investirem tempo e trabalho para obter resultados favoráveis no uso do SAD em seguida.

Após essa interpretação e ponderação, fica claro que a proposta de adoção de um SAD para auditoria interna com RBC deve ser vista como uma possibilidade extremamente interessante para as organizações que tem trabalhos de auditoria no seu dia-a-dia. O que deve ser posto como aspecto realmente determinante para a obtenção de resultados inquestionáveis é o planejamento bem detalhado da construção inicial da base de casos e a obtenção de um grande envolvimento (que pode ser entendida como participação efetiva) de toda a equipe de usuários na implantação.

Mesmo com essa ressalva, a redução de tempo nas pesquisas, aliada as melhorias na precisão e na qualidade das pesquisas indicam que a proposta de utilização da técnica do RBC na construção de SAD para trabalhos de auditoria interna é perfeitamente viável e vantajosa e os investimentos necessários para o desenvolvimento do SAD e para a construção e ajustes no BC serão rapidamente recuperados com os melhores resultados obtidos durante os trabalhos do dia-a-dia.

8.1 DIFICULDADES ENCONTRADAS

Apesar do sucesso da realização dos testes, algumas dificuldades foram verificadas e precisam ser observadas no planejamento do sistema:

- A construção do BC deve ter origem no banco de dados (BD) do sistema transacional caso ele exista, mas não será suficiente, pois as informações mais relevantes para a tomada de decisões não estão disponíveis nesse tipo de BD.
- A atualização dessas informações não existentes no BC deve ser feita antecipadamente de forma planejada, por pessoas que acreditem no uso do SAD e se possível com o apoio do nível estratégico da organização.
- A resistência dos usuários deve ser encarada como um dos problemas mais sérios, pois os resultados tendem a melhorar com o uso e, portanto não adianta implantar o SAD e deixar seu uso restrito a alguns usuários.

Essas dificuldades são justamente aquelas apontadas por KELLER (1991) para o desenvolvimento de SE. Portanto, esses obstáculos já eram previsíveis e devem ser superados com planejamento e persistência.

8.2 PESQUISAS FUTURAS

Essa pesquisa apresentou diversos resultados que podem dar continuidade a novas pesquisas. A metodologia para definição da melhor métrica para cálculo de similaridade (DMMS-RBC) proposta pode ser aperfeiçoada conforme descrito no capítulo 4.8.

Além disso, as pesquisas do SAD com RBC também podem prosseguir com o incremento de outros tipos de informações não implementadas nessa pesquisa como campos multivalorados, campo imagem ou de textos longos (tipo *memo*). Isso ampliará as possibilidades já que os sistemas multimídia atuais armazenam muito mais informações do que os campos tradicionais podem conter. Outra possibilidade de incremento apontada nas pesquisas que pode ser avaliada é a combinação de técnicas com características diferentes (mas que podem ser complementares) como o RBC e redes neurais, por exemplo, que possibilitará o aproveitamento das vantagens de cada uma das técnicas, provavelmente elevando exponencialmente a qualidade dos resultados obtidos.

9 REFERÊNCIAS BIBLIOGRÁFICAS

AAMODT, A; PLAZA, E. *Case-based reasoning: foundational issues, methodological variations, and system approaches*. Artificial Intelligence Communications, IOS Press, Vol. 7 - Pág. 39-54. 1994.

AL-ANI, I.; COOLEY, R. E.; AWAD, E. M. *From Decision Support To Expert Systems*. Proceedings Of The Conference On The 1987 ACM SIGBDP-SIGCPR Conference. Março 1987.

ALLEN, B. P. *Case-Based Reasoning: Business Applications* Communications of The ACM Volume 37, Março 1994

ALTER, S. L. *Information Systems: A Management Perspective*. EUA Menlo Park, Benjamin & Cummings, 2A. ED. 1996

ANGEHRN, A. A.; DUTTA, S. *Case-Based Decision Support*. Communications of the ACM, Vol. 41, Março 1998

ARAÚJO, I. *Introdução à Auditoria Operacional. 1ª. Edição*. FGV Editora – Rio de Janeiro. 2001.

BANDINI, S.; MANZONI, S. *Application of fuzzy indexing and retrieval in case based reasoning for design* - Symposium on Applied Computing Proceedings of the 2001 ACM Symposium on Applied Computing 2001 USA

BARONE, D. (Org) *Sociedades Artificiais. 1a. Edição*. Editora Bookman – Porto Alegre. 2003.

BARRETO, A. S.; BUENO, T. C.; HOESCHL, H. C. *Aplicando Raciocínio Baseado em Casos para Representar o Conhecimento de Decisões Tributárias em Sistemas de Recuperação de Informação*. First Ibero-American Workshop on Automated Deduction and Artificial Intelligence. 2002

BEN-BASSAT, M.; BENIAMINY, I.; JOSEPH, D. *Can Model-Based and Case-Based Expert Systems Operate Together?* IEEE Proceedings International Test Conference 1998.

BENCH-CAPON, T. J. M. *Some Observations on Modelling Case Based Reasoning With Formal Argument Models*. ACM Proceedings of the Seventh International Conference on Artificial intelligence and Law. Junho 1999

BERGMANN, R.; ALTHOFF, K. – *Methodology for building CBR applications. Case base reasoning technology: from foundations to applications*. Editora Springer – Berlin. 1998

BINDER, F. V. *Sistemas de Apoio à Decisão*. Editora Érica, São Paulo, 1994.

BITTENCOURT, G. *Inteligência Artificial – Ferramentas e Teorias – 1ª*. Edição. Trabalho apresentado na Escola de Computação da UNICAMP. Campinas(SP). 1996

BITTENCOURT, G. *Inteligência Computacional*. Material disponibilizado na internet pelo Grupo de Inteligência Artificial do Departamento de Automação e Sistemas da Universidade Federal de Santa Catarina, 1998. Disponível em: <http://www.lcmi.ufsc.br/gia/softcomp>. (Último acesso em 19/10/2005).

BRASIL. Decreto N°. 3.591, de 06 de setembro de 2000 – Dispõe sobre o Sistema de Controle Interno do Poder Executivo Federal e dá outras providências. Disponível em: <http://www.mre.gov.br/ciset/dec3591.htm> (Último acesso em 19/10/2005).

CARDOSO, J. *Considerações ao Projeto de Interfaces de Bibliotecas Digitais*. PUC-RS – 1999.

CASSARRO, A. *Sistemas de Informações para Tomada de Decisões. 3ª*. Edição. Editora Pioneira – São Paulo. 2001

CERCONE, N.; AN, A.; CHAN, C. *Rule-Induction and Case-Based Reasoning: Hybrid Architectures Appear Advantageous*. IEEE Transactions on Knowledge and Data Engineering, Vol. 11, n°. 1, Jan/Fev 1999.

CHI, R. H.; KIANG, M. Y. *An integrated approach of rule-based and case-based reasoning for decision support* Proceedings of the 19°. ACM - Computer Science Conference, Abril 1999.

COSTA, Marcelo T. C. *Uma arquitetura baseada em agentes para suporte ao ensino à distância*. Florianópolis: UFSC, 1999. (Tese de doutorado em Engenharia de Produção)

DAGOSTINI, T. C.; WANGENHEIM, C. G. V.; MATTOS, E. S.; HOESCHI, H. C.; BARCIA, R. M. *JurisConsulta: Retrieval in Jurisprudencial Text Bases using Juridical Terminology*. Proceedings of the seventh international conference on Artificial intelligence and law. Junho 1999.

DAVENPORT, T.; PRUSAK, L. *Conhecimento Empresarial*. Rio de Janeiro: Campus, 1999.

DEMO, P. *Avaliação Qualitativa*. São Paulo: Cortez, 1991.

DIZ, M. L. B.; LÓPEZ, J. C. Y.; RODRÍGUEZ, J. M. C. A Case-Based Reasoning System to Assist the Internal Control Process of a Firm. Proceedings of the Mixed-Initiative Case-Based Reasoning Workshop held at the 2002 European Conference on Case-Based Reasoning.

DOYLE, J.; DEAN, T.; et al. *Strategic Directions in Artificial Intelligence*. ACM Computing Surveys, Vol. 28, N°. 4, Dezembro de 1996.

DUTTA, S.; WIERENGA, B.; DALEBOUT, A. *Case-Based Reasoning Systems: From Automation to Decision-Aiding and Stimulation*. IEEE Transactions on Knowledge and Data Engineering, Vol. 9, N°. 6, Nov/Dez 1997.

EHRlich, P., *Modelo Quantitativos de Apoio às Decisões* - RAE - Revista de Administração de Empresas, v. 36, n.1, p. 33-41, jan/fev/mar - 1996

ESCRITÓRIO DO AUDITOR-GERAL DO CANADÁ, *Glossário de termos de auditoria: parte integrante do manual de auditoria integrada*. Tradução do Tribunal de Contas da Bahia. Salvador, TCE, 1995.

FALSARELLA, O.; CHAVES, E. *Sistemas de Informação e Sistemas de Apoio à Decisão*. Artigo disponível na internet <http://www.chaves.com.br/TEXTSELF/COMPUT/sad.htm>, (Último acesso em 19/10/2005).

