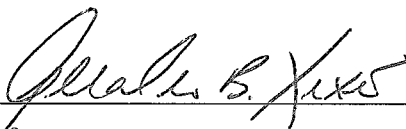


AVALIAÇÃO DA QUALIDADE DE DADOS PELA NÃO CONFORMIDADE

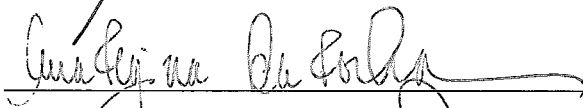
Selma Foline Crespio de Pinho

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

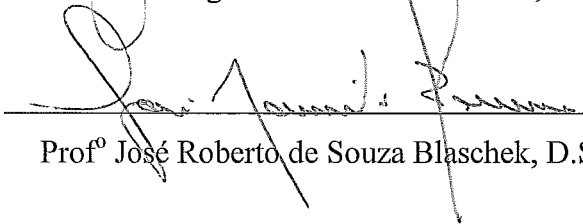
Aprovado por:



Prof^o Geraldo Bonorino Xexéo, D.Sc.



Prof^a Ana Regina Cavalcanti da Rocha, D.Sc.



Prof^o José Roberto de Souza Blaschek, D.Sc.

RIO DE JANEIRO, RJ - BRASIL
ABRIL DE 2001

PINHO, SELMA FOLIGNE CRESPIO

*Avaliação da Qualidade de Dados pela
Não Conformidade* [Rio de Janeiro] 2001

X, 110 p. 29,7 cm (COPPE/UFRJ, M.Sc.,
Engenharia de Sistemas e Computação, 2001)

Tese – Universidade Federal do Rio de
Janeiro, COPPE

1. Avaliação da Qualidade de Dados
2. Não Conformidade

I. COPPE/UFRJ II. TÍTULO (série)

Dedico esta conquista ao meu amado papai, Geraldo Foligne (*in memoriam*), e ao meu filho Leonardo.

Ao meu pai, meu porto seguro, que partiu durante esta minha jornada, levando um pouco do brilho que a vida tem. Quando perdemos alguém que amamos parece que uma parte de nós vai junto. A ele dedico também a minha gratidão, pois mesmo não estando mais entre nós, tenho certeza, nunca me abandonou.

Ao meu filhinho Leonardo, que me ensina o verdadeiro sentido da palavra amor. Que cresce forte e saudável e a cada dia me surpreende com coisas novas, me fazendo ficar maravilhada com a maneira com que ele vai descobrindo o mundo.

A esses dois que me ensinaram que o ciclo da vida tem que se cumprir, onde perdas irreparáveis e surpresas maravilhosas fazem parte da nossa história de vida.

Agradecimentos

Ao professor Geraldo Xexéo, meu orientador, pela confiança, estímulo e paciência a mim dedicado, mas sobretudo pelo exemplo de profissionalismo durante o desenvolvimento deste trabalho.

Aos professores Ana Regina e Blaschek pela honra de tê-los em minha banca examinadora.

Ao Centro de Análises de Sistemas Navais (CASNAV - Marinha do Brasil) por tornar possível este trabalho e acreditar no seu resultado.

Aos amigos do CASNAV Coronel Menezes, Fernando, Capitão-de-Fragata Garnier, Capitão-de-Corveta Mury e Sandra que empenharam seu trabalho me auxiliando durante os momentos em que precisei.

Aos funcionários da COPPE – Programa de Engenharia de Sistemas e Computação, especialmente, Solange, Cláudia, Sueli, Mercedes, Lúcia, Marli, Frederico, Sonia e Tia Gercina, que tornaram essa minha experiência acadêmica mais fácil e alegre.

À Maria Clícia Steling de Castro, professora da UFJF, que com uma carinhosa dedicação e competência muito me auxiliou na conclusão deste trabalho.

Aos meus companheiros Alexandre Soares, Ana Mirtes e Moises, que juntos passamos por tantos momentos alegres e outros difíceis, onde o entusiasmo, a força e a palavra amiga foram a grande fonte de energia que me ajudou a prosseguir nessa caminhada.

À minha mãe Jovelina e minha irmã Josélia pela base sólida e carinho que sempre estão prontas para me oferecer.

À Dona Mercedes e Sr. Delmo que se preocupam comigo como se eu fosse uma verdadeira filha, me dando suporte em todos os momentos.

Ao meu marido, Francisco, preciosa pessoa que Deus colocou no meu caminho, que muito me incentivou ao longo desta caminhada.

A Deus que sempre me indica o caminho certo a seguir e também por me ensinar que somos todos anjos de uma asa só. Onde nesse mundo só podemos voar quando abraçados uns aos outros.

Resumo da Tese apresentada a COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc)

AVALIAÇÃO DA QUALIDADE DE DADOS PELA NÃO CONFORMIDADE

Selma Foligne Crespino de Pinho

Abril / 2001

Orientador: Prof^o. Geraldo Bonorino Xexéo

Programa: Engenharia de Sistemas e Computação

Este trabalho apresenta um modelo compatível com a norma ISO 9126 para avaliação da qualidade de dados baseada em não conformidades. O modelo proposto possui três etapas. A primeira etapa utiliza a opinião de especialistas para estabelecer pesos que expressam a importância de cada característica de qualidade de dados. Assim, são capturadas a percepção e a expectativa de qualidade do especialista. A segunda etapa avalia como o banco de dados satisfaz a essa expectativa, usando um relacionamento pré-definido entre as características de qualidade e os tipos de não conformidades ocorridas durante a utilização do sistema. A terceira etapa, através de gráficos e relatórios, permite o acompanhamento da qualidade dos dados do sistema durante a sua existência. Para validar o modelo e demonstrar a sua utilização, o protótipo AQUA foi desenvolvido e utilizado em um caso real.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc)

DATA QUALITY EVALUATION
BY NON-CONFORMITY

Selma Foligne Crespino de Pinho

April / 2001

Advisor: Geraldo Bonorino Xexéo

Department: Computing and Systems Engineering

This work presents an ISO 9126 compatible model for data quality evaluation based on non-conformities. The proposed model has three steps. The first step uses expert opinion to establish weights that convey the importance of each data quality characteristic, capturing the expert's perception and expectation about data quality. The second step measures how the database satisfies the user, using a predefined relationship between quality characteristics and types of non-conformities found using the system. The third step uses graphs and reports to allow the monitoring of the data quality of a system during its existence. To validate the model and demonstrate its use, the AQUA prototype was developed and used in a real case.

1 INTRODUÇÃO.....	1
2 QUALIDADE	4
2.1 Enfoques sobre Qualidade.....	4
2.1.1 Qualidade do Produto	6
2.2 Qualidade de Dados	10
2.3 Avaliação da Qualidade pela Não Conformidade	13
3 CARACTERÍSTICAS DE QUALIDADE.....	15
3.1 Visão do Usuário	16
3.2 Processo de Validação das Características de Qualidade.....	17
3.3 Aplicação do Modelo Rocha	19
3.4 Características para Avaliação da Qualidade de Dados	20
3.5 Tipos de Erros Associados às Características	29
4 AVALIAÇÃO DA QUALIDADE DE DADOS PELA NÃO CONFORMIDADE.....	34
4.1 Modelo Matemático do Critério de Avaliação	34
4.2 Etapas da Avaliação da Qualidade de Dados pela Não Conformidade.....	52
5 UMA EXPERIÊNCIA DE UTILIZAÇÃO DA AVALIAÇÃO DA QUALIDADE DE DADOS PELA NÃO CONFORMIDADE	54
5.1 Implementação do Protótipo da Ferramenta para Avaliação da Qualidade de Dados pela Não Conformidade.....	54
5.1.1 Diagrama de Casos de Uso	55
5.1.2 Diagrama de Telas	58
5.1.3 Modelo Lógico dos Dado	59
5.2 Etapas da Experiência de Utilização da Avaliação	60
6 CONCLUSÕES	66
6.1 Perspectivas Futuras	67

REFERÊNCIAS BIBLIOGRÁFICAS	69
APÊNDICE I - Formulário de Validação das Características de Qualidade	77
APÊNDICE II - Formulário para Cadastro do Erro ou Não Conformidade	
Ocorrido	87
APÊNDICE III - Instrumento para Classificação das Características Necessárias	
para Avaliar a Qualidade de Dados	90
APÊNDICE IV – Manual para Utilização da Ferramenta AQUA	94
APÊNDICE V – Dicionário de Dados	107

Índice de Figuras

Figura 2.1: Estrutura do Modelo Rocha para Avaliação da Qualidade de Software	10
Figura 4.1: Esquema Geral do Modelo Rocha.....	36
Figura 4.2: Gráfico de Desempenho	40
Figura 4.3: Gráfico dos Índices de Qualidade	52
Figura 5.1: Diagrama de Casos de Uso.....	55
Figura 5.2: Diagrama de Telas.....	58
Figura 5.3: Modelo Lógico dos Dados	59
Figura 5.4: Relatório de Qualidade do Sistema CAM.....	65
Figura 5.5: Gráfico de Desempenho do Sistema CAM	65
Figura IV.1: Tela Principal da Ferramenta AQUA.....	97
Figura IV.2: Barra de Menu.....	97
Figura IV.3: Menu de Cadastros	98
Figura IV.4: Cadastro de Erros	98
Figura IV.5: Cadastro de Versão	99
Figura IV.6: Incluir Característica de Qualidade.....	100
Figura IV.7: Incluir Tipo de Problema	100
Figura IV.8: Menu Avaliação	101
Figura IV.9: Peso das Características	101
Figura IV.10: Tempo de Utilização	102
Figura IV.11: Tolerância.....	103
Figura IV.12: Menu Resultados.....	103
Figura IV.13: Gráfico de Desempenho	104
Figura IV.14: Relatório.....	105

Índice de Tabelas

Tabela 2.1: Características de Qualidade de Software.....	8
Tabela 2.2: Subcaracterísticas de Qualidade de Software	9
Tabela 3.1: Escala de Valores de Importância.....	17
Tabela 4.1: Escala de Pesos	39
Tabela 4.2: Correspondência dos Pesos.....	40
Tabela 4.3: Ocorrência de Erros Associados às Linhas	41
Tabela 4.4: Valores de cada Linha do Gráfico de Desempenho.....	41
Tabela 5.1: Não Conformidades ocorridas no Sistema CAM.....	61
Tabela 5.2: Peso das Características de Qualidade do Sistema CAM	62
Tabela I.1: Escala de Valores.....	78
Tabela II.1: Escala de Valores	88
Tabela II.2: Cadastro do Erro ou Não Conformidade.....	89
Tabela III.1: Escala de Valores.....	91
Tabela IV.1: Escala de Valores.....	99

Capítulo 1

INTRODUÇÃO

Nas empresas de manufatura ou de serviços não é mais possível negar que a qualidade é uma das principais preocupações dos seus responsáveis.

Para que a qualidade esteja presente em uma organização é preciso sobretudo vontade política de seus dirigentes e, para que isso aconteça, é necessário que estejam despertados para este propósito, proporcionando alicerce, orientação e apoio a todas as pessoas envolvidas com o produto. As atitudes começam de cima e, naturalmente, irão se disseminar para cada nível inferior na estrutura da organização (HRADESKY, 1989).

Buscando explorar as oportunidades do mercado, os sistemas de informações exercem um papel crucial para uma organização, pois apresentam visões do mundo real para que as pessoas na organização criem produtos ou adotem decisões acertadamente.

Sendo assim, as informações são elementos de alta prioridade que devem receber cuidados especiais. Assegurar a qualidade dos dados é uma medida para evitar que decisões errôneas possam ser adotadas e conseqüentemente perdas financeiras possam acontecer (BALDWIN e BOWEN, 1999). Ela evita também que dados incorretos influenciem negativamente na formação da opinião do cliente em relação à qualidade do produto ou do serviço.

Na chamada *era das organizações baseadas em conhecimento* ou *era do capitalismo intelectual* (BELASCO *et al.*, 1994) (PETERS, 1993), a informação é um recurso organizacional crítico e a preservação da memória organizacional um fator com prioridade alta. Assim, adotar uma estratégia para garantir a qualidade dos dados é fundamental. Nesta direção, a utilização de uma avaliação da qualidade de dados é uma prática que pode ser realizada de duas formas (WILLSHIRE e MEYEN, 1997):

- (i) Avaliação quantitativa (objetiva) do banco de dados. Nesta forma de avaliação são utilizados os indicadores objetivos para medir a qualidade da representação e da estrutura do banco de dados;

(ii) Avaliação qualitativa (subjetiva) do banco de dados. Nela são utilizados os indicadores subjetivos para julgar a qualidade do banco de dados e sua aptidão de uso. Neste tipo de avaliação os indicadores subjetivos são as características de qualidade que foram definidas, objetivando capturar a percepção e a expectativa do usuário sobre qualidade de dados.

Qualidade é um conceito complexo, porque possui significados diversos para diferentes pessoas. Portanto, gerar uma medida de qualidade aceitável para todos, não é trivial. Para isso é preciso definir as características de qualidade que interessam e, então, decidir como a avaliação da qualidade será feita através delas (KITCHENHAM *et al.*, 1996).

Neste trabalho identificamos o conjunto de características que é importante para determinar a qualidade dos dados armazenados. Cada característica é julgada por um avaliador, onde é estimado um grau de importância para cada uma delas, segundo a sua percepção ou o seu nível de entendimento da questão proposta.

Sendo a aptidão de uso e a qualidade da informação fornecida pelo banco de dados os alvos de uma avaliação qualitativa (subjetiva) de banco de dados (WILLSHIRE e MEYER, 1997), a visão do avaliador não é baseada no conhecimento dos aspectos de elaboração ou implementação do banco de dados que está avaliando. Ela é baseada no seu entendimento sobre a qualidade dos dados armazenados. Esse entendimento pode ser conseguido pela utilização dos serviços oferecidos pelo sistema (aplicação), pelos serviços oferecidos pelo banco de dados e pela informação que os dados geram. As informações geradas evidenciam o nível de qualidade do conteúdo armazenado.

Desta forma, a adoção de um processo de avaliação da qualidade dos dados vêm de encontro à necessidade de preservar a qualidade dos dados armazenados. Esses dados armazenados podem ser influenciados pelas não conformidades, gerando assim a degradação da qualidade. O conceito estatístico de não conformidade é definido da seguinte forma: Uma falha ou não conformidade de um produto é um ponto específico que não satisfaz a uma ou mais especificações do produto que foram feitas pelo usuário (MONTOMERY, 1996), (LOURENÇO FILHO, 1982) e (KUME, 1985).

Esta tese se propõe a realizar uma avaliação da qualidade dos dados pela não conformidade. Para isso, analisa as características subjetivas de qualidade de dados importantes para os usuários, considerando a forma como as não conformidades

ocorridas afetam a qualidade dos dados armazenados. É elaborada também, uma ferramenta que automaticamente valida e realiza a avaliação proposta.

O resultado obtido é a avaliação de como o sistema atual satisfaz às expectativas de qualidade dos usuários, através da taxa de qualidade de dados. Desta forma, é possível, de acordo com a política de qualidade de dados da organização, realizar ações de investigação das prováveis fontes de erros que degradam a qualidade dos dados armazenados.

Esta tese está organizada da seguinte forma.

O **Capítulo 2** discorre sobre qualidade abordando os aspectos necessários para se alcançar a qualidade e ressalta a sua importância para uma organização.

No **Capítulo 3**, estão relacionadas às características de qualidade de dados que foram identificadas como necessárias para avaliar qualitativamente um banco de dados. Além disso, relacionamos os tipos de erros associados a cada uma delas. O capítulo descreve ainda como a visão do usuário é formada e como a pesquisa de campo realizada conduziu à abordagem dos aspectos subjetivos de qualidade de dados importantes para os usuários.

O **Capítulo 4** descreve o modelo matemático desenvolvido que realiza o processo de obtenção do índice de qualidade dos dados armazenados a partir do cadastro das não conformidades ocorridas. Estão também descritas, as etapas necessárias para realizar integralmente a avaliação da qualidade de dados proposta.

No **Capítulo 5**, efetua-se uma experiência de uso da avaliação da qualidade de dados proposta, assim como, descreve a ferramenta automática desenvolvida, denominada AQUA, que foi utilizada para realizar o experimento.

No **Capítulo 6**, são apresentadas as conclusões deste trabalho e perspectivas futuras.

Capítulo 2

QUALIDADE

Considerar que a qualidade é uma meta a ser alcançada com uma perfeição absoluta e definitiva seria despende esforço para algo inatingível. A qualidade é algo factível, relativo, substancialmente dinâmico e evolutivo, amoldada aos objetivos a serem atingidos (BELCHIOR, 1997).

Em todos os negócios a correta utilização das informações é a chave do sucesso. Desta forma, os dados são recursos que as organizações dispõem para conduzir e gerenciar corretamente as transações entre empresas e indivíduos (STOKER, 2000). Assim, assegurar a qualidade dos dados é uma medida para evitar que decisões errôneas possam ser adotadas e, conseqüentemente, perdas financeiras possam acontecer (BALDWIN e BOWEN, 1999).

Neste capítulo abordamos a visão geral dos aspectos necessários para se alcançar a qualidade e ressaltamos a sua importância para uma organização.

2.1 Enfoques sobre Qualidade

Uma organização obtém vantagens decisivas em termos de concorrência quando tem funcionários e um processo de qualidade.

Para obter, então, estas vantagens é preciso reverter algumas visões como, por exemplo, a visão do fatalismo. Deve-se abolir discursos como: fluxos mais ou menos controlados, defeitos ou panes, fatalidade sobre modificações de última hora no produto. Durante muito tempo, as organizações limitaram-se a administrar da melhor maneira possível essas fatalidades que eram consideradas inerentes, não apenas ao sistema, mas aos homens também.

Para se obter qualidade é necessário executar uma mudança radical de postura, abandonar as atitudes fatalistas, passando a ter como objetivo a satisfação do cliente. É preciso criar uma dinâmica de progresso, transformar a resposta a um dado problema em uma oportunidade para se alcançar a confiança e a satisfação do cliente e torná-los armas poderosas, vantagens estratégicas decisivas no mercado. A qualidade, portanto, tem uma função de ser um diferencial para aumentar a participação no mercado e ter a fidelidade dos clientes (CHOW, 2000).

Segundo FALCONI (1992), a qualidade só é alcançada quando todos os componentes da empresa também atingiram bons padrões, onde cada pessoa se sente responsável pelo sucesso buscado e usa de todo o seu potencial mental em benefício da empresa.

A meta de qualidade não pode ser conseguida através de decreto, mas sim construída passo a passo, com determinação e visão a longo prazo. Para isso é preciso modificar uma cultura ainda muito arraigada em nossas organizações, que é o preconceito de que a qualidade custa caro. O custo com modificações, retoques, refugos, devoluções, reclamações ou crises alcançam um volume de negócio que leva os diretores das organizações a reflexões (STOKER, 2000).

Formalmente a qualidade pode ser definida de diversas maneiras. A seguir citamos algumas definições dadas por vários autores e entidades.

"Qualidade ... a gente sabe o que é, e, ao mesmo tempo, não sabe. Isso é contraditório. Mas algumas coisas são melhores que outras, ou seja, têm mais qualidade. Porém, se a gente tenta definir qualidade, isolando-as das coisas que a possuem, então já não há mais o que falar" (TSUKUMO *et al.*, 1997).

TAUSWORTHE (1995) e GENTLEMAN (1996) colocam que a qualidade pode variar com o contexto e ao longo do tempo, pois as pessoas mudam de opinião em relação a um objeto ou a uma questão. Sendo assim, a qualidade depende da perspectiva do avaliador. Ela representa a opinião de um grupo de indivíduos, que utilizam o produto.

Para SCHNEIDEWIND (1993) a responsabilidade sobre a obtenção da qualidade, deve ser apoiada na idéia de observar se os atributos ou características desejados foram alcançados. Esses atributos devem ser combinados de forma clara e definida, a fim de que deixem de ser alcançados somente por conta de uma intuição.

HUMPHREY (1995) define que a qualidade final é composta de várias fases anteriores que foram concluídas com alta qualidade. Para isso cada resultado intermediário é considerado como concluído somente após terem sido bem examinados.

Segundo WANG *et al* (1993) a qualidade deve servir de tema centralizador para se alcançar uma mudança em nossa cultura empresarial. Esse objetivo é alcançado pela prática diária do aperfeiçoamento, do desenvolvimento do *know-how* e dos recursos humanos e pela procura da satisfação eficaz e eficiente dos clientes bem servidos.

Já ISHIKAWA (1984) menciona que a qualidade começa pela educação e acaba na educação. Uma empresa que progride em qualidade é uma empresa que aprende a aprender.

A norma ISO 8402 (1994) define a qualidade como a "Totalidade de características de uma entidade que lhe confere a capacidade de satisfazer às necessidades explícitas e implícitas". Além disso, coloca que necessidades explícitas são aquelas expressas na definição de requisitos propostos pelo produtor. Elas são compostas pelas condições de utilização do produto, seus objetivos, funções e desempenho esperado. Já as necessidades implícitas, são aquelas que, embora não expressas nos documentos do produtor, são necessárias para o usuário.

Numa definição de qualidade aplicada a software, ROCHA (1987) define qualidade de software como sendo um conjunto de propriedades a serem satisfeitas, em determinado grau, de modo que o software satisfaça às necessidades de seus usuários.

A demanda por qualidade de forma geral tem motivado a comunidade de software a desenvolver modelos específicos para avaliar a qualidade de um produto de software, conforme descrito na próxima seção.

2.1.1 Qualidade do Produto

A qualidade de um produto de software é resultante das atividades realizadas no processo de desenvolvimento do produto. Assim, o modo como as atividades são desenvolvidas é fundamental para a obtenção de um produto de qualidade.

A qualidade da equipe de desenvolvimento e o curto tempo disponível para a equipe trabalhar no desenvolvimento do produto são fatores que podem fazer com que o

fato de se ter um bom método de desenvolvimento, não garante que o produto seja de qualidade (STRIGINI, 1996).

Para avaliar a qualidade de um produto de software é preciso verificar, através de técnicas e atividades operacionais o quanto os requisitos que compõem este software são atendidos. Tais requisitos, de uma maneira geral, são a expressão das necessidades dos usuários, explicitadas em termos quantitativos ou qualitativos, e têm por objetivo definir as características de um software.

Assim, algumas abordagens surgiram para definir as etapas necessárias para se avaliar a qualidade de um produto de software, como a norma ISO 9126 (1991) e o modelo ROCHA (1983). Estas duas abordagens estão definidas a seguir.

Norma ISO 9126

A norma ISO 9126 (1991) fornece seis características subdivididas em subcaracterísticas que possibilitam avaliar a capacidade de um produto de software em satisfazer aos itens que foram especificados como necessários. Além disso, estas características possibilitam avaliar àquelas necessidades implícitas que devem ser identificadas e definidas, em conformidade com a definição de qualidade feita na norma ISO 8402 (1994)

As definições das características e subcaracterísticas que avaliam a qualidade de um software, descritas em anexo da norma ISO 9126, estão na Tabela 2.1 e 2.2, respectivamente.

