

Flávio Humberto Cabral Nunes  
Orientador: Arnaldo de Albuquerque Araújo  
Co-orientador: Luiz Antônio Cruz Souza

# **Desenvolvimento de Aplicações Hipermídia para Gerenciamento de Documentos Multimídia e Preservação de Acervos Digitais: Um Estudo de Caso no Cecor**

Dissertação apresentada ao Curso de Pós-graduação em Ciência da Computação da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

Belo Horizonte  
17 de Novembro de 2005

## Resumo

O advento da digitalização transformou o significado de captura, controle, entrega e uso dos recursos de informação na sociedade. Indivíduos, instituições e comunidades são capazes de publicar e disseminar informação sobre seus patrimônios culturais e criar suas próprias bibliotecas digitais, que aumentam o potencial para diversidade, pluralidade de vozes e fortalecimento dessas comunidades. Sistemas hipermídia estão emergindo como uma nova classe de sistemas de informação complexos. Esses sistemas permitem as pessoas criar, anotar, ligar e compartilhar informações de uma grande variedade de mídias, como texto, gráficos, áudio, vídeo, animação e programas. Entretanto, muitas pessoas tendem a pensar que, diferentemente da informação analógica, a informação digital existirá para sempre e não percebem a fragilidade dos documentos digitais. Este trabalho apresenta uma abordagem metodológica para o desenvolvimento de sistemas de informação hipermídia para gerenciamento de acervos multimídia. O trabalho também foca no problema da preservação digital. Um estudo é feito para determinar um fluxo de trabalho (workflow) de digitalização que permita gerar documentos digitais de alta qualidade, tentando minimizar os custos de redigitalização.

## Abstract

The advent of digitization has transformed the meaning of acquirement, control, deliver and use of information resources in society. Individuals, institutions and communities are able to publish and disseminate information about their cultural heritage and create their own digital libraries, which increases the potential for diversity, plurality of voices and empowerment of these communities. These systems allow people to create, note, link and share information of several media, like text, graphics, audio, video, animation and programs. However, most people tend to think that (unlike analog information) digital information will last forever, yet fail to realize the fragility of digital works. This work presents a methodological approach to hypermedia information system development to manage multimedia collections. This work also focuses the digital preservation problem. A digitalization workflow is studied to determine parameters to generate high quality digital documents and minimize the redigitalization costs.

## Résumé

L'arrivée de la numérotisation a transformé la signification de la capture, contrôle, la distribution et utilisation des ressources de l'information dans la société. Les individus, les établissements et les communautés peuvent éditer et diffuser l'information au sujet de leur acquis culturel et créer leurs propres bibliothèques numériques, qui augmente le potentiel pour la diversité, la pluralité de voix et l'habilitation de ces communautés. Ces systèmes permettent à des personnes de créer, noter, lier et à l'information de part de plusieurs médias, comme le texte, de graphiques, d'acoustique, de vidéo, d'animation et de programmes. Cependant, la plupart des personnes tendent à penser que (à la différence de l'information analogue) l'information numérique durera pour toujours, pourtant ne réalisent pas la fragilité des travaux numériques. Ce travail présente une approche méthodologique au développement de système d'information de hypermedia pour contrôler des collections de multimédia. Ce travail focalise également le problème numérique de conservation. Un déroulement des opérations de numérisation est étudié pour déterminer des paramètres pour produire des documents numériques de haute qualité et pour réduire au minimum les coûts de redigitalization.

Aos meus pais  
*Maria Gasparina e Maurício*  
que são para mim  
exemplo de vida ...

## Agradecimentos

Agradeço ao professor Arnaldo de Albuquerque Araújo pela ajuda sem a qual seria impossível a conclusão deste trabalho.

Ao professor Luiz Antônio Cruz Souza pelas oportunidades e apoio para o desenvolvimento do trabalho.

Ao André Veloso Junqueira e ao Fábio por terem me auxiliado na implementação.

Ao pessoal do NPDI pela convivência, em especial: Camillo, Carlos Eduardo, Cecília, Clarissa, Daniel Wu, David Menoti, David e Gláucia.

Ao pessoal do LACICOR e do Cedor, em especial: Alexandre Leão, Alessandra Rosatto, Marco Aurélio, Mário Anacleto, Márcia Figueiredo, Moema e Selma Otília.

Aos amigos Ademir, Adriano Guedes, Breno, Cláudio, Leonardo, Mário Sérgio e Sérgio Villaça.

À minha família que sempre esteve ao meu lado e que jamais esquecerei, em especial: meu pai, Maria Gasparina e Maurício, meus irmãos, Fabiano e Frederico, e à minha tia Márcia.

E por fim, mas não menos importante, a Deus pela vida e pelas pessoas maravilhosas que colocastes em minha vida.

À todos vocês, o meu muito obrigado.

# Conteúdo

Resumo	1
Abstract	1
Résumé	1
Agradecimentos	1
Lista de Figuras	1
Lista de Tabelas	1
Bibliografia	1
<b>1 Introdução</b>	<b>1</b>
1.1 Tecnologia Digital e Gerenciamento de Acervos . . . . .	1
1.2 Objetivos . . . . .	3
1.3 Estrutura da Dissertação . . . . .	4
<b>2 Digitalização de Acervos</b>	<b>6</b>
2.1 Terminologia Básica . . . . .	6
2.1.1 Imagem Digital . . . . .	6
2.1.2 Resolução . . . . .	7
2.1.3 Profundidade de <i>Bit</i> . . . . .	8
2.1.4 Compressão . . . . .	9
2.1.5 Formatos de Arquivo . . . . .	10
2.1.6 Metadados . . . . .	11
2.2 Preservação Digital . . . . .	12
2.2.1 Introdução . . . . .	12
2.2.2 Autenticidade e Segurança de Coleções Digitais . . . . .	15
2.2.3 Estratégias de Preservação Digital . . . . .	15
2.3 Cadeia de Digitalização . . . . .	18
2.3.1 Captura da Imagem . . . . .	19
2.3.2 Armazenamento . . . . .	21

---

2.3.3	Visualização . . . . .	24
2.4	Projetos de Digitalização de Acervos . . . . .	25
2.4.1	Arquivo Público Mineiro (APM) . . . . .	26
2.4.2	Projeto Portinari . . . . .	26
2.4.3	Projeto NARCISSE . . . . .	27
2.5	Considerações Finais . . . . .	28
<b>3</b>	<b>Gerenciamento de Informação Multimídia</b>	<b>29</b>
3.1	Multimídia . . . . .	30
3.2	Sistemas de Informação Multimídia . . . . .	31
3.3	Sistemas Gerenciadores de Banco de Dados . . . . .	33
3.4	Multimídia e Sistemas de Bancos de Dados . . . . .	34
3.5	Indexação e Consulta . . . . .	34
3.5.1	Consulta Baseada em Texto . . . . .	34
3.5.2	Consulta Baseada no Conteúdo . . . . .	35
3.6	Considerações Finais . . . . .	37
<b>4</b>	<b>Projeto e Desenvolvimento de Aplicações Hipermídia</b>	<b>38</b>
4.1	Desenvolvimento de <i>Software</i> . . . . .	39
4.2	Desenvolvimento de Aplicações Hipermídia . . . . .	40
4.3	Desorientação e Sobrecarga Cognitiva . . . . .	43
4.4	Métodos de Desenvolvimento de Aplicações Hipermídia . . . . .	44
4.4.1	Processo para Aplicativos Extensíveis Interativos (PRAXIS) . . . . .	44
4.4.2	<i>Hypertext Design Methodology</i> (HDM) . . . . .	45
4.4.3	<i>Object-Oriented Hypertext Design Method</i> (OOHDM) . . . . .	47
4.5	Considerações Finais . . . . .	49
<b>5</b>	<b>Metodologia de Desenvolvimento</b>	<b>50</b>
5.1	Cecor . . . . .	50
5.1.1	O Acervo . . . . .	50
5.1.2	Processo de Documentação . . . . .	53
5.2	Processo de Digitalização do Acervo . . . . .	53
5.2.1	Parâmetros de Digitalização . . . . .	53
5.2.2	Infraestrutura de <i>Hardware</i> . . . . .	56
5.2.3	<i>Workflow</i> de Digitalização . . . . .	56
5.3	Desenvolvimento do Sistema de Informação Hipermídia . . . . .	61
5.3.1	Levantamento de Requisitos . . . . .	62
5.3.2	Desenho Conceitual . . . . .	64
5.3.3	Desenho Navegacional . . . . .	67
5.3.4	Desenho de Interfaces de Usuário . . . . .	69
5.3.5	Implementação e Testes . . . . .	70
5.4	Considerações Finais . . . . .	76



**6 Conclusão**

**77**

# Lista de Figuras

2.1	Imagem digital, onde o bit 0 indica ausência de luz (preto) e o bit 1 indica presença de luz (branco) (Fonte: [Cornell, 2000]). . . . .	7
2.2	Imagem apresentada em várias resoluções: 600dpi, 300dpi, 150dpi 60dpi e 30dpi (Fonte: [Besser and Trant, 1995]). . . . .	8
2.3	Exemplos de imagens com os três tipos de profundidade de bit: bitonal, tons de cinza e colorida (Fonte: [Bishoff, 2003]). . . . .	9
2.4	Cadeia de digitalização. . . . .	19
2.5	Tipos de <i>scanners</i> . . . . .	21
2.6	Tecnologias de monitores: CRT e LCD. . . . .	24
3.1	Atividades dos sistemas de informação: entrada, processamento e saída (Fonte: [Laudon and Laudon, 1999]). . . . .	32
4.1	Hierarquia de projeto do HDM. . . . .	46
5.1	Vista da Escola de Belas Artes onde se encontra o Cecor. . . . .	51
5.2	Amostra do acervo do Cecor: (a) pintura e duas análises feitas sobre ela (raio-x e raio ultravioleta); (b) tipos de documentos (fichários, <i>slides</i> e cadernos de análises). . . . .	52
5.3	Amostra do acervo do Cecor: fotografias de um quadro de Di Cavalcanti, uma amostra extraída de uma obra e uma escultura. . . . .	52
5.4	Processo de documentação. . . . .	53
5.5	Exemplo de documento texto capturado em modo bitonal e em escala de cinzas. Uma mancha no documento ocultou detalhes que somente poderiam ser capturados em escala de cinza e não no modo bitonal (fonte: [Cornell, 2000]).	55
5.6	<i>Workflow</i> de digitalização. . . . .	59
5.7	Diagrama de casos de uso do sistema do Cecor. . . . .	66
5.8	Exemplo de perspectiva: atributo <i>descricao</i> da classe <i>Amostra</i> pode ser representado como texto ou com uma imagem. . . . .	67
5.9	Diagrama de classes. . . . .	68
5.10	Diagrama navegacional para gerenciar os dados do laboratório. . . . .	69
5.11	Alguns componentes de interface utilizados nos modelos de apresentação. .	70
5.12	Modelo de apresentação do formulário de amostras. . . . .	71

---

5.13	Arquitetura de três camadas do sistema Restaure. . . . .	72
5.14	Tela inicial. . . . .	72
5.15	Cadastro de documentos digitais e metadados técnicos. . . . .	73
5.16	Lista de documentos digitais pertencentes a uma obra, no caso a obra de registro 79-04Mb. . . . .	74
5.17	Formulário de entrada de obra. . . . .	75
5.18	Formulário de amostras. . . . .	76

# Lista de Tabelas

2.1	Categorias de metadados e suas funções [Swetland, 2000]. . . . .	13
5.1	Características da câmera digital Nikon D100. . . . .	57
5.2	Características do <i>scanner</i> ScanMaker 9800XL. . . . .	57
5.3	Algumas normas que reprovam uma matriz digital. . . . .	60
5.4	Tipos de relacionamentos entre casos de uso. . . . .	64
5.5	Lista de requisitos candidatos. . . . .	65
5.6	Descrição dos usuários. . . . .	66

# Capítulo 1

## Introdução

### 1.1 Tecnologia Digital e Gerenciamento de Acervos

O homem, desde os seus primórdios, sente uma forte necessidade de registrar, preservar e transmitir todas as formas de conhecimento adquirido. A preservação do conhecimento ao longo da história da humanidade, sempre foi, e continuará sendo, uma atividade fundamental para a evolução e preservação da humanidade [Silva, 1995].

O desenvolvimento cultural e tecnológico do homem moderno faz com que este seja exposto diariamente a um gigantesco volume de informações, sob as mais diversas formas e naturezas, como sons, imagens, textos e vídeos. Estes tipos de dados são extremamente ricos, capazes de expressar informações quantitativas ou qualitativas de forma amigável e com imediata percepção pelo usuário [David, 1997].

Entretanto, as instituições mantenedoras de arquivos públicos ou outros acervos documentais enfrentam diversos problemas, decorrentes geralmente do grande acúmulo de documentos e de sua fragilidade, destacando-se os riscos decorrentes da degradação dos originais devido a sua manipulação direta e frequente e da dificuldade de acesso às informações por parte dos pesquisadores e do público em geral [APM, 1995]. Muitas vezes, o trabalho de catalogação e consulta é feito manualmente, limitando o universo de consultas aos acervos e sendo também um trabalho bastante cansativo e demorado.

A tecnologia digital surge como uma possibilidade de romper esse compromisso, permitindo dar amplo acesso às cópias digitais de alta qualidade dos documentos originais, ao mesmo tempo resguardando-os da manipulação desnecessária. A tecnologia de imagens digitais permite a geração de cópias idênticas ao original. Enquanto os dados analógicos estão sujeitos às imperfeições do mundo físico, que impedem a fidelidade da replicação, cada cópia digital é um clone, indistinguível do original. Técnicas de processamento digital de imagens (PDI), como controle de brilho, contraste e realce de bordas, permitem melhorar a qualidade dos documentos apresentados aos usuários ou realçar alguns aspectos dos mesmos sem alterar o documento digitalizado original [Gonzalez and Woods, 1992].

Mais ainda, sistemas computadorizados têm um potencial enorme de facilitar a própria tarefa de organização e descrição dos acervos. Através de ferramentas de gestão documental e fluxo de trabalho, é possível multiplicar a capacidade de trabalho de historiadores, arquivistas e técnicos de conservação. Sistemas bem engendrados permitem a esses profissionais automatizar e gerir de forma racional suas complexas atividades.

Do lado dos consulentes, o uso da tecnologia digital permite operações de busca muito mais rápidas e sofisticadas do que é possível com os instrumentos de pesquisa em papel, como inventários e índices. Sistemas gerenciadores de bancos de dados (SGBD) permitem indexar, armazenar, recuperar e gerenciar grandes volumes de dados e fornecem diversas funções importantes ao compartilhar informações concorrentemente, incluindo consultas, suporte de transações, recuperação, segurança e estruturas de armazenamento para acesso de alto desempenho, controle de versão e restrições de integridade de dados. Essas características são muito importantes para gerenciar grandes volumes de informação multimídia eficientemente em ambientes interconectados [Khoshafian and Baker, 1996].

As redes de computadores, em particular a *Internet*, adicionam a possibilidade da consulta remota, expandindo dramaticamente o universo de pesquisadores com acesso ao acervo [Tanenbaum, 1996]. A *Internet* trouxe a possibilidade de acesso e manipulação de informações armazenadas em qualquer lugar do mundo utilizando computadores com poucos recursos de *hardware* e *software* [Paula Filho, 2000]. Com esse meio de comunicação, é possível levar os documentos ao um conjunto bem mais amplo de usuários.

Entretanto, muitas pessoas tendem a pensar que, diferentemente da informação analógica, a informação digital existirá para sempre e não percebem a fragilidade dos documentos digitais. Grandes quantidades de informação digital podem ser perdidas pela deterioração dos discos magnéticos em que elas estão armazenadas. Mas o problema de deterioração das mídias de armazenamento tem impacto menor do que os problemas das rápidas mudanças nos dispositivos de armazenamento e formatos de arquivo [Besser, 2000]. Não adianta preservar os suportes, se as tecnologias de visualização dos dados não forem preservadas, pois dados intactos inacessíveis são dados perdidos.

A informação digital possui facilidades de modificação que comprometem a segurança de sua integridade. Mesmo dados armazenados em mídias não alteráveis, como CD-ROM's e CD-R's, são passíveis de sofrerem adulterações [Byers, 2003]. Enquanto no meio analógico, até a mais perfeita falsificação pode ser detectada, no meio digital isso não é factível. As operações de tratamento de imagens podem modificá-las sem deixar nenhum vestígio de adulteração. Estabelecer a autenticidade é de particular relevância para ambientes digitais, devido à facilidade em que os dados podem ser alterados [Aschenbrenner, 2001].

Dessa forma, é de suma importância considerar esses aspectos ao se iniciar projetos de digitalização de acervos. A tecnologia digital se mostra atraente. Mas se não forem tomadas as devidas precauções, todo o trabalho pode ser perdido.

Este trabalho propõe uma metodologia para projetos de digitalização de acervos. Essa metodologia segue duas linhas. Na primeira, são considerados os aspectos de captura, armazenamento e acesso aos documentos. A segunda linha trata do desenvolvimento de aplicações hipermídia para gerenciar os documentos permitindo que eles possam ser acessados por usuários.

## 1.2 Objetivos

Os objetivos deste trabalho são:

- Realizar um estudo sobre o cenário atual de projetos de digitalização de acervos.

Nesse estudo, pretende-se identificar os parâmetros de qualidade adotados nestes projetos. De posse desses dados, pretende-se propor os parâmetros para a metodologia a ser proposta.

- Identificar as políticas de preservação de dados digitais utilizadas por arquivos e bibliotecas.
- Pesquisar o cenário de gerenciamento de dados multimídia e desenvolvimento de sistemas de informação multimídia. Neste tópico, procura-se definir multimídia e seus principais tipos, fazer o levantamento das características de um SGBD e seus principais requisitos para aplicações multimídia e identificar os principais tipos de consultas envolvendo dados multimídia. Também, serão pesquisados processos de desenvolvimento de *software* voltados para aplicações multimídia e hipermídia.
- Estudar e implementar um fluxo de trabalho de digitalização para os documentos do Cecor, enfatizando o aspecto da preservação digital.
- Estudar as principais metodologias de desenvolvimento de *software* para *Web*.
- Implementar uma aplicação hipermídia que gerencie os documentos digitais que serão criados com base nas metodologias de desenvolvimento estudadas.

### 1.3 Estrutura da Dissertação

Este capítulo apresentou um cenário do gerenciamento de documentos em bibliotecas e arquivos. A problemática desse gerenciamento foi mostrada e a tecnologia digital foi indicada para solucionar esses problemas. As vantagens e desvantagens da tecnologia digital foram descritas. Os objetivos do presente trabalho também foram apresentados. Os demais capítulos desta dissertação estão estruturados da seguinte forma:

- O Capítulo 2 introduz os principais conceitos sobre projetos de digitalização, os tipos de metadados coletados, a infraestrutura da cadeia de digitalização. O capítulo é



finalizado com uma introdução aos conceitos de preservação digital e suas principais políticas, discutindo a problemática de se assegurar a autenticidade dos documentos e apresentando três exemplos de projetos de digitalização.

- O Capítulo 3 contextualiza a multimídia, apresentando seus conceitos, aplicações, gerenciamento, indexação e tipos de consultas.
- O Capítulo 4 define o que é hipermídia e hipertexto e discute sobre alguns métodos de desenvolvimento de aplicações hipermídia.
- O Capítulo 5 apresenta a metodologia de desenvolvimento do sistema de informação para gerenciamento do acervo digital do Cecor. Neste capítulo, são descritos o processo de desenvolvimento, o processo de digitalização e a implantação da aplicação.
- No Capítulo 6, são apresentadas as conclusões do trabalho.

# Capítulo 2

## Digitalização de Acervos

Este capítulo trata de conceitos e métodos importantes em projetos de digitalização. A Seção 2.1 introduz alguns conceitos básicos sobre imagens digitais e digitalização. Na Seção 2.2, é discutida a problemática da preservação digital e seu impacto na integridade e autenticidade dos documentos, apresentando algumas estratégias para prolongar a vida útil dos documentos digitais. A Seção 2.3 descreve a infraestrutura básica para captura, armazenamento e apresentação das imagens. Por fim, na Seção 2.4, são mostrados três exemplos de projetos de digitalização de acervos.

### 2.1 Terminologia Básica

#### 2.1.1 Imagem Digital

Uma imagem é um figura, fotografia, visão ou outra forma que apresenta uma representação visual de um objeto ou cena [Niblack, 1986]. Entretanto, uma imagem digital é uma matriz onde os índices das linhas e colunas identificam um ponto na imagem e o valor do elemento da matriz identifica a informação de cor ou brilho naquele ponto. Os elementos de tal arranjo digital são chamados de elementos da imagem, *picture elements*, *pixels* ou *pels*, com os dois últimos sendo comumente usados como abreviação de “*picture elements*” [Gonzalez and Woods, 1992].