FENTON, N. E.; PFLEEGER, S. L. *Software Metrics – A Rigorous & Practical Approach*. 2a. Edição. PWS Publishing Company, Boston, 1997.

FREITAS, H. *A informação como ferramenta gerencial*. Ortiz, Porto Alegre, 1993.

FREITAS, H.; BECKER, J. L.; KLADIS, C. M.; HOPPEN, N. *Informação e Decisão: Sistema de Apoio e seu Impacto*. Ortiz, Porto Alegre, 1997.

GIL, A. C. *Como Elaborar Projetos de Pesquisa*. Atlas. São Paulo. 1999.

GONZÁLES, P. *Airobotics – A Look Into the Future*. ACM Proceedings of the third international conference on Industrial and engineering applications of artificial intelligence and expert systems - Volume 2. Junho, 1990

GONZALEZ, A.J.; XU, L.; GUPTA, U.M. *Validation Techniques for Case-Based Reasoning Systems*, IEEE Transactions on Systems, Man, and Cybernetics- Part A: Systems and Humans, Vol. 28, No. 4, pag. 465-477, July 1998

GREENSPAN, M.A., GODIN, G., SHU, C. *Nearest Neighbor Search Through Function Minimization*, 13th Canadian Conference on Computational Geometry, Waterloo, Canada, August 2001, pp. 157-160.

HARDAWAY, D. E.; WILL, R. P. *A Review of Barriers to Expert System Diffusion*. Proceedings of the 1990 ACM SIGBDP conference on Trends and directions in expert systems. Setembro, 1990.

HOPLIN, H. P. *Reducing Managerial Risk with Expert Systems*. Proceedings of the conference on The 1987 ACM SIGBDP-SIGCPR Conference. Março 1987.

IGLEZAKIS, I.; ROTH-BERGHOFFER, T.; ANDERSON, C. E. *The Application of Case Properties in Maintaining Case-Based Reasoning Systems* Proceedings of the 9th German Workshop on Case—Based Reasoning, GWCBR 2001, Baden-Baden, Alemanha

JACKSON, P. *Introduction to Expert Systems*. 3^a. Edição. Addison-Wesley - England, 1999

JEDLITSCHKA, A.; ALTHOFF, K. *Using Case-Based Reasoning for User Modeling in an Experience Management System* IESE – Institut Experimentelles Software Engineering Accepted for publication in Proceedings

of the Workshop “Adaptivität und Benutzermodellierung in interaktiven Softwaresystemen” 2001.

JÚLIO, M., *Um Estudo de Métricas de Similaridade em Sistemas Baseados em Casos Aplicados à Área de Saúde*. Dissertação de Mestrado apresentada ao Instituto de Computação da UNICAMP – Fevereiro-2005.

JUND, S. *Auditoria – Conceitos, Normas, Técnicas e Procedimentos*. 4^a. Edição. Editora Impetus – Rio de Janeiro. 2002

KANITZ, S. *O parecer do auditor – 1^a. Edição*. Editora McGraw-Hill do Brasil – São Paulo. 1979 em JUND, S. *Auditoria – Conceitos, Normas, Técnicas e Procedimentos*. 4^a. Edição. Editora Impetus – Rio de Janeiro. 2002

KELLER, R. *Tecnologia de Sistemas Especialistas – 1^a. Edição*. Editora Makron Books – São Paulo 1991

KLUWER, G., FORBUS, K. *MAC/FAC: A model of similarity-based retrieval*. In Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society (pp. 504--509) - 1991.

KOLODNER, J. *Case-Based Reasoning*. Morgan Kaufman Publishers, 668p. 1993

KOLODNER, J.; MARK, W. *Case-Based Reasoning*. IEEE Expert Intelligent Systems & Their Applications, Out 1992.

KOSLOSKY, Marco A. N., *Aprendizagem Baseada em Casos Um Ambiente Para Ensino de Lógica de Programação*. Florianópolis, 1999. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal de Santa Catarina.

LAUDON, K. C.; LAUDON, J. P. *Gerenciamento de Sistemas de Informação*. 3^a. Edição. Editora LTC – Rio de Janeiro. 1999

LAUDON, K. C.; LAUDON, J. P. *Sistemas de Informação*. 4^a. Edição. Editora LTC – Rio de Janeiro. 1999

LEAKE, D. B. (ed) *Case-Based Reasoning – Experiences, Lessons & Future Directions*. 1^a. Edição. American Association for Artificial Intelligence. EUA. 1996

LENZ, M.; BURKHARD, H. *Retrieval one Such?* In Proceedings of the XPS-95 Workshop on Case-Based Reasoning, Germany, 1995.

LORENZI, F. Uso da metodologia de raciocínio baseado em casos na investigação de irregularidades nas internações hospitalares. Porto Alegre: CPGCC da UFRGS, 1998 (Dissertação de mestrado) em BARONE, D. (Org) *Sociedades Artificiais. 1a. Edição*. Editora Bookman – Porto Alegre. 2003.

MANAGO, M., ALTHOFF, K.; TRAPHÖNER, R., Induction and reasoning from cases. In: ECML – European Conference on Machine Learning, Workshop on Intelligent Learning Architectures. Vienna, April 1993.

MANTARAS, R., PLAZA, E., Case-Based Reasoning, State of the Art in Machine Learning. Setembro-1995.

MCCARTHY, J.; HAYES, P.J. *Some Philosophical Problems from Standpoint os Artificial Intelligence*. Publicado no livro: Machine Intelligence, 4. Edimburgh University Press, Edimburgh. 1969.

MENDES, E.; MOSLEY, N.; WATSON, I. *A Comparison of Case-Based Reasoning Approaches to Web Hypermedia Project Cost Estimation* ACM International World Wide Web Conference Proceedings of the eleventh international conference on World Wide Web. 2002.

OLIVEIRA, P.H. *Amostragem Básica – Aplicação em Auditoria*. Editora Ciência Moderna, Rio de Janeiro, 2004.

POWER, D.J. *A Brief History of Decision Support Systems*. DSSResources.com, World Wide Web, <http://dssresources.com/history/dsshhistory.html>, 2002 (Último acesso em 19/10/2005).

PRESSMAN, R. *Engenharia de Software*. Editora Makron Books, São Paulo, 1995.

QUIGLEY, E. J.; DEBONS, A. *Interrogative Theory of Information and Knowledge* Proceedings of the 1999 ACM SIGCPR conference on Computer personnel research New Orleans, Louisiana, United States Pages: Pages: 4 – 10. 1999

REIS, L.A.; CARGNIN, M. L. SDDEP – *Uma Aplicação na Área Médica Utilizando Raciocínio Baseado em Casos*. Florianópolis, 1997. Relatório de Estágio (Graduação em Ciência da Computação) – Centro de Ciências Tecnológicas, Universidade Federal de Santa Catarina.

ROTH-BERGHOFER, T.; IGLEZAKIS, I. *Six Steps in Case-Based Reasoning: Towards a Maintenance Methodology for Case-Based Reasoning Systems*. Proceedings of the 9th German Workshop on Case—Based Reasoning, GWCBR 2001, Baden-Baden, Alemanha

RUSSELL, S.; NORVIG, P. *Artificial Intelligence: a Modern Approach*. New Jersey: Prentice-Hall, 1995.

SAGE, A. P. *Decision Support Systems Engineering*. John Wiley & Sons, Inc., New York, 1991.

SANTI, P. *Introdução à Auditoria*. 1ª. Edição. Editora Atlas S.A. – São Paulo. 1988.

SCHREINER JÚNIOR., S. *Sistema de Apoio à Decisão para o Controle do Tráfego em Interseções Isoladas*. Projeto Final de Graduação em Engenharia Civil – UnB. 2001. Disponível: http://www.ceftru.unb.br/arquivos/sigeti_artigo.pdf (Último acesso em 19/10/2005)

SHEPPERD, M.; KADODA, G. *Comparing software prediction techniques using simulation*; Software Engineering, IEEE Transactions on, Volume: 27 Issue: 11, Nov 2001

SHUGUANG, L.; QING, J.; GEORGE, C. *Combining Case-based and Model-based Reasoning: A Formal Specification*. IEEE Proceedings of the Seventh Asian-Pacific Software Engineering Conference – ASPEC 2000.

SILVA, E. L.; MENEZES, E. M. *Metodologia da Pesquisa e Elaboração de Dissertação*. 3ª. Edição revisada e atualizada. Florianópolis, 2001.

SOVAT, R. B.; ALUÍSIO, S. M.; CARVALHO, A. C. P. L. F. *RaBeCa: A Hybrid Case-based Reasoning Development Environment*

Proceedings of the 13th International Conference on Tools with Artificial Intelligence, 7-9 Nov 2001. Pag. 61 -68

SOVAT, R. B.; CARVALHO, A. C. P. L. F. *A Case-Based Reasoning Development Environment* IEEE Computational Intelligence and Multimedia Applications, 2001. ICCIMA 2001. Proceedings. Fourth International Conference on 2001 Pag. 374 – 378.