CARACTERÍSTICAS	DESCRIÇÃO
Funcionalidade	Evidencia que o conjunto de funções atende às necessidades explícitas e implícitas para a finalidade a que se destina o produto.
Confiabilidade	Evidencia que o desempenho se mantém ao longo do tempo e em condições estabelecidas.
Usabilidade	Evidencia a facilidade para a utilização do software.
Eficiência	Evidencia que os recursos e os tempos envolvidos são compatíveis com o nível de desempenho requerido para o produto.
Manutenibilidade	Evidencia que há facilidades para correções, atualizações e alterações.
Portabilidade	Evidencia que é possível utilizar o produto em diversas plataformas com pequeno esforço de adaptação.

Tabela 2.1 – Características de Qualidade de Software

CARACTERÍSTICAS	SUBCARACTERÍSTICAS	DESCRIÇÃO
Funcionalidade	Adequação	Produto de software ter o conjunto de funções que são apropriadas para as tarefas exigidas
	Acurácia	Geração de resultados e efeitos corretos
	Interoperabilidade	Capacidade de interagir com outros produtos
	Conformidade	Estar de acordo com normas, convenções e regulamentações.
Confiabilidade	Segurança de Acesso	Capacidade de evitar acesso não autorizado a programas e dados
	Maturidade	Medido pela baixa frequência de falhas do software
	Tolerância a Falhas	Manter o nível de desempenho em caso de falha
Usabilidade	Recuperabilidade	Utilização de recursos que possibilitam restabelecer e restaurar dados após falha
	Inteligibilidade	Facilidade oferecida ao usuário em entender os conceitos utilizados
	Aprensibilidade	Facilidade de aprendizado
Eficiência	Operacionalidade	Facilidade de operar e controlar o produto de software
	Comportamento em relação ao tempo	Tempo de resposta (de processamento) das funções solicitadas
	Comportamento em relação aos recursos	Quantidade de recursos utilizados para a execução das tarefas

Manutenibilidade	Analisabilidade	Facilidade de diagnosticar deficiências e causas de falhas
	Modificabilidade	Facilidade de modificação e remoção de defeitos
	Estabilidade	Ausência de riscos de efeitos inesperados associados a alterações feitas
	Testabilidade	Facilidade de o produto ser testado
Portabilidade	Adaptabilidade	Capacidade de ser adaptado a ambientes diferentes
	Capacidade para ser instalado	Facilidade para executar a instalação em ambientes específicos
	Conformidade	Acordo com padrões ou convenções de portabilidade
	Capacidade para substituir	Facilidade em substituir outro software

Tabela 2.2 – Subcaracterísticas de Qualidade de Software

Modelo Rocha

No modelo ROCHA (1983) para avaliação da qualidade de um produto de software são utilizados os conceitos de objetivos de qualidade, fatores de qualidade, critérios, processos de avaliação, medida e medidas agregadas. Cada um destes conceitos está descrito a seguir.

Os objetivos de qualidade são as propriedades gerais que o produto deve possuir.

Já os fatores de qualidade determinam a qualidade do produto sob o ponto de vista dos diferentes usuários em diversos níveis. Esses fatores possibilitam que os objetivos de qualidade sejam atingidos, e quando necessário, podem ser compostos por subfatores.

Os critérios são os atributos primitivos que podem efetivamente ser mensurados. Isso não seria possível através dos objetivos e fatores de qualidade.

Os processos de avaliação determinam quais são as métricas necessárias para averiguar se um determinado critério está, e em que grau, presente num produto. Cada atributo primitivo (critério) tem associado a ele um processo de avaliação.

A medida está relacionada a presença de um critério em um produto. Tendo sido determinada qual a métrica a ser usada, é possível então, efetivamente medir o grau de presença de um critério no produto.

As medidas agregadas estão relacionadas aos fatores de um produto. Após a medição dos critérios é possível agregá-los e obter como resultado, uma medida que indicará o grau de presença, no produto, de um determinado fator.

A seguir a Figura 2.1 mostra a estrutura do modelo Rocha que permite a execução da avaliação da qualidade de software.

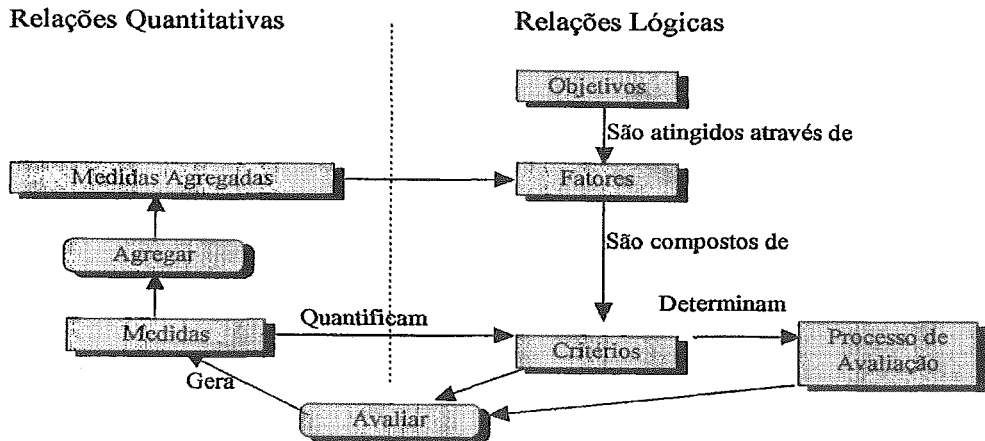


Figura 2.1 - Estrutura do Modelo Rocha para Avaliação da Qualidade de Software

Baseada na estrutura do modelo Rocha, esta tese define os objetivos, fatores e seus respectivos subfatores, critérios e processo de avaliação no contexto de banco de dados (a aplicação do modelo Rocha está descrita na Seção 3.3). A finalidade é verificar como o produto banco de dados satisfaz à expectativa de qualidade dos usuários, além de obter uma medida da qualidade dos dados armazenados.

A próxima seção descreve as motivações para preservar a qualidade dos dados da organização, assim como faz uma revisão da pesquisa de qualidade de dados mostrando as abordagens que surgiram e os tipos de avaliação de qualidade de dados existentes.

2.2 Qualidade de Dados

No ambiente de negócios competitivos atual, é necessário que a informação seja disponibilizada rapidamente. Deste modo, a qualidade dos dados da organização é um aspecto crítico a ser observado. O sucesso em garantir a lealdade do cliente é

determinado pela habilidade com que um negócio se comunica, efetivamente e precisamente com os seus clientes (CHOW, 2000). Assim, a baixa qualidade dos dados pode causar problemas sérios, como conduzir a uma decisão e ação errada, acarretando prejuízos financeiros à organização. Ressalta-se também, que dados incorretos podem influenciar negativamente na formação da opinião do cliente em relação à qualidade do produto ou do serviço (BALDWIN e BOWEN, 1999), (MARCO, 2000).

Os avanços tecnológicos melhoram a capacidade das organizações em adquirir, armazenar, processar e distribuir dados. Assim, os dados compartilhados pela organização são usados por pessoas diferentes e de modos diferentes e com isso problemas com produção de dados se tornam mais prováveis e mais importantes.

Para garantir a qualidade dos dados não basta empregar ferramentas, mas também se deve exigir que seja feito um trabalho com os usuários, a fim de evitar a introdução de dados inadequados. Assim como, é preciso se preocupar com software defeituoso que pode resultar em atualização inexata ou incompatível de dados. Na verdade, problemas com qualidade de dados surgem de muitas fontes e por uma grande variedade de razões (ADELMAN *et al*, 2001).

A fim de revermos a evolução, ao longo dos últimos anos, da pesquisa de qualidade de dados, fornecemos uma visão histórica do assunto, relatando os três principais tipos de abordagens que surgiram, que são: teórica, empírica e estocástica. Cada uma destas abordagens está definida a seguir.

A abordagem teórica enfoca como os dados podem se tornar deficientes durante o processo de sua produção. Por exemplo, WAND e WANG (1996) definiram as dimensões de qualidade dos dados usando conceitos ontológicos, baseados nos problemas que acontecem no mapeamento dos dados do mundo real para os sistemas de informação.

Esse estudo vem da observação do fato de que o desenvolvimento e o uso da informação envolvem duas transformações: a transformação de representação, e a transformação de interpretação. Esta abordagem se baseia que uma deficiência de dados pode acontecer durante a transformação de representação e/ou transformação de interpretação, gerando assim, uma falta de conformidade entre a visão do mundo real e aquela obtida do sistema de informação.

Na abordagem empírica são capturados os atributos de qualidade de dados que são importantes para os usuários. São analisados os dados colecionados pelos usuários, para determinar as características que irão avaliar se os dados são adequados às suas

tarefas ou não. Esta abordagem é usada quando a qualidade de dados está baseada na experiência ou no entendimento sobre quais são os atributos importantes do ponto de vista dos usuários (STRONG e MILLER, 1995), (WANG *et al.*, 1995), (HUH *et al.*, 1990), (WANG *et al.*, 1993).

A abordagem estocástica incorpora um conjunto de controles internos nos sistemas de informações para aumentar a habilidade desses sistemas em prevenir, descobrir e eliminar erros [BALDWIN e BOWEN, 1999].

Tipos de Avaliação da Qualidade de Dados

A maioria dos esforços feitos para avaliar a qualidade tende a focar na precisão do dado armazenado. Entretanto, para usuários de dados com experiências (pessoas ou grupos que têm experiências em usar dados organizacionais para adotar decisões), o conceito de qualidade de dados é mais amplo, pois deve-se capturar também os aspectos de qualidade que são importantes para os usuários (REDMAN, 1995). É preciso captar as diferentes necessidades dos usuários, observando as diferentes formas como os dados serão usados (HEALY, 2000).

Deve-se fazer uma definição precisa dos atributos de qualidade que interessam aos usuários. Neste sentido, é preciso considerar para se fazer a seleção e definição dos atributos de qualidade os seguintes pontos: as exigências dos usuários, o contexto funcional, as recomendações da literatura e as sugestões de especialistas (WILLSHIRE e MEYEN, 1997).

Para alcançar uma solução para os problemas de qualidade de dados, a organização deve ter: um bom modelo dos dados e uma boa compreensão das exigências de qualidade feitas pelos usuários, a fim de conseguir uma avaliação (taxa) da qualidade de dados atual e identificar áreas potenciais de deficiências de dados (REDMAN, 1995).

Existem dois tipos de avaliações que podem ser feitas em um banco de dados, que são a avaliação quantitativa e avaliação qualitativa (WILLSHIRE e MEYEN, 1997).

A *avaliação quantitativa (objetiva) da qualidade* é normalmente realizada por ferramentas automatizadas (software) que utilizam indicadores objetivos para medir a qualidade da representação e da estrutura do banco de dados.

São observados, por exemplo, valores do domínio, presença de valor, entre outros fatores. Este tipo de avaliação está baseado no ponto de vista do desenvolvimento tradicional, pois considera que a precisão e a objetividade são as duas dimensões necessárias. Porém, usuários de dados vêem que elas não são suficientes para que os dados sejam considerados como de alta qualidade.

A *avaliação qualitativa (subjetiva) da qualidade* pode ser realizada por especialistas ou usuários do banco de dados, ou ainda em conjunto com ferramentas automatizadas, que utilizam os indicadores subjetivos para julgar a qualidade atual do banco de dados e sua aptidão de uso. Neste tipo de avaliação os indicadores subjetivos são as características de qualidade que foram definidas, objetivando capturar a percepção e a expectativa do usuário sobre qualidade de dados.

É preciso ressaltar que as exigências de qualidade mudam através do tempo, e as soluções encontradas anteriormente podem também necessitar de mudanças. Entretanto, todo o investimento feito na busca da qualidade de dados tem o retorno garantido, sendo assim, a qualidade não é um problema e sim uma solução, uma solução à necessidade permanente de redução de custos e de ganho de competitividade.

Para que se tenha então, condições de alcançar uma solução para os problemas de qualidade de dados é necessária a adoção de um processo de avaliação da qualidade. Em complemento ao tipo de avaliação adotado, alguma outra técnica existente poderá ser utilizada. Na seção a seguir será mostrado como o conceito estatístico de não conformidades pode complementar a formação do processo de avaliação da qualidade dos dados armazenados.

2.3 Avaliação da Qualidade pela Não Conformidade

Desde 1700 onde, nas indústrias, a qualidade era fortemente determinada pelo esforço individual, e passando pelos anos de 1800 onde Frederick W. Taylor introduziu os princípios de gerenciamento científico, podemos chegar a um conceito estatístico de qualidade mais moderno (MONTGOMERY, 1996):

" A Qualidade é inversamente proporcional à variabilidade "

Embora a qualidade uniforme seja um dos objetivos almejados, sabe-se que a variabilidade é um fato natural. Um processo está sob controle quando a variabilidade

existente é devida ao acaso, isto é, as causas são aleatórias. Quando a variabilidade é anormal, as alterações nas características de qualidade são notórias, tornando o processo fora de controle. Essas causas são identificáveis e devem ser eliminadas (LOURENÇO FILHO, 1982).

Visando reduzir a variabilidade dos produtos, surgiu o controle estatístico de qualidade, que é uma técnica que busca continuamente a melhoria do processo de manufatura de um produto (MONTGOMERY, 1996).

Um produto deve reunir as especificações que foram feitas pelos usuários. Cada ponto específico que não satisfaz à especificação resulta em um defeito ou não conformidade. Um item não conforme é aquele que não satisfaz a uma ou mais especificações do produto.

Dependendo da natureza ou severidade, é possível um produto possuir algumas não conformidades, e mesmo assim, não ser classificado como não conforme. Assim, seria inadequado considerar que o produto é perfeito (sem defeitos) ou imperfeito (com defeitos). Convém avaliar a frequência com que ocorrem os defeitos em cada unidade (LOURENÇO FILHO, 1982).

Em produtos complexos é comum ocorrerem muitos tipos diferentes de não conformidades. Nem todos esses tipos são igualmente importantes. Para classificar a não conformidade de acordo com a sua gravidade utiliza-se o sistema de deméritos.

Um esquema de deméritos pode ser definido em classes como mostrado a seguir (LOURENÇO FILHO, 1982) e (MONTGOMERY, 1996):

Classe A - defeitos graves ou críticos: são aqueles que prejudicam a função essencial, impedindo assim, a utilização do produto;

Classe B - defeitos sérios: são aqueles que reduzem a eficiência do produto; e

Classe C - defeitos menores ou irregulares: são aqueles que, sem alterar o desempenho do produto, constituem imperfeições.

Cada classe tem um peso que é definido de acordo com o problema específico para o qual esta técnica foi adotada. O resultado deste procedimento é a média ponderada das não conformidades ocorridas no produto.

Assim, uma avaliação da qualidade pela não conformidade utiliza todas as não conformidades ocorridas, considerando o peso de cada uma delas, a fim de verificar como a qualidade do produto avaliado foi degradada.

Capítulo 3

CARACTERÍSTICAS DE QUALIDADE

Qualidade é um atributo associado a algo, não sendo possível defini-la universalmente. Deve, portanto, ser instanciada para um determinado produto. Qualidade é, também, um conceito multidimensional, realizando-se por intermédio de um conjunto de características (BELCHIOR, 1992b).

A norma ISO 8402 (1994) define qualidade como a "Totalidade de características de uma entidade que lhe confere a capacidade de satisfazer às necessidades explícitas e implícitas" e a norma ISO 9126 (1991) fornece o significado de característica de qualidade, como sendo a referência básica à qualidade de um produto, utilizada em uma avaliação.

Este trabalho aborda os aspectos subjetivos de qualidade de dados importantes para os usuários, não estando limitado somente às características envolvidas com produção e armazenamento de dados. Está orientado para um conjunto de características que estendem o conceito de qualidade de dados além da visão do desenvolvimento tradicional. O resultado é um processo que captura as perspectivas de qualidade de dados dos usuários.

Assim, neste capítulo, estão relacionadas as características de qualidade de dados, que foram identificadas como necessárias para avaliar qualitativamente (de forma subjetiva) um banco de dados e os tipos de erros associados a cada uma delas. Além disso, descrevemos como a visão do usuário é formada e como uma pesquisa de campo nos conduziu à abordagem dos aspectos subjetivos de qualidade de dados importantes para os usuários.

3.1 Visão do Usuário

O enfoque deste trabalho é baseado na forma como o usuário vê o banco de dados. Isto é, através da sua opinião e percepção, formada com o uso cotidiano do produto, é que as características de qualidade são avaliadas.

Uma avaliação qualitativa (subjéitiva) de banco de dados tem como característica considerar que o avaliador não conhece aspectos da elaboração ou implementação do produto que está avaliando (WILLSHIRE e MEYEN, 1997).

O entendimento do usuário sobre a qualidade dos dados armazenados que utiliza pode ser conseguido através dos seguintes fatores:

- ◆ Pela utilização dos serviços oferecidos pelo sistema (aplicação);
- ◆ Pela utilização dos serviços do banco de dados; e
- ◆ Pela informação que os dados geram através dos serviços mencionados nos itens anteriores, evidenciando assim, o nível de qualidade do conteúdo armazenado.

Consideramos, a aptidão de uso e a qualidade da informação fornecida pelo banco de dados os alvos desta avaliação. Assim, uma lista inicial de características é fornecida para vários tipos de avaliadores para que selecionem aquelas que são importantes para a organização na ocasião da avaliação qualitativa. Essas características selecionadas recebem um peso, a fim de expressar o seu grau de importância para o usuário dentro do contexto atual. Esse grau pode variar de 0 a 4, conforme a escala de valores mostrada na Tabela 3.1.

Baseado nas características subjétivas de qualidade de dados, que foram selecionadas pelo avaliador, é possível prosseguir com o processo de avaliação, e verificar como o sistema atual satisfaz a essas características (expectativas do usuário). Este processo permite determinar sua qualidade atual e identificar áreas potenciais de deficiência de dados. Sendo assim, esta tese se propõe a avaliar a qualidade de banco de dados através das características qualitativas (subjétivas), assim como elaborar uma ferramenta para que essa avaliação seja feita automaticamente

Escala	Equivalência de Importância
0	Nenhuma
1	Baixa
2	Moderada
3	Alta
4	Total

Tabela 3.1 - Escala de Valores de Importância

3.2 Processo de Validação das Características de Qualidade

Para obter uma lista inicial de características, observamos a aplicabilidade de várias características para a avaliação da qualidade dos dados. A definição do conjunto de características foi fundamentada em pesquisa e buscas na literatura sobre qualidade de dados e em trabalhos anteriores realizados na COPPE/UFRJ em avaliação da qualidade.

Para alcançar a adequação necessária, uma pesquisa de campo foi realizada consultando 27 (vinte e sete) especialistas de instituições acadêmicas, militares e empresariais. Essa pesquisa de campo foi realizada observando os aspectos a seguir.

Especialistas Participantes: Foi observada a atuação desses profissionais dentro da organização. Era necessário que eles atuassem como consumidores de dados que, com perspectivas diversas, usem regularmente os dados para adotar decisões.

Para identificar os especialistas mais adequados para participar da pesquisa de campo, foram observados os seguintes elementos do seu perfil profissional: grau de escolaridade, experiência ou as atividades (cargos) que já exerceu ou exerce na área de computação e tempo de experiência.

Para cada grupo de perfil profissional relacionado abaixo, será mencionado a quantidade de especialistas que se identificaram com cada um desses grupos.

Grupo I:

Grau de Escolaridade : Mestrado em Computação

Atividade na área de computação : Gerente de Projeto

Tempo de Experiência : de 10 a 15 anos

Quantidade de Especialistas : 8

Grupo II:

Grau de Escolaridade : Doutorado em Computação

Atividade na área de computação : Gerente de Projeto

Tempo de Experiência : de 10 a 18 anos

Quantidade de Especialistas : 3

Grupo III:

Grau de Escolaridade : Doutorado em Computação

Atividade na área de computação : Professor Universitário

Tempo de Experiência : de 10 a 20 anos

Quantidade de Especialistas : 2

Grupo IV:

Grau de Escolaridade : Especialização (pós-graduação)

Atividade na área de computação : Analista de Sistemas

Tempo de Experiência : de 5 a 12 anos

Quantidade de Especialistas : 14

Instrumento da Pesquisa: Foi utilizado o Formulário de Validação das Características de Qualidade (Apêndice I), para validar o conjunto de características de presença importante no processo de avaliação da qualidade de dados.

Procedimento: Baseando-se na sua opinião e/ou experiência, cada um dos 27 especialistas atribuiu para cada característica um peso, que pode variar de 0 a 4, segundo a escala apresentada na Tabela 3.1, para expressar a importância de cada uma das 31 características que formavam o conjunto sugerido inicialmente. O propósito era eliminar as características de qualidade que não estavam relacionadas à qualidade de dados segundo a opinião da maioria dos especialistas. Esse conjunto inicial de características pode ser encontrado no Apêndice I.

A inclusão de uma nova característica de qualidade acontecia sempre que algum especialista identificava a ausência dessa característica da lista inicial. Para registrar esses casos é preenchido o Cadastro para Sugestão de Nova Característica que consta do Apêndice I.

No caso da inclusão de nova característica, novos tipos de erros associados a ela são cadastrados. Desta forma, na ocasião do Cadastro de Erros, encontrado no Apêndice II, o usuário poderá relacionar o erro ocorrido a algum tipo de erro cadastrado, que está

previamente associado a alguma características de qualidade. A lista dos tipos de erros associados a cada característica de qualidade é encontrada na seção 3.5.

Análise dos Resultados: Conhecendo o peso que cada um dos 27 especialistas atribuiu para cada uma das 31 características, foi adotado o seguinte critério para exclusão:

$$\text{Score}_{c_i} = \sum_1^{\text{ne}} P_{c_i} / \text{ne}$$

Onde : Score_{c_i} é a peso médio calculado para a característica c_i , considerando os pesos atribuídos por cada especialista;

ne é a quantidade de especialistas que atribuiu peso para as características de qualidade; e

P_{c_i} é o peso que cada especialista atribuiu para a característica c_i .

Considerando que $\text{Score}_{c_i} \in [0,4]$, a característica c_i foi excluída da lista se $\text{Score}_{c_i} \leq 2$.

O resultado desta primeira pesquisa de campo foi a obtenção de um conjunto de características capaz de suportar as expectativas de qualidade dos usuários em relação ao produto de banco de dados que está sendo avaliado. Esse conjunto final de características qualidade de dados que foram validadas pode ser encontrado na Seção 3.4.

Na próxima seção será descrito como a avaliação da qualidade de dados proposta nesta tese baseou-se na estrutura do modelo Rocha para definir seus elementos.

3.3 Aplicação do Modelo Rocha

No modelo Rocha os elementos objetivos de qualidade, fatores de qualidade, critérios, processos de avaliação, medidas e medidas agregadas são conceituados, conforme visto na seção 2.1.2, objetivando a definição da qualidade de produtos de software.

A seguir será mostrado como os elementos do modelo Rocha, que é compatível com a norma ISO 9126, influenciaram na definição dos elementos da avaliação da qualidade de dados pela não conformidade proposta nesta tese.

Para efeito de padronização e buscando a uniformidade de terminologia com o modelo Rocha as características serão chamadas de *fatores* e, as subcaracterísticas de *subfatores*, que serão organizadas segundo os *objetivos* de qualidade utilizabilidade, confiabilidade conceitual e confiabilidade da representação, propostos por ROCHA (1983). Todos esses elementos são definidos, ocorrendo as devidas adaptações, para que a abordagem seja qualidade de dados.

Para determinar o *processo de avaliação*, propomos utilizar um *critério* único baseado no número de falhas ou não conformidades que ocorreram durante a utilização do produto de banco de dados.