Na digitalização, a imagem sofre dois processos: amostragem e quantização. No processo de amostragem, a imagem contínua é convertida em um conjunto de coordenadas discretizadas. A quantização codifica os valores contínuos de um sinal em intervalos discretos. Em geral, os equipamentos só medem 8 *bits*, correspondentes a 256 níveis de cinza, mas a tecnologia atual chega a 24 *bits*. O número de *bits* por *pixel* correspondem ao número de cores ou tons de cinza que podem ser representados. A Figura 2.1 ilustra uma imagem digital que foi amostrada em uma matriz  $8 \times 8$  e quantizada utilizando apenas 1 *bit*, correspondente a 2 níveis de cinza (0 - preto, 1 - branco).

1	1	1	1	1	1	1	1	1
1	0	0	0	1	1	0	0	0
1	1	0	1	1	1	1	0	1
1	1	0	1	1	1	1	0	1
1	1	0	1	1	1	1	0	1
1	1	0	0	0	0	0	0	1
1	1	0	1	1	1	1	0	1
1	1	0	1	1	1	1	0	1
1	1	0	1	1	1	1	0	1
1	0	0	0	1	1	0	0	0
1	1	1	1	1	1	1	1	1

Figura 2.1: Imagem digital, onde o bit 0 indica ausência de luz (preto) e o bit 1 indica presença de luz (branco) (Fonte: [Cornell, 2000]).

### 2.1.2 Resolução

Resolução é a habilidade para distinguir os detalhes espaciais de uma imagem. Ela é uma medida de clareza ou detalhe e pode se referir tanto a um arquivo de imagem como um dispositivo, como um monitor ou uma impressora. Resolução de um arquivo de imagem é frequentemente expresso por suas dimensões em *pixels*, tal como  $1000 \times 2000$  *pixels*. Por exemplo, uma matriz semelhante,  $640 \times 480$  *pixels*, é usada para caracterizar o *display* de monitores. Resolução de impressão é mais comumente expressa em termos de pontos por polegada (*dots per inch* - dpi) [Besser and Trant, 1995]. Quanto maior a resolução de uma imagem, maior é a sua qualidade, respeitando certos limites, uma vez que uma cópia nunca terá qualidade superior ao seu original. Entretanto, o tamanho do arquivo também cresce com o aumento da resolução. Mais informações sobre resolução podem ser encontradas

em [Valle Jr., 2003]. A Figura 2.2 apresenta uma imagem em cinco resoluções diferentes: 600dpi, 300dpi, 150dpi, 60dpi e 30dpi. Aqui, pode-se perceber que a qualidade diminui com a diminuição da resolução.



Figura 2.2: Imagem apresentada em várias resoluções: 600dpi, 300dpi, 150dpi, 60dpi e 30dpi (Fonte: [Besser and Trant, 1995]).

### 2.1.3 Profundidade de *Bit*

Profundidade de *bit* é uma indicação das qualidades tonais de uma imagem. Ela mede o número de cores (ou níveis de cinza nas imagens em escala de cinza) disponíveis para representar o valor de cor ou tom de cinza no trabalho original. Todo *pixel* amostrado é associado a um valor que corresponde a cor ou tom de cinza que ele representa [Bishoff, 2003]. Quanto maior a profundidade de *bit*, maior o número de cores ou tons de cinza disponíveis (mais informação capturada) e maior o tamanho do arquivo.

Segundo Cornell [Cornell, 2000], os tipos de profundidade de *bit* são os seguintes:

- **Imagem bitonal** - representada por *pixels* consistindo de 1 *bpit* cada, onde dois tons podem ser representados (tipicamente preto e branco), usando os valores 0 para preto e 1 para branco ou vice-versa.
- **Imagem em tons de cinza** - composta de *pixels* representados por múltiplos *bits* de informação, tipicamente estendendo de 2 a 8 *bits* ou mais.
- **Imagem colorida** - representada por uma profundidade de bit que vai de 8 até 24 ou mais. Com uma imagem de 24 *bits*, os *bits* são frequentemente divididos em três grupos: 8 para vermelho, 8 para verde e 8 para azul. Combinações destes *bits* são

utilizadas para representar outras cores. Uma imagem de 24 *bits* oferece 16,7 milhões de cores.

Figura 2.3 ilustra imagens com diferentes tipos de profundidade de *bit* apresentadas acima.

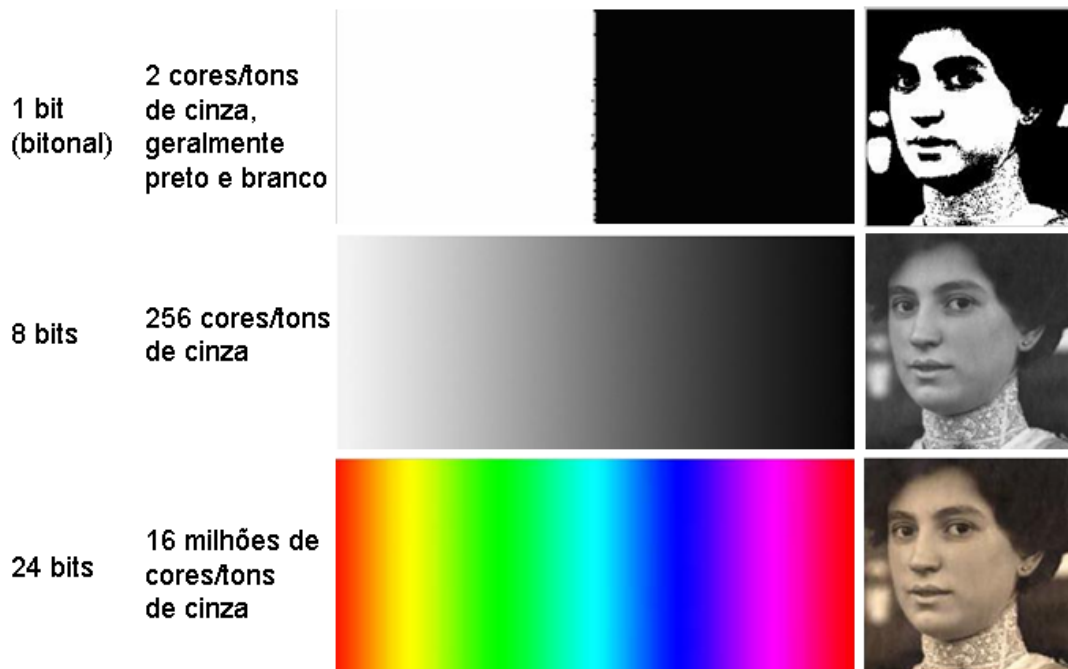


Figura 2.3: Exemplos de imagens com os três tipos de profundidade de bit: bitonal, tons de cinza e colorida (Fonte: [Bishoff, 2003]).

### 2.1.4 Compressão

A compressão de imagens é o processo de reduzir o tamanho dos arquivos de imagens através de métodos, tais como abreviação de informação repetida ou eliminação de informação que é difícil de ser vista pelo olho humano [Besser and Trant, 1995].

Nem todas as imagens respondem à compressão da mesma maneira. Quando uma imagem é comprimida, tipos particulares de características visuais, tais como variações tonais sutis, podem produzir artefatos ou efeitos visuais não esperados. Alguns esquemas de compressão, como o JPEG (ver Seção 2.1.5), permitem o usuário definir um grau de

perda. Através de testes cuidadosos, um balanço entre qualidade visual e tamanho de arquivo pode ser alcançado.

Os métodos de compressão se dividem em dois tipos: métodos sem perdas e métodos com perdas. Esquemas sem perdas, tal como TIFF (ver Seção 2.1.5), abreviam o código binário sem descartar nenhuma informação. Dessa forma, quando a imagem é descompactada, ela é *bit a bit* idêntica a original. Esquemas com perdas, como JPEG, utilizam um misto de média e descarte da informação mais insignificante, baseando-se na compreensão de percepção visual. Entretanto, pode ser extremamente difícil detectar os efeitos da compressão com perdas e a imagem pode ser considerada “visualmente sem perdas”. A compressão sem perdas é usada mais frequentemente em digitalização bitonal de material textual. Compressão com perdas é tipicamente usada com imagens coloridas e, em particular, imagens de tom contínuo, onde a mera abreviação não resulta em uma apreciável diminuição de espaço de armazenamento [Cornell, 2000].

### 2.1.5 Formatos de Arquivo

Formatos de arquivo consistem dos *bits* que compreendem a imagem e a informação de cabeçalho de como ler e interpretar o arquivo. Formatos de arquivo variam em termos de resolução, profundidade de *bit*, capacidades de cor e suporte para compressão e metadados [Cornell, 2000].

Em [California, 2001], são apresentados os principais formatos utilizados para projetos de digitalização:

- **Tagged Image File Format (TIFF)** - Formato de armazenamento de 24 *bits* usado pelo Adobe PhotoShop [Adobe, 2004] e outros editores de mapas de *bits*, este formato TIFF pode ser usado para armazenar imagens coloridas. Esse formato também é adequado para documentos textuais bitonais; ele fornece um alto nível de detalhes (acima de 600dpi ou  $4800 \times 6600$  *pixels* para uma página) [TIFF, 1992]. Este formato utiliza um esquema de compressão sem perdas.

- ***Joint Photographic Experts Group (JPEG)*** - Formato de 24 *bits* com compressão com perdas que são usados para apresentação em monitores e impressão. JPEG é suportado pela maioria de plataformas de computador e pelos navegadores de *Internet*. Com compressão com perdas, a qualidade das imagens dos arquivos compactados é reduzida quando comparada com o arquivo original e não pode ser restaurada, exceto através do próprio original [JPEG, 2004]. A vantagem é que o tamanho do arquivo é muito menor e a qualidade é aceitável na maioria dos casos.
- ***Graphic Interchange Format (GIF)*** - Formato de 8 *bits* com compressão sem perdas que é adequado para *display* de imagens em baixa resolução [GIF, 2004]. GIF é frequentemente usado para imagens *thumbnails* e versões de documentos de texto e é suportado pela maioria das plataformas e navegadores *Web*.
- ***Portable Network Graphics (PNG)*** - Formato de arquivo extensível para imagens armazenadas com compressão sem perdas. PNG provê uma substituição livre de patente para o GIF e pode também substituir muitos dos usos comuns do TIFF [PNG, 2004].
- ***Portable Document Format (PDF)*** - Formato que fornece uma forma conveniente de ver e imprimir imagens em alta resolução e pode ser usada para agrupar arquivos em capítulos e livros. PDF pode prover ferramentas de navegação adicionais, tais como hiperligações entre páginas dentro de um documento e de um documento PDF para outro. Embora seja um formato proprietário da Adobe [Adobe, 2004], a sua especificação e o *software* de visualização são distribuídos livremente. *Plug-ins* estão disponíveis para a maioria dos navegadores *Web* para permitir a leitura de arquivos PDF sem o uso de um visualizador externo [PDF, 2004].

### 2.1.6 Metadados

O termo metadado significa, literalmente, dado acerca do dado, e é utilizado para denotar os atributos associados a um documento digital. Os metadados adicionam valor ao do-

cumento, fornecendo informações importantes acerca do seu conteúdo, formato e história administrativa. Uma definição ampla de metadado seria “qualquer conhecimento a respeito de um objeto de informação, em qualquer nível de agregação destes”. Nesse contexto, objeto de informação é qualquer artefato que possa ser manipulado por uma pessoa ou sistema como uma entidade discreta. O objeto pode compreender um único item ou pode agregar muitos itens [Swetland, 2000]. Por exemplo, um objeto pode ser uma fotografia ou um documento de texto.

A criação de metadados de qualidade é um componente chave para o gerenciamento e preservação responsável de arquivos digitais produzidos [Bishoff, 2003]. Metadados extensivos são a melhor forma de minimizar os riscos de um objeto digital se tornar inacessível. Vários tipos de metadados que parecem não ter importância hoje, podem ser críticos para o acesso a esses arquivos no futuro. Uma boa estratégia pode ser salvar qualquer metadado que não necessite de grande espaço de armazenamento ou seja de fácil recuperação e que alguém indique que possa eventualmente ser importante [Besser, 2000].

Várias perspectivas de metadados se tornaram importantes no desenvolvimento de sistemas de informação digital, mas elas levam a um conceito muito amplo de metadado. Para entender este conceito melhor, é útil dividir em categorias - administrativa, descritiva, preservação, uso e metadados técnicos - que refletem aspectos chave da funcionalidade de metadados. A Tabela 2.1 define cada uma destas categorias de metadados e dá alguns exemplos de funções comuns que podem ser utilizadas em sistemas de informação digital [Swetland, 2000].

## 2.2 Preservação Digital

### 2.2.1 Introdução

Hoje, pode-se armazenar imagens detalhadas de centenas de livros, compostas a partir de minúsculos fragmentos, no limitado espaço de um disco ótico. Esta capacidade de registrar e armazenar dá origem a um dos dilemas centrais da história registrada: a capa-



Tabela 2.1: Categorias de metadados e suas funções [Swetland, 2000].

<b>Tipo</b>	<b>Definição</b>	<b>Exemplos</b>
<b>Administrativo</b>	Metadados usados no gerenciamento e administração de recursos de informação	<ul style="list-style-type: none"> <li>• Rastreamento de direitos e reprodução</li> <li>• Informação de aquisição</li> <li>• Documentação de requisitos de acesso legal</li> <li>• Informação de localização</li> <li>• Critério de seleção para digitalização</li> <li>• Controle de versão e diferenciação entre objetos de informação semelhantes</li> <li>• Auditorias criadas por sistemas de arquivamento</li> </ul>
<b>Descritivo</b>	Metadados usados para descrever ou identificar recursos de informação	<ul style="list-style-type: none"> <li>• Metadados para sistemas de armazenamento gerados por criadores de registros</li> <li>• Instrumentos de pesquisa</li> <li>• Índices especializados</li> <li>• Relacionamentos de hiperligação entre recursos</li> <li>• Anotações feitas por usuários</li> <li>• Registros de catálogos</li> </ul>
<b>Preservação</b>	Metadados relacionados com gerenciamento de recursos de informação	<ul style="list-style-type: none"> <li>• Documentação de ações tomadas para preservar versões físicas e digitais dos recursos, como refrescamento e migração de dados</li> <li>• Documentação da condição física dos recursos</li> </ul>
<b>Técnico</b>	Metadados relacionados a como funções do sistema ou metadados se comportam	<ul style="list-style-type: none"> <li>• Dados de autenticação e segurança, por exemplo, chaves de encriptação, senhas</li> <li>• Documentação de <i>hardware</i> e <i>software</i></li> <li>• Informação de digitalização, por exemplo, formatos, razão de compressão, rotinas de escala</li> <li>• Rastreamento de tempos de resposta de sistema</li> </ul>
<b>Uso</b>	Metadados relacionados ao nível e tipo de uso dos recursos de informação	<ul style="list-style-type: none"> <li>• Informações de reuso de conteúdo e de múltiplas versões</li> <li>• Registros de exibição</li> <li>• Rastreamento de uso e usuário</li> </ul>

cidade de registrar informações aumentou exponencialmente ao longo do tempo, enquanto que a longevidade dos meios utilizados para armazená-la decresceu de modo equivalente [Conway, 2001].

As fitas magnéticas têm uma vida útil estimada em 30 anos. Algumas tecnologias de disco óptico prometem durabilidade de 100 anos desde que guardados de forma adequada, mas tal expectativa é irrelevante quando a tecnologia do sistema pode estar em uso por no máximo 20 ou 30 anos [Van Bogart, 1995].

A necessidade de manter os dados “frescos” e codificados em formatos de arquivos que possam ser lidos pelos sistemas *hardware* e *software* disponíveis é mais importante do que a própria durabilidade da mídia [Smith, 1999]. Ler e compreender a informação na forma digital requer equipamento e *software*, os quais estão mudando constantemente e podem não estar disponíveis dentro de uma década após sua introdução. Mudanças rápidas na forma de gravação da informação, em formatos de armazenamento e nas tecnologias ameaçam o acesso à informação na era digital [Waters and Garret, 1996].

A única forma de superar essa dificuldade é aceitar que os dados digitais, sem intervenção humana, freqüente, não sobrevivem. É preciso a cada poucos anos (5 ou 10), reavaliar o cenário tecnológico e tomar as medidas (ver Seção 2.2.3) para garantir a sobrevivência do acervo nas novas condições.

Para [Waugh et al., 2000], a preservação digital possui três níveis: física, funcional e organizacional. A preservação física envolve preservar os *bits* que formam a informação digital contra a deterioração da mídia e contra a obsolescência dos leitores da mídia. Preservação funcional é a preservação de algumas ou todas as funções do ambiente de *software* original. Simplesmente preservar os *bits*, sem assegurar que tais *bits* possam ser convertidos para um formato que possam ser lidos pelos usuários é um trabalho inútil. É importante notar que preservar as funções não implica em preservar a aplicação como um todo. Preservação organizacional é a preservação de informação suficiente para permitir a uma organização usar a informação preservada em seus negócios. Os requisitos para preservação organizacional variam de organização para organização, mas tipicamente é necessário ser capaz de encontrar a informação, entender seu contexto, estar confiante de sua autenticidade e saber

sua propriedade.

### 2.2.2 Autenticidade e Segurança de Coleções Digitais

Estabelecer a autenticidade é de grande relevância para ambientes digitais, devido à facilidade com que os dados podem ser alterados. Um usuário futuro deve estar seguro de que um objeto digital não esteve sujeito à mudanças não autorizadas, acidentais e nem deliberadas. Portanto, é crucial cuidar da preservação intelectual, onde endereça a integridade e autenticidade da informação como ela foi gravada originalmente [Aschenbrenner, 2001].

Para garantir que a informação não seja corrompida sem conhecimento ou alterada maliciosamente, um documento deve ser autenticado. A técnica usada deve ser fácil de usar e durável por longos períodos de tempo. Tecnologias emergentes, tais como assinatura digital<sup>1</sup> [Bloemers, 2001], marca d'água digital<sup>2</sup> [Katzenbeisser, 2001] ou carimbo de tempo digital<sup>3</sup> [Maniatis et al., 2001], podem superar esse desafio. Estes métodos determinam a autoria, protegem contra cópias não autorizadas e provam a existência de documentos em um tempo específico, respectivamente.

### 2.2.3 Estratégias de Preservação Digital

Muitas estratégias de preservação têm sido propostas. As principais são: refrescamento, migração, emulação, reconstrução, preservação da tecnologia e arqueologia digital.

#### Refrescamento

O refrescamento consiste na transferência dos arquivos de um suporte para outro, sem alteração do seu conteúdo. Ele é feito tanto pelos efeitos da obsolescência, quanto pela degradação física das mídias [Waters and Garret, 1996].

O envelhecimento das mídias digitais precisa ser acompanhado cuidadosamente, porque ao contrário do que acontece com as gravações analógicas, que vão perdendo a qualidade

---

<sup>1</sup>do inglês *digital signaturing*.

<sup>2</sup>do inglês *digital watermarking*.

<sup>3</sup>do inglês *digital time-stamping*.

progressiva e perceptivelmente, as mídias digitais vão sofrendo degradações de sinal a princípio aceitáveis pelos seus mecanismos de leitura e, portanto, invisíveis para o operador - até o dia em que a leitura falha completamente e não é mais possível recuperar os dados.

### **Migração**

Quando ocorre problemas de compatibilidade entre duas versões diferentes ou o *software* necessário para interpretar o documento não existir mais, o refrescamento não resolve mais o problema de assegurar a longevidade dos documentos digitais. Neste caso, a migração pode superar esses problemas. Migração é a transferência periódica dos materiais digitais de uma configuração de *hardware/software* para outra, ou de uma geração de tecnologia de computadores para uma geração subsequente [Waters and Garret, 1996].

A migração tem a vantagem de eliminar a necessidade de manter a aplicação original. Entretanto, a migração não pode ser feita se o conhecimento do formato dos dados original estiver perdido [Waugh et al., 2000].

Mas como qualquer processo de tradução, a migração sujeita os dados a uma certa distorção de significado. É frequente, por exemplo, que no processo de conversão de textos de um formato para outro, parte da formatação original seja perdida ou alterada [Valle Jr., 2003].

### **Emulação**

O processo de migração apresenta dificuldades para preservação dos dados digitais. Além dos erros de tradução a que ela sujeita os dados, a migração é um processo cuja necessidade é impossível de prever com antecedência, uma vez que as mudanças de paradigma da indústria de informática seguem uma agenda completamente estranha aos interesses das instituições de custódia. Mais do que isso, a migração é um procedimento que envolve um certo senso de urgência, pois se um documento não é convertido em tempo hábil, após uma mudança de paradigma, pode ser impossível fazê-lo mais tarde, quando seu *software* ou *hardware* já se houverem tornado obsoletos [Rothenberg, 1998].