SPRAGUE JÚNIOR, R.; WATSON, H. J. *Sistema de Apoio à Decisão – Colocando a Teoria em Prática* – Editora Campus, 1991.

STAIR, R. *Princípios de Sistemas de Informação*. 2ª. Edição. Editora LTC – Rio de Janeiro. 1998

TANAKA, T., SUEDA, N. *Combining Strict Matching and Similarity Assessment for Retrieving Appropriate Cases Efficiently*. In P. Anick et al. (eds.) Case Based Reasoning and Information Retrieval. AAAI Spring Symposium Series, 1993

TESKEY, F. N. *Enriched Knowledge Representations for Information Retrieval* Proceedings of the 10th annual international ACM SIGIR conference on Research and development in information retrieval , 1987 Pages: 43 - 44

TSATSOULIS, C.; CHENG, Q.; WEI, H. *Integrating Case-Based Reasoning and Decision Theory*. IEEE Expert Intelligent Systems & Their Applications, Jul/Ago 1997.

TURING, A. M. *Computer Machinery and Intelligence*. Oxford University Press on behalf of MIND - The Journal of the Mind Association -, vol. LIX, no. 236, pp. 433-60, 1950

TURNER, R., Case-Based and Schema-Based Reasoning of Problem Solving, Proceedings of a Workshop on Case-Based Reasoning, Florida, 1989.

VIDAL, A. G. *et al. Trabalho – Sistemas Especialistas* – Universidade de São Paulo – Departamento de Administração Novembro de 2000.

VELOSO, M., MUÑOZ-AVILA, H. e BERGMANN, R. Case-based planning: selected methods and systems. AI Communications 9, pg 128-137, 1996.

WANGENHEIM, C.; WANGENHEIM, A. *Raciocínio Baseado em Casos*. 1ª. Edição. Editora Manole – Barueri SP. 2003

WATSON, I. *Applying case-based reasoning: techniques for enterprise systems*. Morgan Kaufmann Publishes, 1997.

WATSON, I. *Case-Based Reasoning Development Tools: A Review*. 1995. Disponível: <http://online.loyno.edu/cisa494/papers/Watson.html> (Última acesso em 19/10/2005)

WESS, S.; ALTHOFF, K.; DERWAND, D., Using kd-trees to improve the retrieval step in case-based reasoning. In S. Wess, K.-D. Althoff, and M.M. Richter, editors, *Topics in Case-Based Reasoning*, number 837 in *Lecture Notes in AI*, pages 167--81, Berlin, 1993. Springer-Verla. <http://citeseer.ist.psu.edu/wess93using.html>

WILKE, W.; BERGMANN, R. *Techniques and Knowledge Used for Adaptation during Cased-Based Problem Solving*. In *Proceedings of the 11ª. Internacional Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, 1998.

WILLIAMS, J. *When Expert Systems are Wrong*. *Proceedings of the 1990 ACM SIGBDP Conference on Trends and Directions in Expert Systems*. 1990.

WYSK, R. B. *Expert Systems in the Context of Decision Support Related Interventions*. *Proceedings of the 1990 ACM SIGBDP Conference on Trends and Directions in Expert Systems*. 1990.

YANG, D.; VASARHELYI, M. *The Application of Expert Systems in Accounting* – Publicado no livro: *Artificial Intelligence in Accounting and Auditing Volume IV: Towards New Paradigms International Accounting Network*. Markus Wiener Publishers 1994. Disponível: <http://accounting.rutgers.edu/raw/publisher/applica4.htm>.

YOURDON, E. *Análise Estruturada Moderna*. Editora Campus - Rio de Janeiro, 1992.

ZELEZNIKOW, J.; STRANIERI, A.; HUNTER, D. *Beyond rule based reasoning-the meaning and use of cases*. *IEEE Artificial Intelligence for*

Applications, 1995. Proceedings., 11th Conference on , 20-23 Feb 1995.

10 GLOSSÁRIO

- **Accountability** – é a relação em que o delegante transfere responsabilidade para o delegado, que a aceita e assume o compromisso de informar ao delegante como ele, delegado, está desempenhando as ações inerente à responsabilidade que lhe foi conferida. Fonte (ARAÚJO, 2001)
- **Break-even** - Momento (tempo) em que a soma dos valores de retorno obtidos pelo investimento se igualam ao valor total investido.
- **Controle** - É o processo no qual as atividades de uma organização são conduzidas de acordo com um plano de ação desejado, ou seja, é o processo estabelecido que possibilita à administração dirigir e acompanhar suas atividades, envolvendo: políticas, objetivos, normas, padrões, procedimentos, a delegação de autoridade, a identificação de desvios na estrutura; e, se necessário, ação corretiva. Fonte: Procedimentos de Auditoria, Instituto dos Auditores Internos do Brasil ("Livro Branco").
- **Economia** – É a capacidade de fazer, gastando pouco. É executar uma atividade ao menor custo possível, ou seja, *gastar menos*.
- **Eficiência** – É a capacidade de fazer as coisas direito. É apresentar um desempenho satisfatório sem desperdícios, ou seja, *gastar bem*.
- **Eficácia** – É a capacidade de fazer as coisas certas. É alcançar os objetivos ou metas previstas, ou seja, *gastar sabiamente*.
- **Risco** - Grau de incerteza da rentabilidade (retorno) de um investimento. Ex.: afirmar que um investimento é de alto risco significa que temos pouca chance de prever com precisão a rentabilidade deste investimento. Em contrapartida, esse investimento oferece possibilidade de retorno superior a um investimento conservador.

- **Risco de Auditoria** - Segundo SANTI (1988) Risco de Auditoria é a possibilidade de o auditor, por desconhecimento, deixar de modificar apropriadamente sua opinião sobre demonstrações financeiras materialmente incorretas pelo efeito de erros ou irregularidades existentes, mas não detectadas pelo seu exame. Pode também ser definido como o risco aceito pelo auditor sobre a possibilidade de não detectar um erro ou uma fraude. Ou ainda: é a possibilidade de o auditor vir a emitir uma opinião tecnicamente inadequada sobre demonstrações contábeis significativamente incorretas.
- **Stackholders** – Todos os envolvidos em um processo, um trabalho ou um projeto, estando diretamente ou indiretamente relacionado com ele.
- **Teste de Turing** – Teste a inabilidade de distinção entre respostas de um computador e de um ser humano.

11 ANEXOS

11.1 RESULTADOS DETALHADOS DA DEFINIÇÃO DA MÉTRICA DE SIMILARIDADE ADOTADA

Caso	Casamento Simples	Casamento Ponderado	Vizinho Mais Próximo	Vizinho Mais Próx.Ponder	Distância Euclidiana	Apontada pelo especialista
1	23,07692	25,38461	51,43151	50,09670	69,23771	50,00000
2	11,53846	15,76923	46,74573	44,28368	64,76715	
3	61,53846	72,50000	78,60059	85,89560	86,62494	90,00000
4	23,07692	30,76923	51,98673	54,72302	69,00735	58,80000
5	23,07692	37,69230	55,98571	63,54117	72,70070	67,80000
6	23,07692	24,23076	41,19017	45,15078	62,69707	
7	15,38461	22,69230	42,71948	48,31278	60,56680	49,50000
8	19,23076	23,26923	52,51687	54,96892	69,23749	47,00000
9	11,53846	15,76923	38,62294	36,24729	57,51006	
10	19,23076	35,00000	54,05746	62,36274	69,93951	64,20000
11	7,69230	11,53846	40,07295	37,53620	58,56047	
12	11,53846	13,46153	46,29730	41,04124	64,49186	
13	7,69230	11,92307	32,39010	33,96839	52,40536	
14	15,38461	22,30769	42,91621	46,64168	61,60957	
15	7,69230	7,50000	40,79453	34,35605	57,93007	
16	11,53846	21,53846	46,41497	48,77811	62,90334	
17	15,38461	23,26923	49,58377	49,69606	65,04166	
18	3,84615	6,15384	39,82202	34,67726	57,65928	47,50000
19	11,53846	21,53846	47,51433	50,06187	63,98299	
20	3,84615	5,76923	40,25713	34,56989	58,48231	
21	11,53846	21,53846	48,05065	50,37961	64,73183	
22	15,38461	27,30769	51,89681	56,14884	67,63746	58,70000
23	15,38461	23,26923	49,59334	48,32762	67,08404	45,80000
24	11,53846	21,53846	47,67472	50,06440	64,64982	44,30000
25	11,53846	21,53846	47,50120	49,83016	64,61046	
26	15,38461	25,38461	51,89681	54,22576	67,63746	58,00000
27	15,38461	25,38461	50,87020	53,32062	67,49888	59,00000
28	11,53846	21,53846	46,91839	48,05434	64,34360	
29	11,53846	21,53846	47,94944	50,33912	64,72566	
30	15,38461	26,15384	47,68662	50,21070	64,34462	
31	23,07692	27,69230	51,11346	51,42094	67,15937	
32	15,38461	26,15384	47,58541	50,17022	64,34255	
33	15,38461	26,15384	47,78783	50,25119	64,35083	
34	23,07692	35,38461	46,19083	54,80777	62,57078	58,60000
35	15,38461	26,15384	47,85795	47,85795	65,67971	
36	15,38461	26,15384	47,78783	50,25119	64,35083	
37	15,38461	26,15384	48,23607	50,92499	64,40558	
38	15,38461	26,15384	47,78783	50,25119	64,35083	
39	15,38461	27,69230	47,12758	51,60333	64,40064	
40	15,38461	26,15384	47,98228	50,40675	64,54794	
41	7,69230	15,76923	42,99937	42,93153	59,38252	
42	15,38461	26,15384	47,98228	50,40675	64,54794	
43	15,38461	26,15384	47,88106	50,36626	64,54175	
44	15,38461	26,15384	47,98228	50,40675	64,54794	
45	19,23076	31,92307	51,82843	56,17598	67,46148	61,00000
46	15,38461	25,38461	46,70584	48,74401	63,07822	
47	23,07692	31,53846	54,64798	54,88621	70,12090	59,20000
48	19,23076	27,69230	51,62600	51,86424	67,45358	
49	23,07692	29,23076	54,61746	53,06082	70,11140	58,20000
50	26,92307	43,46153	59,01895	67,11593	73,23522	70,00000
51	19,23076	31,92307	52,15082	56,40164	67,91044	60,20000
52	11,53846	21,53846	47,71064	49,91959	64,15466	
53	11,53846	21,53846	47,71064	49,91959	64,15466	
54	15,38461	27,30769	51,55680	55,68882	67,08528	56,90000
55	26,92307	44,61538	62,27108	72,17234	75,07743	71,00000
56	11,53846	21,53846	47,60943	49,87911	64,14843	
57	11,53846	21,53846	47,71064	49,91959	64,15466	
58	11,53846	21,53846	47,71064	49,91959	64,15466	
59	11,53846	21,53846	47,60943	49,87911	64,14843	