Neste sentido, uma lista de tipos de erros associados às características, que está abordada na Seção 3.5, foi identificada a fim de quantificar os erros ocorridos que afetam a qualidade dos dados. Assim o critério adotado, ao invés de positivo, é baseado nas falhas do produto. O modelo matemático do processo de avaliação está descrito no Capítulo IV.

O resultado da avaliação é a *medida* do quanto a qualidade dos dados armazenados satisfaz às expectativas de qualidade dos usuários.

3.4 Características para Avaliação da Qualidade de Dados

Conforme mencionado anteriormente, a aptidão de uso e a qualidade da informação fornecida pelo banco de dados, são os alvos de uma avaliação qualitativa. Portanto, esta seção define a lista final dos fatores e subfatores qualitativos (subjetivos) de qualidade de dados que foram coletados e validados conforme descrito na Seção 3.2.

No final da definição de cada subfator serão mencionados trabalhos e/ou estudos que também utilizaram o item como um indicador subjetivo de qualidade. Além disso, informamos um outro nome, caso o trabalho ou estudo mencionado não tenha utilizado o item com o mesmo nome adotado nesta tese. Ressalta-se também, que o subfator em outros trabalhos ou estudos pode ter sido utilizado para abordar outro contexto, não tendo sido definido especificamente para contemplar qualidade de dados. Observamos no entanto, que a aparição de um subfator de qualidade em diferentes contextos é um

indicativo de sua generalidade, podendo então, ser aplicado num outro contexto específico.

OBJETIVO : UTILIZABILIDADE

É o objetivo que reúne as diversas características que devem estar presentes para se utilizar o produto com facilidade. Sendo ele utilizado por diferentes usuários, é preciso que as tarefas sejam realizadas de forma fácil, eficiente e lucrativa.

Consideramos que este objetivo é atingido através dos fatores: adequabilidade, eficiência, aplicabilidade e implementabilidade que estão descritos a seguir.

FATOR : Adequabilidade

Refere-se à capacidade dos serviços oferecidos pelo sistema ou pelo banco de dados de serem adequados às perspectivas dos usuários em ter disponível um dado atualizado, para ser utilizado conforme as suas necessidades. É composto pelos subfatores: disponibilidade da informação e idade dos dados.

Subfator : Disponibilidade da Informação

Refere-se a capacidade, oferecida pelo sistema ou pelo banco de dados, de disponibilizar os dados em todos os locais onde poderão ser úteis, independente do local onde estão armazenados.

(CARVALHO, 1997) ; (LIMA, 1999) ; (STRONG *et al*, 1994):
Recuperabilidade.

Subfator : Idade dos Dados

Refere-se ao dado estar atualizado, expressando até que ponto a idade dos dados (tempo que está armazenado) é apropriada para a realização das tarefas da organização.

(STRONG *et al*, 1994).

FATOR : Eficiência

Refere-se à forma como os serviços oferecidos pelo sistema ou serviços do banco de dados utilizam os recursos disponíveis. É composto pelos subfatores: oportunidade e eficiência de execução.

Subfator: Oportunidade

Refere-se à capacidade da informação, gerada pela manipulação do conteúdo dos dados, ser fornecida com a rapidez necessária para que seja útil.

(CAMPOS, 1994 a); (CARVALHO, 1997); (LIMA, 1999); (BALLOU e PAZER, 1985).

Subfator: Eficiência de Execução

Representa a velocidade de acesso que os serviços oferecidos pelo sistema ou serviços do banco de dados necessitam para manipular os dados (recuperação/gravação) ao realizar um processamento.

(BALLOU e PAZER, 1985): Velocidade de Acesso; (ISO 9126, 1991): Comportamento em Relação ao Tempo.

FATOR: Aplicabilidade

Características que expressam até que ponto os dados são úteis para as tarefas da organização. É composto pelos subfatores: relevância e utilidade.

Subfator: Relevância

Refere-se à importância da utilização dos serviços do banco de dados para realizar as tarefas a que se propõe.

(OLIVEIRA, 1995); (WANG *et al*, 1993).

Subfator: Utilidade

Expressa a contribuição que os serviços do banco de dados oferecem à organização através da sua utilização.

(OLIVEIRA, 1995); (WANG *et al*, 1993): Aplicabilidade; (BALLOU e PAZER, 1985) e (TAN e BENSAT, 1990): Pertinência.

FATOR: Rentabilidade

Características que representam o valor adicionado (financeiros e sociais) obtido com a utilização das informações geradas pelos serviços do sistema ou serviços do banco de dados. Até que ponto os dados são benéficos e provêm vantagens com o uso

deles. É composto pelos subfatores: lucratividade, benefício no trabalho do usuário e competitividade.

Subfator: Lucratividade

Refere-se ao aumento de produtividade conseguido através da utilização dos dados, para alcançar os objetivos da organização.

(CAMPOS, 1994 a); (CAMPOS, 1994 b); (CARVALHO, 1997); (VALLE *et al*, 1997); (OLIVEIRA, 1995); (LIMA, 1999)

Subfator: Benefício no Trabalho do Usuário

Característica da utilização dos serviços do sistema do banco de dados oferecer simplificação e melhoria nas condições de trabalho dos seus usuários.

(CARVALHO, 1997): Apoio a Trabalhos de Rotina; (OLIVEIRA, 1995).

Subfator: Competitividade

Capacidade da utilização dos dados proporcionar vantagens competitivas com relação aos concorrentes do mercado, somando valor às execuções das tarefas da organização.

(CAMPOS, 1994 a); (CAMPOS, 1994 b); (CARVALHO, 1997); (VALLE *et al*, 1997); (OLIVEIRA, 1995); (LIMA, 1999); (STRONG *et al*, 1994): Valor Adicionado.

OBJETIVO: CONFIABILIDADE CONCEITUAL

São as características que expressam se os requisitos que compõem as necessidades dos usuários em relação aos dados, foram incorporadas ao produto, tornando-o então, do ponto de vista do seu conteúdo, confiável e útil para os seus usuários.

Os seguintes fatores possibilitam que a confiabilidade conceitual seja atingida: fidedignidade, integridade, confiabilidade, funcionalidade e legibilidade, conforme descritos a seguir.

FATOR: Fidedignidade

Características que indicam a existência dos requisitos que permitirão atingir os propósitos que foram especificados pelos usuários. É composto pelos subfatores: quantidade de dados adequada, acurácia, completitude e cobertura.

Subfator: Quantidade de Dados Adequada

Refere-se à quantidade ou volume de dados disponíveis ser apropriada para realizar as tarefas da organização.

(WANG *et al*, 1993); (ISO 9126, 1991).

Subfator: Acurácia

Capacidade dos serviços oferecidos pelo sistema ou pelo banco de dados realizarem a manipulação dos dados gerando resultados corretos ou conforme esperados.

(ISO 9126, 1991).

Subfator: Completitude

Refere-se à existência de todos os dados necessários para que os serviços oferecidos pelo sistema ou serviços do banco de dados gerem informações úteis e necessárias, a fim de satisfazer completamente às exigências especificadas pelos usuários.

(CAMPOS, 1994 a); (CAMPOS, 1994 b); (CARVALHO, 1997); (VALLE *et al*, 1997); (OLIVEIRA, 1995); (LIMA, 1999).

Subfator: Cobertura

Refere-se à característica que evidencia que o âmbito (escopo) da informação contida nos dados é suficientemente profundo, a fim de cobrir todo o escopo necessário exigido pelas tarefas da organização. Este subfator está relacionado com a forma como o usuário vê a importância do conteúdo global dos dados.

(HOXMEIER, 1997), (McGEE e WANG, 1993) e (WANG *et al*, 1993):
Âmbito da Informação.

FATOR: Integridade

Refere-se à capacidade de preservação dos dados durante um processamento normal ou em situações hostis, assegurando que os dados permaneçam corretos, seguros e livres de erros. É composto pelos subfatores: robustez, precisão do dado, consistência, sinalização e atributabilidade.

Subfator: Robustez

Refere-se à existência de serviços do sistema capazes de controlar situações anormais, reagindo a elas sem perda do controle, não inserindo erro à base de dados.

(CAMPOS, 1994 a); (CAMPOS, 1994 b); (CARVALHO, 1997); (VALLE *et al*, 1997); (OLIVEIRA, 1995); (LIMA, 1999).

Subfator: Precisão do Dado

Capacidade do dado representar com exatidão, de forma perfeita, correta, clara e concisa o seu significado no mundo real.

(McGEE e WANG, 1993) e (WANG *et al*, 1993)

Subfator: Consistência

Refere-se ao valor do dado ser o mesmo em toda a base de dados, não possuindo conteúdo contraditório ou significados diferentes.

(CAMPOS, 1994 a); (CAMPOS, 1994 b); (CARVALHO, 1997); (VALLE *et al*, 1997); (OLIVEIRA, 1995); (LIMA, 1999).

Subfator: Sinalização

Refere-se à existência de serviços oferecidos pelo sistema que possibilitem a identificação fácil de erros, provendo alertas para entrada de dados que estão fora dos padrões especificados ou para qualquer manipulação que indique problemas para a qualidade dos dados.

(CARVALHO, 1997); (STRONG *et al*, 1994): Identificação Fácil de Erros.

Subfator: Atributabilidade

Capacidade de algum serviço oferecido pelo sistema oferecer mecanismo que registre a identificação do autor de qualquer manipulação de dado, assim como a data e de onde foi realizada a operação.

Este subfator foi coletado através de contato com especialistas.

FATOR: Funcionalidade

Refere-se à existência de funcionalidades que tornam o produto confiável sob o ponto de vista de seu uso, gerando assim, uma confiabilidade no conteúdo do dado. Este fator é composto pelos seguintes subfatores: recuperabilidade, flexibilidade, interoperabilidade e segurança.

Subfator: Recuperabilidade

Refere-se à existência de serviços oferecidos pela aplicação ou pelo banco de dados que possibilitem, em caso de falha, recuperar os dados diretamente afetados, em tempo e esforço viáveis.

(HOXMEIER, 1997), (McGEE e WANG, 1993) e (WANG *et al*, 1993): Recuperação em Caso de Erro; (ISO 9126, 1991).

Subfator: Flexibilidade

Característica que evidencia a capacidade dos dados serem manipulados e administrados facilmente, possibilitando a sua expansão, adaptação e agregação com outros dados.

(STRONG *et al*, 1994): Facilmente Manipulado.

Subfator: Interoperabilidade

Capacidade de interação com outras bases de dados, a fim de complementar informações e tecnologias.

(STRONG *et al*, 1994): Facilmente Integrado; (ISO 9126, 1991).

Subfator: Segurança

Capacidade de garantir a segurança dos dados armazenados sob o ponto de vista do seu uso, evitando acesso não autorizado, acidental ou deliberado, protegendo-se de violações e garantindo a privacidade dos dados.

(ISO 9126, 1991).

FATOR: Legibilidade

Capacidade da informação gerada pela manipulação dos dados ser clara e sem ambigüidade para que seja facilmente compreendida.

É uma compreensão detalhada dos dados e não somente uma inspeção superficial deles. É composta pelos subfatores: facilidade de compreensão da informação e adequabilidade da informação.

Subfator: Facilidade de Compreensão da Informação

Capacidade da informação gerada ser precisa e objetiva, ou seja, que não coloque o usuário em situações de dúvida.

(VALLE *et al*, 1997): Precisão da Informação; (STRONG *et al*, 1994): Facilidade de Compreensão.

Subfator: Adequabilidade da Informação

Capacidade da informação gerada ser adequada a todos os seus usuários, levando em consideração aspectos sociais e culturais compatíveis com os diferentes níveis funcionais dentro da organização.

(VALLE *et al*, 1997).

OBJETIVO: CONFIABILIDADE DA REPRESENTAÇÃO

Este objetivo é composto pelos fatores que reúnem as facilidades de compreensão e manipulação do produto que será utilizado por diferentes tipos de usuários. Neste sentido, a forma do produto é o aspecto abordado neste objetivo. Os seus fatores são: uniformidade e manipulabilidade.

FATOR: Uniformidade

Refere-se à capacidade dos dados estarem apresentados em um mesmo formato, estando eles consistentemente representados e formatados.

Esse fator não possui subfatores.

(HOXMEIER, 1997), (McGEE e WANG, 1993) e (WANG *et al*, 1993):
Representação Concisa.

FATOR: Manipulabilidade

Refere-se à capacidade de poder analisar os dados ou a sua documentação, possibilitando encontrar com facilidade as informações desejadas. É composto pelos subfatores: disponibilidade de documentação e rastreabilidade.

Subfator: Disponibilidade de Documentação

Refere-se à existência de uma documentação atualizada e disponível dos dados, tornando-os assim, bem documentados, verificáveis e facilmente associados a uma fonte, facilitando a sua localização e verificação pelos usuários.

(CAMPOS, 1994 a); (CAMPOS, 1994 b); (CARVALHO, 1997); (VALLE *et al*, 1997); (OLIVEIRA, 1995); (LIMA, 1999); (STRONG *et al*, 1994): Bem documentado.

Subfator: Rastreabilidade

Refere-se à existência de serviços oferecidos pelo sistema que possibilitam que a busca da informação seja feita através da seqüência de agregação de detalhes de um determinado aspecto, desde a sua visão mais geral até a mais detalhada, e vice-versa.

(CAMPOS, 1994 a); (CAMPOS, 1994 b); (CARVALHO, 1997); (VALLE *et al*, 1997); (OLIVEIRA, 1995); (LIMA, 1999); (WANG *et al*, 1993): Localização.

3.5 Tipos de Erros Associados às Características

Toda não conformidade ou erro ocorrido deve ser cadastrado, para que a avaliação meça o índice de qualidade dos dados armazenados, baseando-se na forma como cada característica de qualidade foi afetada por esses erros.

Para realizar esse cadastro é utilizado o Formulário para Cadastro do Erro ou Não Conformidade Ocorrido, onde além dos campos versão, data do cadastro, funcionário, gravidade do erro e de uma descrição livre que é feita para detalhar o problema ocorrido, é preciso que seja escolhido um tipo de erro.

Tipo de erro é uma descrição resumida de um tipo de problema que pode ocorrer, e que conseqüentemente afeta a uma característica de qualidade de dados. Assim, é necessário que os tipos de erro possíveis de acontecer sejam identificados, para que o usuário, ao encontrar uma não conformidade do banco de dados, possa realizar o cadastro desta ocorrência.

Durante a pesquisa de campo realizada junto a 27 especialistas, descrita na seção 3.2, uma lista inicial dos possíveis tipos de erros associados às características foi também analisada por eles. A inclusão de um novo tipo de erro se realizou sempre que algum especialista identificava a sua ausência da lista inicial. Nesse momento é atribuído também um peso, a fim de expressar o grau de importância desse novo tipo de erro incluído. Ressalta-se, que a validação do conjunto de tipos de erros foi um processo diferenciado da validação do conjunto de características, pois a lista inicial sugerida de tipos de erros não foi oriunda de uma pesquisa formal como foi o processo de origem da lista inicial de características de qualidade.

A seguir, os tipos de erros, que foram coletados e validados, estão relacionados e associados às características que afetam:

Característica: Disponibilidade da Informação

Tipo de Erro Associado:

- Dado indisponível.

Característica: Idade dos Dados

Tipo de Erro Associado:

- Dado armazenado sem atualização por muito tempo.

Característica: Oportunidade

Tipo de Erro Associado:

- Demora no fornecimento de resposta que necessita ser utilizada.

Característica: Eficiência de Execução

Tipos de Erros Associados:

- Demora na recuperação de um dado;
- Demora na gravação de um dado; e
- Demora na execução de uma tarefa.

Característica: Relevância

Tipo de Erro Associado:

- Serviço oferecido não está sendo realizado de forma correta ou completa.

Característica: Utilidade

Tipo de Erro Associado:

- Funcionalidade não é importante para as tarefas da organização.

Característica: Lucratividade

Tipo de Erro Associado:

• Dificuldade na utilização dos dados gerou queda de produtividade (perda financeira).

Característica: Benefício no Trabalho do Usuário

Tipo de Erro Associado:

- Utilização dos serviços oferecidos dificulta o trabalho dos usuários.

Característica: Competitividade

Tipo de Erro Associado:

• Utilização dos dados não auxilia a organização a obter vantagens competitivas no mercado.

Característica: Quantidade de Dados Adequada

Tipo de Erro Associado:

- Quantidade de dados insuficiente.

Característica: Acurácia

Tipo de Erro Associado:

- Manipulação do dado resulta em informação incorreta.

Característica: Completitude

Tipos de Erros Associados:

- Falta de campo para cadastramento; e
- Falta de informação.

Característica: Cobertura

Tipo de Erro Associado:

- Informação oferecida pelo dado não é suficiente.

Característica: Robustez

Tipo de Erro Associado:

• Inserção de erro na base de dados após ocorrer situação anormal de funcionamento do sistema.

Característica: Precisão do Dado

Tipo de Erro Associado:

• O dado armazenado não representa corretamente o seu significado no mundo real.

Característica: Consistência

Tipo de Erro Associado:

- Dado com dois ou mais valores diferentes armazenados na base de dados.

Característica: Sinalização

Tipos de Erros Associados:

- Falta de alerta (aviso) para indicar entrada de dado incorreta ou não conforme;
- Falta de alerta (aviso) para indicar manipulação de dados não permitida.

Característica: Atributabilidade

Tipo de Erro Associado:

- Ausência de registro da autoria de alteração ou manipulação de dados.

Característica: Recuperabilidade

Tipos de Erros Associados:

- Ausência de mecanismo, em caso de falhas, para recuperar os dados afetados;
- Demora ou dificuldade, em caso de falhas, para recuperar dados afetados.

Característica: Flexibilidade

Tipo de Erro Associado:

- Dificuldade na manipulação (expansão, adaptação ou agregação) dos dados.

Característica: Interoperabilidade

Tipo de Erro Associado:

- Impossibilidade de interagir com outra base de dados.

Característica: Segurança

Tipos de Erros Associados:

- Ausência de mecanismo de segurança para controle de acesso ao sistema; e
- Falta de definição do escopo de acesso aos dados, permitido para cada tipo de usuário.

Característica: Facilidade de Compreensão da Informação

Tipo de Erro Associado:

- Informação pouco objetiva e gerando dúvidas.

Característica: Adequabilidade da Informação

Tipo de Erro Associado:

- Entendimento exigido pela informação é incompatível com os usuários.

Característica: Uniformidade

Tipo de Erro Associado:

- Apresentação dos dados insatisfatória.

Característica: Disponibilidade de Documentação

Tipos de Erros Associados:

- Falta de documentação, a fim de auxiliar na verificação e localização dos dados; e
- Falta de documentação que possibilite a associação do dado à sua fonte.

Característica: Rastreabilidade

Tipo de Erro Associado:

- Ausência de mecanismo que possibilite rastrear uma informação a fim de localizá-la.

Capítulo 4

AVALIAÇÃO DA QUALIDADE DE DADOS PELA NÃO CONFORMIDADE

Produtos que lidam com alvos fundamentados em ambigüidades e abstrações, num mercado competitivo, estão destinados a extinção. Convicções como “menos que o perfeito é bom o bastante” (YORDON, 1995) não são mais aceitáveis na atual era da qualidade.

Para se alcançar uma solução para os problemas de qualidade de dados, é necessário realizar avaliações, onde um acompanhamento da taxa de qualidade permitirá identificar áreas potenciais de deficiência de dados (REDMAN, 1995).

Neste capítulo, está descrito o modelo matemático desenvolvido que realiza o processo de obtenção de um índice de qualidade dos dados armazenados a partir do cadastro das não conformidades ocorridas. Estão também descritas as etapas necessárias para realizar integralmente a avaliação da qualidade de dados.

4.1 Modelo Matemático do Critério de Avaliação

O modelo matemático, desenvolvido exclusivamente para ser aplicado ao propósito deste trabalho, baseou-se no Controle Estatístico da Qualidade que sugere a utilização do Sistema de Deméritos para observar como as não conformidades ocorridas auxiliam na busca contínua da melhoria do processo de manufatura de um produto (MONTGOMERY, 1996), (KUME, 1985) e (LOURENÇO FILHO, 1982).

O Sistema de Deméritos é utilizado para classificar a não conformidade de acordo com a sua gravidade, pois seria inadequado considerar que um produto é perfeito

(sem defeitos) ou imperfeito (com defeitos). Convém avaliar a frequência com que ocorrem os defeitos em cada unidade.

Em produtos complexos é comum ocorrer muitos diferentes tipos de não conformidades. Como nem todos esses tipos são igualmente importantes, é possível diferenciá-los através de um esquema de deméritos.

Uma definição de esquema de deméritos é definida a seguir (LOURENÇO FILHO, 1982) e (MONTGOMERY, 1996) :

Classe A: Defeitos Graves ou Críticos: Aqueles que prejudicam a função essencial, impedindo assim, a utilização do produto.

Classe B: Defeitos Sérios: Aqueles que reduzem a eficiência do produto.

Classe C: Defeitos Menores ou Irregulares: Aqueles que, sem alterar o desempenho do produto, constituem imperfeições.

Associado a cada classe existe um peso que é definido de acordo com o problema específico para o qual esta técnica foi adotada.

O modelo matemático desenvolvido se divide basicamente em seis passos que expressam a idéia geral de formação do critério de avaliação. Deste modo, este modelo além de utilizar o sistema de deméritos aplica-o à uma estrutura hierárquica, conforme descrito a seguir.

Passo 1 : Idéia Geral da Formação do Critério

Sendo um item *não conforme* aquele que não satisfaz a uma ou mais especificações do produto (LOURENÇO FILHO, 1982), sempre que uma característica de qualidade é afetada por alguma não conformidade, o produto de banco de dados tem a sua qualidade geral degradada.

O esquema geral de ligação dos tipos de erro às características e das características ao produto de banco de dados é mostrado na Figura 4.1.

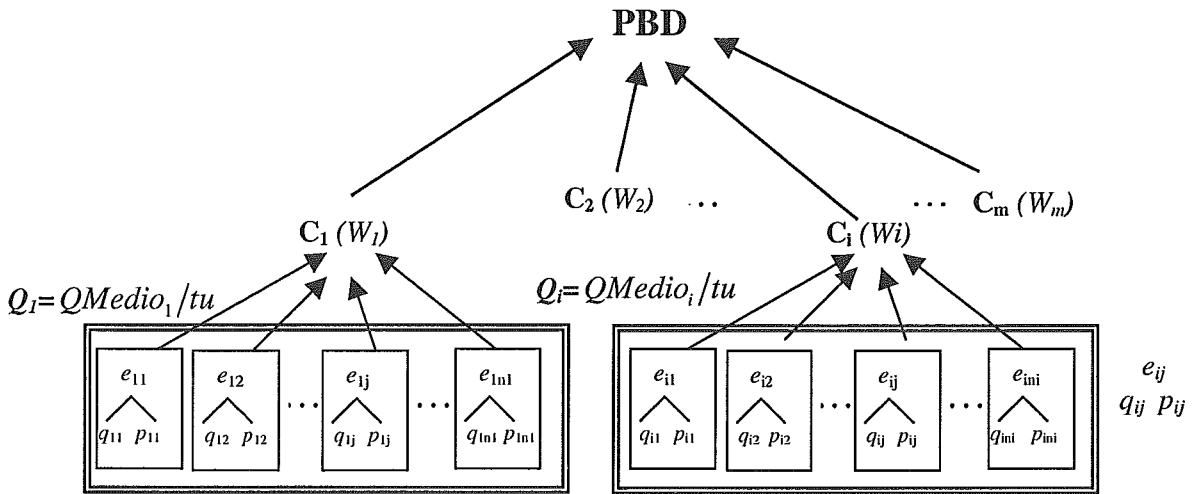


Figura 4.1 – Esquema Geral do Modelo de Avaliação

Através da Figura 4.1 é possível observar que as ocorrências de erro Q_i afetam a característica C_i associada, que por sua vez afeta a qualidade dos dados armazenados no produto de banco de dados (PBD).