Uma outra estratégia que pode ser utilizada em conjunto ou em substituição a migração

é a emulação. Esta abordagem se baseia na emulação da tecnologia obsoleta em um novo sistema. Assim, é possível ver o documento digital em seu ambiente original recriando sua funcionalidade [Aschenbrenner, 2001].

Como os objetos digitais são dependentes de tecnologia, a emulação poder ser feita em dois níveis: *software* ou *hardware*. Na tentativa de torná-los independentes da tecnologia e permitir o seu uso por novas tecnologias melhores, defende-se que os objetos digitais sejam armazenados simplesmente como um fluxo de *bits*<sup>4</sup> e com as especificações detalhadas sobre o ambiente técnico no qual ele possa ser recriado no futuro [Russell, 1999].

Apesar da emulação ser utilizada em larga escala dentro da indústria de computação para prolongar a vida de aplicações legadas, há desafios práticos significantes no uso da emulação para preservar informação digital durante longos períodos [Waugh et al., 2000]. Para emular um sistema de informação antiquado, é necessário especificar exaustivamente o *hardware* do sistema original. Portanto, este método pode requerer uma extensiva participação dos fabricantes. Muitos fabricantes podem relutar em fornecer todas as especificações para os desenvolvedores de *software* porque algumas das tecnologias podem ainda estar em uso em sistemas avançados que eles desenvolvem [Heminger and Robertson, 2000].

## Reconstrução

A reconstrução é um estratégia que consiste em preservar os dados dos documentos e informações que possibilitem recriar as funções das ferramentas de *software/hardware*. Ela se diferencia da emulação, pois o ambiente original não é reconstruído. Com base nas informações preservadas, pode ser criar uma nova aplicação no ambiente de trabalho atual que consiga acessar os documentos digitais.

## Preservação da Tecnologia

A forma mais óbvia de assegurar que o objeto está preservado quando ele é criado e também preservar o ambiente usado para criar e usar o recurso, isto é, preservar o ambiente de *software* e *hardware* que foi usado para acessar o recurso quando ele foi criado.

---

<sup>4</sup>Fluxo de *bits*: do inglês *bitstream*.

Preservação do ambiente técnico pela conservação das cópias do *software* e *hardware* é referenciado como “preservação da tecnologia”. Para alguns objetos digitais, isto pode ser a melhor solução - pelo menos em um tempo curto - porque assegura que o material está acessível através da preservação tanto das ferramentas como do próprio objeto. Entretanto, para um tempo muito longo, esta abordagem é problemática. Por exemplo, aspectos de espaço e manutenção do *hardware* e os custos podem tornar esta solução impraticável [Russell, 1999].

### Arqueologia Digital

Refreshamento, migração, emulação e reconstrução são medidas preventivas de conservação da informação digital. Quando as mídias originais já se degradaram além da legibilidade em seus dispositivos originais, ou a tecnologia de interpretação dos arquivos já foi há muito esquecida, é preciso tomar ações corretivas para preservar, tanto quanto possível, os registros digitais remanescentes. Esse esforço, numa analogia à delicadeza dos métodos - e à imprevisibilidade dos resultados - é chamado de arqueologia digital [Valle Jr., 2003]. A arqueologia digital inclui métodos e procedimentos para resgatar conteúdo de mídias danificadas ou de ambientes de *software* e *hardware* obsoletos ou danificados [Cornell, 2000].

Contudo, a não ser que os registros sejam realmente valiosos, os custos desses procedimentos podem se revelar um impedimento insuperável. Como na conservação de materiais convencionais, as medidas preventivas são muito preferíveis às corretivas, e têm um custo muito mais baixo [Valle Jr., 2003].

## 2.3 Cadeia de Digitalização

Em projetos de digitalização de acervos, é necessária uma infra-estrutura de *hardware* e *software* que permita que todas as atividades sejam executadas de forma correta para a geração e acesso de documentos digitais de ótima qualidade. Deve-se estar atento não só à captura dos documentos, mas também ao armazenamento, gerenciamento e acesso ao acervo digital. Ao conjunto dessas atividades, dá-se o nome de cadeia de digitalização. A

Figura 2.4 apresenta um esquema de cadeia de digitalização com suas etapas e componentes.

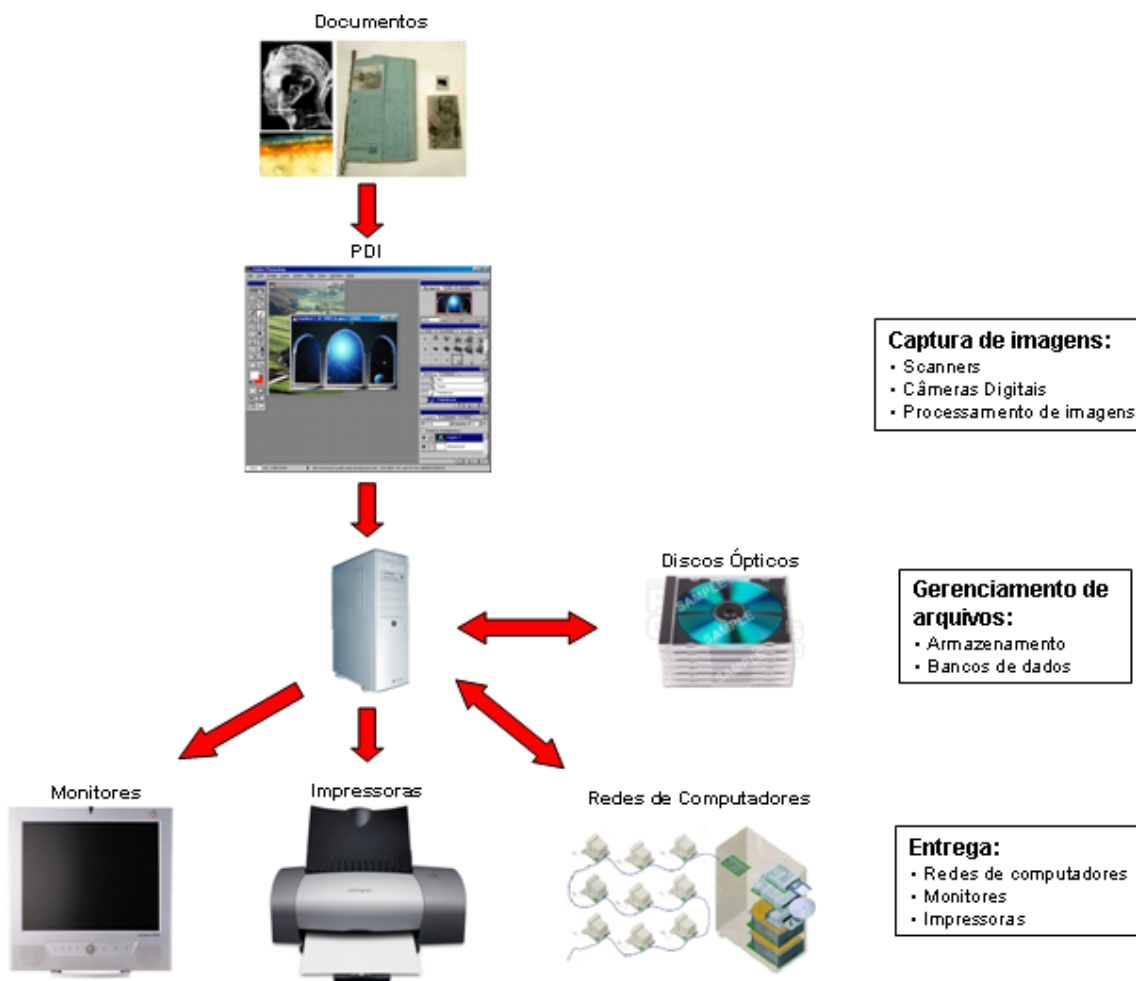


Figura 2.4: Cadeia de digitalização.

Decisões técnicas de infraestrutura requerem planejamento cuidadoso porque as tecnologias mudam rapidamente. A melhor forma de minimizar o impacto de depreciação e obsolescência é através de avaliações cuidadosas e evitar a adoção de soluções proprietárias [Cornell, 2000].

### 2.3.1 Captura da Imagem

A captura da imagem trata da aquisição inicial ou conversão de um documento ou objeto para a forma digital, tipicamente com um *scanner* ou uma câmera digital. Pode haver dois

ou mais arquivos ou etapas de processamento de imagens aplicadas à imagem inicial, que podem alterar, adicionar ou extrair dados [Cornell, 2000].

### Tipos de Documentos

Antes de iniciar o processo de captura das imagens, é necessário identificar os tipos de documentos que serão digitalizados. Com base nos tipos de documentos e suas características, é possível determinar os parâmetros de captura mais adequados ao tipo de acervo existente. A captura deve levar em consideração os processos técnicos envolvidos na conversão da representação analógica para digital e os atributos dos documentos fonte: tamanho físico, nível de detalhe e presença de cor [Cornell, 2000]. Os principais tipos de documentos são [Nara, 1998]:

- Documentos textuais - incluem registros originais, fotocópias ou microfilme.
- Fotografias - incluem fotografias impressas em preto e branco ou em cores, negativos ou transparências.
- Mapas, Plantas e Registros Grandes - registros originais ou microfilme.
- Registros Gráficos - desenhos, ilustrações artísticas ou similares.

### Equipamentos

A captura pode ser efetuada por dois tipos de equipamentos: *scanners* ou câmeras digitais. *Scanners* operam iluminando o objeto ou documento a ser digitalizado e direcionando a luz refletida para elementos fotosensíveis. Na maioria dos *scanners*, o sensor é um circuito integrado sensível à luz conhecido como *charged coupled device* (CCD). O CCD converte os níveis de luz em sinais eletrônicos que são processados em uma imagem digital. Ao se escolher um *scanner*, deve-se considerar se o equipamento é compatível com os documentos a serem digitalizados, se os documentos gerados possuem os requisitos de qualidade necessários e se as especificações do *scanner* são padronizadas e fáceis de interpretar [Cornell, 2000]. A Figura 2.5 apresenta alguns tipos de *scanners*.



Figura 2.5: Tipos de *scanners*.

As câmeras digitais combinam um *scanner* com uma câmera óptica para formar uma ferramenta que pode produzir imagens de alta qualidade. Embora mais lentas do que os *scanners*, as câmeras digitais são mais adaptáveis a uma grande quantidade de documentos e objetos. Materiais muito frágeis podem ser seguramente capturados, embora necessitem de luz externa para que não sejam danificados pelo *flash* da máquina [Cornell, 2000]. Outra vantagem das câmeras digitais em relação aos *scanners* é que elas podem ser utilizadas para fotografar imagens de tamanhos maiores do que os tamanhos suportados pelos *scanners* [Besser and Trant, 1995]. Em geral, as câmeras digitais apresentam *zoom* digital. Este recurso aproxima o assunto da foto por meio da ampliação de uma área selecionada do CCD. Entretanto, o *zoom* digital deve ser evitado para fins de preservação porque ele cria informação que não existe no documento original utilizando um processo de interpolação. O *zoom* óptico aumenta a imagem antes de gravá-la como *pixels* através de lentes. O *zoom* óptico provê uma qualidade de imagem melhor que o *zoom* digital sem perda de dados da imagem e sem a criação de informação inexistente.

### 2.3.2 Armazenamento

Depois que as imagens foram capturadas, é necessário armazená-las em algum lugar. Como pré-requisito, espaço deve ser fornecido. Assim, é necessária uma solução robusta e de alta capacidade de armazenamento. Um *trade-off* entre capacidade, tempo de acesso e

custo deve ser levado em consideração de acordo com os recursos e metas do projeto de digitalização [Aschenbrenner, 2001].

Tipicamente, o componente de cadeia de digitalização que recebe mais atenção é a captura, porque ela interage diretamente com o objeto que está sendo digitalizado e tem grande influência na qualidade e na fidelidade da imagem resultante. Pouca atenção é dispensada às mídias de armazenamento. Isto é preocupante, pois escolhas pobres em tecnologia de armazenamento podem diminuir a eficiência no momento de visualizar o resultado e podem comprometer o trabalho com a perda ou corrupção dos dados [Cornell, 2000].

Como um sistema que armazena grandes quantidades de imagens aumenta em tamanho e funcionalidade, é crítico para a plataforma de tecnologia entregar o desempenho e a confiabilidade requeridos. A capacidade de armazenamento também deve ser escalável para se adaptar ao rápido crescimento e deve se adaptar aos vários tipos de mídia que serão armazenados [Sun, 2002].

### **Discos Magnéticos**

O espaço de armazenamento é quase exclusivamente baseado em discos magnéticos. Estes discos se caracterizam por acoplarem a mídia dos dados aos dispositivos de leitura/gravação em uma unidade, permitindo grandes velocidades de transferência de dados, confiabilidade em operação contínua e dispensabilidade da montagem.

Mas tais dispositivos, como qualquer outro dispositivo elétrico, estão sujeitos a falhas. Por isso, é necessário fazer *backups* regulares para que se tenha uma cópia em caso de perda dos discos rígidos.

### **Discos Ópticos**

Os discos ópticos são mídias muito utilizadas em projetos de digitalização de acervos. Os discos ópticos tornaram-se a mídia terciária mais popular, por uma série de razões. Seu procedimento de montagem é extremamente rápido e simples, podendo ser executado pelo próprio usuário, sem necessidade de treinamento especial. Os discos permitem acesso direto, isto é, a localização imediata dos dados desejados sem a necessidade de lentas

operações de avanço e retrocesso e - em comparação com fitas magnéticas, disquetes ou discos rígidos - são suportes muito mais resistentes [Valle Jr., 2003].

Mas essa resistência é limitada. Qualquer coisa na superfície óptica que impeça o *laser* de focalizar na camada de dados pode resultar em perda de dados. Impressões digitais, manchas, arranhões, sujeiras, cisco, solventes, umidade ou qualquer material estranho podem interferir com a capacidade de leitura de dados do *laser*. Eles também podem interferir na capacidade do *laser* acompanhar a trilha de dados do disco. Arranhões afetam os discos de forma diferente, dependendo do lado do disco afetado, da severidade e a direção do arranhão e o tipo de disco [Byers, 2003].

Os principais representantes dos discos ópticos são os CD's<sup>5</sup> e os DVD's<sup>6</sup>. Os CD's podem chegar a 700MB de capacidade, enquanto os DVD's variam de 4,7GB a 17,1GB. Mesmo apresentando grandes capacidades de armazenamento, os DVD's ainda não possuem um formato padronizado. Na área de armazenamento de dados, a grande rivalidade é protagonizada entre os padrões DVD-R/RW e DVD+R/RW. Por isso, é necessário tomar cuidado com o padrão em que são salvos os dados, pois eles podem não ser lidos pelo dispositivo de DVD disponível na máquina em uso. Felizmente, já começam a aparecer leitores/gravadores de DVD que suportam todos os formatos [Lopes et al., 2003].

### **Fitas Magnéticas**

As fitas magnéticas são outro dispositivo utilizado para *backup* de dados. Quando comparadas aos discos ópticos e óptico-magnéticos, as vantagens das fitas são expectativa de vida e a confiabilidade na retenção dos dados ao longo de sua vida útil. Suas desvantagens são o acesso seqüencial (as fitas requerem um moroso avanço e retrocesso para que sejam acessados os dados desejados), a necessidade de treinar o operador ou usuário para sua manipulação correta, o elevado custo dos dispositivos de leitura/gravação e a maior fragilidade [Valle Jr., 2003].

---

<sup>5</sup> *Compact Disc*.

<sup>6</sup> *Digital Versatile Disc* ou *Digital Video Disc*.

### 2.3.3 Visualização

A visualização compreende o processo de permitir o acesso e visualizar as imagens digitais pelos usuários. Os principais componentes desta etapa são os monitores, impressoras e redes de computadores.

#### Monitores

Um monitor serve como uma janela para os usuários visualizarem a coleção de imagens digitais [Cornell, 2000]. Esses dispositivos são atualmente o ponto mais fraco da cadeia de digitalização [Besser and Trant, 1995]. Embora a razão preço/desempenho dos monitores tenha melhorado enormemente, mesmo os produtos tecnicamente mais avançados ainda requerem compromissos significantes [Cornell, 2000]. As duas principais tecnologias empregadas nos monitores são o CRT (Tubos de Raios Catódicos<sup>7</sup>) e o LCD (*Display* de Cristal Líquido<sup>8</sup>). A Figura 2.6 mostra os dois tipos de tecnologias de monitores.



Figura 2.6: Tecnologias de monitores: CRT e LCD.

Ao se escolher um tipo de monitor, deve-se levar em consideração as características da aplicação. Deve-se também ficar atento a certos parâmetros. A resolução máxima é importante para CRT e LCD. No caso do CRT, atenção especial também para *dot pitch*

---

<sup>7</sup>CRT - do inglês *Cathode Ray Tube*

<sup>8</sup>LCD - do inglês *Liquid Crystal Display*

e a taxa de atualização da imagem na tela (*refresh*). Conexões atualizadas e recursos multimídia valorizam os LCD's [Carvalho et al., 2004].

### **Impressoras**

Atualmente, a impressão preto e branco é dominada por duas tecnologias: impressoras jato de tinta, que respinga tinta líquida através de pequenas extremidades no papel, e impressoras a *laser*, que usa uma fonte de luz para criar cargas em um tambor fotocondutivo, permitindo atrair partículas de tinta seca para fundir no papel. Impressoras jato de tinta têm se tornado bastante baratas, mas são mais lentas que as *laser*. Impressoras a *laser* podem produzir em torno de 100 páginas por minuto a 600dpi. Ambas tecnologias têm adotado cor, mas são muito mais caras e lentas [Cornell, 2000].

### **Internet**

Redes de computadores são as auto-pistas por onde trafegam, em âmbito mundial, informações eletrônicas dos mais variados tipos, incluindo textos, figuras, sons e imagens. Para o mundo globalizado - o mundo das redes de computadores - não existem fronteiras entre países, assim como também não há distinção de raça, sexo, cor ou nacionalidade entre pessoas.

Graças à *Internet*, um cidadão comum, a um custo muito baixo, poderá ter acesso a informações localizadas nos mais distantes pontos do planeta e também gerar e distribuir informações em larga escala, no âmbito mundial. Isto até poucos anos atrás só era possível a uma grande organização através dos meios de comunicação convencionais [Pereira, 2004]. Dessa forma, a Internet possui um grande poder para democratizar o acesso e disseminar informação para uma quantidade bem ampla de usuários.

## **2.4 Projetos de Digitalização de Acervos**

Com tantas possibilidades vislumbradas pelas tecnologia digital, vários projetos têm sido propostos e implementados para digitalizar e disponibilizar acervos de muitas instituições.

Nesta seção, são apresentados três exemplos: o Arquivo Público Mineiro, o Projeto Portinari e o Projeto NARCISSE.

### 2.4.1 Arquivo Público Mineiro (APM)

O Arquivo Público Mineiro é uma instituição centenária que recolhe e conserva importantes documentos administrativos abrangendo os principais aspectos da história mineira, garantindo aos cidadãos total acesso aos mesmos. O APM tem trabalhado intensamente para digitalizar o seu acervo e o Núcleo de Processamento Digital de Informações (NPDI) tem tido um importante papel neste processo através de alunos de iniciação científica e pós-graduação.

Na primeira tentativa, [Andrade, 1998, Andrade and Araújo, 2000] estudou a aplicação de sistemas de informação multimídia para automação de acervos. Como resultado, um protótipo foi implementado, englobando a coleção do presidente Arthur Bernardes. Este trabalho visou mostrar a aplicabilidade dos sistemas de informação multimídia para o gerenciamento de acervos digitais.

A segunda tentativa expandiu o protótipo para outras coleções. Neste trabalho, vários estudos foram feitos sobre procedimentos que assegurassem a preservação digital dos documentos digitais produzidos. Outros aspectos considerados foram a questão da gestão documental e o fluxo de trabalho para catalogação e geração de materiais digitais de boa qualidade [Valle Jr., 2003].

Os trabalhos não param por aí. Depois de definidos e implementados os procedimentos para captura e catalogação dos materiais digitais, a parceria APM/NPDI está trabalhando no desenvolvimento de novas aplicações sobre esse acervo digital, como métodos de pesquisa específicos baseados no conteúdo dos documentos.

### 2.4.2 Projeto Portinari

O Projeto Portinari [Lanzelotte et al., 1993] começou em 1979 com a catalogação de 4500 trabalhos e 25000 documentos sobre o trabalho do pintor brasileiro Cândido Portinari. Os

trabalhos, que incluem pinturas, desenhos e textos, foram fotografados em cor e preto e branco. Esses documentos foram digitalizados mais tarde.