Caso	Casamento Simples	Casamento Ponderado	Vizinho Mais Próximo	Vizinho Mais Próx.Ponder	Distância Euclidiana	Apontada pelo especialista
60	11,53846	21,53846	48,12360	50,57708	65,04139	
61	15,38461	23,07692	51,96975	52,11554	67,93377	
62	30,76923	46,34615	66,73262	74,64157	78,33814	80,00000
63	11,53846	21,53846	48,32603	50,65805	65,04958	44,60000
64	11,53846	21,53846	48,32603	50,65805	65,04958	44,90000
65	11,53846	21,53846	48,22481	50,61757	65,04344	45,00000
66	11,53846	21,53846	48,32603	50,65805	65,04958	45,50000
67	26,92307	36,34615	59,04031	64,64157	73,26413	68,00000
68	11,53846	21,53846	48,32603	50,65805	65,04958	46,00000
69	11,53846	21,53846	48,32603	50,65805	65,04958	46,50000
70	15,38461	24,61538	50,52261	52,46361	67,64945	
71	26,92307	43,84615	61,95986	71,65389	75,69343	76,00000
72	15,38461	24,61538	50,52261	52,46361	67,64945	48,00000
73	15,38461	24,61538	50,52261	52,46361	67,64945	48,50000
74	15,38461	24,61538	50,32018	52,38264	67,64157	49,00000
75	23,07692	36,15384	57,39074	63,17790	72,99327	62,10000
76	15,38461	24,61538	50,42140	52,42312	67,64354	
77	84,61538	88,26923	86,19588	91,26829	92,29680	94,00000
78	15,38461	24,61538	50,42140	52,42312	67,64354	
79	15,38461	24,61538	50,52261	52,46361	67,64945	
80	11,53846	21,53846	50,86209	52,65648	68,08803	
81	11,53846	21,53846	50,86209	52,65648	68,08803	
82	11,53846	21,53846	50,86209	52,65648	68,08803	
83	11,53846	21,53846	50,86209	52,65648	68,08803	
84	11,53846	21,53846	50,86209	52,65648	68,08803	
85	11,53846	21,53846	50,86209	52,65648	68,08803	
86	11,53846	21,53846	50,86209	52,65648	68,08803	
87	11,53846	21,53846	50,76087	52,61599	68,08216	
88	11,53846	21,53846	50,86209	52,65648	68,08803	
89	11,53846	21,53846	50,86209	52,65648	68,08803	
90	11,53846	21,53846	48,46999	50,43009	64,89549	
91	11,53846	21,53846	48,46999	50,43009	64,89549	
92	11,53846	21,53846	48,46999	50,43009	64,89549	
93	11,53846	21,53846	48,26756	50,34912	64,88728	
94	11,53846	21,53846	48,26756	50,34912	64,88728	
95	11,53846	21,53846	46,22241	48,67251	62,20049	
96	11,53846	21,53846	46,22241	48,67251	62,20049	
97	11,53846	21,53846	46,69517	49,00344	62,78075	
98	11,53846	21,53846	46,79638	49,04393	62,78712	
99	11,53846	21,53846	46,20293	48,65013	62,35434	
100	15,38461	29,23076	51,23368	57,12546	67,73090	60,00000
101	15,38461	25,38461	50,75426	54,29101	67,49549	59,50000
102	11,53846	21,53846	47,91924	51,19959	64,82215	
103	11,53846	21,53846	47,37437	51,21882	64,71741	
104	19,23076	27,30769	51,14720	54,98334	67,49902	58,05000
105	11,53846	21,53846	45,65317	47,91576	64,36201	
106	7,69230	15,76923	41,01337	43,08115	60,02437	
107	15,38461	18,65384	47,80208	44,49105	66,05669	
108	11,53846	21,53846	44,96732	48,10937	63,14600	
109	7,69230	11,92307	34,82076	31,81586	52,78993	
110	11,53846	21,53846	46,99225	48,28422	64,28637	
111	11,53846	21,53846	47,09347	48,32470	64,29258	
112	11,53846	21,53846	47,09347	48,32470	64,29258	
113	11,53846	21,53846	46,89104	48,24373	64,28430	
114	15,38461	23,26923	49,06275	49,85939	66,53626	
115	19,23076	24,61538	53,72865	50,97877	69,87638	
116	69,23076	79,42307	86,84267	91,66602	92,55666	97,00000
117	11,53846	21,53846	47,81643	49,10839	64,42358	
118	11,53846	21,53846	47,91764	49,14888	64,42979	
119	11,53846	23,46153	46,79079	50,71500	63,78360	
120	11,53846	21,53846	46,80578	49,28338	64,31923	
121	7,69230	15,76923	40,38826	41,08889	58,22704	
122	11,53846	17,50000	46,80730	43,42101	63,05525	
123	15,38461	23,07692	45,10525	46,64993	62,68221	
124	11,53846	21,53846	44,01914	44,97452	60,39757	
125	19,23076	27,69230	49,65269	50,50187	66,72055	
126	19,23076	29,23076	46,46661	51,47086	62,68792	
127	3,84615	6,15384	29,56944	27,14676	48,20748	
128	15,38461	29,23076	39,87656	49,93258	58,73313	
129	11,53846	21,53846	42,72167	43,81904	59,54796	
130	11,53846	21,53846	45,66612	46,92502	62,39118	