Na Figura 4.1 as variáveis têm os respectivos significados:

PBD - produto de banco de dados;

C_i - *i*-ésima característica de qualidade do produto de banco de dados;

W_i - peso da característica C_i ;

e_{ij} - tipo de erro j associado à característica i , onde cada tipo de erro tem:

q_{ij} - quantidade média ponderada dos erros ocorridos do tipo de erro e_{ij} , considerando a gravidade das ocorrências dos erros; e

p_{ij} - peso do tipo de erro e_{ij} .

$Q_1, Q_2, \dots, Q_i, \dots, Q_m$ - taxa média (erro/min) de ocorrências de erros que afetam $C_1, C_2, \dots, C_i, \dots, C_m$, respectivamente.

Sendo $Q_i = \frac{QMedio_i}{tu}$, onde $QMedio_i$ é a quantidade média de ocorrências de erros que afetam a C_i , e tu é o tempo acumulado de utilizações do produto até o momento da avaliação (expresso em minutos).

Considerando que a *utilização* é dada pelo tempo de uso do produto que está sendo avaliado, esse valor pode ser obtido, por exemplo, através do controle de acesso a uma seção *logout - login*. Sendo tu o acúmulo desses valores até o momento da avaliação, então tu é constante para todo $QMedio_i$.

Passo 2: Cálculo de q_{ij}

A variável q_{ij} corresponde a quantidade média ponderada dos erros ocorridos do tipo de erro e_{ij} . Para isso é preciso para cada tipo de erro, verificar as ocorrências de erros juntamente com o grau que expressa a sua gravidade. Uma ocorrência de erro pode ter uma gravidade diferente de outra mesmo que pertença ao mesmo tipo de erro. Esse grau pode variar de 0 a 4, conforme escala de valores mostrada na Tabela 3.1. Assim, temos que:

$$q_{ij} = \frac{\sum_{k=0}^4 o_{ijk} \cdot g_k}{\sum_{k=0}^4 g_k}$$

onde:

O_{ijk} corresponde a ocorrência do erro e_{ij} com o grau de gravidade g_k e

g_k corresponde ao grau que expressa a gravidade da ocorrência do erro, podendo k variar de 0 até 4.

Passo 3: Cálculo de Q_i - Taxa Média de Ocorrência de Erros que afetam a Característica C_i

Baseado no conjunto de erros cadastrados é possível identificar quais foram as características afetadas por eles. Assim, a taxa média de ocorrência de erros que afetam as características, é calculada da seguinte forma:

$$QMedio_i = \frac{q_{i1} \cdot p_{i1} + q_{i2} \cdot p_{i2} + \dots + q_{ij} \cdot p_{ij}}{p_{i1} + p_{i2} + \dots + p_{ij}}$$

$$QMedio_i = \frac{\sum_{j=1}^n q_{ij} \cdot p_{ij}}{\sum_{j=1}^n p_{ij}}$$

$$Q_i = \frac{QMedio_i}{tu}$$

Onde n é igual a quantidade de tipos de erro associados a C_i .

Passo 4: Cálculo de X - Quantidade Média de vezes que o PBD foi afetado

Considerando as taxas médias de ocorrência de erros que afetaram a cada característica de qualidade, realizado no passo anterior, o cálculo da quantidade média de vezes que o PBD foi afetado pelas não conformidades relacionadas às características, é realizado da seguinte forma:

$$X = \frac{Q_1 \cdot W_1 + Q_2 \cdot W_2 + \dots + Q_m \cdot W_m}{W_1 + W_2 + \dots + W_m}$$

$$X = \frac{\sum_{k=1}^m Q_k \cdot W_k}{\sum_{k=1}^m W_k}$$

onde:

W_i é igual do peso da Característica C_i , e

m corresponde a quantidade de características de qualidade que afetam ao banco de dados.

Conforme descrito na Seção 2.3, o sistema estatístico de deméritos utiliza uma escala de pesos que é determinada de acordo com o problema a ser resolvido (MONTOMERY, 1996).

Vários experimentos e simulações foram realizados, onde se pode observar que uma escala exponencialmente crescente deveria ser utilizada para representar os pesos das características de qualidade. Desta forma é possível obter uma diferença expressiva entre o peso dado a uma característica pouco importante e uma característica imprescindível.

A escala dos pesos das características de qualidade adotada neste trabalho está descrita na Tabela 4.1.

Peso	Significado
0	Indica que a característica que está sendo apresentada não tem nenhuma importância.
1	Indica que a característica que está sendo apresentada tem pouca importância.
3	Indica que a característica que está sendo apresentada tem importância em algumas circunstâncias, mas em outras não.
9	Indica que a característica que está sendo apresentada é muito importante.
27	Indica de maneira absoluta que não há dúvidas que a característica que está sendo apresentada é imprescindível.

Tabela 4.1 – Escala de Pesos

Desta forma, através dos pesos estipulados pela Tabela 4.1, a ocorrência de 27 (vinte e sete) erros que afetam a uma característica pouco importante (peso 1), equivalem à ocorrência de um (um) erro que afeta uma característica imprescindível (peso 27). Essa distribuição não linear dos pesos garante um índice que expressa a degradação da qualidade dos dados mais eficientemente.

Assim, o peso atribuído pelo usuário a cada característica (que pode variar de 0 a 4), é modificado conforme a Tabela 4.2 de correspondência.

Peso Atribuído pelo avaliador	Peso correspondente utilizado no modelo
0	0
1	1
2	3
3	9
4	27

Tabela 4.2 – Correspondência dos Pesos

Passo 5: Associação do X calculado a um índice (percentual) de qualidade

Somente com X calculado, média de vezes que o produto de banco de dados foi afetado, não é possível saber qual é a qualidade dos dados armazenados. É preciso, então, calcular o percentual de qualidade correspondente.

Buscando um índice de qualidade dos dados armazenados que pode variar de 0 a 100 %, o gráfico de desempenho proposto foi dividido em cinco faixas cujo tamanho equivale a vinte pontos percentuais: [0 , 20] , [20 , 40], [40 , 60], [60 , 80] e [80 , 100], conforme mostrado na Figura 4.2.

Passo 5.1: Montagem do Gráfico de Qualidade

A cada $X \in [0 , \infty]$ devemos associar um $d = [0 , 100 \%$], de forma que $d(X = 0) = 100\%$ e $d(X = \infty) = 0\%$. Isto significa que, quanto maior for o valor de X, quantidade média de não conformidades que afetam ao PBD, pior será o desempenho.

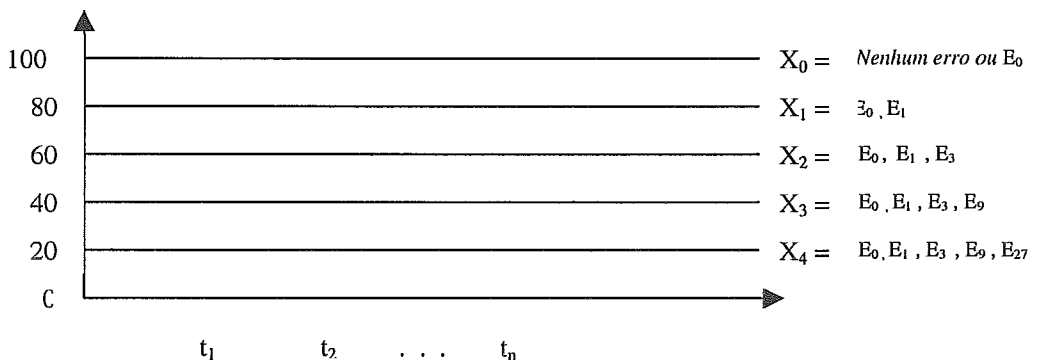


Figura 4.2 – Gráfico de Desempenho

Onde t_n corresponde ao momento da avaliação e E_i indica a ocorrência de todos os tipos de erro, cada um exatamente uma vez, que afeta a característica que tem peso i , conforme discriminado pela Tabela 4.3.

Erro	Significado
E_0	Valor da ocorrência de todos os tipos de erros, cada um exatamente uma vez, associados às características que têm peso 0 (característica que não tem nenhuma importância)
E_1	Valor da ocorrência de todos os tipos de erros, cada um exatamente uma vez, associados às características que têm peso 1 (característica que tem pouca importância)
E_3	Valor da ocorrência de todos os tipos de erros, cada um exatamente uma vez, associados às características que têm peso 3 (característica que tem importância moderada)
E_9	Valor da ocorrência de todos os tipos de erros, cada um exatamente uma vez, associados às características que têm peso 9 (característica muito importante)
E_{27}	Valor da ocorrência de todos os tipos de erros, cada um exatamente uma vez, associados às características que têm peso 27 (característica imprescindível)

Tabela 4.3 – Ocorrência de Erros Associados às Linhas

Representando cada linha um conjunto de valores, a montagem do gráfico de desempenho é indicada na Tabela 4.4.

Linha	Corresponde ao d (percentual de qualidade)	Significado
X_0	100	Nenhuma ocorrência de erro ou, ocorreu 1 vez cada tipo de erro associados às características que têm peso.
X_1	80	Ocorreu 1 vez cada tipo de erro, associados às características que têm pesos 0 e 1
X_2	60	Ocorreu 1 vez cada tipo de erro, associados às características que têm pesos 0, 1 e 3
X_3	40	Ocorreu 1 vez cada tipo de erro, associados às características que têm pesos 0, 1, 3 e 9.
X_4	20	Ocorreu 1 vez cada tipo de erro, associados às características que têm pesos 0, 1, 3, 9 e 27

Tabela 4.4 – Valores de cada Linha do Gráfico de Desempenho

Passo 5.2: Cálculo do d correspondente ao X calculado

Quando o X calculado for igual a X_0, X_1, X_2, X_3 ou X_4 o percentual de qualidade d corresponde exatamente a uma das cinco linhas principais (100 %, 80 %, 60 %, 40 % ou 20%). Buscando então, o d correspondente ao X calculado, os passos que são descritos a seguir são necessários:

Passo 5.2.1 – Cálculo das cinco linhas

Passo 5.2.1.1 - Linha X_0 que corresponde a 100 %

Para calcular a linha X_0 , que corresponde ao percentual de desempenho de 100%, é preciso considerar todas as características que tem peso igual a 0 (zero) e realizar os seguintes cálculos:

a) Cálculo do $Q_{[1..m]}$ das $C_{[1..m]}$ considerando que todos os tipos de erros das características ocorreram apenas uma vez durante tu – tempo acumulado de utilização:

Sendo m a quantidade de características com peso 0 e que $m \geq 1$, para cada C_i calcula-se o Q_i (taxa média de erros que afetam C_i) da seguinte forma:

$$Q_{Medio_i} = \frac{1 \cdot p_{i1} + 1 \cdot p_{i2} + \dots + 1 \cdot p_{ij}}{p_{i1} + p_{i2} + \dots + p_{ij}}$$

$$Q_{Medio_i} = \frac{\sum_{j=1}^n 1 \cdot p_{ij}}{\sum_{j=1}^n p_{ij}}$$

$$Q_i = \frac{Q_{Medio_i}}{tu}$$

onde: n corresponde a quantidade de tipo de erro associado à característica C_i e p_{ij} é igual ao peso do tipo de erro e_{ij} .

b) Cálculo do X_0 Parcial:

Com os valores de $Q_{[1..m]}$, o valor de X_0 Parcial é determinado como:

$$X_0 \text{ Parcial} = \frac{Q_1 \cdot W_1 + Q_2 \cdot W_2 + \dots + Q_m \cdot W_m}{W_1 + W_2 + \dots + W_m}$$

$$X_0 \text{ Parcial} = \frac{\sum_{j=1}^i Q_j \cdot W_j}{\sum_{j=1}^i W_j}$$

onde: Q_j é a quantidade média de erros que afetam à característica C_j , W_j é o peso da j -ésima característica e i é igual quantidade de características com peso W_j .

c) Valor de X_0 :

$$X_0 = X_0 \text{ Parcial.}$$

Passo 5.2.1.2 – Cálculo da Linha X_1 que corresponde a 80 %

Tendo sido calculado primeiramente o valor de X_0 (características com peso igual a 0), é preciso calcular o valor de X_1 Parcial considerando somente as características que têm peso igual a 1 (um):

a) Cálculo do $Q_{[1..m]}$ das $C_{[1..m]}$ considerando que todos os tipos de erros das características ocorreu apenas uma vez durante tu – tempo acumulado de utilização:

Sendo m a quantidade de características com peso 1 e que $m \geq 1$, para cada C_i calcula-se o Q_i (taxa média de erros que afetam C_i) da seguinte forma:

$$Q_{\text{Medio}_i} = \frac{1 \cdot p_{i1} + 1 \cdot p_{i2} + \dots + 1 \cdot p_{ij}}{p_{i1} + p_{i2} + \dots + p_{ij}}$$

$$QMedio_i = \frac{\sum_{j=1}^n 1 \cdot p_{ij}}{\sum_{j=1}^n p_{ij}}$$

$$Q_i = \frac{QMedio_i}{tu}$$

onde: n é a quantidade de tipo de erro associado à característica C_i e p_{ij} é o peso do tipo de erro e_{ij} .

b) *Calculo do X_1 Parcial:*

Com os valores de $Q_{[1..m]}$, o valor de X_1 Parcial é determinado como:

$$X_1Parcial = \frac{Q_1 \cdot W_1 + Q_2 \cdot W_2 + \dots + Q_m \cdot W_m}{W_1 + W_2 + \dots + W_m}$$

$$X_1Parcial = \frac{\sum_{j=1}^i Q_j \cdot W_j}{\sum_{j=1}^i W_j}$$

onde: Q_j é a quantidade média de erros que afetam à característica C_j , W_j é o peso da j -ésima característica e i é igual a quantidade de características com peso W_j .

c) *Valor de X_1 :*

$$X_1 = X_0 + X_1Parcial$$

Passo 5.2.1.3 – Cálculo da Linha X_2 que corresponde a 60 %

Tendo sido calculado anteriormente os valores de X_0 (características com peso igual a 0) e de X_1 (características com peso igual a 1) é preciso calcular o valor de X_2 Parcial considerando somente as características que têm peso igual a 3 (três):

a) Cálculo do $Q_{[1..m]}$ das $C_{[1..m]}$ considerando que todos os tipos de erros das características ocorreu apenas uma vez durante tu – tempo acumulado de utilização:

Sendo m a quantidade de características com peso 3 e que $m \geq 1$, para cada C_i calcula-se o Q_i (taxa média de erros que afetam C_i) da seguinte forma:

$$Q_{Medio_i} = \frac{1 \cdot p_{i1} + 1 \cdot p_{i2} + \dots + 1 \cdot p_{ij}}{p_{i1} + p_{i2} + \dots + p_{ij}}$$

$$Q_{Medio_i} = \frac{\sum_{j=1}^n 1 \cdot p_{ij}}{\sum_{j=1}^n p_{ij}}$$

$$Q_i = \frac{Q_{Medio_i}}{tu}$$

onde: n significa a quantidade de tipo de erro associado à característica C_i e p_{ij} é o peso do tipo de erro e_{ij} .

b) Calculo do X_2 Parcial:

Com os valores de $Q_{[1..m]}$, o valor de X_2 Parcial é determinado como:

$$X_2 \text{ Parcial} = \frac{Q_1 \cdot W_1 + Q_2 \cdot W_2 + \dots + Q_m \cdot W_m}{W_1 + W_2 + \dots + W_m}$$

$$X_2 \text{ Parcial} = \frac{\sum_{j=1}^i Q_j \cdot W_j}{\sum_{j=1}^i W_j}$$

onde: Q_j é a quantidade média de erros que afetam à característica C_j , W_j é o peso da j -ésima característica e i é igual a quantidade de características com peso W_j .

c) Valor de X_2 .

$$X_2 = X_1 + X_2\text{Parcial}$$

Passo 5.2.1.4 – Cálculo da Linha X_3 que corresponde a 40 %

Tendo sido calculado anteriormente os valores de X_0 (características com peso igual a 0), de X_1 (características com peso igual a 1) e de X_2 (características com peso igual a 3) é preciso calcular o valor de $X_3\text{Parcial}$ considerando somente as características que têm peso igual a 9 (nove):

a) Cálculo do $Q_{[1..m]}$ das $C_{[1..m]}$ considerando que todos os tipos de erros das características ocorreu apenas uma vez durante tu – tempo acumulado de utilização:

Sendo m a quantidade de características com peso 9 e que $m \geq 1$, para cada C_i calcula-se o Q_i (taxa média de erros que afetam C_i) da seguinte forma:

$$Q\text{Medio}_i = \frac{1 \cdot p_{i1} + 1 \cdot p_{i2} + \dots + 1 \cdot p_{ij}}{p_{i1} + p_{i2} + \dots + p_{ij}}$$

$$Q\text{Medio}_i = \frac{\sum_{j=1}^n 1 \cdot p_{ij}}{\sum_{j=1}^n p_{ij}}$$

$$Q_i = \frac{Q\text{Medio}_i}{tu}$$

onde: n é a quantidade de tipo de erro associado à característica C_i e p_{ij} significa o peso do tipo de erro e_{ij} .

b) Cálculo do $X_3\text{Parcial}$:

Com os valores de $Q_{[1..m]}$, o valor de $X_3\text{Parcial}$ é determinado como :

$$X_3Parcial = \frac{Q_1.W_1 + Q_2.W_2 + \dots + Q_m.W_m}{W_1 + W_2 + \dots + W_m}$$

$$X_3Parcial = \frac{\sum_{j=1}^i Q_j.W_j}{\sum_{j=1}^i W_j}$$

onde: Q_j é a quantidade média de erros que afetam a característica C_j , W_j é o peso da j -ésima característica e i corresponde a quantidade de características com peso W_j .

c) **Valor de X_3 :**

$$X_3 = X_2 + X_3Parcial$$

Passo 5.2.1.5 – Cálculo da Linha X_4 que corresponde a 20 %

Tendo sido calculado anteriormente os valores de X_0 (características com peso igual a 0), de X_1 (características com peso igual a 1), de X_2 (características com peso igual a 3) e de X_3 (características com peso igual a 9), é preciso calcular o valor de $X_4Parcial$ levando em consideração somente as características que têm peso igual a 27 (vinte e sete):

a) *Cálculo do $Q_{[1..m]}$ das $C_{[1..m]}$ considerando que todos os tipos de erros das características ocorreu apenas uma 1 vez durante tu – tempo acumulado de utilização:*

Sendo m a quantidade de características com peso 27 e que $m \geq 1$, para cada C_i calcula-se o Q_i (taxa média de erros que afetam C_i) da seguinte forma:

$$QMedio_i = \frac{1.p_{i1} + 1.p_{i2} + \dots + 1.p_{ij}}{p_{i1} + p_{i2} + \dots + p_{ij}}$$

$$QMedio_i = \frac{\sum_{j=1}^n 1.p_{ij}}{\sum_{j=1}^n p_{ij}}$$

$$Q_i = \frac{QMedio_i}{tu}$$

onde: n é a quantidade de tipo de erro associado a característica C_i e p_{ij} é igual ao peso do tipo de erro e_{ij} .

b) *Calculo do X_4 Parcial:*

Com os valores de $Q_{1..m}$, o valor de X_4 Parcial é determinado como:

$$X_4 \text{ Parcial} = \frac{Q_1.W_1 + Q_2.W_2 + \dots + Q_m.W_m}{W_1 + W_2 + \dots + W_m}$$

$$X_4 \text{ Parcial} = \frac{\sum_{j=1}^i Q_j.W_j}{\sum_{j=1}^i W_j}$$

onde: Q_j é a quantidade média de erros que afetam à característica C_j , W_j é o peso da j -ésima característica e i equivale a quantidade de características com peso W_j .

c) *Valor de X_4 :*

$$X_4 = X_3 + X_4 \text{ Parcial}$$

Nota:

Considerando que todos os tipos de erros das características ocorreram uma vez em um tu (tempo de utilização) acumulado até o momento da avaliação, todos os Q_{Medio_i} (valor médio de erros que afetam a C_i) serão sempre iguais a 1. Da mesma forma, o valor do Q_i (taxa média de erros que afetam a C_i) será sempre igual a $1/tu$. Sendo o valor parcial de $X_{[0..4]}$ a média de todos os Q_i , conseqüentemente, com exceção do X_0 , os valores parciais para $X_{[1..4]}$ serão sempre iguais a $1/tu$. Isso implica que as cinco linhas principais podem ser calculadas como:

$X_0 \text{ Parcial} = 0$; ($X_0 \text{ Parcial}$ é um caso especial, pois todos os pesos são iguais a 0)

$$X_0 = X_0 \text{ Parcial} \therefore \boxed{X_0 = 0}$$

$$X_1 = X_0 + X_{1Parcial} \therefore X_1 = 0 + \frac{1}{tu} \therefore X_1 = \frac{1}{tu}$$

$$X_2 = X_1 + X_{2Parcial} \therefore X_2 = \frac{1}{tu} + \frac{1}{tu} \therefore X_2 = 2 \cdot \frac{1}{tu}$$

$$X_3 = X_2 + X_{3Parcial} \therefore X_3 = 2 \cdot \frac{1}{tu} + \frac{1}{tu} \therefore X_3 = 3 \cdot \frac{1}{tu}$$

$$X_4 = X_3 + X_{4Parcial} \therefore X_4 = 3 \cdot \frac{1}{tu} + \frac{1}{tu} \therefore X_4 = 4 \cdot \frac{1}{tu}$$

Caso o X calculado esteja em algum intervalo, por exemplo, entre X_0 e X_1 , é preciso interpolar esse ponto para que se conheça exatamente qual é o valor do d correspondente. Neste sentido, os seguintes passos são necessários:

Passo 5.2.2 – Interpolação do X calculado

Podemos utilizar a fórmula geral para interpolação linear no cálculo do d , percentual de desempenho, correspondente ao X calculado.

$$d = \frac{(x - x_a)}{(x_b - x_a)} \cdot (d_b - d_a) + d_a$$

⇒ Se o $X \in [0, X_1]$, então o índice de qualidade está entre 100 e 80 %, assim:

$$x_a = 0 \quad ; \quad d_a = 100$$

$$x_b = X_1 \quad ; \quad d_b = 80$$

$$d = \frac{(X - 0)}{(X_1 - 0)} \cdot (80 - 100) + 100$$

$$d = \frac{-20X}{X_1} + 100$$

⇒ Se o $X \in [X_1, X_2]$, então o índice de qualidade está entre 80 e 60 %, assim:

$$x_a = X_1 \quad ; \quad d_a = 80$$

$$x_b = X_2 \quad ; \quad d_b = 60$$

$$d = \frac{(X - X_1)}{(X_2 - X_1)} \cdot (60 - 80) + 80$$

$$d = \frac{-20(X - X_1)}{(X_2 - X_1)} + 80$$

⇒ Se o $X \in [X_2, X_3]$, então o índice de qualidade está entre 60 e 40 %, assim:

$$x_a = X_2 \quad ; \quad d_a = 60$$

$$x_b = X_3 \quad ; \quad d_b = 40$$

$$d = \frac{(X - X_2)}{(X_3 - X_2)} \cdot (40 - 60) + 60$$

$$d = \frac{-20(X - X_2)}{(X_3 - X_2)} + 60$$

⇒ Se o $X \in [X_3, X_4]$, então o índice de qualidade está entre 40 e 20 %, assim:

$$x_a = X_3 \quad ; \quad d_a = 40$$

$$x_b = X_4 \quad ; \quad d_b = 20$$

$$d = \frac{(X - X_3)}{(X_4 - X_3)} \cdot (20 - 40) + 40$$

$$d = \frac{-20(X - X_3)}{(X_4 - X_3)} + 40$$

⇒ Se o $X \in [X_4, \infty]$, então o índice de qualidade está entre 20 e 0 %, assim:

$$d = \frac{20 \cdot X_4 - X_3}{X - X_3}$$

Passo 6: Resultado

Com o percentual de qualidade d calculado, o gráfico de qualidade de uma determinada versão do produto de banco de dados poderá indicar, além da qualidade atual e da qualidade calculada em avaliações anteriores, o momento em que o índice de qualidade ficou abaixo do nível tolerável (inferior a 60%), conforme ilustra a Figura 4.3. Junto com essa informação pode ser emitido um relatório onde constam todos os erros ocorridos que levaram a qualidade do produto para a Zona de Baixo Desempenho – ZBD. Este fato possibilita realizar investigações para a realização de ações corretivas.