Inicialmente, o projeto se dividiu em três fases. Na primeira fase, compreendeu-se um trabalho de localização e documentação do material documentário. Nessa tarefa, os documentos foram listados e fotografados “in loco” em cores e em preto e branco. A segunda fase envolveu pesquisa e catalogação, indexação, estabelecendo vários vocabulários controlados e tesouros do material coletado. A terceira fase se preocupou com a disseminação da informação [Lanzelotte et al., 1993].

Nesta última fase, uma aplicação hipermídia [Portinari, ] foi desenvolvida contendo toda a informação resultante da catalogação. Nesta aplicação, as obras de Portinari podem ser exploradas por tema, data e técnicas, assim como as relações entre as obras e os eventos em que foram exibidas [Rossi, 1996].

### 2.4.3 Projeto NARCISSE

Em 1990, um projeto europeu intitulado NARCISSE (*Network of Art Research Computer Image Systems in Europe*) foi lançado para construir um banco de dados multicultural para gerenciar documentação laboratorial de museus relacionada com materiais de pinturas.

Em 2001, a base de dados do NARCISSE foi transferida para o sistema EROS (*European Research Open System*). O sistema EROS é um sistema de pesquisa multicultural aberto para recuperação de imagens por conteúdo dedicado a troca de informações de conservação e restauração entre instituições culturais. Este processo envolveu um intenso trabalho de correção de dados armazenados incorretamente, normalização destes dados e a criação de listas controladas de termos relevantes para cada campo. Atualmente, o sistema possui mais de 150 mil fotografias e filmes de raio-x digitalizados em altas resoluções [Lahanier et al., 2002].

## 2.5 Considerações Finais

Este capítulo tratou de questões relevantes em projetos de digitalização de acervos. Os principais parâmetros para geração de documentos digitais foram apresentados. Uma cadeia de digitalização básica e seus principais componentes também foram discutidos.

A preservação digital é outro aspecto de grande importância. Se ela não é considerada, é bastante provável que todo o trabalho, que é bastante caro, seja perdido em pouco tempo. Em caso de risco de obsolescência ou perda dos documentos, existem estratégias que podem retardar esse processo. Como a preservação digital é ainda uma área nova, não se sabe ainda os possíveis problemas que as estratégias de preservação podem impor.



# Capítulo 3

## Gerenciamento de Informação

### Multimídia

Os sistemas de informação estão transformando a maneira como o trabalho é conduzido e como os produtos e serviços são produzidos. Os sistemas de informação também estão dando aos indivíduos novas ferramentas para melhorar suas vidas e suas comunidades [Laudon and Laudon, 1999]. Além disso, convive-se com uma série de formas de comunicação e informação, como a imagem, o som, o vídeo e o texto, capazes de expressar informações quantitativas ou qualitativas de forma amigável e com imediata percepção do usuário. Este capítulo discute sobre sistemas de informação, informação multimídia e as ferramentas necessárias para o seu gerenciamento. A Seção 3.1 define o conceito multimídia e apresenta as diferentes mídias. Na Seção 3.2, os sistemas de informação são definidos e é feita uma discussão sobre tais sistemas estendidos para o gerenciamento de informação multimídia. Conceitos sobre sistemas gerenciadores de banco de dados (SGDB) são introduzidos na Seção 3.3. A Seção 3.4 apresenta os principais requisitos para SGDB's multimídia. Por fim, a Seção 3.5 apresenta alguns tipos de indexação e consulta que podem ser feitos sobre a informação multimídia.

## 3.1 Multimídia

Por multimídia entende-se todos os programas e sistemas em que a comunicação entre homem e computador se dá através de múltiplos meios de representação de informação, como som e imagem animada, além da imagem estática já usada nos aplicativos gráficos [Paula Filho, 2000].

Os principais tipos de dados multimídia são [Barros, 2003, Khoshafian and Baker, 1996, Paula Filho, 2000]:

- **Texto:** É a forma mais básica e simples de se representar dados em um computador. Um texto em um computador pode estar em dois formatos. No formato ASCII, o texto não possui nenhum tipo de formatação enquanto num formato estruturado (*Word*<sup>1</sup>, *WordPerfect*<sup>2</sup>, HTML<sup>3</sup>) é possível apresentar o texto formatado, tornando a leitura mais agradável.
- **Imagens:** Imagens digitalizadas são sequências de *pixels* que representam uma região na tela do monitor do usuário. *Pixels* são números interpretados que permitem exibir na tela um pequeno ponto com valores diferentes de luminância, cor e contraste. *Pixels* podem ser simples, como 0's e 1's (preto e branco), ou podem conter 8, 16 ou 24 *bits* por *pixel*, representando imagens coloridas de alta resolução.
- **Áudio:** A principal característica que o som apresenta e que não se encontram no texto e nas imagens é que o som possui característica temporal. A maioria dos formatos de som (WAV<sup>4</sup>, AIFF<sup>5</sup>, SND<sup>6</sup>) armazena a informação sonora na forma de sua respectiva onda. Já o formato MIDI<sup>7</sup>, mais indicado para armazenar informações

---

<sup>1</sup> *Word* - processador de textos da *Microsoft*.

<sup>2</sup> *WordPerfect* - processador de texto.

<sup>3</sup> HTML - *HyperText Markup Language*.

<sup>4</sup> WAV - *Waveform Audio*.

<sup>5</sup> AIFF - *Audio Interchange File Format*.

<sup>6</sup> SND - *Sound Format*.

<sup>7</sup> MIDI - *Musical Instrument Digital Interface*.

sonoras oriundas de instrumentos musicais armazena uma seqüência de notas equivalente a que é tocada no instrumento.

- Vídeo: É sem dúvida nenhuma a forma mais rica de se apresentar um conteúdo. Num computador, o vídeo é armazenado de forma muito parecida com a de um rolo de filme, ou seja, uma seqüência de quadros.
- Gráficos: É a maneira de se representar dados graficamente. Existem duas formas de armazenamento de imagens em um computador. A maioria das imagens é armazenada na forma de mapa de bits, mas alguns aplicativos mais sofisticados utilizam imagens vetoriais que são formadas a partir de primitivas gráficas (ponto, reta e círculo).

## 3.2 Sistemas de Informação Multimídia

Um sistema de informação pode ser definido como um conjunto de componentes inter-relacionados trabalhando juntos para coletar, recuperar, processar, armazenar e distribuir informação com a finalidade de facilitar o planejamento, o controle, a coordenação, a análise e o processo decisório em empresas e outras organizações. Os sistemas de informação essencialmente transformam a informação em uma forma utilizável para a coordenação de fluxo de trabalho de uma instituição, ajudando as pessoas a tomar decisões, analisar e visualizar assuntos complexos e resolver outros tipos de problemas. Os sistemas de informação fazem isso através de um ciclo de três atividades básicas: entrada, processamento e saída. A Figura 3.1 mostra as três atividades e sua interação.

A entrada (ou *input*) envolve a captação ou coleta de fontes de dados brutos de dentro da organização ou de seu ambiente externo. O processamento envolve a conversão dessa entrada bruta em uma forma mais útil e apropriada. A saída (ou *output*) envolve a transferência da informação processada às pessoas ou atividades que a usarão. Os sistemas de informação também armazenam informação sob várias formas, até que ela seja necessária para o processamento ou a saída. A realimentação (ou *feedback*) é a saída que retorna

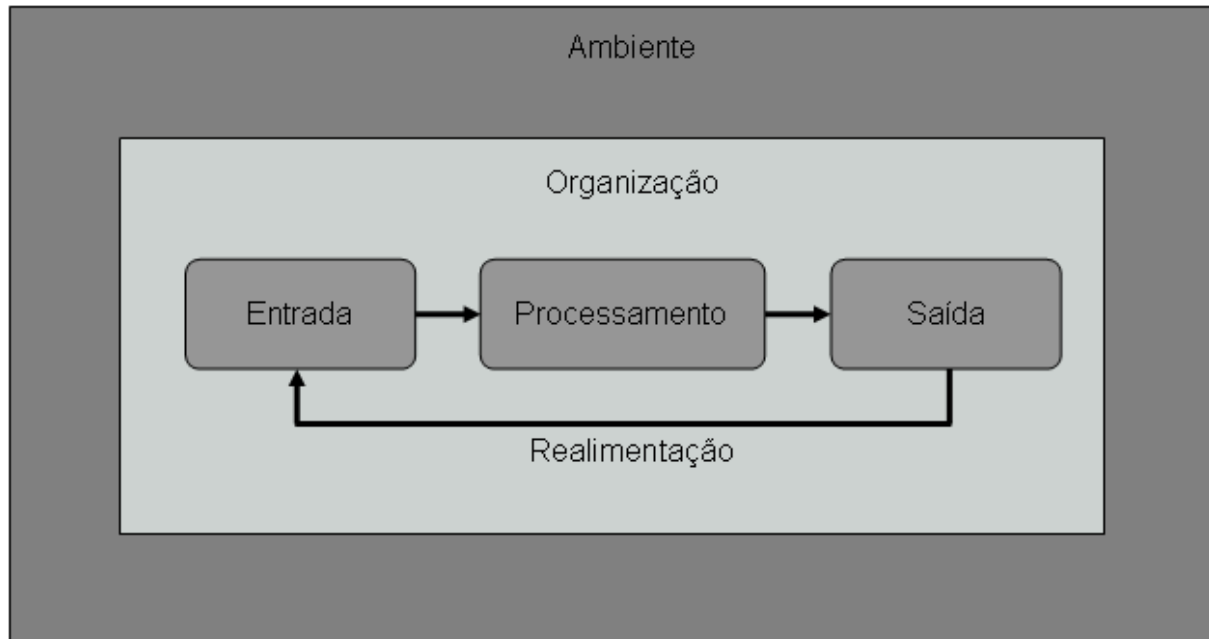


Figura 3.1: Atividades dos sistemas de informação: entrada, processamento e saída (Fonte: [Laudon and Laudon, 1999]).

aos membros adequados da organização para ajudá-los a refinar ou corrigir os dados de entrada [Laudon and Laudon, 1999].

Inicialmente, os sistemas de informação trabalhavam somente com informações textuais. Com a evolução dos sistemas de computador, a informação multimídia passou a ser tratada em tais sistemas, os chamados sistemas de informação multimídia. Devido à natureza heterogênea dos dados multimídia, sistemas que pretendem armazenar, transportar, exibir e gerenciar tais dados devem ter consideravelmente mais capacidades do que os sistemas de informação tradicionais [Ghafoor, 1995].

Os principais objetivos de um sistema de informação multimídia são organizar, apresentar e sincronizar dados multimídia para ambientes interativos [David, 1997]. Organização, sincronização e interação são elementos críticos em ambientes multimídia.

### 3.3 Sistemas Gerenciadores de Banco de Dados

Em todos os sistemas de informação, os dados devem ser organizados e estruturados para que possam ser usados com eficácia. No entanto, a menos que as informações possam ser facilmente processadas e acessadas, o sistema não pode alcançar seu propósito [Laudon and Laudon, 1999]. Assim, é necessário o uso de sistemas gerenciadores de bancos de dados para que seja possível ao sistema de informação cumprir o seu papel.

O desenvolvimento dos sistemas de gerência de bancos de dados (SGBD) foi de fundamental importância para o incremento do uso da informática, principalmente em aplicações com grande volume de dados, com ênfase nas áreas administrativa e de negócios.

Os SGBD's foram inicialmente projetados, a partir da segunda metade do século XX, para armazenar e manipular grandes volumes de dados alfanuméricos, bem formatados e comportados, voltados predominantemente para aplicações financeiras e comerciais [Melo, 1988].

Uma característica fundamental da abordagem de banco de dados é que ela fornece algum nível de abstração de dados, escondendo detalhes do armazenamento dos dados que não são necessários para a maioria dos usuários de banco de dados. Um modelo de dados - uma coleção de conceitos que podem ser usados para descrever a estrutura de um banco de dados - fornece o ferramental necessário para alcançar essa abstração [Elmasri and Navathe, 2000]. Os principais modelos são: o modelo hierárquico, o modelo de rede, o modelo relacional e o modelo orientado a objetos. Os dois primeiros praticamente não são mais utilizados.

O modelo de dados relacional está sedimentado em base matemática e definido com bastante rigor e formalismo, dominando atualmente a implementação de SGBD's disponíveis comercialmente. A maioria dos sistemas baseados no modelo de dados relacional suportam basicamente todo tipo de dado clássico constituído de cadeias não muito extensas de números, letras e caracteres e são restritos quando se trata de administrar outros tipos de dados [Siberschatz et al., 1999].

Bancos de dados orientados a objetos foram propostos para sanar as deficiências do mo-

delo relacional para modelar aplicações complexas, como bancos de dados de engenharia, experimentos científicos, telecomunicações, sistemas de informação geográfica e multimídia. Essas novas aplicações têm requisitos e características que diferem das aplicações tradicionais, tais como estruturas mais complexas, transações de longa duração, novos tipos de dados e grandes itens textuais. Outra razão para a criação dos bancos de dados orientados a objeto é o aumento do uso de linguagens de programação orientadas a objeto no desenvolvimento de aplicações de *software* [Elmasri and Navathe, 2000].

## 3.4 Multimídia e Sistemas de Bancos de Dados

Um SGBD multimídia provê um ambiente favorável ao uso e gerência de bancos de dados multimídia. Portanto, ele deve suportar os vários tipos de dados multimídia, em adição ao atendimento de facilidades providas pelas funções dos SGBD tradicionais, tais como definição e criação de um banco de dados, recuperação de dados, acesso a dados, independência de dados, privacidade, integração, controle de integridade, controle de versão e suporte a concorrência [Apers et al., 1997, Prabhakaran, 1997]. Assim, ele deve atender a todos os requisitos impostos pela informação multimídia, estendendo os requisitos de um SGBD tradicional.

A arquitetura de sistema de banco de dados padrão consiste de módulos para processamento de consultas, gerenciamento de transações, gerenciamento de *buffer*, gerenciamento de arquivos, recuperação e segurança. Implementações podem diferir dependendo do modelo de dados, mas a natureza dos módulos é basicamente a mesma [Grosky, 1997].

## 3.5 Indexação e Consulta

### 3.5.1 Consulta Baseada em Texto

Recuperação textual tem sempre sido a característica chave em aplicações de negócios e sistemas de bibliotecas e, embora muito trabalho tenha sido feito, ainda são necessários

alguns melhoramentos, especialmente nas seguintes questões [Elmasri and Navathe, 2000]:

- Indexação de frases: Melhorias substanciais podem ser percebidos se descritores de frases (em oposição a termos de indexação de uma palavra) forem atribuídos a documentos e usados em consultas, fornecidas tal que estes descritores sejam bons indicadores de informação e conteúdo de documentos necessários.
- Uso de tesouros: Uma razão para a pobre resposta dos sistemas correntes é que o vocabulário do usuário difere do vocabulário usado para indexar os documentos. Uma solução é usar um tesouro para expandir a consulta do usuário com termos relacionados. O problema, então, é encontrar um tesouro para o domínio de interesse.
- Resolver ambigüidade: Uma das razões para a baixa precisão (a razão do número de itens relevantes recuperados para o número total de itens recuperados) nos sistemas de recuperação de informação textual é que palavras têm múltiplos significados. Uma maneira de resolver a ambigüidade é usar um dicionário *on-line*; outro é comparar os contextos onde as duas palavras ocorrem.

Segundo [Wielinga et al., 2001], existe um número de formas diferentes para recuperar informação textual:

- Busca de palavras-chave com vocabulário livre;
- Busca de palavras-chave com vocabulário fechado;
- Busca baseada em tesouros, onde não somente o vocabulário é fechado, mas também hierárquico e outras relações podem ser consideradas no processo de busca.

### 3.5.2 Consulta Baseada no Conteúdo

Um objeto visual é associado a dois tipos de informação: informações textuais descritivas (metadados) e características contidas nos objetos (características visuais - *visual features*) [Rui and Huang, 1999].

Nos últimos anos, melhorias em tecnologia de *hardware* têm permitido processar, armazenar e recuperar grandes quantidades de dados no formato multimídia. Primeiramente, tentou-se gerenciar bancos de dados de imagens com base na descrição textual feita por um operador humano. Essa abordagem não captura a riqueza do conteúdo visual das imagens que não podem ser facilmente expressadas em palavras. Por essa razão, pesquisas têm sido feitas focando na extração automática do conteúdo visual de imagens para possibilitar a indexação e recuperação [Cinque et al., 2001].

A necessidade de encontrar uma imagem desejada de uma coleção é compartilhada por muitos profissionais, incluindo jornalistas, engenheiros e historiadores de arte. Como os requisitos de usuários de imagens podem variar consideravelmente, pode ser útil caracterizar consultas de imagens em três níveis de abstração [Eakins and Graham, 1999]:

- Primitivo: Compreende características como cor, textura, forma ou localização espacial dos elementos da imagem. Um exemplo de consulta pode ser “encontrar imagens com a cor azul na parte superior e vermelho na parte inferior”.
- Lógico: Compreende recuperação por características derivadas (conhecidas como lógicas), envolvendo algum grau de inferência sobre a identidade dos objetos descritos na imagem. Por exemplo, “encontrar uma figura da Torre Eiffel”.
- Abstrato: Compreende recuperação por atributos abstratos, envolvendo uma quantidade significativa de raciocínio de alto nível sobre o significado e propósito de objetos e cenas descritas.

A maioria dos trabalhos desenvolvidos tratam do nível primitivo devido à complexidade dos demais níveis de abstração.

Os métodos de busca baseados em conteúdo costumam ser mais bem sucedidos em acervos especializados, em que as características visuais das imagens provocam uma classificação significativa. Assim, em um acervo que contenha apenas moedas, as características de textura, forma e cor são bons classificadores, permitindo ao usuário localizar itens de interesse. Em acervos muito variados, entretanto, esses métodos não são bem sucedidos,



pois um grande número de imagens não relacionadas irá apresentar características extraídas similares [Bimbo, 1999]. Uma opção interessante seria unir os dois tipos de busca, textual e baseada no conteúdo, para gerar mecanismos de busca mais precisos.

## 3.6 Considerações Finais

O conceito de multimídia pode variar de autor para autor, mas é inegável a dimensão que ela alcançou com a evolução dos computadores. Agora, usuários podem interagir das mais diversas formas com as máquinas. Entretanto, é necessário uma arquitetura de *hardware/software* que suporte o processamento requerido pela aplicação multimídia utilizada.

Este capítulo apresentou uma definição para multimídia. Também foi feita uma breve discussão sobre os sistemas gerenciadores de bancos de dados (SGBD's). Para suportar os dados multimídias, os SGBD's devem possuir características específicas para o gerenciamento de dados multimídia em adição às características tradicionais.

# Capítulo 4

## Projeto e Desenvolvimento de Aplicações Hipermedia

Atualmente, o uso da WWW (*World Wide Web*) e aplicações hipermedia cresceu enormemente. Muitas organizações, institutos e outros têm implementado um grande número de sítios *Web* para os mais diversos fins e objetivos [Kyaw and Boldyreff, 1998]. Entretanto, as peculiaridades da hipermedia exigem o desenvolvimento de novos métodos específicos para suas características [Garzotto et al., 1993].

Este capítulo discute sobre o desenvolvimento de aplicações hipermedia e técnicas de engenharia de *software* tradicionais e específicas para aplicações hipermedia. A Seção 4.1 apresenta o processo de desenvolvimento de software, suas vantagens e alguns conceitos. A Seção 4.2 discute aspectos sobre o desenvolvimento de software hipermedia. Na Seção 4.3, os principais problemas de sistemas hipermedia são descritos. Alguns métodos específicos para aplicações hipermedia são apresentados na Seção 4.4. Finalmente, na Seção 4.5, são feitas algumas considerações finais.

## 4.1 Desenvolvimento de *Software*

O desenvolvimento de sistemas de *software* é uma indústria relativamente jovem e ainda não alcançou o nível de maturidade encontrado em outros ramos da indústria tradicional. Conseqüentemente, produtos desenvolvidos usando tecnologia de *software* frequentemente sofrem de uma falta de práticas estabelecidas para o seu desenvolvimento e exploração como produto comercial [Jacobson et al., 1997].

Atualmente, a tendência do *software* é na direção de grandes sistemas complexos. Isto é devido, em parte, ao fato de que os computadores tornam-se mais poderosos a cada ano, levando os usuários a esperar mais deles. Esta tendência também é influenciada pela expansão do uso da *Internet* para a troca de todo o tipo de informação - desde texto simples até texto formatados, imagens, diagramas e multimídia [Jacobson et al., 1999].