Caso	Casamento Simples	Casamento Ponderado	Vizinho Mais Próximo	Vizinho Mais Próx.Ponder	Distância Euclidiana	Apontada pelo especialista
131	11,53846	21,53846	45,66612	46,92502	62,39118	
132	57,69230	70,96153	75,16241	81,76777	85,60146	87,50000
133	11,53846	21,53846	45,46369	46,84405	62,38264	
134	15,38461	27,30769	51,24354	54,19987	67,64053	58,00000
135	42,30769	49,61538	67,09788	70,48110	79,69642	68,00000
136	11,53846	21,53846	45,56490	46,88454	62,38477	
137	11,53846	21,53846	45,66612	46,92502	62,39118	
138	11,53846	21,53846	45,46369	46,84405	62,38264	
139	19,23076	35,00000	53,35842	60,38656	68,27803	66,20000
140	38,46153	44,03846	62,43481	65,49965	75,20804	67,40000
141	11,53846	21,53846	44,59560	45,64041	61,00776	
142	11,53846	17,50000	44,49439	41,56146	61,00121	
143	11,53846	21,53846	44,59560	45,64041	61,00776	
144	30,76923	35,00000	57,45022	58,25242	71,69005	63,60000
145	11,53846	21,53846	44,49439	45,59992	61,00121	
146	15,38461	25,38461	48,44176	49,48656	64,08246	
147	15,38461	25,38461	48,44176	49,48656	64,08246	
148	11,53846	15,38461	44,59560	39,48656	61,00776	
149	15,38461	25,38461	48,34054	49,44607	64,07622	
150	15,38461	25,38461	49,48663	50,74041	65,36689	
151	11,53846	21,15384	45,64048	46,50964	62,35555	
152	15,38461	25,38461	49,28420	50,65943	65,35874	
153	11,53846	19,61538	45,53926	44,93069	62,34914	
154	15,38461	25,38461	49,38542	50,69992	65,36077	
155	15,38461	25,38461	49,38542	50,69992	65,36077	
156	19,23076	35,00000	52,53913	59,77582	67,54439	67,00000
157	11,53846	21,53846	44,84683	46,31428	61,58745	
158	23,07692	30,57692	52,53913	55,35274	67,54439	57,20000
159	11,53846	21,53846	44,84683	46,31428	61,58745	
160	11,53846	13,46153	42,02562	36,23116	59,22132	
161	11,53846	21,53846	44,74561	46,27379	61,58096	
162	11,53846	21,53846	44,16947	45,50146	60,67938	
163	11,53846	21,53846	44,16947	45,50146	60,67938	
164	11,53846	21,53846	44,06826	45,46097	60,67279	
165	11,53846	21,53846	44,16947	45,50146	60,67938	
166	11,53846	21,53846	44,44420	45,77618	60,72787	
167	19,23076	25,00000	50,73216	47,95483	66,47164	
168	11,53846	21,53846	44,34299	45,73570	60,72129	
169	11,53846	21,53846	44,44420	45,77618	60,72787	
170	11,53846	21,53846	44,44420	45,77618	60,72787	
171	11,53846	21,53846	45,92981	47,72085	63,13226	
172	11,53846	21,53846	46,13223	47,80182	63,14070	
173	15,38461	15,57692	46,13223	41,84028	63,14070	
174	11,53846	21,53846	46,13223	47,80182	63,14070	
175	11,53846	21,53846	46,13223	47,80182	63,14070	
176	15,38461	21,15384	49,46890	46,75477	66,51526	
177	11,53846	21,53846	46,92588	48,38180	63,85941	
178	11,53846	21,53846	46,92588	48,38180	63,85941	
179	11,53846	21,53846	45,84497	47,55755	62,68340	
180	11,53846	21,53846	45,94618	47,59804	62,68977	
181	11,53846	21,53846	45,94618	47,59804	62,68977	
182	11,53846	21,53846	45,84497	47,55755	62,68340	
183	23,07692	32,69230	52,94805	58,26858	67,80288	58,10000
184	11,53846	21,53846	45,94618	47,59804	62,68977	
185	11,53846	21,53846	45,74375	47,51707	62,68127	
186	11,53846	21,53846	45,46903	47,24234	62,66562	
187	11,53846	21,53846	45,57024	47,28283	62,66774	
188	11,53846	21,53846	45,96732	48,02636	62,69652	
189	3,84615	6,15384	37,87794	31,89821	56,19622	
190	11,53846	21,53846	45,67146	47,32331	62,67412	
191	15,38461	29,23076	49,51761	55,01562	65,67085	60,00000
192	11,53846	21,53846	41,84575	43,01111	58,78297	
193	11,53846	23,46153	45,67146	49,24639	62,67412	
194	11,53846	21,53846	45,67146	47,32331	62,67412	
195	7,69230	15,76923	41,72409	41,51360	59,52000	
196	69,23076	73,65384	78,12336	83,13738	87,63747	85,00000

11.2 QUESTIONÁRIO DE RECUPERAÇÃO DE CASOS

Questionário

Após usar o aplicativo disponibilizado para testes, responda as perguntas abaixo, sempre em relação ao quesito facilidade de uso:

1) Na sua avaliação, qual das três rotinas de preenchimento de dados você considerou como a melhor:

Rotina A Rotina B Rotina C

2) Na sua avaliação, qual das três rotinas de preenchimento de dados você considerou como a pior:

Rotina A Rotina B Rotina C

3) Na sua avaliação, qual das três rotinas de preenchimento de dados você entende que apresenta o nível de facilidade de uso intermediário:

Rotina A Rotina B Rotina C

11.3 ER PARA UM SAD EM AUDITORIA INTERNA – REGISTRO DAS OBSERVAÇÕES

- Participante/observador: Alexandre Carvalho Dantas
- Período de observação: 04/07/2004 a 10/07/2004
- Etapa analisada: Levantamento de dados na pré-auditoria operacional
- Documentos analisados: Manual de Auditoria do Banco do Brasil S.A.
- Sistemas analisados: GDA – Gestão de Dados da Auditoria – sistema transacional utilizado diariamente pelo auditores do Banco do Brasil S.A.

Observações principais: Rotina: Para o levantamento de dados na pré-auditoria de cada risco/objeto a ser auditado o auditor deve levantar pelo menos os seguintes dados:

a) Identificação do Objeto de Auditoria

b) Ambiente Externo

- Legislação aplicável;
- Identificação do público-alvo;
- Fatores econômicos que afetam a comercialização e operacionalização;
- Atuação da concorrência;
- Posição no mercado;
- Atuação dos Órgãos reguladores (CVM, Banco Central, TCU, Secretaria Federal de Controle Interno, Auditoria Externa);
- Tendências do mercado - oportunidades e ameaças que podem influenciar o desempenho esperado;

c) Ambiente Interno

- Planejamento Estratégico - aderência à Estratégia Corporativa
- Normatização Interna
- Informações Econômico-Financeiras
- Visão estratégica do gestor/responsável sobre o objeto da auditoria;
- Metas estabelecidas;
- Itens formadores de receita (tarifas, spread, taxas, etc.);

- Informação e Comunicação
 - Dados de bases corporativas;
 - Dados de auditorias anteriores;
 - Denúncias;
 - Recomendações de auditorias anteriores não implementadas;
 - Processos administrativos internos;
 - Relatórios de auditoria externa e de Órgãos de fiscalização (TCU, BACEN, etc.).
- d) Ambiente Operacional
- Clientes (internos e externos);
 - Responsáveis pela execução e supervisão;
 - Mecanismos de monitoramento;
 - Nível de automação;
 - Pontos fortes e pontos fracos (internos da Organização).

Apesar das informações serem sempre iguais, a forma de pesquisa se altera a cada trabalho devido a diversidade de sistemas internos e externos usados para cada processo. Assim a automatização se mostra impraticável.

2) Funcionalidades básicas necessárias no sistema:

- i) Rotina de Consulta a base de casos;
- ii) Rotina de inclusão de casos no BC;
- iii) Rotina de exclusão de casos obsoletos do BC;
- iv) Rotina de emissão de relatórios;

3) Requisitos não-funcionais desejáveis

- a) Controle de acessos;
- b) Simplicidade de uso;
- c) Possibilidade de implementação gradual dos recursos.

11.4 ER PARA UM SAD EM AUDITORIA INTERNA - NECESSIDADES DO SISTEMA

Identificação do Participante:	
Nome:	
Cargo:	Local de trabalho:

Esse questionário estará dividido em cinco partes. Cada pergunta estará sempre relacionada à parte anteriormente identificada:

Parte 1: Características gerais do sistema:

1.1) O sistema deve restringir o acesso:

- Somente a auditores autorizados especificamente para o sistema
- Somente a funcionários vinculados a auditoria
- Não deve restringir o acesso.

1.2) Os recursos do sistema devem ser implementados mesmo que com isso o seu uso fique mais complexo ?

- Sim, deve-se buscar um sistema bem completo. Se necessário, os funcionários aprenderão o seu uso através de treinamento
- Não, deve-se inicialmente oferecer apenas recursos de fácil entendimento para todos os usuários

Parte 2: Rotina de Consulta a base de casos:

2.1) Indique os campos que devem fazer parte da rotina de consulta a casos de forma não obrigatória:

- Risco/objeto auditado Gestor envolvido Processo auditado
- Unidade auditada Data do trabalho Métodos adotados
- Resultados atingidos Auditor responsável Gerência responsável
- Documentos consultados Instruções internas consultadas
- Particularidades do trabalho Dicas

2.2) Indique os campos que devem fazer parte da rotina de consulta a casos de forma obrigatória:

- Risco/objeto auditado Gestor envolvido Processo auditado
 Unidade auditada Data do trabalho Métodos adotados
 Resultados atingidos Auditor responsável Gerência responsável
 Documentos consultados Instruções internas consultadas
 Particularidades do trabalho Dicas

2.3) A rotina de consulta deve ser restrita:

- A funcionários vinculados a auditoria
 A funcionários vinculados a auditoria autorizados para essa rotina.

2.4) Para execução da consulta, quantos campos devem ser preenchidos (lembrar que a ausência de um número mínimo gera resultados pobres e degrada o desempenho do sistema):

- Não deve ser fixado um número mínimo
 No mínimo dois
 No mínimo três
 No mínimo quatro
 No mínimo cinco

2.5) Após a execução da consulta, quantos casos semelhantes devem ser exibidos:

- Não deve ser fixado um número máximo
 No máximo 100
 No máximo 50
 No máximo 10
 No máximo 3

2.6) Descreva outras necessidades relacionadas a rotina de consulta de casos que você gostaria que o sistema apresentasse.

Parte 3: Rotina de cadastramento de casos no BC;

3.1) A rotina de cadastramento deve ser restrita:

- A funcionários vinculados a auditoria
 A funcionários vinculados a auditoria autorizados especificamente para essa rotina.