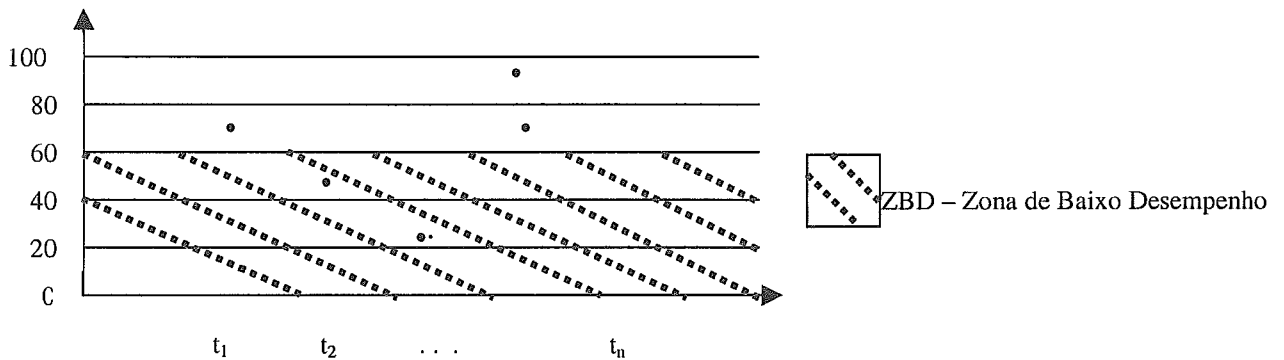


Figura 4.3 – Gráfico dos Índices de Qualidade

Onde T_n representa uma avaliação realizada no dia $dd/mm/aa$, que alcançou o índice de qualidade de $x\%$ durante um tempo de utilização de n minutos.

Sendo o modelo matemático descrito na Seção 4.1, uma das partes que compõem o processo de avaliação, a seguir serão descritas todas as etapas necessárias para realizar integralmente a avaliação da qualidade de dados pela não conformidade proposta nesta tese.

4.2 Etapas da Avaliação da Qualidade de Dados pela Não Conformidade

PRIMEIRA ETAPA: Determinação do Padrão de Qualidade de Dados Esperado (PQDE) pelos avaliadores do banco de dados:

Nesta etapa obtém-se do avaliador do banco de dados, o grau de importância de cada característica de qualidade relacionada na Seção 3.4. Para isso, é utilizado o Instrumento para Classificação das Características Necessárias para Avaliar a Qualidade de Dados, que pode ser visto no Apêndice III.

O peso atribuído a cada característica retrata, então, a expectativa de qualidade do usuário em relação aos dados armazenados.

Desta etapa resulta, o padrão de qualidade de dados esperado - PQDE.

SEGUNDA ETAPA: Avaliação da qualidade de um Banco de Dados, apoiando-se em um PQDE previamente definido:

A partir do padrão de qualidade de dados esperado (PQDE) obtido na primeira etapa, é possível medir o quanto o banco de dados satisfaz ao ideal de qualidade estabelecido.

Para obter o índice (percentual) de qualidade é preciso que todas as não conformidades ocorridas sejam cadastradas. Esse cadastramento é possível de ser realizado através do Formulário para Cadastro do Erro ou Não Conformidade Ocorrido, que se encontra no Apêndice II.

No momento da avaliação é identificado que característica de qualidade uma não conformidade está afetando. Assim, considerando o tempo acumulado de utilização do produto e as não conformidades que degradam a sua qualidade, o *Índice de Qualidade dos Dados Armazenados* é obtido através do modelo de avaliação que foi descrito na Seção 4.1.

TERCEIRA ETAPA: Resultados:

Assim como os índices de qualidade obtidos em avaliações anteriores, é exibido também o índice atual. Para isso, é utilizado um gráfico dividido em faixas percentuais, como o gráfico apresentado na Figura 4.2.

Caso o índice calculado seja inferior a 60%, a qualidade estará posicionada em uma área considerada como de baixo desempenho (ZBD – Zona de Baixo Desempenho). Nesse momento, além de um alerta, é também fornecido para o avaliador um relatório onde constam as características de qualidade e os erros cadastrados associados a elas.

Desta forma, é possível, de acordo com a política de qualidade de dados da organização, realizar ações de investigação das prováveis fontes de erros que levaram a degradação da qualidade dos dados.

Capítulo 5

UMA EXPERIÊNCIA DE UTILIZAÇÃO DA AVALIAÇÃO DA QUALIDADE DE DADOS PELA NÃO CONFORMIDADE

Neste capítulo estão descritas as etapas que são necessárias para realizar o experimento de validação da avaliação da qualidade de dados proposta.

A avaliação foi realizada através da utilização do protótipo da ferramenta automática desenvolvida, denominada AQUA. Os aspectos importantes da implementação desta ferramenta estão descritos na próxima seção.

5.1 Implementação do Protótipo da Ferramenta para Avaliação da Qualidade de Dados pela Não Conformidade

Para implementar a ferramenta automática AQUA utilizamos a linguagem de programação Delphi 5.0 e o banco de dados Paradox (CATÚ, 1997).

Utilizamos nesta seção três representações gráficas para elucidar os aspectos importantes da implementação da ferramenta. Estas representações são o diagrama de casos de uso, o diagrama de telas e o modelo lógico dos dados.

O diagrama de casos de uso modela as interações entre a ferramenta e os seus usuários, exibindo assim, as funcionalidades existentes em AQUA.

Através do diagrama de telas mostramos a hierarquia das telas da ferramenta.

Já a representação gráfica do esquema do banco de dados relacional utilizado é representada pelo modelo lógico dos dados.

As próximas seções ilustram cada uma destas representações gráficas

5.1.1 Diagrama de Casos de Uso

A Figura 5.1 ilustra o diagrama de casos de uso da ferramenta AQUA, onde os atores são os usuários e o avaliador. Observe que as funcionalidades presentes em AQUA são: Cadastra Erros, Cadastra Nova Característica, Cadastra Tipo de Erro Associado, Atribui Peso Características, Exibe Relatório, Exibe Gráfico Desempenho e Realiza Avaliação.

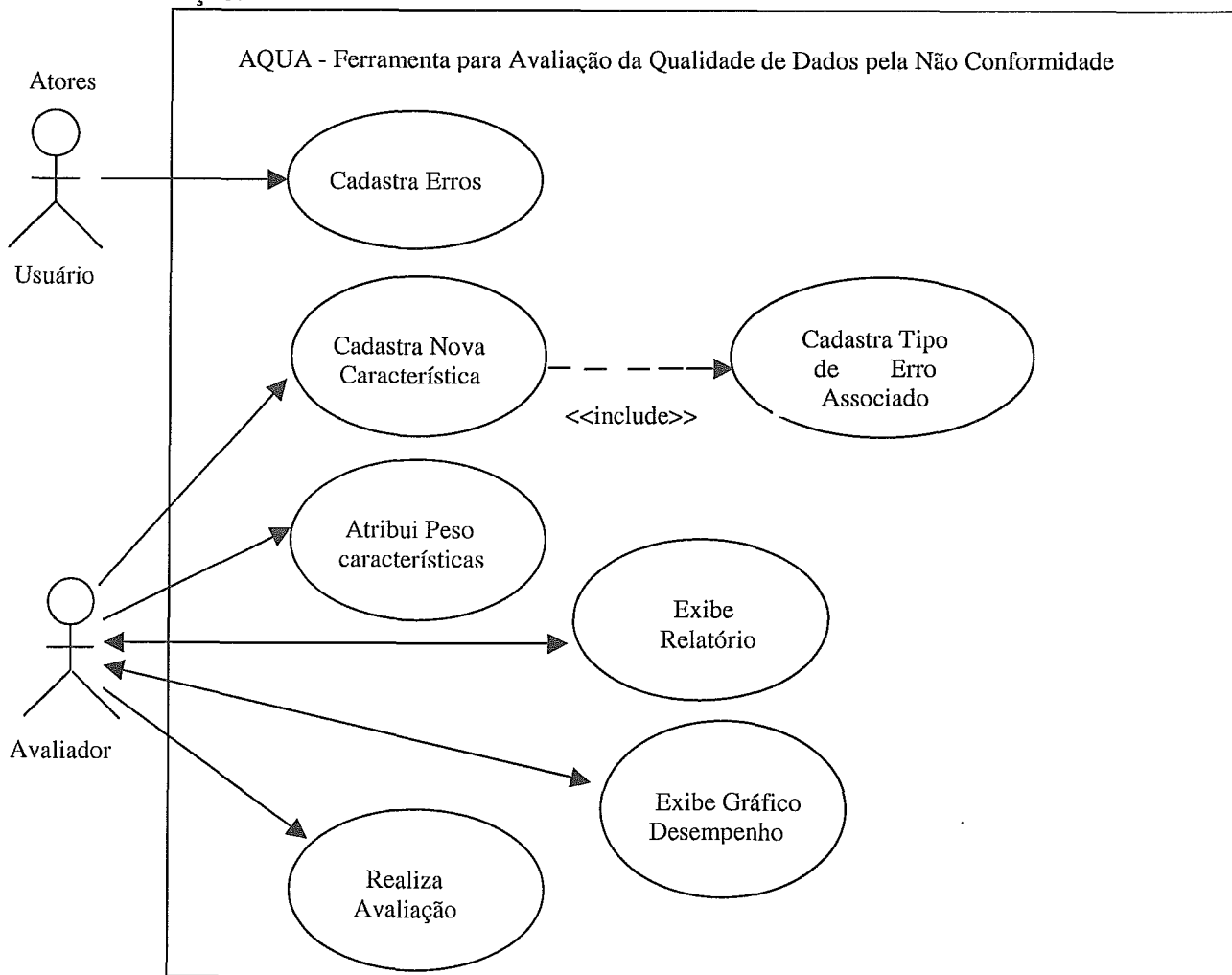


Figura 5.1 – Diagrama de Casos de Uso

As descrições a seguir, detalham a seqüência das transações de cada caso de uso.

Nome : Cadastra Erros
Ator : Usuário
Descrição : Realiza o cadastro dos erros ocorridos com o banco de dados que está sendo avaliado
Eventos : 1. Usuário abre tela de cadastro de erros
2. Usuário preenche os campos necessários para o cadastro: versão do produto, tipo de problema, descrição do problema, gravidade do erro, nome do funcionário e data
3. Usuário confirma os dados digitados
4. Os dados são armazenados na tabela caderro.db

Nome : Atribui Peso Características
Ator : Avaliador
Descrição : Cadastra o peso de cada característica de qualidade
Eventos : 1. Avaliador abre tela de atribuição do peso das características
2. É atribuído um peso para cada característica de qualidade que está armazenada na tabela carac.db
3. Avaliador confirma o dado digitado
5. O dado é armazenado na tabela carac.db

Nome : Cadastra nova característica
Ator : Avaliador
Descrição : Cadastra uma nova característica de qualidade, assim como o(s) tipo(s) de erro(s) associado(s) a ela
Eventos : 1. Avaliador abre tela de cadastro de nova característica
2. Avaliador preenche os campos necessários para o cadastro: nome da característica, descrição da característica e o peso da característica
3. Avaliador confirma os dados digitados
4. Os dados são armazenados na tabela carac.db
6. É executado o caso de uso Cadastra Tipo de Erro Associado
7. Sistema exibe tela de cadastro de tipo de erro
8. Avaliador preenche os campos necessários para o cadastro: o nome, a descrição e o peso do tipo de erro associado(s) a nova característica cadastrada
9. Avaliador confirma os dados digitados
10. Os dados são armazenados na tabela tipoprob.db

Nome : Realiza Avaliação

Ator : Avaliador

Descrição : Realiza a avaliação da qualidade de dados pela não conformidade e exibe os resultados obtidos através de um relatório e de um gráfico de desempenho

Eventos : 1. Avaliador solicita que a avaliação seja realizada
2. Sistema exibe tela de solicitação dos campos: tempo de utilização do produto e data da avaliação
3. Avaliador preenche os dados solicitados
4. Avaliador confirma os dados digitados
5. Sistema exibe tela de solicitação da tolerância a erros gravíssimos
6. Avaliador preenche o dado solicitado
7. Avaliador confirma o dado digitado
8. Sistema realiza a avaliação da qualidade e armazena os resultados obtidos na tabela indicequalidade.db

Nome : Exibe Relatório

Ator : Avaliador

Descrição : Exibe um relatório contendo os resultados obtidos na avaliação corrente

Eventos : 1. Avaliador solicita a exibição do relatório
2. Sistema exibe o relatório solicitado baseado nas tabelas indicequalidade.db e caderro.db

Nome : Exibe Gráfico de Desempenho

Ator : Avaliador

Descrição : Exibe o gráfico de desempenho que contém os índices de qualidade que foram calculados em todas as avaliações realizadas

Eventos : 1. Avaliador solicita a exibição do gráfico de desempenho
2. Sistema exibe o gráfico solicitado baseado na tabela resultados.db

5.1.2 Diagrama de Telas

A Figura 5.2 mostra a hierarquia das telas em AQUA.

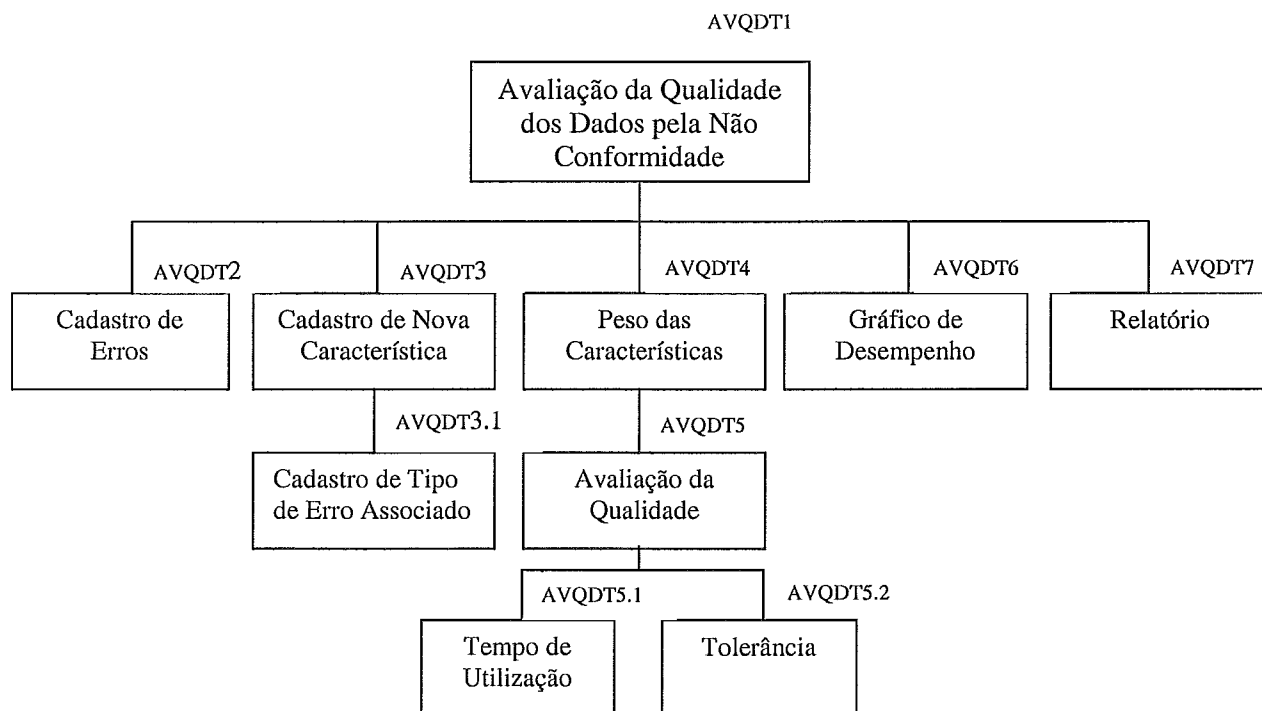


Figura 5.2 – Diagrama de Telas

A sigla das telas é composta da seguinte forma:

AQUA - avaliação da qualidade de dados pela não conformidade

T - componente é uma tela

9 - número da tela, que representa a sua posição na hierarquia

5.1.3 Modelo Lógico dos Dados

A Figura 5.3 apresenta o modelo lógico dos dados utilizados em AQUA.

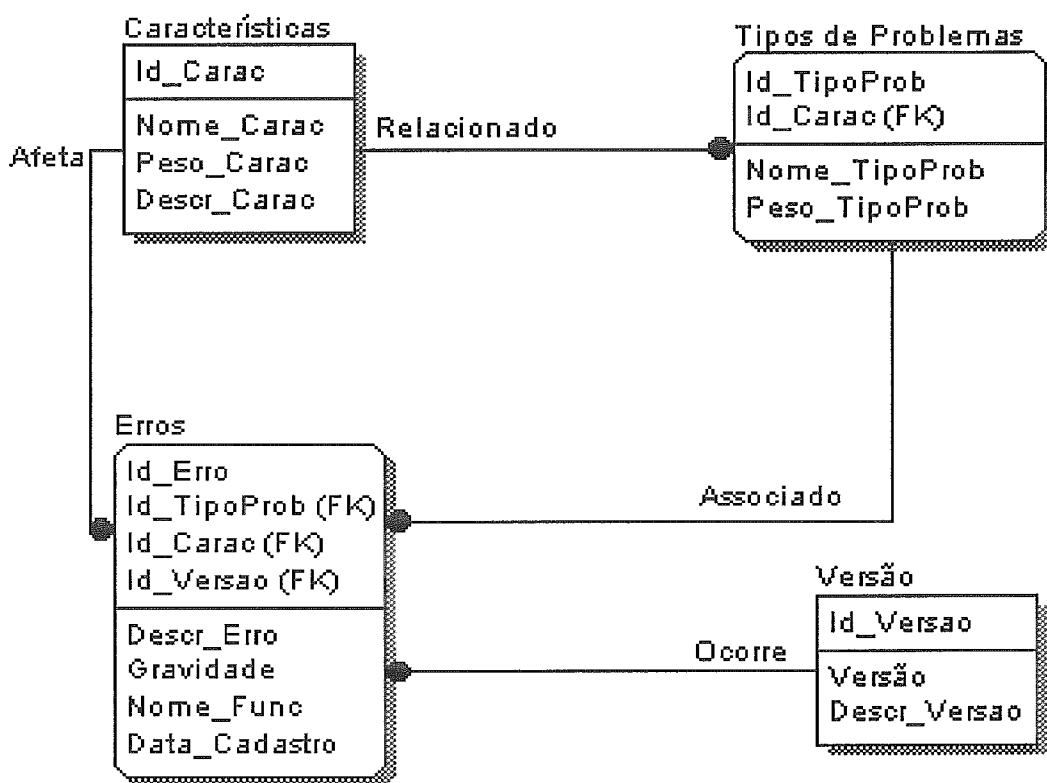


Figura 5.3 – Modelo Lógico dos Dados

O dicionário de dados do modelo físico que foi gerado a partir do modelo lógico mostrado na Figura 5.3, pode ser encontrado no Apêndice V. Da mesma forma, o guia para utilização da ferramenta está descrito no Apêndice IV.

A próxima seção descreve as etapas do experimento realizado. O alvo da avaliação foi o Sistema CAM – Controle de Área Marítima, um sistema de apoio à decisão do Centro de Análises de Sistemas Navais, organização militar da Marinha do Brasil.

5.2 Etapas da Experiência de Utilização da Avaliação

Para avaliar a qualidade dos dados armazenados do Sistema CAM foram realizadas as etapas de cadastro dos erros ou não conformidades, determinação do padrão de qualidade de dados esperado (PQDE), avaliação da qualidade do banco de dados e exibição do resultado. Cada uma dessas etapas está descrita a seguir.

Etapa 1: Cadastro dos erros ou não conformidades

Especialistas navais, usuários do Sistema CAM, cadastraram todos os erros que ocorreram num período de utilização de 4800 minutos (aproximadamente 10 dias). O objetivo foi obter o conjunto de não conformidades que afetam à qualidade dos dados armazenados. Esse conjunto de não conformidades pode ser visto na Tabela 5.1.

A Tabela 5.1 corresponde as não conformidades que ocorreram na versão P1.0. A primeira e a segunda coluna correspondem aos tipos de problemas que ocorreram e a descrição sucinta dos mesmos, respectivamente. A terceira coluna informa os pesos relativos a gravidade (G) de cada erro ocorrido. A quarta e quinta coluna identifica o funcionário (F) que cadastrou o erro e a data (DC) em que o erro ocorreu.

Por medidas de cumprimento ao grau de sigilo das informações contidas nesta tabela, os nomes dos funcionários e os meios navais citados nas descrições dos erros são fictícios.