Durante o desenvolvimento de grandes sistemas, é importante saber como as diferentes etapas do método interagem e como elas preenchem o processo de desenvolvimento como um todo [Jacobson et al., 1997]. Um processo é um conjunto de passos parcialmente ordenados constituídos por atividades, métodos, práticas e transformações, usado para atingir uma meta. Esta meta geralmente está associada a um ou mais resultados concretos finais, que são os produtos da execução do processo [Paula Filho, 2001].

Muitas metodologias têm sido apresentadas na literatura, como cascata, espiral e prototipagem evolutiva. Entretanto, independente da metodologia escolhida, um processo de desenvolvimento de *software* contém três fases: definição, desenvolvimento e manutenção [Pansanato, 1999].

Na fase de definição, o papel de cada elemento num sistema baseado em computador é identificado. Também é feito um planejamento do projeto, estabelecendo o escopo do *software* com análise de riscos, alocação de recursos, estimação dos custos e programação do trabalho. É nesta fase que são definidos o domínio da informação e a função do *software*.

Durante o desenvolvimento, os requisitos do *software* são representados como modelos que descrevem a estrutura de dados, a arquitetura, o procedimento algorítmico e as características de interface. Também fazem parte dessa fase a implementação do sistema e a

realização de testes.

A fase de manutenção concentra-se nas mudanças que estão associadas à correção de erros, adaptações e ampliações, reaplicando os passos das fases de definição e desenvolvimento, mas dentro do contexto do software existente.

## 4.2 Desenvolvimento de Aplicações Hipermídia

Sistemas hipertexto estão emergindo como uma nova classe de sistemas de informação complexos. Estes sistemas permitem as pessoas criar, anotar, ligar e compartilhar informações de uma grande variedade de mídias, como texto, gráficos, áudio, vídeo, animação e programas. Sistemas hipertexto provêm um método novo e não seqüencial de acesso à informação diferentemente de sistemas de informação tradicionais que são sequenciais por natureza [Balasubramanian, 1994].

Projetos hipermídia diferem de projetos de desenvolvimento de *software* tradicionais em várias dimensões. Primeiro, projetos hipermídia podem envolver pessoas com várias habilidades diferentes, como autores, bibliotecários, projetistas de conteúdo, artistas, músicos e programadores. Segundo, o projeto de aplicações hipermídia envolve a captura e organização da estrutura de um domínio complexo e fazê-lo claro e acessível aos usuários. Terceiro, aspectos multimídia das aplicações hipermídia aumentam inúmeras dificuldades. O projeto hipermídia é, portanto, um processo desafiante que atualmente é mais uma arte do que uma ciência. Finalmente, é necessário prototipar e realizar testes mais intensivos com os usuários desses sistemas do que com sistemas tradicionais, porque a tolerância a erros pelos usuários é muito menor [Isakowitz et al., 1995].

O uso do formato hipertexto possui várias vantagens. Algumas são citadas abaixo [Balasubramanian, 1994, Kyaw and Boldyreff, 1998]:

- Formulários hipertexto podem suportar navegação flexível através de referências-cruzadas de informação;
- Mídias eletrônicas podem armazenar grandes quantidades de informação;

- A informação pode ser instantaneamente pesquisada por palavras-chave indexadas ou apresentadas em sistemas hipermídia;
- A informação pode ser facilmente modificada e atualizada.

Da mesma forma que os métodos de desenvolvimento de software tradicionais, as abordagens atuais para o desenvolvimento de aplicações hipermídia podem oferecer uma série ordenada de etapas, além de fornecer novos conceitos e notações convenientes (modelos) para representar o domínio de hipermídia [Pansanato, 1999]. As abordagens para desenvolvimento podem ser classificadas em cinco categorias [Christodoulou et al., 1998]:

- Abordagem Orientada a Objetos (OO) - O modelo OO foca na definição e herança de características comportamentais no forma de operações embutidas dentro de objetos. A abordagem OO é capaz de desenvolver aplicações que requerem estruturas complexas e relacionamentos entre itens de informação.
- Abordagem Entidade-Relacionamento (ER) - O modelo ER é derivado de banco de dados. Mais fraco que OO, em termos de suporte de características e funcionalidades hipermídia avançadas, essa abordagem é melhor para o desenvolvimento e manutenção de sistemas simples e de pequena escala, não incluindo itens de informação complexos.
- Abordagem Baseada em Componentes - Nesta abordagem, o conteúdo da aplicação é hierarquicamente decomposto em componentes. Um componente é definido por seu estado (especificado por uma lista de propriedades) e seu comportamento (operações sobre os estados dos componentes). É um modelo mais simples do que o modelo OO.
- Abordagem Híbrida - Esta abordagem integra dados de várias fontes heterogêneas.
- Outras Abordagens - Esta abordagem se baseia no uso de produtos comerciais, como o *Microsoft FrontPage*<sup>1</sup> ou o *Macromedia DreamWeaver*<sup>2</sup>. Esses produtos são editores

---

<sup>1</sup>*Microsoft Frontpage* - programa da *Microsoft* usado para criar e editar documentos HTML.

<sup>2</sup>*DreamWeaver* - programa da *Macromedia* usado para criar e editar documentos HTML.

de HTML (*Hypertext Markup Language*) e podem incluir ferramentas para gerenciar a estrutura do sítio, mantendo a consistência das ligações e fornecendo conexão com os bancos de dados. Eles são focadas em GUI's (*Graphical User Interface*) atrativas e úteis, mas elas não fornecem uma estruturação eficiente do modelo e do conteúdo da aplicação.

De acordo com [Kyaw and Boldyreff, 1998], os estágios para desenvolvimento de aplicações hipermídia são:

1. Estudo de Viabilidade e Análise de Requisitos - Todos os projetos de desenvolvimento incluem um estudo de viabilidade e uma análise de requisitos para analisar o domínio do problema e identificar os requisitos.
2. Projeto Conceitual - Diferentemente do desenvolvimento convencional, aplicações hipermídia requerem estruturação da informação e conceitualização das estruturas da organizacionais. Isto inclui o projeto de aspectos estruturais da aplicação.
3. Projeto Navegacional - Neste estágio, projetistas têm que prover ligações ricas e flexíveis em muitos caminhos para criar redes semânticas. Muitos modelos navegacionais diferentes podem ser construídos de um modelo conceitual para preencher as necessidades de conjuntos diferentes de usuários.
4. Projeto de Interface de Usuário - Esta fase pode ser considerada como um projeto detalhado dos modelos navegacionais para apresentar a informação aos usuários finais. Neste estágio, os projetistas identificam e constroem janelas, formulários, menus, *templates*, âncoras, objetos da apresentação, ligações e outros recursos *Web* para obter uma estrutura consistente em toda aplicação.
5. Implementação - Nesta fase, o projeto é implementado.
6. Testes e Manutenção - Como em engenharia de *software*, aplicações hipermídia requerem testes rigorosos para assegurar que eles podem ser usados por vários tipos de

usuários. O principal objetivo de aplicar métodos de projeto às aplicações hipermídia é ser capaz de mantê-las e atualizá-las facilmente.

### 4.3 Desorientação e Sobrecarga Cognitiva

O hipertexto se baseia na habilidade de produzir itens de informação complexos e ricamente conectados. Entretanto, pode também se tornar um sistema complexo de nodos desordenados, confundindo autores e usuários. De acordo com Conklin, a desorientação e a sobrecarga cognitiva são os dois problemas mais desafiantes relacionados com hipertexto. Ele disse que esses dois problemas podem limitar a utilidade do hipertexto [Conklin, 1987].

O problema da desorientação ou “ficar perdido no espaço” surge da necessidade de saber onde alguém está na rede, de onde ele veio e como chegar a outro lugar na rede [Balasubramanian, 1994]. Para minimizar a desorientação, os autores devem fornecer recursos para permitir aos leitores identificar a sua posição corrente em relação à estrutura global, reconstruir o caminho que o levou a esta posição, e distinguir entre as diferentes opções para mover-se a partir desta posição. Por exemplo, a manutenção do histórico de navegação (isto é, o caminho percorrido pelo usuário) auxilia o leitor a reconstruir o caminho até a sua posição atual [Pansanato, 1999].

A sobrecarga cognitiva é o esforço e concentração adicionais necessários para manter várias tarefas ou trilhas ao mesmo tempo. Como a capacidade humana de processamento de informação é limitada, cada esforço adicional para a leitura reduz os recursos mentais disponíveis para a compreensão [Pansanato, 1999]. Todos os sistemas hipertexto fornecem a utilidade básica de seguir uma ligação unidirecional até um nó alvo. Entretanto, o verdadeiro potencial do hipertexto não pode ser alcançado por essa abordagem apenas [Balasubramanian, 1994]. Os autores devem fornecer recursos para reduzir o esforço dos leitores na tomada de decisões de navegação, cobrindo uma certa quantidade de movimentos possíveis sem impor uma série de ações complicadas. As interfaces também devem permitir uma fácil adaptação por parte dos leitores, evitando que elas aumentem a carga cognitiva sobre os leitores. Para minimizar o esforço de adaptação, os autores devem considerar

cuidadosamente quais tarefas são indispensáveis e quais devem ser evitadas através de um projeto de interface apropriado.

## 4.4 Métodos de Desenvolvimento de Aplicações Hiper-mídia

Muitos métodos têm sido propostos para descrever o processo de desenvolvimento de aplicações hiper-mídia. Esta seção apresenta alguns métodos utilizados para o desenvolvimento de sistemas para *Web*. Desses, apenas o PRAXIS é de uso geral, sendo os demais específicos para aplicações hiper-mídia.

### 4.4.1 Processo para Aplicativos Extensíveis Interativos (PRAXIS)

O PRAXIS (sigla de PProcesso para Aplicativos eXtensíveis InterativoS) é um processo desenhado para suportar projetos realizados individualmente ou por pequenas equipes, com duração de seis meses a um ano. O PRAXIS propõe um ciclo de vida composto por fases que produzem um conjunto precisamente definido de artefatos (documentos e modelos). Para construir cada um desses artefatos, o usuário do processo precisa exercitar um conjunto de práticas recomendáveis da engenharia de *software*. Na construção desses artefatos, o usuário do processo é guiado por padrões e auxiliado pelos modelos de documentos e exemplos constantes do material de apoio [Paula Filho, 2001]. O PRAXIS é baseado na tecnologia orientada a objetos e sua notação de análise e desenho é a UML (Unified Modeling Language) [Rumbaugh et al., 1999]. O PRAXIS é dividido em quatro fases: Concepção, Elaboração, Construção e Transição.

Na fase de Concepção, são identificadas as necessidades dos usuários e os conceitos da aplicação são analisados o suficiente para justificar a especificação de um produto de *software*, resultando em uma proposta de especificação. Esta fase possui apenas uma atividade, a Ativação. O objetivo da Ativação é verificar se o cliente tem necessidades de



negócio suficientes para justificar estudos detalhados de especificação de um produto de *software*, que seriam então realizados na fase de Elaboração.

Na fase de Elaboração, a especificação do produto é detalhada o suficiente para modelar conceitualmente o domínio do problema, validar os requisitos em termos desse modelo conceitual e permitir um planejamento acurado da fase de construção. Contém duas iterações:

- Levantamento de Requisitos: visa à captura das necessidades dos usuários em relação ao produto, expressas na linguagem desses usuários;
- Análise dos Requisitos: confecciona um modelo conceitual do produto, que é usado para validar os requisitos levantados e para planejar a evolução posterior do projeto.

Uma versão operacional do produto é desenvolvida na fase de Construção, atendendo os requisitos especificados. Divide-se em três iterações: Desenho Implementável, Liberações e Testes Alfa. Durante o Desenho Implementável, os componentes internos e externos de um produto de *software* são definidos, em nível suficiente para decidir as principais questões de arquitetura e tecnologia e para permitir o planejamento detalhado das liberações. Em cada Liberação é implementado um subconjunto de funções do produto que será avaliado pelos usuários. Os Testes Alfa são testes que são feitos no ambiente do desenvolvedor para a identificação de possíveis problemas.

A Transição é a fase na qual o produto é colocado à disposição de uma comunidade de usuários para testes finais, treinamento e uso inicial. Suas iterações são: Testes Beta e Operação Piloto. Os Testes Beta são testes feitos no ambiente dos usuários e a Operação Piloto é a operação experimental do produto em instalação piloto, com a resolução de eventuais problemas através de processo de manutenção.

#### **4.4.2 *Hypertext Design Methodology* (HDM)**

O *Hypertext Design Model* (HDM) [Garzotto et al., 1993] é um modelo para descrever aplicações hipermídia. Uma aplicação (hiperdocumento) no HDM consiste de estruturas de informação chamadas entidades, que denotam um objeto físico ou conceitual do domínio,

e são agrupadas em tipos de entidades (*entity types*). Por exemplo, “motor elétrico” é uma entidade do tipo “equipamento”.

As entidades são compostas por componentes que são pedaços de informação. Um componente pode ter uma ou mais unidades. Essas unidades correspondem ao conceito de perspectivas. As perspectivas são as aparências que a informação pode ter. Por exemplo, uma informação pode ser visualizada em português e inglês, ou utilizando mídias diferentes como texto, gráfico, imagem, vídeo, áudio, etc. A Figura 4.1 mostra a hierarquia de projeto do HDM.

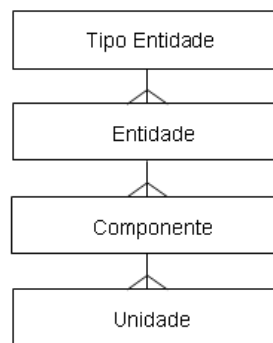


Figura 4.1: Hierarquia de projeto do HDM.

HDM diferencia três tipos de ligações: estrutural, aplicação e perspectiva. Ligações estruturais conectam componentes e ligações de perspectiva ligam unidades. Estas ligações podem ser automaticamente derivadas da estrutura das entidades. Ligações de aplicação são definidas pelo autor conectando componentes e entidades de mesmo ou diferentes tipos.

O modelo HDM faz uma distinção entre as noções de esquema (*schema*) e instância. Um esquema é uma coleção de definições de tipos (entidades e elos de aplicação) que descrevem uma aplicação em nível global; uma instância de um esquema é uma coleção de entidades, componentes, unidades e elos que satisfazem as definições do esquema. Essa característica pode permitir a reutilização do mesmo esquema para diferentes aplicações do mesmo domínio.

[Garzotto et al., 1995] distinguem as seguintes dimensões na análise de aplicações hiper-mídia: conteúdo, estrutura, apresentação, dinâmica e interação. O conteúdo endereça

as partes da informação, enquanto a estrutura é a organização do conteúdo. A apresentação define como o conteúdo da aplicação e suas funções serão mostradas aos usuários. A interação para o HDM é funcionalmente dinâmica operada nos elementos apresentacionais. Em outros métodos, a interação é considerada como parte da dinâmica e a apresentação como uma combinação dos dois fatores. As estruturas de acesso definidas pelo HDM são: índices, roteiros e coleções. Os Índices conectam um nó de coleção a cada membro da coleção. Um roteiro conecta nós de coleção em uma sequência linear com cada membro conectado ao próximo e ao anterior. Em coleções circulares, o último elemento se conecta ao primeiro. Coleções são índices ou roteiros que permitem referências cruzadas dos nós da coleção.

#### **4.4.3 *Object-Oriented Hypertext Design Method (OOHDM)***

O *Object-Oriented Hypermedia Design Method* (OODHM) [Rossi, 1996] é um método para construção de aplicações hipermídia, compreendendo quatro atividades diferentes: modelagem conceitual (ou análise de domínio), projeto navegacional, projeto da interface abstrata e implementação. Esse processo suporta uma mistura de estilos de desenvolvimento incremental, iterativo e baseado em protótipo. Cada atividade produz ou enriquece um modelo orientado a objetos enfocando um determinado interesse de projeto.

O principal objetivo do desenho conceitual é analisar a estrutura global da aplicação usando modelos orientados a objetos. Ele inclui o desenho de classes conceituais, hierarquias de generalização/especialização, agregação, instâncias de objeto, abstrações e subclasses. Durante essa etapa, poucas informações detalhadas sobre os usuários são consideradas.

Na etapa de desenho navegacional, os projetistas desenvolvem todas as estruturas navegacionais da aplicação que refletem os requisitos e tarefas dos usuários. Ele utiliza contextos navegacionais e notações de ligação para desempenhar navegação estruturada das classes da etapa anterior. Como em outros métodos, nós, ligações, índices e roteiros são usados para navegar entre nós. No OOHDM, os nós são considerados como esquemas conceituais.

No desenho navegacional, conjuntos diferentes de modelos navegacionais são desenvolvidos pelo mesmo desenho conceitual para expressar diferentes conjuntos de usuários e tarefas. Em outras palavras, caminhos de navegação diferentes podem ser construídos se a aplicação for usada por mais de um perfil de usuário.

Durante desenho navegacional, os projetistas consideram todas as maneiras possíveis de navegação requerida pelos diferentes usuários. Os modelos navegacionais podem herdar do desenho conceitual ou podem ser desenvolvidos independentemente. As ligações navegacionais podem ser estáticas ou dinâmicas. Quando estiver projetando ligações dinâmicas, o projetista tem que definir as transformações que ocorrem quando as ligações cruzarem. Uma vez que os desenhos conceituais e navegacionais estiverem completos, a estrutura da aplicação pode ser visualizada. Manutenção iterativa pode ser feita para corrigir erros dentro de nós e ligações.

O desenho de interface abstrata é a fase onde os modelos de interface são desenhados com texto e objetos multimídia, como imagens, animações e áudio. Objetos de interface são mapeados com objetos navegacionais para produzir a aparência perceptível. Pela integração dos objetos navegacionais e objetos de interface, os projetistas podem produzir muitos modelos de interface para o mesmo objeto navegacional para usabilidade.

O método usa cartões *Abstract Data View* (ADV) [Carneiro et al., 1994] dos modelos orientado a objetos para representar objetos de interface. ADV's são modelos formais, orientados a objetos, de objetos de interface (menus, campos de texto, botões, etc) que permitem especificar as metáforas de interface e descrever suas propriedades estáticas (relacionamentos com os objetos navegacionais) e dinâmicas (comportamento frente a eventos externos), de uma forma independente de implementação. Isto inclui o desenho de interface detalhado para cada objeto em percepções diferentes, incluindo sua aparência e eventos de usuários relacionados.

Na implementação, o produto do desenho navegacional e de interface pode ser implementado com ferramentas de desenvolvimento hipermídia atualmente disponíveis, como o

*Director*<sup>3</sup>, *toolbook* HTML, CGI<sup>4</sup>, etc.

## 4.5 Considerações Finais

Neste capítulo, foram discutidos alguns conceitos de Engenharia de *Software*, considerando também aplicações hipermídia e seus problemas, e foram apresentados algumas metodologias para projeto de aplicações hipermídia existentes na literatura. A navegação é um dos principais aspectos do paradigma hipermídia. As metodologias tradicionais de desenvolvimento de *software* não dão atenção especial a essa característica, inserindo seu projeto dentro da fase de desenho. Entretanto, os métodos direcionados para aplicações hipermídia quebram a fase de desenho em outras fase, onde uma delas é inteiramente dedicada ao projeto da estrutura navegacional do sistema.

Os métodos apresentados serviram de embasamento teórico para definição do método a ser adotado para o desenvolvimento da aplicação hipermídia que gerenciará o acervo digital deste trabalho. A descrição do método é apresentada no Capítulo 5.

---

<sup>3</sup>*Director* - *software* da *Macromedia*.

<sup>4</sup>CGI - *Common Gateway Interface*, é um conjunto de regras que descrevem como o servidor *Web* se comunica com outra parte do *software* na mesma máquina e como a outra parte do *software* (o programa CGI) se comunica com o servidor *Web*.

# Capítulo 5

## Metodologia de Desenvolvimento

Este capítulo apresenta a metodologia adotada para o projeto de digitalização do acervo do Cecor e do desenvolvimento do sistema para o gerenciamento deste acervo. A Seção 5.1 apresenta a instituição alvo deste trabalho, seu acervo e o processo de documentação que origina tal coleção. Seção 5.2 descreve os parâmetros, a infraestrutura e o *workflow* de digitalização desenvolvidos para esse projeto. A Seção 5.3 apresenta processo seguido para o desenvolvimento do sistema de informação hipermídia implementado para gerenciar o acervo digital.