3.2) Indique os campos que devem ser obrigatórios no cadastramento de novos casos:

- Risco/objeto auditado Gestor envolvido Processo auditado
 Unidade auditada Data do trabalho Métodos adotados
 Resultados atingidos Auditor responsável Gerência responsável
 Documentos consultados Instruções internas consultadas
 Particularidades do trabalho Dicas

3.3) É conveniente que exista algum tipo de controle sobre o cadastramento para impedir a inclusão de casos pouco relevantes ou repetidos. Você tem alguma sugestão de como isso pode ser feito de forma prática ?

3.4) Os cadastramentos realizados devem manter registro sobre o responsável e data?

- Sim
 Não

3.5) Se sim, por quanto tempo ?

3.7) Deve haver limite máximo de data de ocorrência de um caso para que ele seja cadastrado ?

() Sim () Não

3.8) Se sim, por quanto tempo ?

3.9) Descreva outras necessidades relacionadas a rotina de cadastramento de casos que você gostaria que o sistema apresentasse.

Parte 4 – Exclusão de casos

4.1) A rotina de exclusão deve ser restrita:

() A funcionários vinculados a auditoria

() A funcionários vinculados a auditoria autorizados para essa rotina.

4.2) É conveniente que exista algum tipo de controle sobre a exclusão (além do controle de usuário) para impedir exclusões indevidas. Você tem alguma sugestão de como isso pode ser feito de forma prática ?

4.3) As exclusões realizadas devem manter registro sobre o responsável, data, etc.?

() Sim () Não

4.4) Se sim, por quanto tempo ?

4.5) Deve haver limite máximo de data de ocorrência para permanência de um caso na base de casos ?

() Sim

() Não

4.6) Se sim, por quanto tempo ?

4.7) Se sim, a exclusão deve ser automática quando o prazo for atingido ?

4.8) Descreva outras necessidades relacionadas à rotina de exclusão de casos que você gostaria que o sistema apresentasse.

Parte 5 - Rotina de emissão de relatórios;

5.1) Descreva quais os tipos de relatórios que o sistema deve oferecer, informando as informações que devem ser exibidas, quem pode utilizá-lo, etc.

11.5 LISTA DE EVENTOS DO SISTEMA AUDIT-AJUD

Todos os acontecimentos externos que geram algum tipo de ação de resposta em um sistema podem ser chamados de eventos.

Nessa versão inicial do AUDIT-AJUD, pode-se listar os seguintes eventos:

- Usuário cadastra novos usuários no controle de senhas.
- Usuário cadastra os parâmetros de trabalho.
- Usuário importa base de casos
- Usuário cadastra casos na base de casos.
- Usuário cadastra Auditores
- Usuário cadastra resultados obtidos
- Usuário configura pesos e atributos
- Usuário informa dados do Novo Caso a ser pesquisado
- Usuário obtém casos similares
- Usuário retém novos casos no BC

11.6 DIAGRAMA DE CONTEXTO DO SISTEMA AUDIT-AJUD

O diagrama de contexto é uma representação gráfica das trocas de informações entre o sistema e o ambiente. No AUDIT-AJUD o diagrama de contexto fica da seguinte forma:

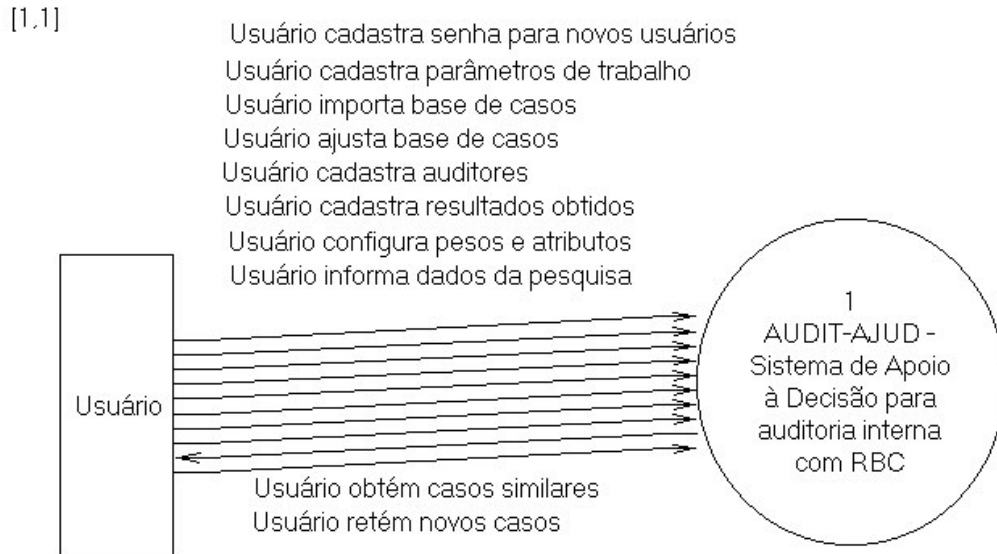


Figura 26: Diagrama de Contexto do AUDIT-AJUD gerado pelo CASE Studio®

11.7 DIAGRAMAS DE FLUXO DE DADOS (DFD) DO SISTEMA AUDIT-AJUD

O Diagrama de Fluxo de Dados é uma forma gráfica de representação, usada para descrever como os dados fluem pelos processos de um sistema. Ele mostra os componentes funcionais e as interfaces de dados entre eles, representando assim todo o funcionamento do aplicativo.

O conjunto de DFDs apresentado a seguir é chamado de DFD Preliminar e nele cada evento da lista de eventos representa um processo do sistema.

[1.1]