TIPO DE PROBLEMA	DESCRIÇÃO	G	F	DC
Inserção de erro na base de dados após ocorrer situação anormal de funcionamento do sistema	Por causa de um problema no fornecimento de energia elétrica, houve a interrupção da distribuição de uma área fixa e os dados dos meios amigos foram perdidos	3	Func1	06/01/2001
Falta de alerta (aviso) para indicar entrada de dado incorreta ou não conforme	Foi possível cadastrar uma velocidade de reação negativa para os meios amigos	3	Func1	07/01/2001
Falta de campo para cadastramento	Não foi possível cadastrar o nome do comandante dos meios amigos	1	Func2	09/01/2001
Dado armazenado sem atualização por muito tempo	A largura de varredura dos meios amigos não está sendo atualizada conforme o plano em vigência, assim uma distribuição foi inadequada por causa deste dado estar sem a atualização correta	4	Func3	12/01/2001
Demora no fornecimento de resposta que necessita ser utilizada	A distribuição dos meios navais, quando utilizando dados do navio mãe, está demorando muito	3	Func1	08/02/2001
Manipulação do dado resulta em informação incorreta	A aeronave HXX com largura de varredura 0 não poderia ser classificada como patrulhador	4	Func3	14/02/2001
Falta de alerta (aviso) para indicar entrada de dado incorreta ou não conforme	Foi possível cadastrar probabilidade > 1.0	2	Func2	23/02/2001
Ausência de registro da autoria de alteração ou manipulação de dados	A velocidade de reação da FXX foi cadastrada como 45 nós e não foi possível identificar o autor deste cadastro	2	Func2	02/03/2001
Dado indisponível	Falta informação do navio mãe durante a distribuição	2	Func2	02/03/2001
Impossibilidade de interagir com outra base de dados	Não foi possível receber a informação lat/long do sistema Mercúrio, pois o CAM a utiliza em minutos e o Mercúrio como string	3	Func1	06/03/2001
Manipulação do dado resulta em informação incorreta	A aeronave HYY e seu navio mãe, não podem ser distribuídos como interceptadores separadamente	4	Func3	06/03/2001
Dado com dois ou mais valores diferentes armazenados na base de dados	O campo Tempo Morto, que é exclusivo de aeronaves, equivale ao campo Preparação que existe para todos os meios amigos	2	Func1	09/03/2001
Falta de informação	Falta a informação de qual foi o perímetro protegido	2	Func2	02/11/2000
Demora na execução de uma tarefa	Levou muito tempo para recuperar as informações de todas as distribuições feitas para o Planejamento 045/2000	3	Func4	02/04/2001
Apresentação dos dados insatisfatória	Os dados do meio inimigo deveriam ser exibidos também na apresentação da área máxima e de forma destacada	2	Func4	23/03/2001
Informação pouco objetiva e gerando dúvidas	O nome do responsável pelo planejamento é do Planejador Tático ou do Oficial responsável pela utilização do Sistema CAM?	1	Func1	20/03/2001
Manipulação do dado resulta em informação incorreta	O meio AXX não pode ser distribuído separadamente das aeronaves que estão baseadas nele	4	Func2	15/03/2001
O dado armazenado não representa corretamente o seu significado no mundo real	O campo distância mínima navegada não reflete completamente os valores de gap que deve ser considerado	3	Func2	23/03/2001

Tabela 5.1 – Não Conformidades ocorridas no Sistema CAM

Etapa 2: Determinação do padrão de qualidade de dados esperado (PQDE)

Um especialista naval, responsável pela avaliação, analisou o conjunto de características de qualidade, cuja lista pode ser encontrada no Capítulo 3. Foi atribuído a cada uma delas o grau de importância da sua presença na avaliação da qualidade dos dados do Sistema CAM. A atribuição do grau de importância tem como objetivo capturar a expectativa de qualidade dos dados. Os pesos atribuídos estão exibidos na Tabela 5.2.

CARACTERÍSTICA	PESO
Disponibilidade da Informação	4
Idade dos Dados	3
Oportunidade	2
Eficiência de Execução	2
Relevância	2
Utilidade	2
Lucratividade	3
Benefício no Trabalho do Usuário	3
Competitividade	4
Quantidade de Dados Adequada	0
Acurácia	3
Completitude	2
Cobertura	3
Robustez	2
Precisão do Dado	4
Consistência	2
Sinalização	3
Atributabilidade	0
Recuperabilidade	3
Flexibilidade	1
Interoperabilidade	2
Segurança	4
Facilidade de Compreensão da Informação	0
Adequabilidade da Informação	2
Uniformidade	3
Disponibilidade de Documentação	2
Rastreabilidade	1

Tabela 5.2 – Peso das Características de Qualidade do Sistema CAM

Etapa 3: Avaliação da qualidade do banco de dados

Para realizar a avaliação proposta no Capítulo 4, foi considerado o PQDE obtido na etapa 2, assim como todas as não conformidades cadastradas. Desta forma, ao final da execução da avaliação, foi gerado o índice de qualidade dos dados armazenados.

Etapa 4: Exibição do resultado

O índice de qualidade aferido na etapa 3 representa como as não conformidades ocorridas no Sistema CAM afetaram às características de qualidade. Conseqüentemente, isto acarretou a degradação da qualidade geral dos dados do sistema.

A exibição do resultado obtido pode ser analisado com dois tipos de representação: o relatório dos erros ocorridos no sistema e o gráfico de desempenho. Tanto o relatório quanto o gráfico estão descritos a seguir.

No relatório dos erros ocorridos no sistema é mostrado o índice de qualidade que foi calculado, as não conformidades cadastradas e o tempo de utilização do produto. Deste modo, é possível que a organização, quando necessário, adote ações de investigação e/ou correção nas prováveis fontes de erro que degradaram a qualidade dos dados armazenados. Este relatório está ilustrado na Figura 5.4.

Neste relatório podemos observar que o índice de qualidade dos dados foi de 96,23%. Além dessa informação, o relatório também disponibiliza o dia em que foi realizada a avaliação, o tempo de utilização do produto e os erros que degradaram a qualidade dos dados armazenados no Sistema CAM.

O gráfico de desempenho mostra os índices de qualidade dos dados aferidos em todas as avaliações já realizadas. Deste modo, o gráfico exibido na Figura 5.5 mostra que o índice de qualidade de 96,23% foi o resultado da avaliação do dia 03/04/2001. Nesta avaliação foram utilizados os dados das Tabelas 5.1 e 5.2.

Assim, o gráfico de desempenho mostra o resultado da avaliação mais recente, assim como, o resultado de todas as outras avaliações realizadas. Desta forma, é possível que seja feito um acompanhamento da evolução da qualidade dos dados ao longo do tempo.

Avaliação da Qualidade de Dados pela Não Conformidade – Erros que degradaram o produto

Versão: P1.0 Data da Avaliação: 03/04/01 Utilização (min): 4800 Ind. Qualidade: 96.2334 %

Tipo Problema : Inserção de erro na base de dados após ocorrer situação anormal de funcionamento do sistema

Descrição : Por causa de um problema no fornecimento de energia elétrica, houve a interrupção da distribuição de uma área fixa e os dados dos meios amigos foram perdidos

Carac. Afetada : Robustez

Tipo Problema : Falta de alerta (aviso) para indicar entrada de dado incorreta ou não conforme

Descrição : Foi possível cadastrar uma velocidade de reação negativa para os meios amigos

Carac. Afetada : Sinalização

Tipo Problema : Demora na execução de uma tarefa

Descrição : Levou muito tempo para recuperar as informações de todas as distribuições feitas para o Planejamento 045/2000

Carac. Afetada : Eficiência de Execução

Tipo Problema : Falta de campo para cadastramento

Descrição : Não foi possível cadastrar o nome do comandante dos meios amigos

Carac. Afetada : Completitude

Tipo Problema : Dado armazenado sem atualização por muito tempo

Descrição : A largura de varredura dos meios amigos não está sendo atualizada conforme o plano em vigência, assim uma distribuição foi inadequada por causa deste dado estar sem a atualização correta

Carac. Afetada : Idade dos Dados

Tipo Problema : Manipulação do dado resulta em informação incorreta

Descrição : A aeronave P95 com largura de varredura 0 não poderia ser classificada como patrulhador

Carac. Afetada : Acurácia

Tipo Problema : Falta de alerta (aviso) para indicar entrada de dado incorreta ou não conforme

Descrição : Foi possível cadastrar probabilidade > 1.0

Carac. Afetada : Sinalização

Tipo Problema : Demora no fornecimento de resposta que necessita ser utilizada

Descrição : A distribuição dos meios navais, quando utilizando dados do navio mãe, está demorando muito

Carac. Afetada : Oportunidade

Tipo Problema : Informação pouco objetiva e gerando dúvidas

Descrição : O nome do responsável pelo planejamento é do Planejador Tático ou do Oficial Responsável pela utilização do Sistema CAM ?

Carac. Afetada : Facilidade de Compreensão da Informação

Tipo Problema : Ausência de registro da autoria de alteração/manipulação de dados

Descrição : A velocidade de reação da F43 foi cadastrada como 45 nós e não foi possível identificar o autor deste cadastro

Carac. Afetada : Atributabilidade

Tipo Problema : Apresentação dos dados insatisfatória

Descrição : Os dados do meio inimigo deveriam ser exibidos também na apresentação da área máxima e de forma destacada

Carac. Afetada : Uniformidade

Tipo Problema : Manipulação do dado resulta em informação incorreta

Descrição : O meio A11 não pode ser distribuição separadamente das aeronaves que estão baseados nele

Carac. Afetada : Acurácia

Tipo Problema : O dado armazenado não representa corretamente o seu significado no mundo real

Descrição : O campo distância mínima navegada não reflete completamente os valores de gap que devem ser considerados

Carac. Afetada : Precisão do dado

Tipo Problema : Dado indisponível

Descrição : Falta informação do navio mãe durante a distribuição

Carac. Afetada : Disponibilidade da Informação

Tipo Problema : Impossibilidade de interagir com outra base de dados	
Descrição : Não foi possível receber a informação lat/long do sistema mercúrio, pois o CAM a utiliza em minutos e o Mercúrio como string	
Carac. Afetada : Interoperabilidade	

Tipo Problema : Manipulação do dado resulta em informação incorreta	
Descrição : A aeronave SH3 e seu navio mãe, não podem ser distribuídos como interceptadores separadamente	
Carac. Afetada : Acurácia	

Tipo Problema : Dado com dois ou mais valores diferentes armazenado na base de dados	
Descrição : O campo Tempo Morto, que é exclusivo de aeronaves, equivale ao campo Preparação que existe para todos os meios amigos	
Carac. Afetada : Consistência	

Tipo Problema : Falta de Informação	
Descrição : Falta a informação de qual foi o perímetro protegido	
Carac. Afetada : Completitude	

Figura 5.4 - Relatório de Qualidade Sistema CAM

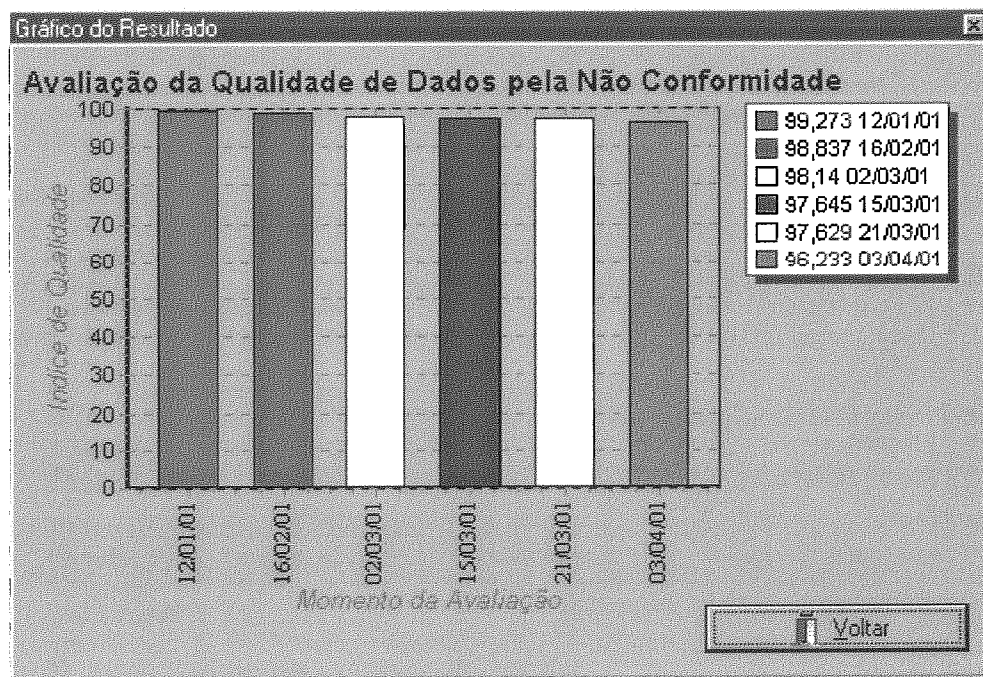


Figura 5.5 – Gráfico de Desempenho Sistema CAM

Capítulo 6

CONCLUSÕES

Neste trabalho foi discutido como a qualidade dos dados de uma organização auxiliam a tomada de decisões acertadas, favorecendo assim, a obtenção da diferenciação dentre outros concorrentes no mercado (*Capítulo 2*).

Discorreu-se ainda, sobre o papel das características de qualidade como sendo um elemento de referência básica à qualidade de um produto. As características de qualidade de dados utilizadas neste trabalho foram coletadas e definidas após pesquisa e buscas na literatura. Elas alcançaram a adequação necessária através de pesquisa de campo realizada junto a especialistas de instituições acadêmicas, militares e empresariais (*Capítulo 3*).

Desta forma, foi dado o devido suporte para que o modelo matemático do critério de avaliação desenvolvido pudesse ser utilizado durante a execução das etapas que realizam a avaliação da qualidade dos dados (*Capítulo 4*).

Uma experiência de utilização da avaliação proposta foi realizada através da ferramenta desenvolvida, denominada AQUA. Assim, foi possível validar a avaliação e fornecer um exemplo de sua utilização (*Capítulo 5*).

Com o desenvolvimento deste trabalho podemos ressaltar as seguintes contribuições:

- i.* a identificação e a descrição das características de qualidade de dados, necessárias para avaliar qualitativamente um banco de dados, assim como, os tipos de erros associados a cada uma delas. Essas características

foram padronizadas com o Modelo Rocha compatível com a Norma ISO 9126;

- ii. a sistematização da avaliação da qualidade de dados pela não conformidade, definindo os procedimentos necessários para se obter o índice de qualidade dos dados armazenados;
- iii. o desenvolvimento do modelo matemático do critério de avaliação; e
- iv. a implementação de um protótipo da ferramenta AQUA que automatiza a avaliação da qualidade de dados proposta.

Ressalta-se também, a importância da realização da experiência que utilizou um sistema de apoio à decisão da Marinha do Brasil denominado CAM. Nesta utilização do modelo de avaliação proposto, pode-se observar que os resultados obtidos satisfizeram a expectativa dos especialistas navais.

6.1 Perspectivas Futuras

Sendo a teoria dos conjuntos *fuzzy* (nebulosos) capaz de capturar as imprecisões do conhecimento humano (DUBOIS, 1991), é possível vislumbrar que a utilização desta técnica possa ser utilizada para melhor capturar a expectativa de qualidade de dados dos usuários.

Em relação ao protótipo da ferramenta AQUA, as seguintes sugestões podem ser consideradas para dar continuidade a este trabalho.

- Automação do cadastro de erros ou não conformidades;
- Captação automática do tempo de utilização do produto que está sendo avaliado; e
- Associação a um sistema de controle de versões.

A experiência de considerar como os erros, ou não conformidades, ocorridos influenciam na qualidade dos dados armazenados foi gratificante, motivando-nos a prosseguir nessa pesquisa. Assim, acreditamos que nos direcionamos para o caminho certo de preservar esse bem precioso para a organização, que são os seus dados armazenados.

REFERÊNCIAS BIBLIOGRÁFICAS

ADELMAN, S., MARCO, D., MOSS, L., 2001, “When we are cleaning data from a relational source, and some values”, *DM Review Online*, March 2001.

ANDRADE, C. J., 1991, *ANAFOR : Um Analisador de Programas Fortran*. Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro, Brasil.

BAHIA, A. S., 1992, *Avaliação da Complexidade de Software Científico*. Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro, Brasil.

BALDWIN, A. A., BOWEN, P. L., 1999, “Data Quality: A Review and Case Study”, *Journal of Information Systems*.

BALLOU, D. P., PAZER, H. L., 1985, “Modeling Data and Process Quality in Multi-input, Multi-output Information Systems”, *Management Science*, v. 31, n. 2 (Feb), pp. 150-162.

BELASCO, J. A. ., STAYER, R. C., 1994, *O vôo do Búfalo: Decolando para a Excelência, aprendendo a deixar os Empregados Assumirem a Direção*, Rio de Janeiro, Editora Campus, in [BELCHIOR, 1997].

BELCHIOR, A. D., 1992, *Características de Qualidade de Programas*, ES-265/92, COPPE/UFRJ, Rio de Janeiro, Brasil.

BELCHIOR, A. D., 1992, *Controle da Qualidade de Software Financeiro*. Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro, Brasil.

BELCHIOR, D. A., 1997, *Um Modelo Fuzzy para Avaliação da Qualidade de Software*. Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro, Brasil.

BLASCHECK, J. R., 1995, *Planejamento de Projetos*. Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro, Brasil.

BOWEN, P., 1992, *Managing data quality accounting information systems: A stochastic clearing systems approach*. PhD dissertation, University of Tennessee, USA.

BOWEN, P., SCHNEIDER, G. P., FIELDS K. T., 1995, "Managing Data Quality in Client/Server Environment", *IS Audit & Control Journal*, v.4, pp. 29-35.

BOWEN, L.P., FUHRER, A. D., GUESS, M. F., 1998, "Continuously Improving Data Quality in Persistent Databases", *Data Quality Journal*, v. 4, n. 1 (Set).

BRUNSTEIN, I., BUZZINI, R. R., 1996, "TQM: Um Paradigma Visionário para as Médias Empresas Brasileiras: Estudo de Casos". In: *XVII ENEGEP*, Piracicaba, SP.

CAMPOS, F. C., 1994, *Hipermídia na Educação: Paradigmas e Avaliação da Qualidade*. Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro, Brasil.

CAMPOS, G. H. B., 1994, *Metodologia para Avaliação da Qualidade de Software Educacional: Diretrizes para Desenvolvedores e Usuários*. Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro, Brasil.

CANTÚ, M., 1997, *Dominando o Delphi 3: A bíblia*. Makron Books, São Paulo.

CARVALHO, D., 1994, *Requisitos de Qualidade para Software Médico*, Relatório Técnico, Fundação Baiana de Cardiologia, Rio de Janeiro, Brasil.

CARVALHO, D. O., 1997, *Qualidade de Sistemas de Informação Hospitalar*. Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

CHOW, K. T., 2000, "Enterprise Data Quality: Covering all your Bases", *DM Review*, December 2001.

- CLUNIE, C.E., 1987, *Verificação e Validação de Software na fase de Especificação de Requisitos*. Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro, Brasil.
- CLUNIE, C. E., 1997, *Avaliação da Qualidade de Especificações Orientadas a Objeto*. Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro, Brasil.
- CODD, E. F., 1970, “A Relational model of data for large shared data banks”, *Communication of the ACM*, v.13 (Jun), pp. 377-387.
- COAD, P., YOURDON, E., 1996, *Análise baseada em objetos*. 2. ed. Rio de Janeiro, Campus.
- COELHO, A . C. B., 1997, *Atributos de Qualidade para uma Abordagem Sociotécnica na Aquisição de Software*. Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro, Brasil.
- COMMERLATO, C. A., 1994, *Uma Ferramenta para Seleção de Módulos Reutilizáveis*. Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro, Brasil.
- DATE, C.J., 1995. *Introduction to Database Systems*, Sixth Edition. Addison-Wesley.
- DUBOIS, D., 1991, *Fuzzy sets in approximate reasoning, Part 1: Inference with possibility distributions*, Fuzzy Sets and Systems, IFSA, Special Memorial Volume:25 years of fuzzy sets, North-Holland – Amsterdam , in (BELCHIOR, 1997).
- DUNN, R. H., ULLMAN, R. S., 1994, *TQM for Computer Software*. 2 ed. New York, McGraw-Hill.
- DYER, M., 1992, *The Cleanroom Approach to Quality Software Development*. New York, John Wiley and Sons.
- FALCONI C. V., 1992, *TQC - Controle da Qualidade Total (no estilo japonês)*, UFMG, Minas Gerais, Brasil.

FLYNN, B. B., SCHROEDER, R. G., SAKAKIBARA S., 1994, “A framework for quality management research and an associated measurement instrument”, *Journal of Operations Management*, v. 11, pp. 339-366.

FURLAN, J. D., 1998, *Modelagem de objetos através da UML - the unified modeling language*. 1. ed., São Paulo, Makron.

GENTLEMAN, W. M., 1996, “Is Software quality is a perception, how do we measure it?”. *The 6th I.C. on Software Quality*, Ottawa, October, in (BELCHIOR, 1997).

HEALY, M., 2000, “The Importance of Customer Data Quality “, *DM Review Direct*, August 2000.

HETZEL, B., 1988, *The Complete Guide to Software Testing*. 2 ed. New York, QED Information Science.

HOXMEIER, J. A., 1997, “A Framework for Assessing Database Quality”, [//osm7.cs.byu.edu/ER97/workshop4/jh.html](http://osm7.cs.byu.edu/ER97/workshop4/jh.html).

HRADESKY, J. L., 1989, *Aperfeiçoamento da Qualidade e da produtividade: Guia prático para a implementação do CEP: Controle Estatístico de Processos*. São Paulo, McGraw-Hill.

HUH, Y. U., KELLER, F. R., REDMAN, T. C., WATKINS, A. R., 1990, “Data Quality”, *Information & Software Technology*, v. 32, n. 8 (Oct), pp. 559-565.

HUMPHREY, W. S., 1995, *A discipline for Software Engineering*, Addison-Wesley Publishing Company, in (BELCHIOR, 1997).

ISHIKAWA, KAOM, 1984, *La gestion de la qualité: outils et applications pratiques*. Paris, Bordas.

ISHIKAWA, K., 1990, *Introduction to Quality Control*. Japan, 3A Corporation.

ISO 9126, 1991, *Information technology - Software product evaluation- Quality characteristics and guidelines for their use*, ISO/IEC.

ISO 8402, 1994, *Quality Vocabulary*, ISO DIS, in (BELCHIOR, 1997).

JOHNSON, J.R., LEITH, R.A., NETER, J., 1981, "Characteristics of errors in accounts receivable and Inventory Audits", *Accounting Review*, v. 56, pp. 270-293.

KITCHENHAM, B. *et al.*, 1996, *Software Quality : The Elusive Target*, IEEE Software, January, in (BELCHIOR, 1997).

KUME, H., 1985, *Statistical Methods for Quality Improvement*, Japan.

LAUDON, K. C., 1986, "Data Quality and due process large inter-organizational record systems", *Communications of the ACM*, v. 29, n. 1, pp 4-11.

LEVITIN, A V., REDMAN, T C., 1993, "A model of the data (life) cycles with application to quality", *Information & Software Technology*, v. 35, n. 4 (Apr), pp 217-223.

LIMA, K. V., 1999, *Uma Aplicação em Telemedicina*. Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro, Brasil.

LOURENÇO FILHO, R. de C. B., 1982, *Controle Estatístico de Qualidade*. Rio de Janeiro, Brasil, LTC.

MARCO, D., 2000, "Meta Data and Data Administration: Implementing Data Quality Through Meta Data, Part 1", *DM Review Direct*, November 2000.

McGEE, A. M., WANG, R. Y., 1993, "Total Data Quality Management (TDQM): Zero Defect Data Capture". *The Chief information Officer (CIO) - Perspectives Conference*, Tucson, Arizona, USA, March.