### 5.1 Cecor

#### 5.1.1 O Acervo

O Cecor - Centro de Conservação e Restauração de Bens Culturais Móveis - é um órgão vinculado à Escola de Belas Artes (Figura 5.1) da Universidade Federal de Minas Gerais (UFMG). A instituição desenvolve pesquisas, estudos e trabalhos em pintura de cavalete, escultura policromada e obras sobre papel, além de atuar no ensino, na pesquisa, na preservação do patrimônio e conseqüentemente na formação de especialistas em Conservação/Restauração. A instituição restaura peças de instituições públicas e privadas, inclusive pessoas físicas. Muitas obras são atribuídas a artistas famosos, como

Cândido Portinari, Alberto da Veiga Guignard, Aleijadinho, Di Cavalcanti, entre outros [Nunes et al., 2002, Cecor, 2004].



Figura 5.1: Vista da Escola de Belas Artes onde se encontra o Cecor.

Durante a restauração, é executado um minucioso processo de documentação de todas as informações relativas às peças em restauração. Desde 1978, a instituição reúne um representativo acervo, composto em sua maioria de *dossiers*, fotografias e *slides* das várias técnicas fotográficas empregadas para o estudo das obras de arte. Esse acervo é considerado um dos mais importantes na América Latina sobre conservação e restauração de obras. A Figura 5.2 apresenta uma amostra do acervo: (a) uma pintura e algumas análises físicas pelas quais ela foi submetida; e (b) fichários, slides e cadernos científicos. A Figura 5.3 ilustra fotografias de obras que foram restauradas pelo Cecor: um quadro de Di Cavalcanti, uma amostra retirada para análise em laboratório e uma escultura.

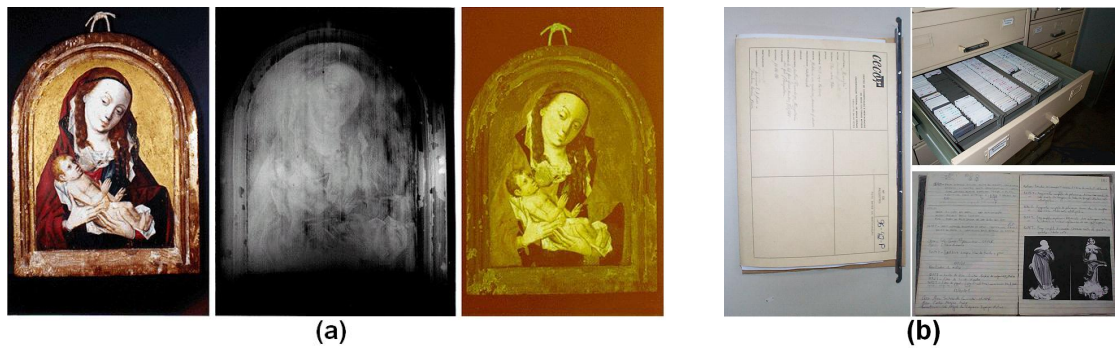


Figura 5.2: Amostra do acervo do Cedor: (a) pintura e duas análises feitas sobre ela (raio-x e raio ultravioleta); (b) tipos de documentos (fichários, *slides* e cadernos de análises).

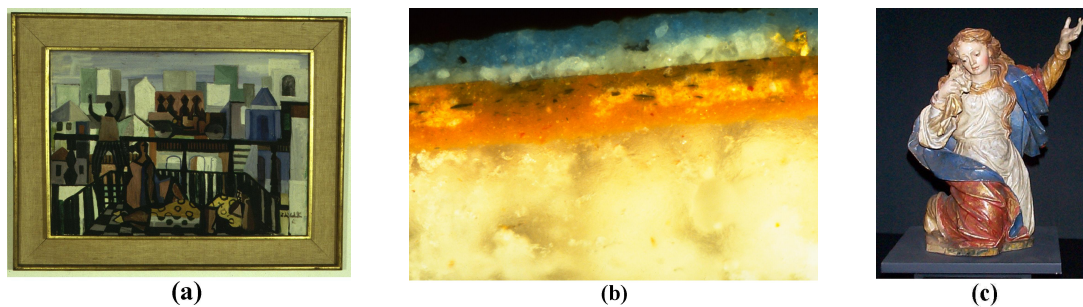


Figura 5.3: Amostra do acervo do Cedor: fotografias de um quadro de Di Cavalcanti, uma amostra extraída de uma obra e uma escultura.



### 5.1.2 Processo de Documentação

Inicialmente, a obra entra no Cecor e é expedido um recibo de entrada. Após a entrada, o estado de conservação de obra é avaliado para se fazer um orçamento do trabalho a ser executado. Esse orçamento é entregue ao proprietário da peça e, caso seja aprovado, os trabalhos de restauração se iniciam. Durante a restauração, vários dados são colhidos, como bibliografia, estado de conservação atual e as intervenções realizadas. Se for necessário, amostras são extraídas da obra para serem analisadas no laboratório da própria instituição. As análises também são todas documentadas. A Figura 5.4 apresenta o fluxo de trabalho simplificado do processo de restauração e documentação executado no Cecor.

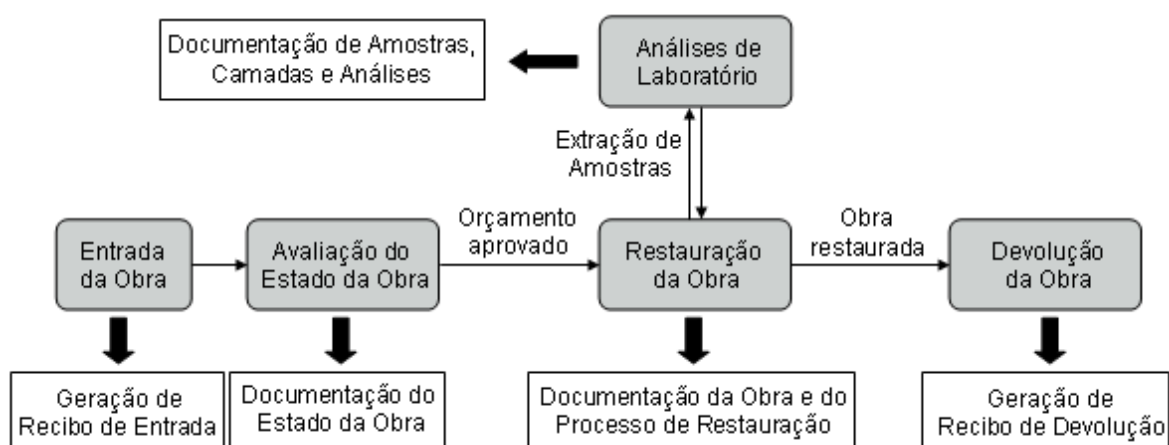


Figura 5.4: Processo de documentação.

## 5.2 Processo de Digitalização do Acervo

### 5.2.1 Parâmetros de Digitalização

Antes de se iniciar o processo de digitalização, é muito importante que a utilização dos documentos esteja definida, quer seja para simples visualização ou para um uso mais complexo. Além disso, deve-se considerar que muitos dos usos podem ainda não ter sido identificados.

Primeiramente, é necessário identificar os tipos de documentos que compõem o acervo da instituição. Os parâmetros podem variar dependendo do documento. Por exemplo, um documento de texto sem figuras inicialmente pode apresentar apenas as cores preto e branco, podendo ser digitalizado como uma imagem bitonal. Mas devido a alguma característica presente nesse documento, como manchas, ele pode ser capturado utilizando escala de cinza para capturar uma quantidade maior de detalhes. Por essa razão, os parâmetros de digitalização devem ser definidos de acordo com a natureza dos documentos. A Figura 5.5 exemplifica essa diferença. Muitos documentos especificam parâmetros de digitalização como os relatórios da *Cornell University* [Cornell, 2000] e do *Getty Institute* [Besser and Trant, 1995]. Tais documentos devem ser usados como guia, pois são documentos de uso geral. Mas para se obter um melhor resultado, as instituições devem executar testes até se alcançar parâmetros que atendam aos seus interesses. No acervo do Cedor, predominam fotografias e *slides*.

O custo mais expressivo no processo de reformatação de uma coleção consiste no esforço de remoção e preparação dos materiais para serem reformatados e sua posterior devolução aos seus depósitos permanentes [Valle Jr., 2003]. Para evitar que todo o processo de digitalização tenha que ser refeito devido aos usos que não foram antecipados, é gerada uma cópia digital que captura todos os detalhes do documento original. Esta imagem é chamada de imagem mestre ou de arquivamento. Tal imagem é capturada em altas resoluções, sem compressão e sem nenhum tipo de intervenção. As altas resoluções possibilitam que todos os detalhes presentes no documento seja capturados. A compressão de imagens adiciona algum tipo de complexidade ao documento digital e também algum tipo de uso futuro pode exigir que a imagem esteja no seu estado sem compressão. Por isso evita-se a utilização da compressão nas imagens mestres. Pelo mesmo motivo não são permitidas intervenções de nenhuma natureza nestas imagens, pois elas poderiam alterar a formatação dos dados do documento digital.

Quando a digitalização é feita através da cópia fotográfica em papel, deve-se considerar que resoluções menores são necessárias. De fato, a resolução de uma cópia fotográfica será sempre inversamente proporcional à sua ampliação, e às vezes menor ainda devido à óptica

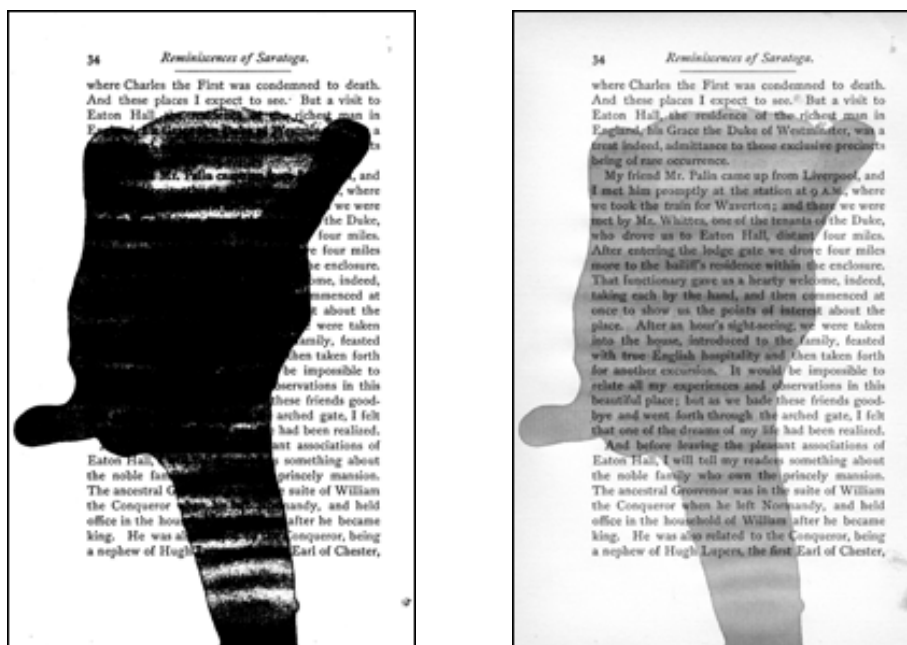


Figura 5.5: Exemplo de documento texto capturado em modo bitonal e em escala de cinzas. Uma mancha no documento ocultou detalhes que somente poderiam ser capturados em escala de cinza e não no modo bitonal (fonte: [Cornell, 2000]).

do ampliador. Existem algumas equações que permitem calcular a resolução ideal para esse tipo de documento. Algumas dessas equações podem ser vistas em [Valle Jr., 2003]. A resolução para documentos fotográficos de formato  $10 \times 15$  cm é algo em torno de 600 dpi.

A resolução de um negativo ou *slide* fotográfico é determinada por fatores complexos, que incluem o tipo de filme, as qualidades ópticas e mecânicas da câmera, e os cuidados e decisões no processamento do filme. Valores típicos situam-se em torno de 1000 a 4000 dpi.

A partir das imagens de arquivamento, são geradas cópias para atender ao tipo de utilização definido para o projeto. No caso do Cedor, principal utilização dos documentos é para simples visualização. As imagens, devido às altas resoluções e ausência de compressão, ocupam muito espaço em disco, o que torna inviável a sua visualização pela *Internet*. Essas cópias são chamadas imagens derivadas. Geralmente, essas imagens é que são disponibilizadas nos sistemas de informação. As imagens mestres são armazenadas em outras mídias,

como CD ou DVD.

### 5.2.2 Infraestrutura de *Hardware*

Em projetos de digitalização, uma escolha que deve ser feita é onde os documentos serão digitalizados: por terceiros (*outsourcing*) ou pela própria instituição (*in-house*) a quem eles pertencem. Ambas as abordagens possuem seus prós e contras que são apresentados por [Bishoff, 2003].

Se a instituição escolher a abordagem *in-house*, é necessário que ela tenha uma infraestrutura de *hardware* que permita a geração de documentos digitais com o nível de qualidade condizente com os objetivos do projeto.

Antes de se decidir quais seriam os parâmetros a serem utilizados para o Cecor, foi feito um levantamento dos equipamentos disponíveis no Cecor. Esse levantamento foi útil para determinar se a digitalização *in-house* era viável e se os documentos a serem gerados poderiam atender aos parâmetros de qualidade definidos.

Existem um *scanner ScanMaker 9800XL* [ScanMaker, 2004] da *Microtek* e uma câmera digital *Nikon D100* [Nikon D100, 2004] como equipamentos de aquisição. São excelentes equipamentos que permitem uma captura de alta qualidade, obedecendo aos parâmetros estipulados para este projeto. As Tabelas 5.1 e 5.2 apresentam alguns características destes dois equipamentos.

Dentre outros equipamentos existentes, o laboratório do Cecor dispõe de duas estações Macintosh, sendo um PowerMac G4, com gravadores de CD/DVD, um PowerBook G4 e um PC Pentium 4 2.4GHz. Para visualização, as máquinas estão equipadas com monitores de 17" e há uma impressora Epson Stylus Pro que possui uma resolução máxima de 2880 × 1440 dpi e um método de impressão que utiliza 7 cores diferentes.

### 5.2.3 *Workflow* de Digitalização

Após definir os parâmetros de digitalização, é necessário planejar e implantar um conjunto de atividades que devem ser seguidas de forma ordenada para gerar documentos digitais

Tabela 5.1: Características da câmera digital Nikon D100.

<i>Pixels</i> efetivos	6.1 MP
Sensibilidade	ISO equivalente de 200 a 1600
Armazenamento	Cartão CompactFlash <sup>TM</sup> (CF) (Tipo I/II) e Microdrive <sup>TM</sup> 512MB/1GB
Monitor de LCD	1.8 polegadas
Interface	USB
Formatos	TIFF/RAW (sem compressão) ou JPEG/RAW (com compressão)

Tabela 5.2: Características do *scanner* ScanMaker 9800XL.

Resolução ótica	1600 dpi
Resolução máxima	1600 × 3200 dpi (12800 × 12800 dpi interpolado)
Interfaces	FireWire, USB e SCSI
Área refletiva	12" × 17" (305 × 431.8 mm)
Modos de varredura	cor, tons de cinza e preto e branco em uma única passada; profundidade de cor em 48 <i>bits</i> ; profundidade de tons de cinza em 16 <i>bits</i> .
Software	ScanWizard Pro
Acessórios	Microtek TMA1600 – acessório opcional para capturar <i>slides</i> e transparências. Ele é montado no lugar da cobertura original do <i>scanner</i> .

de boa qualidade. A esse conjunto de atividades dá-se o nome de *workflow* ou fluxo de digitalização. Um *workflow* bem documentado permite fazer um controle de qualidade do processo e seus produtos. O fluxo de digitalização engloba procedimentos e técnicas para verificar a qualidade, a precisão e a consistência dos produtos finais [Cornell, 2000].

Embora o escaneamento automático seja frequentemente consistente, problemas com a exposição, alinhamento e balanço de cor são suficientes para requerer um componente de controle de qualidade. Sem controle de qualidade, não é possível garantir a integridade e a consistência dos arquivos digitais produzidos. Um fluxo de trabalho deve ser seguido para minimizar as variações entre diferentes operadores e diferentes dispositivos de captura. O operador e o dispositivo, bem como as configurações utilizadas, devem ser registradas para cada escaneamento de forma que seja possível corrigir problemas de imprecisão.

O *workflow* de digitalização proposto neste trabalho agrupa as atividades em quatro fases (ver Figura 5.6). Essas fases contemplam todo o processo que vai desde da remoção dos originais para o local de captura, a captura e armazenamento até a sua devolução e a geração das imagens derivadas.

### **Preparação**

Essa fase compreende tanto a preparação dos documentos para serem digitalizados como a preparação do ambiente de trabalho onde será feita a digitalização. Os documentos são selecionados e levados ao local de digitalização. Os dispositivos, como *scanners* e câmeras digitais, são preparados e calibrados. Se necessário, pode-se fazer testes para avaliar se a captura está sendo feita de acordo com o planejado.

### **Captura**

A captura consiste nas atividades para geração da imagem matriz a partir dos documentos originais. Nesta fase, são definidos todos os passos para se digitalizar um documento, como a forma que deve ser posto sobre o *scanner*, a configuração dos equipamentos, a avaliação do documento digital gerado, etc.

Uma vez que documento esteja colocado corretamente sobre o *scanner*, o operador

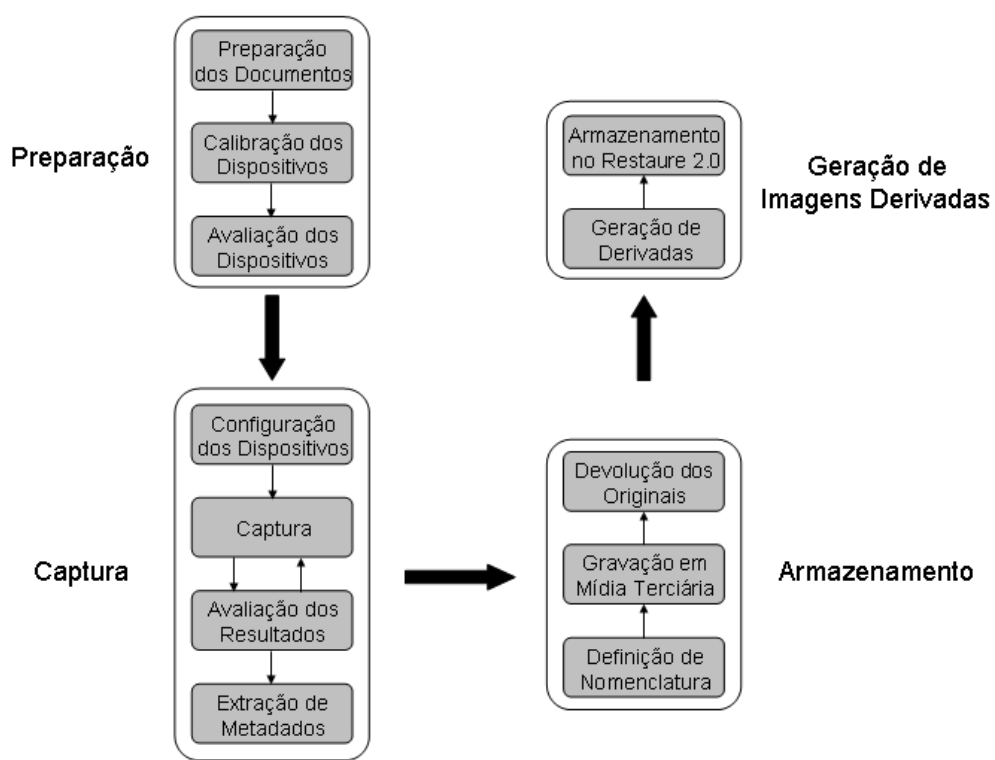


Figura 5.6: *Workflow* de digitalização.

configura o *software* de captura para fazer a captura na resolução e profundidade definidos, dependendo do documento (600 dpi para fotografia e 1600 dpi para *slides*, no caso do Cecor). O operador pode ainda fazer uma pré-visualização do documento antes de fazer a digitalização. Se tudo estiver correto, a captura é feita.

Entretanto, após a captura, a imagem matriz é avaliada por um operador e, somente se tiver cumprindo com todos os padrões definidos durante o planejamento, é aprovada. Para isso, é estabelecida uma lista de normas que reprovam uma matriz (ver Tabela 5.3). Por exemplo, se a imagem não for digitalizada na resolução correta ou há perda de detalhes por causa de sombras, ela é descartada e uma nova matriz é gerada. Somente quando ela cumpre com todas as normas ela é aprovada. A partir daí, um conjunto de metadados são extraídos e armazenados em banco de dados.

Tabela 5.3: Algumas normas que reprovam uma matriz digital.