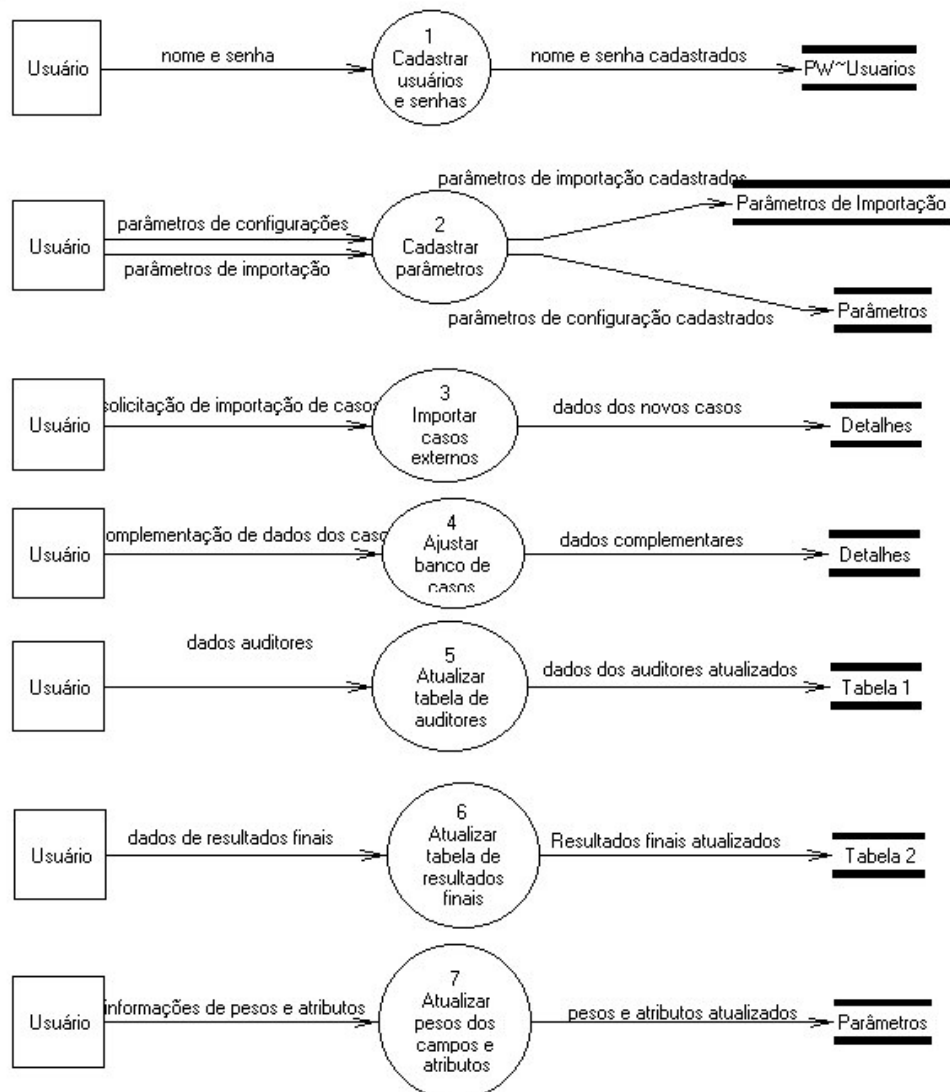


Figura 27: Diagramas de Fluxo de Dados do Cadastro gerados pelo CASE Studio®

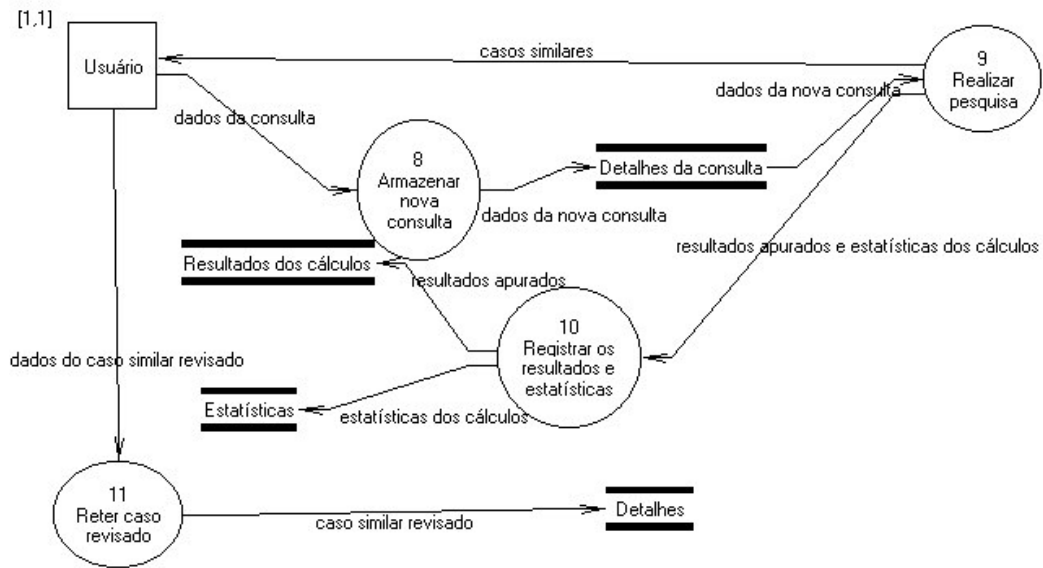


Figura 28: Diagrama de Fluxo de Dados da Rotina de Recuperação de Casos gerado pelo CASE Studio®

11.8 DESCRIÇÃO DOS PROCESSOS DO SISTEMA AUDIT-AJUD

Para facilitar o entendimento do funcionamento do AUDIT-AJUD, temos a seguir uma descrição textual resumida dos principais processos do aplicativo. Para uma descrição mais detalhada de todas as rotinas e recursos disponibilizados no AUDIT-AJUD, consulte o Módulo de Ajuda do Sistema e o Manual do Usuário impresso que está sendo disponibilizado em conjunto com essa pesquisa para consultas.

Processo 1 - Cadastrar usuários e senhas

Nesse processo um usuário administrador acessa a rotina onde o novo usuário informa o nome e senha (repetindo a senha no campo apropriado para confirmação). O sistema verifica se os dados são válidos e cadastra o usuário.

Processo 2 - Cadastrar parâmetros

No processo cadastrar parâmetros, o usuário preenche os diversos valores necessários para o funcionamento do sistema. O processo verifica os campos preenchidos e emite uma mensagem de alerta caso os campos necessários estejam incompletos. Esse processo permite ao sistema realizar os cálculos de similaridade e recuperação de casos e também que as regras de utilização (como obrigatoriedade de campos ou limites de intervalo) sejam definidas.

Processo 3 - Importar casos externos

Nesse processo um usuário transfere dados de uma tabela usada em outros sistemas transacionais de auditoria para o BC do AUDIT-AJUD. A rotina importa os campos coincidentes da tabela de origem para a tabela de destino (que é o BC do AUDIT-AJUD) deixando os campos não coincidentes intactos.

Processo 4 - Ajustar banco de casos

No processo ajustar banco de casos o usuário informa o valor de alguns campos não existentes no arquivo de origem (como a descrição da situação anterior à auditoria, por exemplo). Em seguida o processo testa se todos os

campos obrigatórios estão preenchidos, emitindo uma mensagem de alerta em caso de inconsistência.

Processo 5 - Atualizar tabela de auditores

Nesse processo o usuário informa o nome do auditor e o sistema verifica se os dados são válidos e se o auditor não existe. Caso positivo, cadastra o auditor.

Processo 6 - Atualizar tabela de resultados obtidos

Nesse processo o usuário informa o resultado obtido e o sistema verifica se o resultado obtido é válido e não existe na tabela. Caso positivo, cadastra o resultado obtido.

Processo 7 – Atualizar pesos dos campos e atributos

No processo atualizar pesos dos campos e atributos, o usuário preenche os diversos valores necessários para a realização dos cálculos de similaridade fundamentais para as pesquisas do sistema. O processo verifica os campos preenchidos e emite uma mensagem de alerta caso os campos necessários estejam incompletos/incorretos.

Processo 8 – Armazenar nova consulta

Essa é a rotina através da qual o auditor poderá localizar casos similares existentes na base de casos. Para executá-la o usuário deve preencher em todos os campos, as informações necessárias para que o caso de consulta seja armazenado. O sistema fará em seguida uma checagem dos dados preenchidos, emitindo uma mensagem de alerta em caso de inconsistência.

Processo 9 – Realizar pesquisas

Nesse processo, o usuário apenas informa o número da consulta que deverá ser usada como base para a pesquisa e o processo realiza todos os cálculos. Como o próprio aplicativo alerta, essa rotina pode ser demorada de acordo com o equipamento e o número de registros existentes no BC, devido ao grande volume de cálculos envolvidos.

Processo 10 – Registrar os resultados e estatísticas

Ao termino dos cálculos, todos os resultados são gravados na tabela de Resultados dos Cálculos e Estatísticas automaticamente pelo sistema. Em seguida, será exibida uma consulta com os 10 resultados ordenados por similaridade (do mais similar para o menos similar).

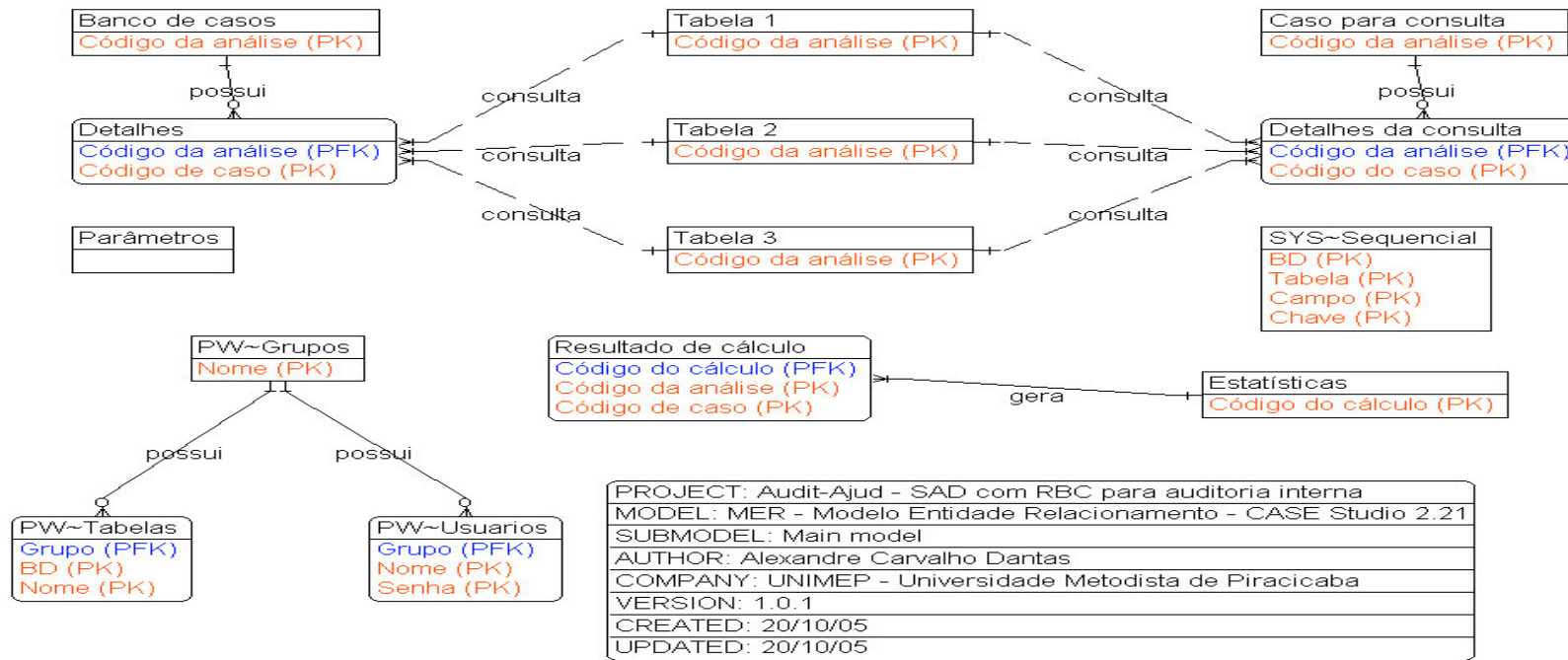
Processo 11 – Reter caso revisado

Em seguida a análise dos resultados, o usuário que tiver autorização pode após revisar o novo caso, atualizar o BC com o armazenamento desses novos registros para melhoria dos dados armazenados no BC.