- MONTGOMERY, D. C., 1996, *Introduction to Statistical Quality Control*, 3 ed. USA.
- MUSA, J., IANINO A., OKUMOTO K., 1990. *Software Reliability: Professional Edition*, New York, McGraw Hill.
- OLIVEIRA, K. M., 1995, *Avaliação da Qualidade de Sistemas Especialistas*. Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro, Brasil.
- PALERMO, S., ROCHA, A . R. C., 1989, *An Experience on Evaluation Software Quality for High Energy Physics*, Computer Physics Communications.
- PASSOS, M. C. J. F., 1991, *Uma Ferramenta para Construção e Avaliação de Gráficos de Estrutura*. Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro, Brasil.
- PETERS, T., 1993, *Rompendo as Barreiras da Administração para Enfrentar a nova realidade*. São Paulo, Editora Harbra, in (BELCHIOR, 1997).
- PHAM, H., 1995, “Software Reliability and Testing”, *IEEE Computer Society Press*, California, USA.
- REDMAN, T. C., 1995, “Improve Data Quality for Competitive Advantage”, *Sloan Management Review*, v. 36, n. 2, pp. 99-109.
- ROCHA, A . R. C. da, 1983, *Um Modelo para Avaliação da Qualidade de Especificações*. Tese de Doutorado, PUC-RJ, Rio de Janeiro, Brasil.
- ROCHA, A . R. C. da, 1987, *Análise e Projeto Estruturado de Sistemas*. Rio de Janeiro, Campus, in (BELCHIOR,1997).
- SCHNEIDEWIND, N. F., 1993, *Report on the IEEE standard for a software quality metrics methodology*, IEEE, Conference on Software Measurements, Montreal, Canada, September, in (BELCHIOR, 1997).

STOKER, S., 2000, "Data Cleansing: The First Step to Knowledge", *DM Review Direct*, February 2000.

STRIGINI, L., 1996, *Limiting the Dangers of Intuitive Decision Making*, IEEE Software, January, in (BELCHIOR, 1997).

STRONG, D., WANG, R.Y., GUARASCIO, M. L., 1994, "Beyond Accuracy: What Data Quality Means to Data Consumers". *TDQM Research Program*, USA, October.

STRONG, D. M., MILLER, S. M., 1995, "Exceptions and Exception Handling", pp. 206-233.

SULLIVAN, L. P., 1986, *The Seven Stage in Company-Wide Quality Control*, Quality Progress.

TAN, J. K. H., BENBASAT, I., 1990, "Processing of Graphical Information: A Decomposition Taxonomy to Match Data Extraction Tasks and Graphical Representations", *Information Systems Research*, v. 1, n. 4 (Apr), pp. 416-439.

TAUSWORTHE, R. C., 1995, "Software quality management through process and product modeling". In: *Annals of Software Engineering*, pp. 119-139, in (BELCHIOR, 1997).

TSICHRITZIS, D. C., LOCHOVSKY, F.H., 1982, *Data Models*. Englewood Cliffs, NJ, Prentice Hall.

TSUKUMO, A. N., REGO, C.M., SALVIANO, C. F., AZEVEDO, G. F., MENEGHETTI, L. K., COSTA, M. C. C., CARVALHO, M. B., COLOMBO, R. M. T, 1997, "Qualidade de Software - Visões de Produto e Processo". In: *II ERI da SBC*, Junho.

VALLE, C., XIMENES, A. A., CAMPOS, G., 1997, "Educação de Pacientes através de Sistemas de Acesso Público", *Revista Brasileira de Informática na Educação*, v. 1, n. 1 (Set).

VALLE, C., 1997, *Sistemas de Acesso Público para a Educação de Pacientes*. Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro, Brasil.

YOURDON, E., 1985, "When Good Enough Software is Best", *IEEE Software*, May.

WAND, Y., WANG, R.Y., 1996, "Anchoring Data Quality Dimensions in Ontological Foundations", *Communications of the ACM*, v. 39, n. 11 (Nov), pp. 86-95.

WANG, R. Y., KON, H. B., MANDICK, S. E., 1993, "Data Quality Requirements Analysis and Modeling". In: *The Proceeding of the 9th International Conference on Data Engineering*, pp. 670-677, Vienna.

WANG, R.Y., KON, H. B., 1992, "Toward Total Data Quality Management (TDQM)", *TDQM Research Program, Sloan School of Management*, June.

WANG, R.Y., REDDY, M P., KON, H. B., 1995, "Toward quality data: An attribute-based approach", *Decision Support Systems*, v. 13, n. 3 (Mar), pp. 349-372.

WEBER, R., 1988. *EDP Auditing*, 2nd Edition, New York, McGraw-Hill.

WILLSHIRE, J. M., MEYEN, D., 1997, "A Process for Improving Data Quality", *Data Quality Journal*, v. 3, n. 1 (Set).



Universidade Federal do Rio de Janeiro

COPPE - Programa de Engenharia de Sistemas e Computação

Apêndice I

FORMULÁRIO DE VALIDAÇÃO DAS CARACTERÍSTICAS DE QUALIDADE

I . Objetivo

Através de uma pesquisa de campo, especialistas validaram o conjunto de características de presença importante na avaliação da qualidade de dados. O resultado desta etapa é o conjunto final de características capaz de suportar todos os erros do produto de banco de dados que está sendo avaliado. A inclusão de uma nova característica de qualidade é realizada sempre que algum especialista identificava a ausência dessa característica da lista inicial. Para registrar esses casos é preenchido o Cadastro para Sugestão de Nova Característica.

Ao final da definição de cada característica de qualidade está citado o *score* obtido por cada uma delas durante a pesquisa de campo. Esse *score* representa a importância da presença da característica de qualidade na avaliação segundo a opinião dos especialistas. Uma característica foi excluída da lista quando não obteve o *score* ≤ 2 , conforme critério descrito na Seção 3.2.

II . Instruções para Preenchimento do Formulário

Atribua valores de 0 a 4, baseado na sua opinião e/ou experiência, às características de acordo com o grau de importância da sua presença na avaliação, segundo a escala apresentada na Tabela I.1. No caso da identificação de alguma característica que não esteja relacionada, preencher o Cadastro para Sugestão de Nova Característica apresentado no final deste formulário.

Escala	Equivalência	Interpretação
0	Presença com Nenhuma Importância	Indica que a presença da característica que está sendo apresentada não tem nenhuma importância.
1	Presença com Importância Baixa	Indica que a presença da característica que está sendo apresentada tem pouca importância.
2	Presença com Importância Moderada	Indica que a presença da característica que está sendo apresentada tem importância moderada.
3	Presença com Importância Alta	Indica que a presença da característica que está sendo apresentada é muito importante.
4	Presença com Total Importância	Indica de maneira absoluta que não há dúvidas que a presença da característica que está sendo apresentada é imprescindível.

Tabela I.1 - Escala de Valores

III . Identificação do Avaliador

Nome do Avaliador:	
Posto/Cargo:	Departamento:

IV. Validação das Características de Qualidade

AVALIAÇÃO DA QUALIDADE DOS DADOS OBJETIVO: UTILIZABILIDADE

1 . Disponibilidade da Informação					
Processo de Avaliação: • Qual a importância de avaliar se os dados estão disponíveis em todos os locais onde poderão ser úteis, independente do local onde estão armazenados?	Valores				
	0	1	2	3	4
Comentários:					
Score : 4					

2 . Idade dos Dados					
• Processo de Avaliação: Qual a importância de avaliar se os dados estão atualizados, isto é, o tempo que estão armazenados é apropriado para a realização das tarefas da organização?	Valores				
	0	1	2	3	4
Comentários:					
Score : 3,2					

3 . Oportunidade					
• Processo de Avaliação: Qual a importância de avaliar se a manipulação dos dados gera uma informação com a rapidez necessária para que seja útil?	Valores				
	0	1	2	3	4
Comentários:					
Score : 3,7					

4 . Eficiência de Execução					
• Processo de Avaliação: Qual a importância de avaliar se a velocidade de acesso da manipulação dos dados é a ideal para executar um processamento?	Valores				
	0	1	2	3	4
Comentários:					
Score : 3,4					

5 . Auxílio em Situação de Erro					
• Processo de Avaliação: Qual a importância de avaliar a existência de serviços que disponibilize informações para seus usuários, apoiando-os em situações de erro, onde a precisão do dado possa ser afetada ?	Valores				
	0	1	2	3	4
Comentários:					
Score : 1,6 - Característica excluída					

6 . Relevância					
• Processo de Avaliação: Qual a importância de avaliar se as tarefas estão sendo feitas de forma correta e completa?	Valores				
	0	1	2	3	4
Comentários:					
Score : 2,3					

7 . Utilidade					
• Processo de Avaliação: Qual a importância de avaliar a contribuição que a utilização dos serviços de banco de dados oferecem à organização?	Valores				
	0	1	2	3	4
Comentários: 2,7					
Score :					

8 . Lucratividade					
• Processo de Avaliação: Qual a importância de avaliar se a utilização dos dados proporciona um aumento de produtividade?	Valores				
	0	1	2	3	4
Comentários:					
Score : 3,5					

9 . Benefício no Trabalho do Usuário					
• Processo de Avaliação: Qual a importância de avaliar se os serviços oferecidos pelo sistema ou pelo banco de dados simplificaram e melhoraram as condições de trabalho dos usuários?	Valores				
	0	1	2	3	4
Comentários:					
Score : 3,2					

10 . Competitividade					
• Processo de Avaliação: Qual a importância de avaliar se a utilização dos dados proporciona vantagens competitivas à organização com relação aos concorrentes do mercado?	Valores				
	0	1	2	3	4
Comentários:					
Score : 3,4					

11 . Viabilidade Tecnológica					
• Processo de Avaliação: Qual a importância de avaliar a existência e a disponibilidade da tecnologia necessária para realizar todas as tarefas envolvidas nesta meta ?	Valores				
	0	1	2	3	4
Comentários:					
Score : 1,2 - Característica Excluída					

12 . Viabilidade de Mão-de-Obra					
• Processo de Avaliação: Qual a importância de avaliar a existência e a disponibilidade e mão-de-obra necessária para realizar todas as tarefas envolvidas nesta meta ?	Valores				
	0	1	2	3	4
Comentários:					
Score : 1,4 - Característica Excluída					

AVALIAÇÃO DA QUALIDADE DOS DADOS
OBJETIVO : CONFIABILIDADE CONCEITUAL

1 . Quantidade de Dados Adequada					
• Processo de Avaliação: Qual a importância de avaliar se a quantidade e/ou volume de dados disponível são apropriados para realizar as tarefas da organização?	Valores				
	0	1	2	3	4
Comentários:					
Score : 2,9					

2 . Acurácia					
• Processo de Avaliação: Qual a importância de avaliar se os serviços oferecidos pelo sistema ou pelo banco de dados realizam a manipulação dos dados gerando resultados corretos ou conforme esperados?	Valores				
	0	1	2	3	4
Comentários:					
Score : 3,8					

3 . Completitude					
• Processo de Avaliação: Qual a importância de avaliar se existem todos os dados necessários, para que sejam geradas todas as informações úteis e necessárias para os usuários?	Valores				
	0	1	2	3	4
Comentários:					
Score : 3,3					

4 . Cobertura					
• Processo de Avaliação: Qual a importância de avaliar se o âmbito (escopo) da informação contida nos dados é suficientemente profundo, a fim de cobrir todo o escopo necessário exigido pelas tarefas da organização?	Valores				
	0	1	2	3	4
Comentários:					
Score : 2,7					

5. Robustez					
• Processo de Avaliação: Qual a importância de avaliar a existência de serviços do sistema capazes de controlar situações anormais, reagindo a elas sem perda do controle, não inserindo erro à base de dados?	Valores				
	0	1	2	3	4
Comentários:					
Score : 3,9					

6. Precisão do Dado					
• Processo de Avaliação: Qual a importância de avaliar se o dado representa com exatidão, de forma perfeita, correta, clara e concisa o seu significado no mundo real?	Valores				
	0	1	2	3	4
Comentários:					
Score : 4,0					

7. Reputação					
• Processo de Avaliação: Qual a importância de avaliar a fonte dos dados?	Valores				
	0	1	2	3	4
Comentários:					
Score : 1,6 - Característica excluída					

8. Consistência					
• Processo de Avaliação: Qual a importância de avaliar se o valor do dado é o mesmo em toda a base de dados, não possuindo conteúdo contraditório ou significados diferentes?	Valores				
	0	1	2	3	4
Comentários:					
Score : 3,4					

9. Sinalização					
• Processo de Avaliação: Qual a importância de avaliar a existência de alertas para entrada de dados que estão fora dos padrões especificados ou para qualquer manipulação que indique problemas para a qualidade dos dados armazenados?	Valores				
	0	1	2	3	4
Comentários:					
Score : 3,7					

10 . Atributabilidade					
• Processo de Avaliação: Qual a importância de avaliar a existência do mecanismo que registra a identificação do autor de qualquer manipulação de dados, assim como a data e de onde foi realizada a operação?	Valores				
	0	1	2	3	4
Comentários:					
Score : 3,2 - Característica incluída durante a pesquisa de campo					

11 . Recuperabilidade					
• Processo de Avaliação: Qual a importância de avaliar a possibilidade, em caso de falha, de se recuperar os dados diretamente afetados, em um tempo e com um esforço viável?	Valores				
	0	1	2	3	4
Comentários:					
Score : 3,1					

12 . Flexibilidade					
• Processo de Avaliação: Qual a importância de avaliar se os dados podem ser manipulados e administrados facilmente, possibilitando a sua expansão, adaptação e agregação com outros dados?	Valores				
	0	1	2	3	4
Comentários:					
Score : 2,8					

13 . Interoperabilidade					
• Processo de Avaliação: Qual a importância de avaliar a capacidade de interação com outras bases de dados, a fim de complementar informações e tecnologias?	Valores				
	0	1	2	3	4
Comentários:					
Score : 2,6					

14 . Segurança					
<p>• Processo de Avaliação: Qual a importância de avaliar a segurança dos dados armazenados sob o ponto de vista do seu uso, evitando acesso não autorizado, acidental ou deliberado, protegendo os dados de violações e garantindo a sua privacidade?</p>	Valores				
	0	1	2	3	4
Comentários:					
Score : 4,0					

15 . Facilidade de Compreensão da Informação					
<p>• Processo de Avaliação: Qual a importância de avaliar se a informação gerada é precisa e objetiva, não colocando o usuário em situações de dúvida?</p>	Valores				
	0	1	2	3	4
Comentários:					
Score : 3,3					

16. Adequabilidade da Informação					
<p>• Processo de Avaliação: Qual a importância de avaliar se a informação gerada é adequada a todos os seus usuários, levando em consideração aspectos sociais e culturais compatíveis com os diferentes níveis funcionais dentro da organização?</p>	Valores				
	0	1	2	3	4
Comentários:					
Score : 2,7					

**AVALIAÇÃO DA QUALIDADE DOS DADOS
OBJETIVO: CONFIABILIDADE CONCEITUAL**

1 . Uniformidade					
<p>• Processo de Avaliação: Qual a importância de avaliar se os dados estão apresentados continuamente em um mesmo formato, estando eles consistentemente representados e formatados?</p>	Valores				
	0	1	2	3	4
Comentários:					
Score : 3,3					

2 . Disponibilidade de Documentação					
<ul style="list-style-type: none"> • Processo de Avaliação: Qual a importância de avaliar a existência de uma documentação atualizada e disponível dos dados, tornando-os assim, bem documentados, verificáveis e facilmente associados a uma fonte, a fim de facilitar a sua localização e verificação pelos usuários? 	Valores				
	0	1	2	3	4
Comentários:					
Score : 3,1					

3 . Rastreabilidade					
<ul style="list-style-type: none"> • Processo de Avaliação: Qual a importância de avaliar a existência do mecanismo de busca da informação que seja feito através da seqüência de agregação de detalhes de um determinado aspecto, desde a sua visão mais geral até a mais detalhada, e vice-versa? 	Valores				
	0	1	2	3	4
Comentários:					
Score : 2,6					

V. Cadastro para Sugestão de Nova Característica

Data:
Característica sugerida:
Justificativa:
Tipos de Erros associados a característica sugerida:
Comentários:



Universidade Federal do Rio de Janeiro

COPPE - Programa de Engenharia de Sistemas e Computação

Apêndice II

FORMULÁRIO PARA CADASTRO DO ERRO OU NÃO CONFORMIDADE OCORRIDO

I . Objetivo

Registrar todas as ocorrências de erro do sistema de banco de dados que está sendo avaliado.

II . Instruções

Preencha os campos da Tabela II.2 baseado nas descrições feitas a seguir.

- *Versão*: Identifica a versão do produto em que o erro ocorreu.
- *Tipo de Problema Associado*: Refere-se ao tipo de problema ao qual o erro ocorrido pertence.
- *Descrição*: Campo que registra uma descrição detalhada do problema ocorrido.
- *Gravidade do Erro*: Refere-se ao valor que expressa a gravidade do erro ocorrido. A importância desse erro pode variar de 0 (zero) a 4 (quatro), conforme escala de valores mostrada na Tabela II.1.
- *Funcionário*: Indica o nome do funcionário que identificou o erro.
- *Data Cadastro*: Refere-se a data em que o erro ocorreu.

Escala	Equivalência de Importância
0	Nenhuma
1	Baixa
2	Moderada
3	Alta
4	Total

Tabela II.1 - Escala de Valores

Versão	
Tipo de Problema Associado	
Descrição	
Gravidade do Erro	
Funcionário	
Data Cadastro	

Tabela II.2 – Cadastro do Erro ou Não Conformidade



Universidade Federal do Rio de Janeiro

COPPE - Programa de Engenharia de Sistemas e Computação

Apêndice III

INSTRUMENTO PARA CLASSIFICAÇÃO DAS CARACTERÍSTICAS NECESSÁRIAS PARA AVALIAR A QUALIDADE DE DADOS

I . Objetivo

Coletar junto aos avaliadores o grau de importância de cada característica na avaliação da Qualidade de Dados.

II . Instruções

Atribua valores de 0 a 4, segundo a escala apresentada na Tabela III.1, às características (descritas no Capítulo III) que são necessárias para avaliar a qualidade de dados, de acordo com o grau de importância na sua opinião e/ou experiência.

Grau de Importância	Explicação
0	Indica que a característica que está sendo apresentada não tem nenhuma importância.
1	Indica que a característica que está sendo apresentada tem pouca importância.
2	Indica que a característica que está sendo apresentada tem importância em algumas circunstâncias, mas nem sempre.
3	Indica que a característica que está sendo apresentada é muito importante.
4	Indica de maneira absoluta que não há dúvidas que a característica que está sendo apresentada é imprescindível.

Tabela III.1 - Escala de Valores

III . Identificação do Avaliador

Nome do Avaliador:

Posto/Cargo:

Departamento:

IV. Classificação das Características Necessárias para Avaliar a Qualidade dos Dados

OBJETIVO: UTILIZABILIDADE

Característica	Importância				
1 . Disponibilidade da Informação	0	1	2	3	4
2 . Idade dos Dados	0	1	2	3	4
3 . Oportunidade	0	1	2	3	4
4 . Eficiência de Execução	0	1	2	3	4
5 . Relevância	0	1	2	3	4
6 . Utilidade	0	1	2	3	4
7 . Lucratividade	0	1	2	3	4
8 . Benefício no Trabalho do Usuário	0	1	2	3	4
9 . Competitividade	0	1	2	3	4

OBJETIVO: CONFIABILIDADE CONCEITUAL

Característica	Importância				
1 . Quantidade de Dados Adequada	0	1	2	3	4
2 . Acurácia	0	1	2	3	4
3 . Completitude	0	1	2	3	4
4 . Cobertura	0	1	2	3	4
5 . Robustez	0	1	2	3	4
6 . Precisão dos Dados	0	1	2	3	4
7 . Consistência	0	1	2	3	4
8 . Sinalização	0	1	2	3	4
9 . Atributabilidade	0	1	2	3	4
10. Recuperabilidade	0	1	2	3	4
11. Flexibilidade	0	1	2	3	4
12. Interoperabilidade	0	1	2	3	4
13. Segurança	0	1	2	3	4
14. Facilidade de Compreensão da Informação	0	1	2	3	4
15. Adequabilidade da Informação	0	1	2	3	4

OBJETIVO: CONFIABILIDADE DA REPRESENTAÇÃO

Característica	Importância				
1 . Uniformidade	0	1	2	3	4
2 . Disponibilidade de Documentação	0	1	2	3	4
3 . Rastreabilidade	0	1	2	3	4



Universidade Federal do Rio de Janeiro

COPPE - Programa de Engenharia de Sistemas e Computação

Apêndice IV

MANUAL PARA UTILIZAÇÃO DA FERRAMENTA AQUA

IV.1 – Objetivo

Este manual tem como propósito orientar os usuários da ferramenta AQUA através da descrição das suas funcionalidades.

IV.2 – Definição do Sistema

A ferramenta AQUA realiza uma avaliação da qualidade de dados pela não conformidade. Para isso, os erros ou não conformidades que ocorreram durante a utilização do produto são considerados, a fim de verificar como cada erro degradou a qualidade dos dados armazenados.

A ferramenta incorpora vários módulos de visam:

- Cadastrar os erros que ocorreram durante a utilização do banco de dados;
- Cadastrar novas características de qualidade de dados e os seus respectivos tipos de erros;
- Facilitar a tarefa do avaliador em atribuir pesos às características de qualidade; e
- Realizar a avaliação da qualidade dos dados armazenados.

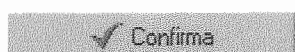
IV.3 – Execução da ferramenta

Inicie o AQUA pelo ícone localizado na área de trabalho.

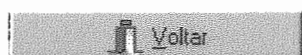
A ferramenta é inicializada aparecendo a tela principal que pode ser vista na Figura IV.1.

IV.4 – Botões de Uso Geral no Sistema

Os botões apresentados abaixo são utilizados em várias telas da ferramenta com o mesmo significado. Para evitar as repetições desses significados, eles serão vistos a seguir.



Executa a ação apresentada na caixa de diálogo.



Retorna à tela anterior.

O conjunto de botões a seguir tem o objetivo de manipular os dados de uma tabela.



O significado de cada botão deste conjunto está descrito a seguir.



Retorna para o primeiro elemento da tabela;



Retorna para o elemento anterior;



Vai para o próximo elemento;



Vai para o último elemento da tabela;



Adiciona um elemento na tabela;



Remove um elemento da tabela;



Permite a edição do elemento, isto é, que alterações possam ser feitas;



Confirma a adição ou alteração do elemento;



Cancela a adição ou alteração do elemento;



Atualiza a tabela.

IV.5 – Tela Principal da Ferramenta

A tela principal da ferramenta é apresentada na Figura IV.1.



Figura IV.1 – Tela Principal da Ferramenta AQUA

Esta tela é composta por:

- ◆ Barra de Menu
- ◆ Símbolo que representa da ferramenta

IV.6 – Barra de Menu

Na barra de menu principal da ferramenta, apresentada na Figura IV.2, estão disponíveis as funções descritas a seguir.

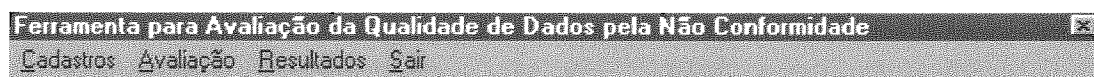


Figura IV.2 – Barra de Menu

- ◆ **Cadastros** - Permite cadastrar um erro ocorrido com o sistema, assim como, cadastrar uma nova característica de qualidade com o seu respectivo tipo de erro.
- ◆ **Avaliação** - Permite iniciar o processo de avaliação da qualidade dos dados. Para isso, disponibiliza a função de atribuição de pesos às características de qualidade e, posteriormente, a função que ativa a realização da avaliação.
- ◆ **Resultados** - Possibilita a exibição dos resultados aferidos na avaliação.
- ◆ **Sair** - Encerra a execução da ferramenta AQUA.