No.	Normas
01	Imagem não tem o tamanho correto.
02	Imagem não tem a resolução correta.
03	Nome do arquivo está incorreto.
04	Formato do arquivo está incorreto.
05	Imagem está no modo correto (por exemplo, imagem colorida capturada como tons de cinza).
06	Perda de detalhe por luz excessiva ou sombras.
07	Ruído excessivo especialmente em áreas escuras ou sombreadas.
08	Imagem muito clara ou muito escura.
09	Valores de tons irregulares
10	Falta de definição/Definição excessiva.
11	Pixelização.
12	Presença de artefatos digitais (tais como muito regular, linhas retas transversais na figura).
13	Padrões de Moiré.
14	Balanço de cor incorreto.
15	Imagem sem variação tonal

## Armazenamento

Depois de geradas as imagens matrizes ou de arquivamento, é necessário armazená-las em algum lugar. As imagens mestres são gravadas em CD. Inicialmente, é importante nomear os arquivos seguindo um padrão. Dessa forma, torna-se mais fácil encontrar os documentos. A nomenclatura das pastas e dos CD's também é um outro aspecto que deve



ser padronizado. Todas as imagens matrizes são salvas em formato TIFF [TIFF, 1992] por ser o formato padrão adotado atualmente.

Depois que as imagens já estão capturadas e nomeadas corretamente, elas são gravadas em mídia terciária. As mídias de DVD possuem maior capacidade. Entretanto, ainda não se existe uma padronização. Vários formatos lutam por espaço no mercado. Isso pode por todo o trabalho a perder se for adotado um formato que pode deixar de ser usado futuramente. Neste caso, o CD se mostrou mais atraente por já estar consolidado no mercado. Cada matriz digital é armazenada em dois CD's que são guardados em locais diferentes para assegurar que um acidente não destrua os dados.

Os documentos originais também são devolvidos aos seus locais de origem. Caso os documentos, por algum motivo, não possam ser digitalizados, isso é registrado. Assim, é possível determinar quais documentos não foram digitalizados e qual foi o motivo.

### **Geração de Imagens Derivadas**

A fase trata da geração das imagens derivadas e seu armazenamento em banco de dados. As imagens derivadas são geradas a partir das matrizes. Como essas imagens são utilizadas para visualização na *Internet*, as imagens matrizes tornam-se inviáveis. Portanto, elas são comprimidas e a resolução delas é diminuída. Para a *Web*, a resolução utilizada é de 72 dpi. O formato de arquivo adotado é o JPEG [JPEG, 2004] porque é um padrão na *Web* e possui uma boa taxa de compressão. Uma vez geradas, as imagens derivadas são armazenadas em banco de dados juntamente com os metadados extraídos na fase de captura.

## **5.3 Desenvolvimento do Sistema de Informação Hiper-mídia**

Como foi visto no Capítulo 4, aplicações hipermídias possuem características que as diferenciam das aplicações convencionais. Devido a isso, muitas metodologias baseadas no paradigma hipermídia foram propostas na literatura. O interessante a ser notado é que não

existe um padrão seguido por essas metodologias. Cada uma delas possui suas próprias particularidades, notações e etapas. Algumas abrangem todo o ciclo de desenvolvimento. Outras dão atenção somente à modelagem da aplicação.

A metodologia seguida neste trabalho utilizou conceitos de várias metodologias, como o conceito de perspectiva do HDM e o fatiamento do RMM. O ciclo de desenvolvimento não foi todo explorado, sendo que a metodologia não contemplou a parte de planejamento e custos. Ela foi dividida nas seguintes etapas: levantamento de requisitos, desenho conceitual, desenho navegacional, desenho de interfaces de usuário e implementação.

### 5.3.1 Levantamento de Requisitos

A primeira fase de um projeto de desenvolvimento de *software* é identificar quais são as funções que a aplicação deve ter, isto é, os requisitos de *software*. Um sistema de *software* existe para servir aos seus usuários. Portanto, para construir um sistema de sucesso, deve-se saber o que os usuários querem e precisam [Jacobson et al., 1999]. Esta fase reúne as atividades que visam a obter o enunciado completo, claro, e preciso dos requisitos de um produto de *software*. Uma boa engenharia de requisitos é um passo essencial para o desenvolvimento de um produto, em qualquer caso [Paula Filho, 2001]. Os requisitos podem ser dos seguintes tipos:

- **Requisitos Explícitos:** requisitos descritos em um documento que arrola os requisitos de um produto;
- **Requisitos Normativos:** requisitos que decorrem de leis, regulamentos, padrões e outros tipos de normas a que o tipo de produto deve obedecer;
- **Requisitos Implícitos:** expectativas dos clientes e usuários, que são cobradas por esses, embora não-documentadas.

Um dos problemas básicos da engenharia de *software* é o levantamento e a documentação dos requisitos dos produtos de software. Quando esse levantamento é bem feito,

os requisitos implícitos são minimizados e os requisitos documentados têm maiores chances de serem corretamente entendidos pelos desenvolvedores [Paula Filho, 2001].

Projetos de *software* sofrem variações devido às diferenças entre o tipo de sistema utilizado, clientes, organização de desenvolvimento, tecnologia, etc. Devido a isso, a captura dos requisitos pode ter vários pontos de partida. Apesar dos diferentes pontos de partida, certas atividades são viáveis na maioria dos casos, como [Jacobson et al., 1999]:

- **Listar requisitos candidatos:** Uma lista é montada com todas as funções que forem consideradas relevantes por clientes e usuários. Esses requisitos são chamados candidatos porque nem todos são implementados na primeira versão do sistema por causa de fatores, como prazo e número de desenvolvedores disponíveis. Alguns podem ser deixados para futuras versões do sistema. Além dos requisitos, também podem ser descritas algumas características dos mesmos, como status, custo de implementação estimado, prioridade e risco.
- **Compreender o contexto do sistema:** Indica todo o conjunto de tarefas que determinam os aspectos relevantes do contexto em que o produto de *software* operará, como a definição do produto, a modelagem do processo de negócios e a engenharia de requisitos de sistema.
- **Capturar requisitos funcionais:** Uma abordagem simples para identificar os requisitos de sistema é baseada em casos de uso. Estes casos de uso capturam tanto os requisitos funcionais como os não-funcionais que são específicos aos casos de uso individuais. Um caso de uso descreve uma interação entre atores como uma sequência de mensagens entre o sistema e um ou mais atores. O termo ator inclui usuários humanos, bem como processos e sistemas de computador. Um caso de uso pode participar em vários relacionamentos, além da associação com atores (Tabela 5.4) [Rumbaugh et al., 1999].
- **Capturar requisitos não-funcionais:** Requisitos não-funcionais especificam propriedades do sistema, tais como limitações de ambiente e implementação, desempe-

no, dependências de plataforma, manutenibilidade e confiabilidade.

Tabela 5.4: Tipos de relacionamentos entre casos de uso.

Relacionamento	Função	Notação
Associação	Comunicação entre um ator e um caso de uso.	—
Extensão	Uma inclusão de comportamento adicional em um caso de uso base.	— «exten» →
Generalização	Um relacionamento entre um caso de uso geral e mais um específico que herda e adiciona características.	—→
Inclusão	Uma inclusão de comportamento adicional em um caso de uso base que descreve explicitamente a inclusão.	— «include» →

Dessa forma, o projeto do sistema para gerenciar o acervo digital do Cecor iniciou-se com entrevistas com o pessoal do Cecor, os usuários da aplicação. Ao fim das entrevistas, foi montada uma lista com todos os requisitos candidatos da aplicação com suas prioridades que é mostrada na Tabela 5.5. Os usuários identificados no levantamento são apresentados na Tabela 5.6.

A Figura 5.7 mostra o diagrama de casos de uso construído para o sistema do Cecor. Apesar de terem sido adiados, os casos de uso e atores para gerência de orçamento são mostrados.

### 5.3.2 Desenho Conceitual

Diagramas de caso de uso correspondem a visão do problema da ótica do cliente. Dessa forma, é necessário traduzir o diagrama para uma representação direcionada para os desenvolvedores da aplicação.

Durante essa fase, é construído um modelo conceitual que represente os elementos do problema. Geralmente, arquivos e bibliotecas possuem em seus acervos documentos nas mais diversas mídias, como vídeo, texto, fotografias, etc. O paradigma orientado a objeto fornece vários conceitos úteis para modelagem de dados multimídia, como encapsulação, herança, visões, entre outras [Khoshafian and Baker, 1996]. O modelo conceitual é constituído de objetos, relações e pacotes.

Tabela 5.5: Lista de requisitos candidatos.

No.	Caso de Uso	Descrição	Status	Prioridade
01	Gestão de Entrada de Obra	Processamento de inclusão, exclusão e alteração de obras.	Implementado	Essencial
02	Emissão de Recibo	Emissão opcional de recibo de entrada e saída da obra para o cliente.	Implementado	Essencial
03	Gestão de Restauração	Processamento de inclusão, exclusão e alteração de obras em restauração.	Implementado	Essencial
04	Emissão de Relatório de Restauração	Emissão opcional de relatório de restauração para o restaurador.	Adiado	Opcional
05	Gestão de Amostras	Processamento de inclusão, exclusão e alteração de amostras.	Implementado	Essencial
06	Emissão de Relatório de Amostra	Emissão opcional de relatório de amostra para o técnico do laboratório.	Adiado	Desejável
07	Gestão de Camadas	Processamento de inclusão, exclusão e alteração de camadas.	Implementado	Essencial
08	Emissão de Relatório de Camada	Emissão opcional de relatório de camada para o técnico do laboratório.	Adiado	Desejável
09	Gestão de Análises Extras	Processamento de inclusão, exclusão e alteração de Análises Extras.	Implementado	Essencial
10	Gestão Cromatografia	Processamento de inclusão, exclusão e alteração de Cromatografia.	Implementado	Essencial
11	Gestão de Estratigrafia	Processamento de inclusão, exclusão e alteração de Estratigrafia.	Implementado	Essencial
12	Gestão de FTIR	Processamento de inclusão, exclusão e alteração de FTIR.	Implementado	Essencial
13	Gestão de PLM	Processamento de inclusão, exclusão e alteração de PLM.	Implementado	Essencial
14	Gestão de Solubilidade	Processamento de inclusão, exclusão e alteração de solubilidade.	Implementado	Essencial
15	Gestão de Orçamento	Processamento de inclusão, exclusão e alteração de orçamento.	Adiado	Opcional
16	Emissão de Proposta	Emissão opcional de proposta de orçamento para o cliente.	Adiado	Opcional
17	Gestão de Usuários	Controle dos usuários que terão acesso ao sistema.	Implementado	Essencial
18	Consulta da Fotografias da Restauração	Consultas avançadas baseadas no conteúdo dos documentos.	Implementado	Desejável

Tabela 5.6: Descrição dos usuários.

No.	Ator	Descrição
01	Administrador	Funcionário responsável por administrar o sistema, como gerenciar usuários.
02	Técnico	Funcionário responsável pela gestão de amostras, camadas e análises.
03	Restaurador	Funcionário responsável pela gestão de restauração e de obra.
04	Secretário	Funcionário responsável pelo controle de orçamentos e emissão da proposta de orçamento.

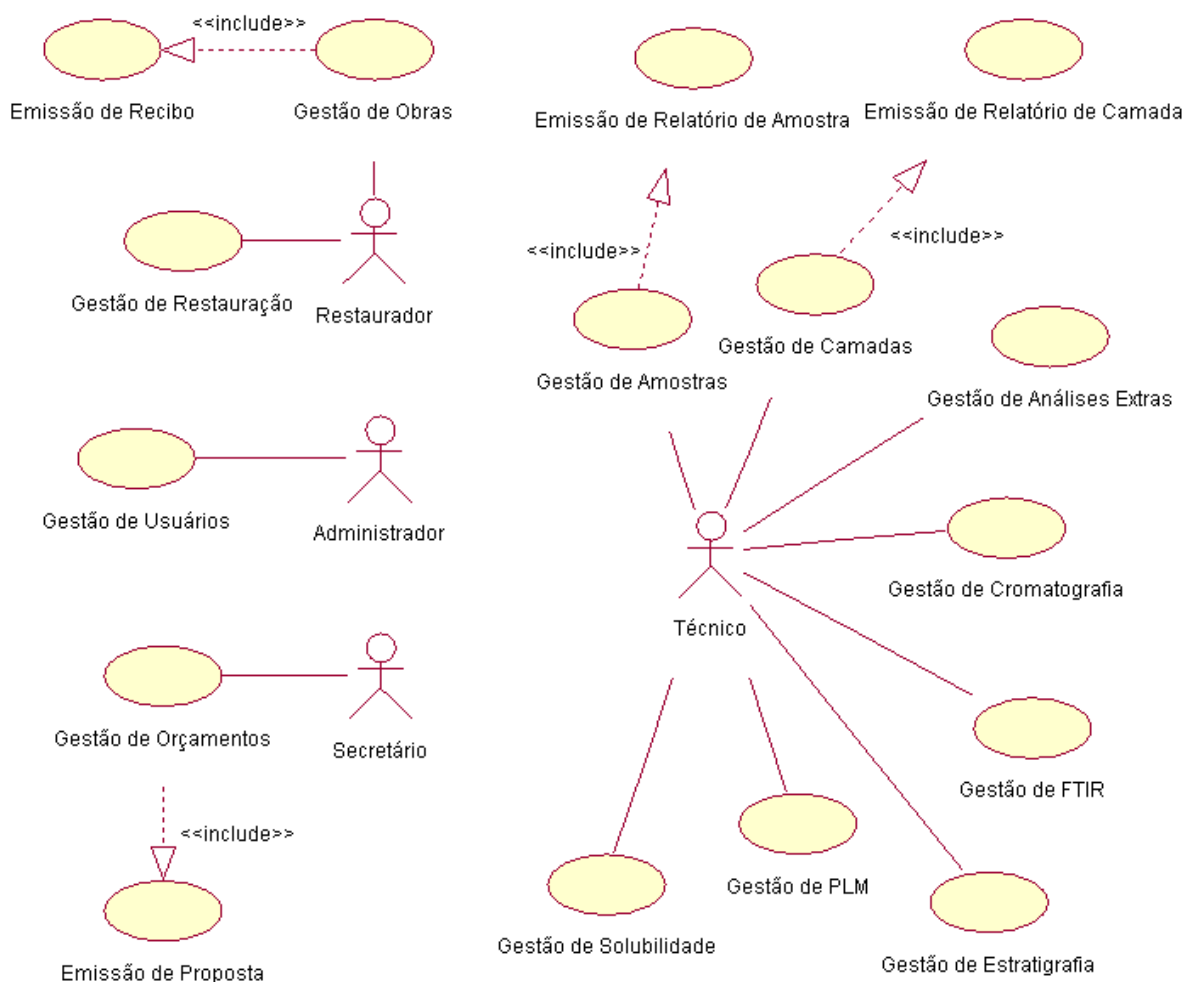


Figura 5.7: Diagrama de casos de uso do sistema do Cecor.

Classes são utilizadas para encapsular conceitos do domínio do problema. Os tipos de relacionamentos utilizados são os relacionamentos de associação, agregação e herança. As classes e outros elementos de modelagem de significado semântico semelhante são agrupados em pacotes.

Os atributos das classes podem ter várias formas de apresentação ou perspectivas, um conceito derivado do HDM [Garzotto et al., 1993] e do OODHM [Rossi, 1996] que amplia a metodologia orientada a objetos. A Figura 5.8 exemplifica o conceito de perspectiva. Nela, o atributo *descricao* da classe *Amostra* possui as perspectivas de texto e imagem.

<b>Amostra</b>
Id: inteiro; Data: data; Objetivo: texto; Descrição: texto, imagem;
inserir(); remover(); atualizar(); listaCamadas();

Figura 5.8: Exemplo de perspectiva: atributo *descricao* da classe *Amostra* pode ser representado como texto ou com uma imagem.

A identificação das classes pode ser feita considerando substantivos singulares do domínio do problema. A partir daí, pode-se determinar suas propriedades e operações. A Figura 5.9 ilustra o diagrama de classes para o Sistema Restaura 2.0. Para simplificar o diagrama, as propriedades e operações foram omitidas.

### 5.3.3 Desenho Navegacional

Os aplicativos hipermídia são projetados para efetuar navegação através de um espaço de informações. Por isto, o projeto da estrutura de navegação de tais aplicativos é a etapa crucial no empreendimento de desenvolvimento. Metodologias convencionais consideram que a estrutura navegacional é modelada na fase de desenho. Entretanto, metodologias de desenvolvimento hipermídia possuem uma fase totalmente dedicada a essa etapa. Durante o desenho navegacional, é projetada a estrutura navegacional da aplicação.

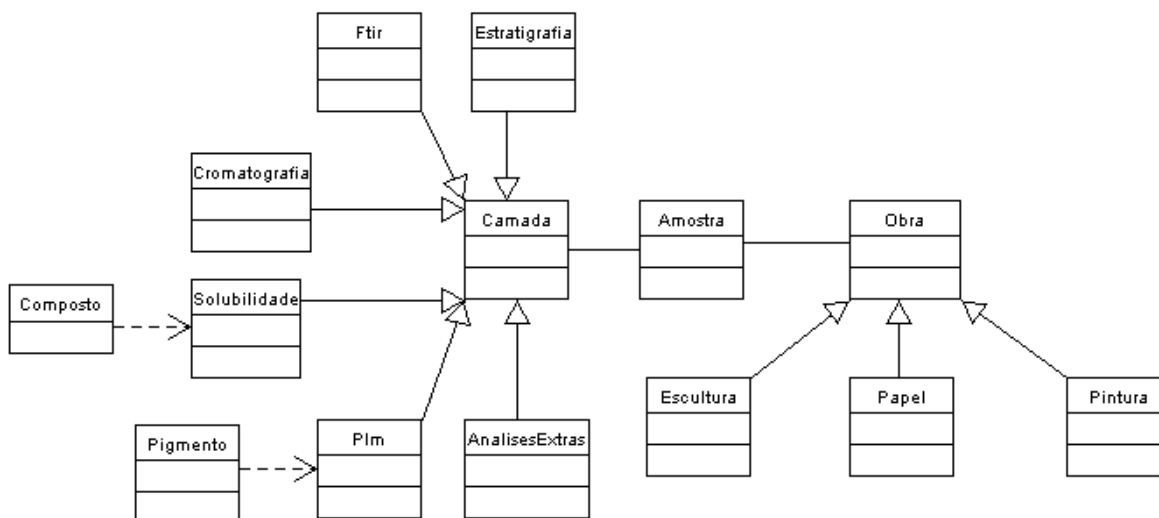


Figura 5.9: Diagrama de classes.

Aplicações hipermídia possuem estruturas chamadas nós e elos. Os nós contêm as informações que serão apresentadas ao usuário e os elos fazem a ligação entre os nós e somente podem ser unidirecionais [Pansanato, 1999].

Um conjunto de nós que desempenham uma função dentro da aplicação podem ser agrupados em contextos. Esses contextos permitem uma espécie de encapsulamento dentro do modelo navegacional. Os nós possuem atributos e funções. Os atributos especificam pedaços de informação que serão exibidos ao usuário. As funções são procedimentos que podem ser executadas dentro do nó. Os nós podem ser mapeamentos dos objetos conceituais ou podem ser partes desses objetos o que se assemelha ao conceito de fatia do RMM (*Relationship Management Methodology*) [Isakowitz et al., 1995]. Uma fatia é um subgrupo de atributos dentro de um objeto. Dois objetos conceituais também podem ser mapeados para um nó apenas.

Os elos podem ser modelados através de estruturas de acesso. As estruturas de acesso funcionam como índices ou dicionários e ajudam o usuário final a encontrar a informação desejada. As estruturas de acesso são caracterizadas por um conjunto de objetos alvo e um conjunto de destino. As principais estruturas de acesso são índices, roteiros e âncoras.

A Figura 5.10 apresenta um modelo navegacional construído para o Sistema Restaure



2.0 que gerencia as informações do laboratório do Cecor. Os retângulos com cantos arredondados são contextos ou conjuntos de nós, como o contexto *camadas* e os outros retângulos representam nós, como o nó *relatorioCamada*. A ligação entre os nós e contextos é feita por estruturas de acesso. No modelo da Figura 5.10, essas estruturas são representadas por âncoras e índices.