11.9 MODELO ENTIDADE-RELACIONAMENTO (MER) DO SISTEMA AUDIT-AJUD

O MER é um modelo baseado na percepção do mundo real, com o objetivo de facilitar o projeto do banco de dados, permitindo a especificação de uma estrutura lógica do banco de dados. O modelo foi proposto inicialmente por Chen em 1977 e tem três conceitos básicos: a entidade, o relacionamento e o atributo. Posteriormente outras extensões foram propostas.

[1.1]



[2.1]

[1.2]

[2.2]

Figura 29: Modelo Entidade Relacionamento do AUDIT-AJUD gerado pelo CASE Studio®

11.10 DICIONÁRIO DE DADOS DO SISTEMA AUDIT-AJUD

O dicionário de dados é como um resumo centralizado de todas os significados das informações usadas na estruturação e construção do sistema. Possui a especificação de arquivos e/ou tabelas, de processos e de fluxo de dados modelados nos DFD.

Pelo grande número de informações existentes, a descrição das tabelas de dados, dos atributos, gráficos, relatórios e telas do aplicativo estão disponíveis no manual de documentação técnica do sistema, disponibilizado para pesquisa.

11.11 RESULTADOS DOS TESTES COMPARATIVOS

As tabelas a seguir listam os resultados de todos os testes realizados para avaliação do AUDIT-AJUD.

Amostra Piloto - Apuração do tempo médio de recuperação de casos

Quantidade analisada	Caso selecionado aleatoriamente	Tempo de recuperação do caso	(tempo - média) ²
1	4	33	239,2832128
2	13	8	97,38685768
3	35	14	10,39671331
4	43	10	58,83653006
5	66	39	460,9085954
6	80	9	73,74395946
7	81	7	101,1087247
8	88	16	3,387831187
9	89	16	2,236915938
10	99	8	90,37974494
11	100	42	598,7212867
12	108	16	3,760575597
13	109	8	99,70561055
14	118	7	104,7683835
15	179	8	83,72322263
16	187	14	11,83320671
17	214	17	0,337228616
18	258	17	0,008686264
19	287	17	0,048659384
20	288	14	10,91625748
21	301	8	85,57244399
22	309	11	41,35644491
23	315	19	1,567005563
24	317	15	8,792546231
25	322	38	418,9710316
26	329	18	0,276123436
27	335	9	67,522346
28	350	8	100,1100267
29	355	62	1977,472562
30	363	17	0,002925472
	Média	18	
	Soma		4753,135659
	Variância		163,9012296

Amostra Final - Apuração do tempo médio de recuperação de casos

Quantidade analisada	Caso selecionado aleatoriamente	Tempo de recuperação do caso
1	7	14
2	8	18
3	15	14
4	20	42
5	21	12
6	37	19
7	44	11
8	46	7
9	48	27
10	53	18
11	67	11
12	79	36
13	101	12
14	105	11
15	126	61
16	141	10
17	159	7
18	168	15
19	169	13
20	173	14
21	191	7
22	200	19
23	208	10
24	209	6

Amostra Final - Apuração do tempo médio de recuperação de casos

Quantidade analisada	Caso selecionado aleatoriamente	Tempo de recuperação do caso
25	210	15
26	212	7
27	222	19
28	223	12
29	224	14
30	230	19
31	246	8
32	254	12
33	255	8
34	262	19
35	265	23
36	277	10
37	279	16
38	289	9
39	295	14
40	305	9
41	306	7
42	307	15
43	312	7
44	314	12
45	316	19
46	324	9
47	325	15
48	326	17
49	339	13
50	341	17
51	356	15
52	357	19
53	360	10
54	362	8
55	368	14
56	380	14
57	385	14
58	388	18
59	390	31
60	399	11
Média		15,2167

Amostra Piloto - Apuração da proporção de precisão na recuperação de casos

Quantidade analisada	Caso selecionado aleatoriamente	Resultado da recuperação
1	6	Adequado
2	12	Adequado
3	38	Adequado
4	50	Adequado
5	61	Inadequado
6	65	Adequado
7	74	Adequado
8	78	Adequado
9	83	Adequado
10	97	Adequado
11	122	Adequado
12	130	Adequado
13	160	Adequado
14	170	Adequado
15	171	Adequado
16	179	Adequado
17	190	Adequado
18	209	Adequado
19	236	Adequado
20	271	Adequado
21	278	Adequado
22	311	Adequado
23	326	Adequado
24	328	Adequado
25	348	Adequado
26	365	Adequado
27	373	Adequado
28	375	Adequado

Amostra Piloto - Apuração da proporção de precisão na recuperação de casos

Quantidade analisada	Caso selecionado aleatoriamente	Resultado da recuperação
29	393	Adequado
30	398	Adequado
	% Adequado	96,67%
	% Inadequado	3,33%
	Variância	3,22%

Amostra Final - Apuração da proporção de precisão na recuperação de casos

Quantidade analisada	Caso selecionado aleatoriamente	Resultado da recuperação
1	23	Adequado
2	41	Adequado
3	45	Adequado
4	47	Adequado
5	55	Adequado
6	63	Adequado
7	68	Adequado
8	73	Adequado
9	85	Adequado
10	99	Adequado
11	102	Inadequado
12	104	Adequado
13	112	Adequado
14	125	Adequado
15	134	Adequado
16	144	Adequado
17	151	Adequado
18	152	Adequado
19	154	Adequado
20	155	Adequado
21	161	Adequado
22	182	Adequado
23	187	Adequado
24	200	Adequado
25	213	Adequado
26	220	Adequado
27	221	Adequado
28	237	Adequado
29	248	Adequado
30	261	Adequado
31	262	Adequado
32	263	Adequado
33	277	Inadequado
34	281	Adequado
35	286	Adequado
36	294	Adequado
37	295	Adequado
38	306	Adequado
39	325	Adequado
40	364	Adequado
41	365	Adequado
42	376	Adequado
43	377	Adequado
44	384	Adequado
45	386	Adequado
	% Adequado	95,56%
	% Inadequado	4,44%

Amostra Piloto - Apuração da proporção de qualidade das informações dos casos recuperados

Quantidade analisada	Caso selecionado aleatoriamente	Resultado da recuperação
1	20	Superior
2	22	Inferior
3	28	Superior
4	60	Igual
5	62	Superior
6	88	Inferior

Amostra Piloto - Apuração da proporção de qualidade das informações dos casos recuperados

Quantidade analisada	Caso selecionado aleatoriamente	Resultado da recuperação
7	89	Superior
8	95	Superior
9	97	Igual
10	100	Superior
11	132	Igual
12	163	Superior
13	187	Superior
14	203	Igual
15	206	Igual
16	225	Superior
17	231	Superior
18	255	Superior
19	259	Inferior
20	260	Superior
21	261	Superior
22	275	Superior
23	284	Superior
24	297	Inferior
25	306	Superior
26	332	Superior
27	343	Inferior
28	370	Superior
29	388	Inferior
30	398	Superior
	% Superior	63,33%
	% Inferior	36,67%
	Variância	23,22%

Amostra Final - Apuração da proporção de qualidade das informações dos casos recuperados

Quantidade analisada	Caso selecionado aleatoriamente	Resultado da recuperação
1	13	Superior
2	21	Inferior
3	24	Superior
4	32	Igual
5	39	Superior
6	40	Inferior
7	41	Superior
8	42	Superior
9	43	Superior
10	49	Superior
11	51	Igual
12	55	Superior
13	57	Superior
14	73	Igual
15	77	Igual
16	80	Superior
17	110	Superior
18	111	Superior
19	129	Inferior
20	135	Superior
21	140	Superior
22	154	Superior
23	155	Superior
24	156	Inferior
25	157	Superior
26	160	Superior
27	168	Inferior
28	169	Superior
29	173	Inferior
30	175	Igual
31	179	Igual
32	183	Superior
33	184	Inferior
34	185	Superior
35	186	Superior

Amostra Final - Apuração da proporção de qualidade das informações dos casos recuperados

Quantidade analisada	Caso selecionado aleatoriamente	Resultado da recuperação
36	187	Superior
37	190	Superior
38	192	Inferior
39	193	Superior
40	202	Inferior
41	204	Superior
42	216	Superior
43	218	Inferior
44	224	Superior
45	227	Inferior
46	254	Igual
47	257	Superior
48	262	Inferior
49	263	Superior
50	273	Superior
51	285	Igual
52	291	Superior
53	301	Inferior
54	307	Inferior
55	308	Inferior
56	324	Superior
57	330	Inferior
58	334	Superior
59	335	Inferior
60	338	Igual
61	340	Superior
62	353	Igual
63	357	Inferior
64	358	Superior
65	359	Superior
66	360	Inferior
67	370	Superior
68	371	Superior
69	374	Inferior
70	375	Superior
71	380	Superior
72	384	Igual
73	390	Superior
74	391	Superior
	% Superior	58,11%
	% Inferior	41,89%