IV.6.1 - Cadastros

Ao selecionar “Cadastros” na barra de menu, a tela apresentada é mostrada na Figura IV.3.

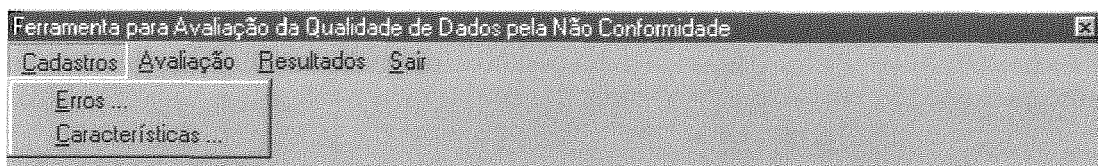


Figura IV.3 – Menu de Cadastros

A opção “Cadastros” disponibiliza outras opções descritas a seguir.

Erros

Opção que permite o cadastro de um erro ocorrido. Após selecioná-la a tela apresentada é mostrada na Figura IV.4.

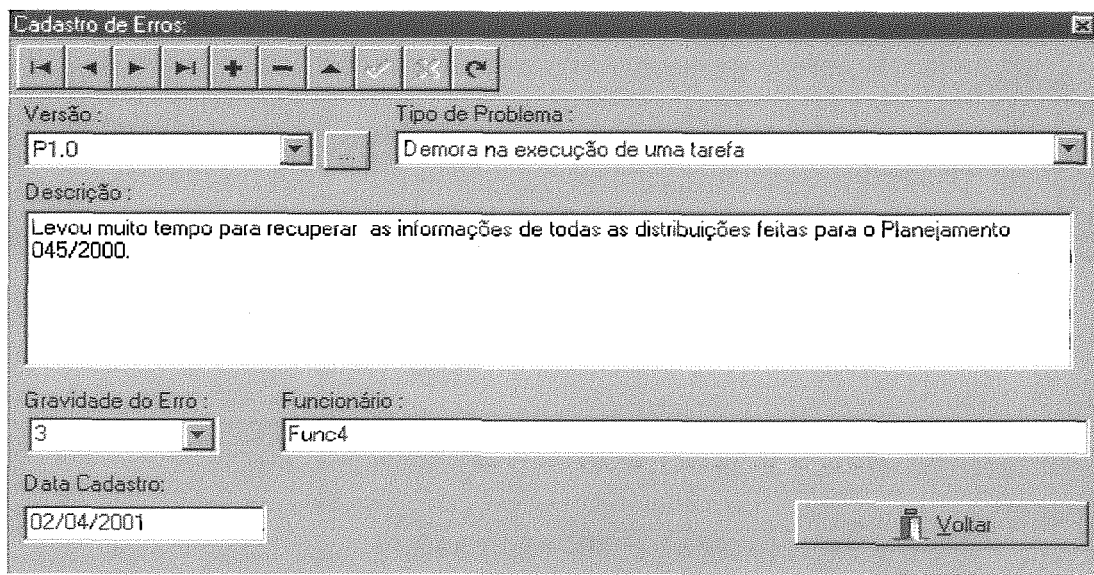



Figura IV.4 – Cadastro de Erros

O preenchimento dos campos descritos a seguir é necessário para realizar o cadastro do erro.

➤ **Versão** - Identifica a versão do produto em que o erro ocorreu. O botão  , ao lado do campo versão, abre uma tela que permite o cadastramento de uma nova versão, conforme exibido na Figura IV.5.

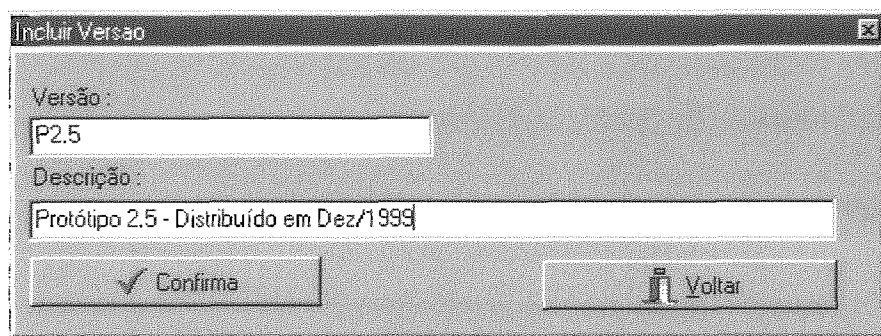


Figura IV.5 – Cadastro de Versão

➤ **Tipo de Problema** - Refere-se ao tipo de problema ao qual o erro ocorrido pertence. As opções desse campo são os elementos da tabela de tipos de problemas (tipoprob.db). Esses tipos são cadastrados no momento da inclusão da característica de qualidade que o tipo de problema está associado.

➤ **Descrição** - Campo que registra uma descrição detalhada do problema ocorrido.

➤ **Gravidade do Erro** - Refere-se ao valor que expressa a gravidade do erro ocorrido. A importância desse erro pode variar de 0 (zero) a 4 (quatro), conforme escala de valores mostrada na Tabela IV.1.

➤ **Funcionário** – Indica o nome do funcionário que identificou o erro.

➤ **Data Cadastro** – Refere-se a data em que o erro ocorreu.

Escala	Equivalência de Importância
0	Nenhuma
1	Baixa
2	Moderada
3	Alta
4	Total

Tabela IV.1 - Escala de Valores

Características

Opção que permite o cadastro de uma característica de qualidade. Após selecioná-la é apresentada a tela mostrada na Figura IV.6.


A janela 'Incluir Característica de Qualidade' possui os seguintes elementos:

- Campos de entrada: 'Característica:' com o texto 'Idade dos Dados' e 'Peso:' com o valor '3' selecionado em uma lista suspensa.
- Área de texto: 'Descrição:' contendo o texto: 'Refere-se ao dado estar atualizado, expressando até que ponto a idade dos dados (tempo que está armazenado) é apropriada para a realização das tarefas da organização.'
- Botões: '✓ Continua', 'Incluir Tipo de Problema Associado' e 'Voltar'.

Figura IV.6 – Incluir Característica de Qualidade

O preenchimento dos campos descritos a seguir é necessário para realizar o cadastro da característica.

- Característica – Refere-se ao nome da característica de qualidade.
- Peso – Indica o grau de importância da presença da característica na avaliação. Esse peso pode variar de 0 (zero) a 4 (quatro), conforme escala de valores mostrada na Tabela IV.1.
- Descrição - Campo que registra uma descrição detalhada da característica de qualidade.

Após a confirmação do cadastro, é necessário que sejam cadastrados os tipos de erros associados à característica. Para isso, o botão  abre a tela mostrada na Figura IV.7.

A janela 'Tipo de Problema Associado à Nova Característica' possui os seguintes elementos:

- Campos de entrada: 'Tipo de Problema:' com o texto 'Dado armazenado sem atualização por muito tempo' e 'Peso:' com o valor '3' selecionado em uma lista suspensa.
- Botão: 'Voltar'.

Figura IV.7 – Incluir Tipo de Problema

O preenchimento dos campos descritos a seguir é necessário para realizar o cadastro do tipo de problema.

➤ Tipo de Problema – Indica o nome do tipo de problema.

➤ Peso - Indica o grau de importância do tipo de problema. Esse peso pode variar de 0 (zero) a 4 (quatro), conforme escala de valores mostrada na Tabela IV.1.

IV.6.2 - Avaliação

Ao selecionar “Avaliação” na barra de menu, a tela mostrada na Figura IV.8 é apresentada.

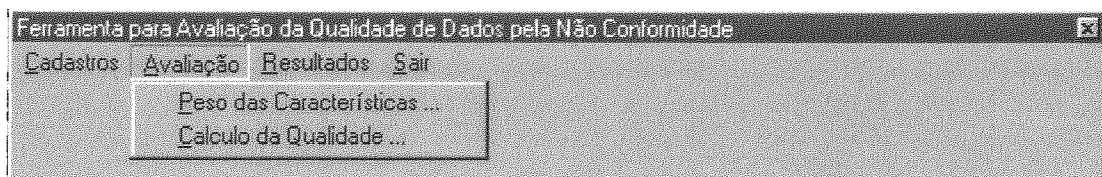


Figura IV.8 – Menu Avaliação

A opção “Avaliação” disponibiliza outras opções descritas a seguir.

Peso das Características

Opção que permite atribuir peso às características de qualidade. Após selecioná-la é apresentada a tela mostrada na Figura IV.9.

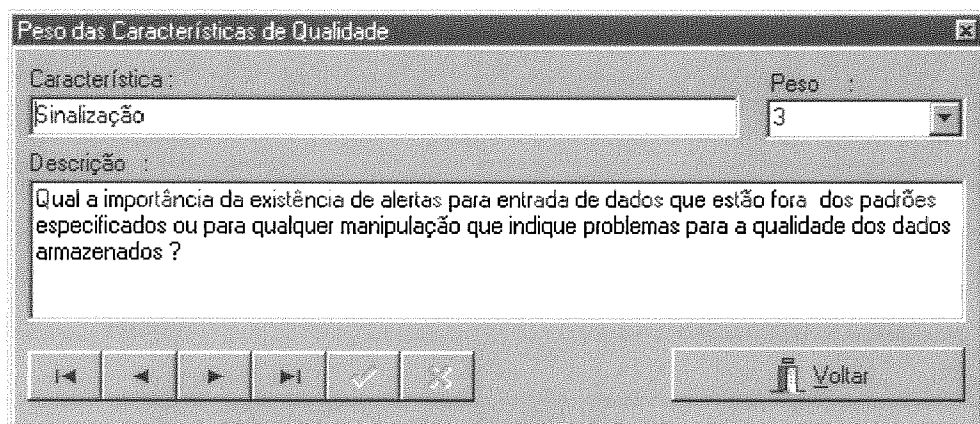


Figura IV.9 – Peso das Características

Os campos descritos a seguir são exibidos (somente para leitura), a fim de auxiliar na atribuição do peso da característica.

- Característica – Refere-se ao nome da característica de qualidade.
- Descrição - Campo que descreve detalhadamente a característica de qualidade.

O preenchimento do campo descrito a seguir é necessário para realizar a atribuição do peso da característica.

- Peso – Indica o grau de importância da presença da característica na avaliação. Esse peso pode variar de 0 (zero) a 4 (quatro), conforme escala de valores mostrada na Tabela IV.1.

Cálculo da Qualidade

Opção que permite iniciar o processo de avaliação da qualidade dos dados. Para isso, é dado início a uma seqüência de telas que visam orientar o usuário a inserir as informações necessárias à execução da avaliação. A primeira tela exibida pode ser vista na Figura IV.10.

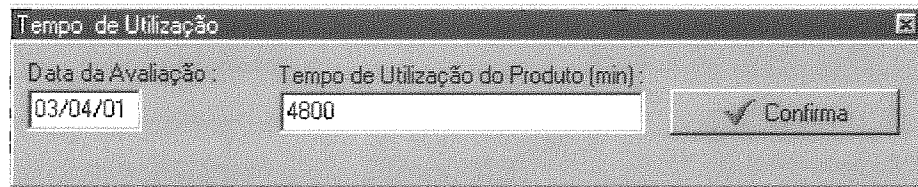


Figura IV.10 – Tempo de Utilização

O preenchimento dos campos descrito a seguir é necessário para prosseguir na avaliação da qualidade.

- Data da Avaliação – Indica a data em que a avaliação da qualidade está sendo realizada.
- Tempo de Utilização do Produto – Refere-se ao tempo acumulado de uso do produto que está sendo avaliado. Esse valor é expresso em minutos.

Dando seqüência à execução da avaliação, a segunda tela apresentada está exibida na Figura IV.11.

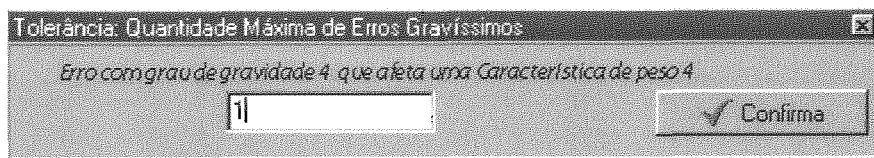


Figura IV.11 – Tolerância

O preenchimento do campo descrito a seguir é necessário para prosseguir na avaliação da qualidade.

➤ Tolerância – Quantidade máxima permitida de ocorrência de erros gravíssimos. Erros gravíssimos são aqueles que tem o grau de gravidade igual a quatro e afetam a uma característica de qualidade com peso igual a 4.

Caso a tolerância permitida seja alcançada, isto é, o cadastro dos erros tenha um número maior, de erros gravíssimos, do que o tolerável, a avaliação não é realizada. Deste modo, o índice de qualidade é igual a 0 (zero) e a ferramenta disponibiliza um relatório contendo todos os erros que degradaram a qualidade dos dados armazenados.

IV.6.3 - Resultados

Ao selecionar “Resultados” na barra de menu, a tela mostrada na Figura IV.12 é apresentada.

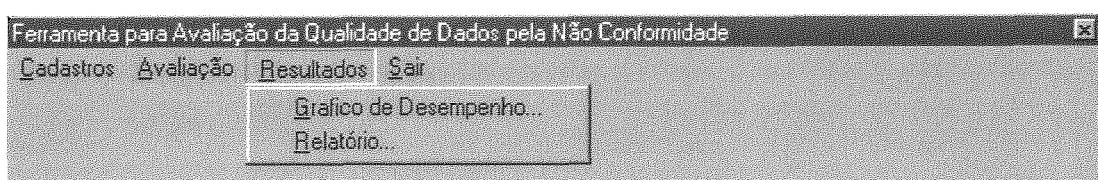


Figura IV.12 – Menu Resultados

A opção “Resultados” disponibiliza outras opções descritas a seguir.

Gráfico de Desempenho

Opção que exibe graficamente os resultados de todas as avaliações da qualidade realizadas. Após selecioná-la é apresentada a tela exibida na Figura IV.13.

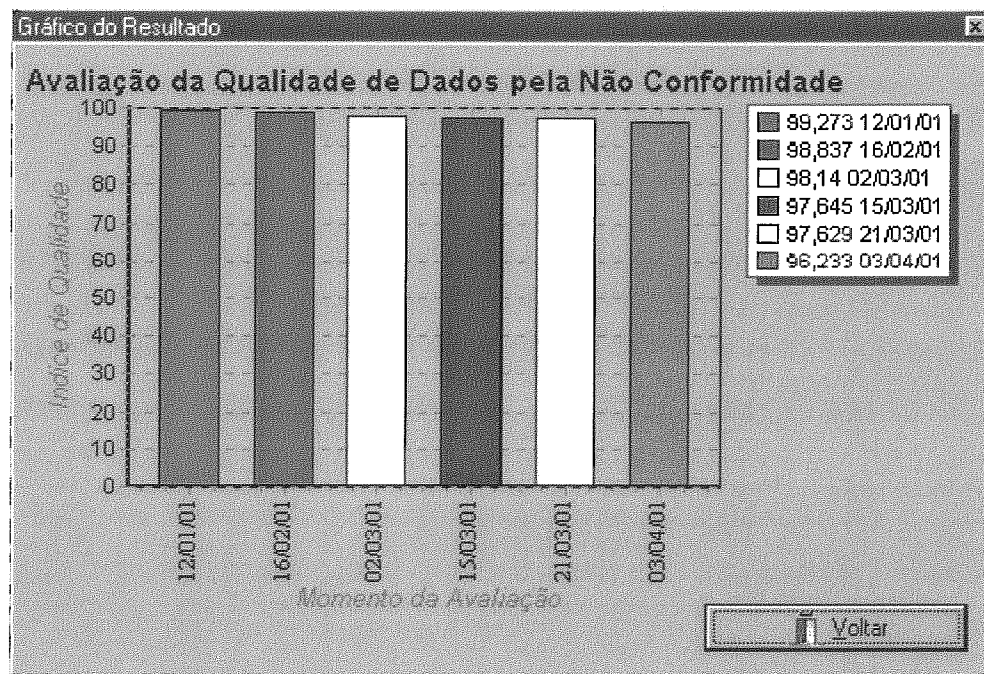


Figura IV.13 – Gráfico de Desempenho

Neste gráfico pode ser observado o resultado da avaliação mais recente, assim como, o resultado de todas as outras avaliações realizadas. Para isso, os seguintes dados são exibidos:

- Momento da Avaliação – Refere-se ao dia em que a avaliação da qualidade foi realizada.
- Índice de Qualidade – Refere-se ao índice (percentual) de qualidade que foi calculado em cada avaliação.

Relatório

Opção que exibe o relatório contendo o resultado da avaliação realizada. Após selecioná-la é apresentada a tela que pode ser vista na Figura IV.14.

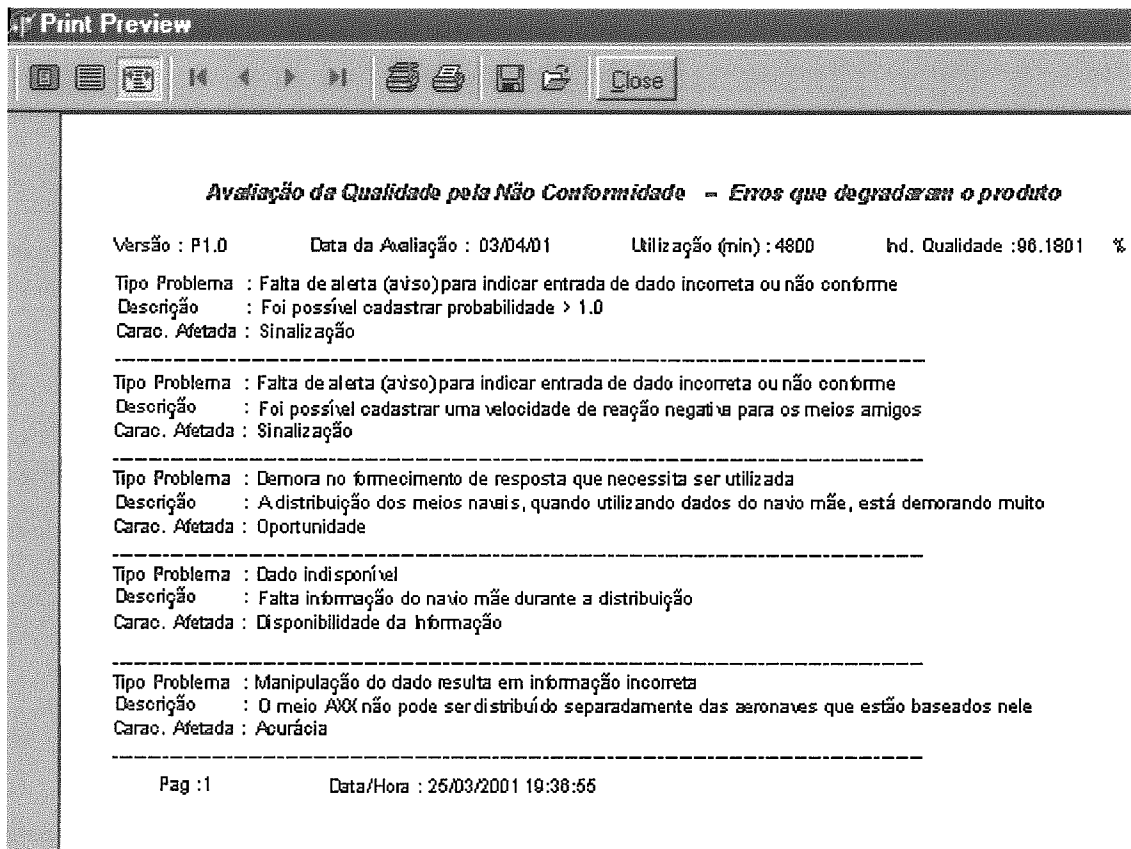


Figura IV.14 – Relatório

Neste relatório pode ser observado o resultado da avaliação mais recente. Para isso, os seguintes dados são exibidos:

- Versão - Identifica a versão do produto que foi avaliado.
- Utilização - Refere-se ao tempo acumulado de uso do produto que foi avaliado. Esse valor é expresso em minutos.
- Data da Avaliação - Indica a data em que a avaliação da qualidade foi realizada.
- Ind. Qualidade - Refere-se ao índice (percentual) de qualidade que foi calculado na avaliação.

- Erros que Degradaram o Produto – Relaciona todos os erros cadastrados que ocorreram durante o tempo de utilização do produto. Para cada erro relacionado tem-se:
- ❖ Tipo Problema - Refere-se ao tipo de problema ao qual o erro ocorrido pertence.
 - ❖ Descrição – Descrição detalhada do erro.
 - ❖ Carac. Afetada – Nome da característica de qualidade que foi afetada pelo erro.

IV.6.4 – Sair

Ao seleccionar “Sair” na barra de menu, a ferramenta AQUA é finalizada.



Universidade Federal do Rio de Janeiro

COPPE - Programa de Engenharia de Sistemas e Computação

Apêndice V
DICIONÁRIO DE DADOS

I. Objetivo

Descrever os atributos do modelo físico oriundo do modelo lógico apresentado no Capítulo 5 .

II. Descrição dos Dados

Tabela: Característica (carac.db)

Campo : Id_Carac (*pk*)

Descrição : Identificador da característica de qualidade

Tamanho : Auto Incremento

Campo : Nome_Carac

Descrição : Nome da característica de qualidade

Tamanho : String (50)

Campo : Peso_Carac

Descrição : Peso atribuído pelo avaliador para a característica de qualidade

Tamanho : Numérico

Campo : Descr_Carac

Descrição : Descrição do significado da característica de qualidade

Tamanho : String (240)

Tabela : Tipos de Problema (tipoprob.db)

Campo : Id_TipoProb (*pk*)

Descrição : Identificador do tipo de problema

Tamanho : Auto Incremento

Campo : Id_Carac (*fk*)

Descrição : Identificador da característica de qualidade que o tipo de problema está associado

Tamanho : Numérico

Campo : Nome_TipoProb

Descrição : Nome do tipo de problema

Tamanho : String (255)

Campo : Peso_TipoProb

Descrição : Peso atribuído pelo avaliador para o tipo de problema

Tamanho : Numérico

Tabela : Erros (Caderros.db)

Campo : Id_Erro (*pk*)

Descrição : Identificador do erro que ocorreu durante a utilização do banco de dados avaliado

Tamanho : Auto Incremento

Campo : Nome_Func

Descrição : Nome do funcionário da organização que identificou e cadastrou o erro

Tamanho : String (50)

Campo : Data_Cadastro

Descrição : Dia, mês e ano da ocorrência do erro

Tamanho : Data

Campo : Id_Versao (*fk*)

Descrição : Identificador da versão do produto que ocorreu o erro

Tamanho : Numérico

Campo : Id_TipoProb (*fk*)

Descrição : Identificador do tipo de problema ao qual o erro ocorrido está associado

Tamanho : Numérico

Campo : Descr_Erro

Descrição : Descrição do erro ocorrido

Tamanho : String (240)

Campo : Gravidade

Descrição : Peso atribuído pelo usuário que expressa a gravidade do erro ocorrido

Tamanho : Numérico

Tabela : Versão (versao.db)

Campo : Id_Versao

Descrição : Identificador da versão do produto que está sendo avaliado

Tamanho : Auto Incremento

Campo : Versao

Descrição : Nome da versão do produto

Tamanho : String (50)

Campo : Descr_Versao

Descrição : Descrição da versão do produto

Tamanho : String (255)