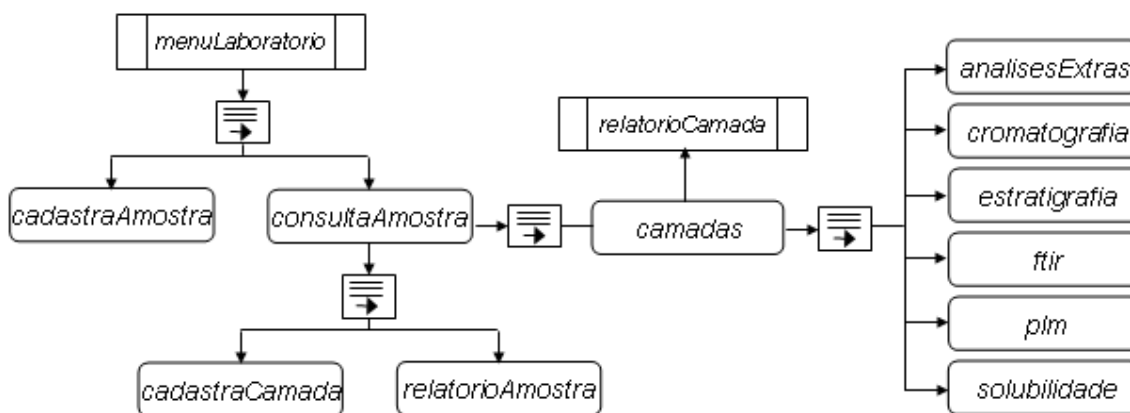


Figura 5.10: Diagrama navegacional para gerenciar os dados do laboratório.

### 5.3.4 Desenho de Interfaces de Usuário

Enquanto o projeto navegacional define qual é a estrutura de navegação de uma aplicação, a tarefa do projeto de interfaces é definir como esta estrutura navegacional é apresentada ao usuário [Baumeister et al., 1999]. Note que a mesma estrutura navegacional pode produzir diferentes projetos de apresentação dependendo, entre outras coisas, de restrições da plataforma alvo e da tecnologia utilizada.

O objetivo da fase de desenho de interface é gerar um modelo visual das interfaces que seja ainda independente do ambiente de implementação. Nesta fase, um modelo de apresentação das interfaces é construído. A maioria das metodologias para projeto hipermídia sugere o desenvolvimento de protótipos de páginas. Neste caso, o modelo construído nesta fase define um modelo de apresentação como uma composição de componentes de interface de usuário, como texto, âncoras, formulários, botões, vídeos, etc. O modelo de apresentação é uma interface de usuário esquemática que provê dicas de posicionamento, cor e tamanho

relativo dos componentes, mas não define a aparência final da página *Web*. Através desse modelo, pode-se tomar decisões a respeito das interfaces sem a necessidade de se desenvolver um protótipo para avaliá-las. A Figura 5.11 ilustra alguns componentes utilizados para compor os modelos.

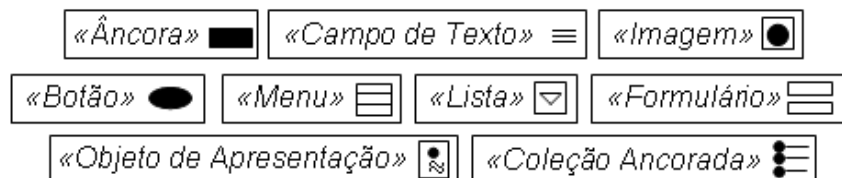


Figura 5.11: Alguns componentes de interface utilizados nos modelos de apresentação.

A Figura 5.12 ilustra o modelo de apresentação do formulário de amostras. Nesta figura, dois modelos de apresentação são combinados na implementação de uma página *Web*. Existe um modelo mais geral para a tela e um modelo para o formulário principal da página.

### 5.3.5 Implementação e Testes

Na implementação, os modelos navegacionais e de interface são traduzidos para artefatos do ambiente de implementação escolhido e os testes são realizados. Os testes verificam o resultado da implementação [Jacobson et al., 1999]. Eles são indispensáveis para avaliar o grau de qualidade de um produto de *software* e seus componentes [Paula Filho, 2001].

Os modelos foram traduzidos para a linguagem PHP (acrônimo de *PHP Hypertext Pre-processor*) [PHP, 2004], para funções executadas no servidor, e *JavaScript* [JavaScript, 2004], para funções executadas na máquina cliente. O Apache [Apache, 2004] é o servidor *Web* utilizado. A arquitetura da aplicação foi dividida em três camadas:

- **Persistência:** camada de interface entre o Sistema Restaura e o sistema gerenciador de banco de dados MySQL [MySQL, 2004].
- **Negócio:** trata as funções relacionadas com o negócio.

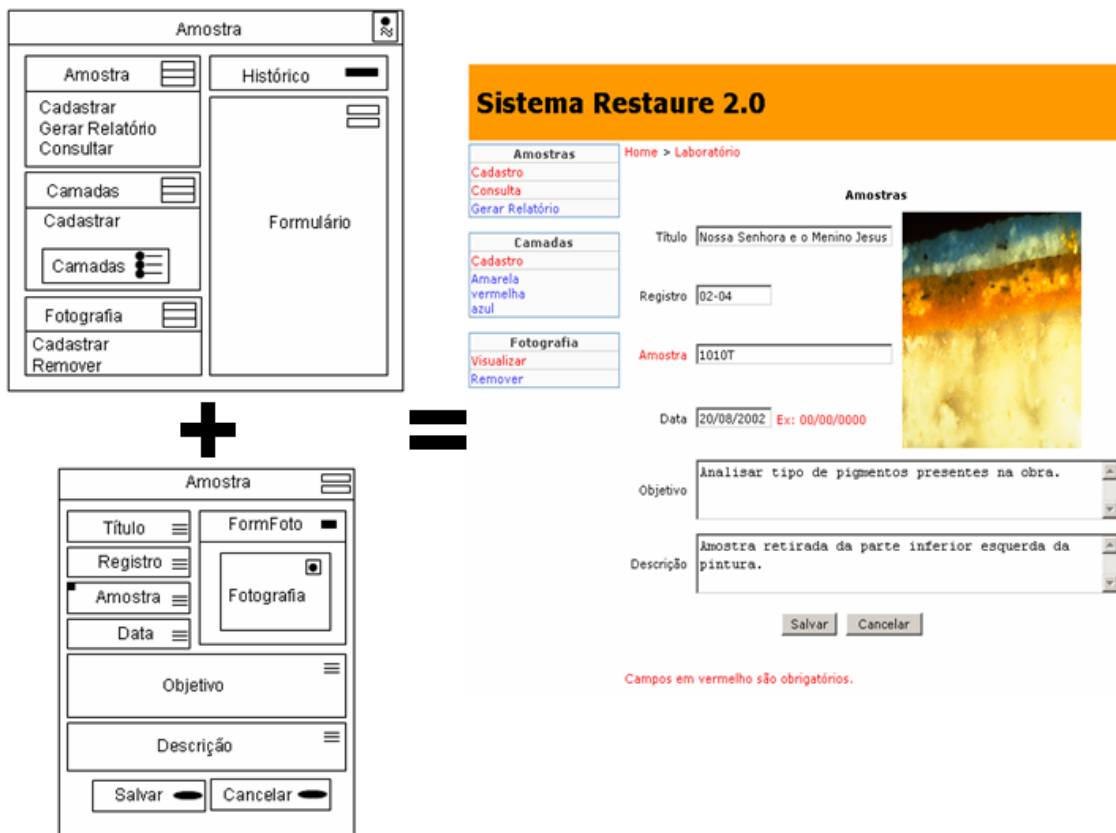


Figura 5.12: Modelo de apresentação do formulário de amostras.

- **Apresentação:** interface entre o Sistema Restaura e os usuários da aplicação. Esta camada foi desenvolvida utilizando a biblioteca *Smarty* [Smarty, 2004], para PHP.

A Figura 5.13 ilustra a arquitetura do sistema. As imagens derivadas, geradas a partir das imagens matrizes, são armazenadas no banco de dados MySQL.

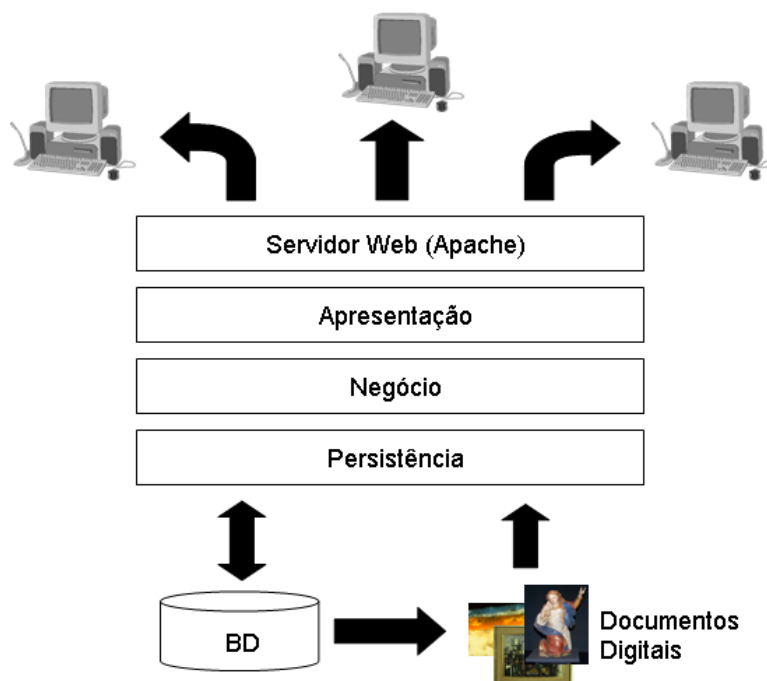


Figura 5.13: Arquitetura de três camadas do sistema Restaura.

As Figuras 5.14, 5.15, 5.16, 5.17 e 5.18 mostram algumas das telas do sistema implementado.

## Sistema Restaura 2.0

[logout](#)

Menu Principal
<a href="#">Laboratório</a>
<a href="#">Restauração</a>
<a href="#">Usuários</a>

Usuário administrador, bem vindo ao Sistema Restaura!

Esta é a versão 2.0. Este sistema é destinado ao cadastro e pesquisa de obras restauradas ou em restauração no Centro de Conservação e Restauração de Bens Culturais Móveis (Cecor).

Para cadastrar, atualizar e pesquisar dados sobre amostras, camadas e análises, clique no link Laboratório no menu ao lado.

Para cadastrar, atualizar e pesquisar obras, clique no link Restauração no menu ao lado.

Figura 5.14: Tela inicial.

## Sistema Restaura 2.0

[logout](#)

**Obras** [Home](#) > [Restauração](#) > [Obra](#) Usuário: administrador

**Cadastro**  
[Consulta](#)

**Restauração**  
[Bibliografia](#)  
[Tecnologia](#)  
[Estado de Conservação](#)  
[Intervenções Anteriores](#)  
[Proposta de Tratamento](#)  
[Intervenções no Cedor](#)  
[Slides](#)

**Fotografias**  
[Cadastro](#)  
[Ver](#)

**Fotografia**

**Registro** 79-02Pa **Número Provisório**

**Título** NOSSA SENHORA DO PATROCÍNIO

**Tipo** Escultura

**Fotografia**

**Informações sobre o Documento**

**Descrição**

**Autor do Documento**

**Dimensão**  **Tipo de Documento**

**Informações sobre a Digitalização**

**Formato da Matriz**  **Resolução da Matriz**

**Tamanho da Matriz**  **Profundidade de Bit**

**Formato da Derivada**  **Resolução da Derivada**

**Hardware**  **Software**

**Operador**  **Data**  Ex: 00/00/0000

**Localização da Matriz Digital**

**CD**  **Pasta**

Figura 5.15: Cadastro de documentos digitais e metadados técnicos.

## Sistema Restaure 2.0

[logout](#)

<b>Obras</b>
<a href="#">Cadastro</a>
<a href="#">Consulta</a>

<b>Restauração</b>
<a href="#">Bibliografia</a>
<a href="#">Tecnologia</a>
<a href="#">Estado de Conservação</a>
<a href="#">Intervenções Anteriores</a>
<a href="#">Proposta de Tratamento</a>
<a href="#">Intervenções no Cedor</a>
<a href="#">Slides</a>

<b>Fotografias</b>
<a href="#">Cadastro</a>
<a href="#">Ver</a>

[Home](#) > [Restauração](#) > [Obra](#)
















**Obra**

**N. Provisório**      **Registro** 79-04Mb      **Tipo** Escultura

**Título** SÃO MIGUEL ARCANJO

**Autor** ANÔNIMO

**Usuário:** administrador

Coleção de Fotografias				
Fotografia	Número	Descrição	Editar	Remover
	1			
	2			
	3			
	4			
	5			

PRIMEIRO
| 1 2 3 4 5 6 7 8 9 |
PRÓXIMO
ÚLTIMO

Figura 5.16: Lista de documentos digitais pertencentes a uma obra, no caso a obra de registro 79-04Mb.

## Sistema Restaura 2.0

[logout](#)

Home > Restauração Usuário: administrador

<b>Obras</b>
<a href="#">Cadastro</a>
<a href="#">Consulta</a>

<b>Recibos</b>
<a href="#">Entrada</a>
Devolução: - Restaurada
- Não Restaurada

### Entrada de Obras

**Número Provisório**

**Título**

**Autor**

**Tipo**  **Dimensões**

**Técnica**  **Época/Estilo**

**Proprietário**

**Nome**

**Rua**  **Número**

**Complemento**  **Bairro**

**Cidade**  **Estado**

**País**  **CEP**  Ex: 00000-000

**Telefone**  Ex: (00)0000-0000

**Recebida por**

**Prazo**

**Situação**

**Restaurador**

**Observação**

Campos em vermelho são obrigatórios.

Figura 5.17: Formulário de entrada de obra.

**Sistema Restaura 2.0** [logout](#)

Home > Laboratório Usuário: administrador

**Amostras**

[Cadastro](#)  
[Consulta](#)  
[Associar Amostra a Obra](#)  
[Gerar Relatório](#)

**Camadas**

[Cadastro](#)  
[Camada de sais.](#)  
[\[REMOVER\]](#)

**Fotografia**

[Visualizar](#)  
[Remover](#)

**Amostras**

**Título**

**Registro**

**Amostra** 1298T

**Data** 31/01/2002 Ex: 00/00/0000

**Objetivo**

Identificar os materiais constituintes da amostra.

**Descrição**

Retirada da parede três. Área interna da capela. Região superior-camada de tintas com sais.

Campos em vermelho são obrigatórios.

Figura 5.18: Formulário de amostras.

## 5.4 Considerações Finais

Este capítulo apresentou o acervo que foi objeto de estudo para este trabalho. O desenvolvimento do projeto seguiu duas linhas. O planejamento e implantação do processo de digitalização dos documentos e o desenvolvimento de um sistema de informação para gerenciamento dos documentos digitais e automatização da documentação do processo de restauração das obras.

O processo de digitalização dividiu-se em três etapas. A primeira etapa se concentrou na definição dos parâmetro de digitalização a serem seguidos. Esses parâmetros consideraram aqueles conceitos apresentados na Seção 2.1. Na segunda etapa, fez-se um levantamento dos recursos de *hardware* disponíveis no Cecor. Através desse estudo, optou-se pela digitalização *in-house* dos documentos, pois esses equipamentos permitiam a captura seguindo os parâmetros determinados na primeira etapa. Finalmente, o processo de digitalização e suas atividades foram definidas e implantadas.

O sistema de informação para gerenciar o acervo digital produzido foi desenvolvido para acesso *Web*. A metodologia de desenvolvimento utilizou conceitos de vários métodos de desenvolvimento hipermídia. Alguns desses métodos são apresentados na Seção 4.4.



# Capítulo 6

## Conclusão

A tecnologia digital apresenta características muito interessantes no que diz respeito à preservação e acesso de acervos permanentes. Entretanto, o uso da tecnologia deve ser feito com cuidado, acompanhado de um bom planejamento. A fragilidade das mídias digitais e a rápida obsolescência das tecnologias de *hardware* e *software* ameaçam a sobrevivência da informação no meio digital. Além de capturar todos os detalhes dos documentos, um plano de preservação deve ser definido para estender a vida útil dos documentos digitais.

Apesar de importante, este trabalho não propôs um plano de preservação. O refrescamento não resolve o problema da obsolescência da tecnologia digital e a migração altera a formatação, podendo provocar possíveis perdas de informação. As demais estratégias possuem custo bastante alto. De todas as estratégias apresentadas, a reconstrução (ver Seção 2.2.3) se mostrou a mais interessante. Com a manutenção de metadados de aplicação, torna-se possível reconstruir o ambiente que foi perdido podendo restabelecer o acesso ao acervo digital. Mas manter um formato de dados obsoleto pode se tornar bastante caro, se tornando até em um fator limitante ao acesso aos documentos. Dessa forma, um estudo mais aprofundado deveria ser feito para detectar prováveis problemas e soluções para contorná-los.

Apesar dos vários problemas da tecnologia digital, a sua adoção é bastante atraente. A possibilidade de entregar documentos de excelente qualidade aos usuários utilizando um

conjunto rico e variado de consultas é algo jamais imaginado no meio analógico. É importante notar que, muitas vezes, a tecnologia não vem substituir os documentos originais, mas sim ser um meio a mais para difusão do conteúdo desses documentos.

Como a *Internet* tende a expandir o número de usuários, ela se apresenta como um excelente meio para disponibilizar um acervo digital. O paradigma hipermídia, muito utilizado na Internet, também se mostra muito atraente. Ele permite que o usuário possa acessar informações em uma ordem não sequencial. Mas o desenvolvimento de aplicações hipermídia possui características que o diferenciam das aplicações tradicionais. A questão da navegação deve ser levada em consideração no processo de desenvolvimento. Também existem problemas específicos desse paradigma que precisam ser tratados, como a sobrecarga cognitiva, que aumenta o esforço de concentração dos usuários, e a desorientação, quando o usuário se sente perdido no “ciberespaço”.

Algumas metodologias de desenvolvimento de aplicações hipermídia foram citadas. Não existem estudos que indiquem qual é a melhor. Isso depende do tipo e da complexidade da aplicação que será desenvolvida. Algumas se baseiam na notação orientada a objetos (OO) enquanto outros se baseiam no modelo Entidade-Relacionamento (ER). A modelagem ER é mais simples do que a OO. Metodologias baseadas no modelo ER são indicadas para projetos mais simples, pois o modelo não abrange tipos de aplicações mais recentes, como aplicações que lidam com dados multimídia. Neste caso, são indicadas as metodologias orientadas a objeto.

A metodologia de desenvolvimento utilizada combinou os principais conceitos de vários métodos de desenvolvimento. Outro aspecto interessante dessas metodologias é que, embora tenham suas peculiaridades, muitas das etapas de desenvolvimento são bem semelhantes como uma fase para desenho da estrutura de navegação e para o desenho de interfaces para usuários. Entretanto, a metodologia seguida pode ser melhor formalizada e melhor especificada. Essa deficiência da metodologia pode ser deixada para trabalhos futuros onde, até os modelos podem ser aprofundados. É importante frisar que ela serviu bem aos propósitos da aplicação desenvolvida neste trabalho.

O fluxo de digitalização apresentado procurou ser uma forma de controle de qualidade

dos documentos gerados. Pelo fato de ser um processo rígido, o trabalho pode se tornar monótono para o operador e sujeito a falhas. O processo foi apenas especificado, sendo interessante automatizá-lo. Assim, pode-se melhorar a qualidade tanto dos documentos quanto do próprio processo, pois ele pode diminuir a taxa de erros por parte do operador.

O Sistema Restaure foi desenvolvido seguindo a metodologia apresentada. Esta aplicação foi desenvolvida para automatizar o gerenciamento do processo de documentação do Cecor. Por enquanto, apenas o módulo restrito aos profissionais do Cecor está disponível. Este módulo é dedicado à alimentação da base de dados e consultas por parte dos restauradores da instituição. Está em desenvolvimento o módulo de pesquisa para o público em geral.

# Bibliografia

- [Adobe, 2004] Adobe (2004). Adobe Systems Incorporated - Homepage. Disponível na WWW: <http://www.adobe.com> (31/mar/2004).
- [Andrade, 1998] Andrade, N. S. (1998). Sistemas de informação multimídia. Dissertação de Mestrado, Fundação João Pinheiro e Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.
- [Andrade and Araújo, 2000] Andrade, N. S. and Araújo, A. A. (2000). Multimídia para acesso a acervos históricos. *Revista iP-Informática Pública - PRODABEL*, 2(1):49–66.
- [Apache, 2004] Apache (2004). Welcome! The apache software foundation. Disponível na WWW: <http://www.apache.org> (03/dez/2004).
- [Apers et al., 1997] Apers, P. M. G., Blanken, H. M., and Houtsma, M. A. W. (1997). *Multimedia Databases in Perspective*. Springer.
- [APM, 1995] APM (1995). Arquivo público mineiro: Projeto de digitalização de imagens. Relatório Técnico, Prodemge, Belo Horizonte.
- [Aschenbrenner, 2001] Aschenbrenner, A. (2001). Long-term preservation of digital material - building an archive to preserve digital cultural heritage from the internet. Dissertação de Mestrado, Vienna University of Technology, Viena, Áustria, Disponível na WWW: [http://www.ifs.tuwien.ac.at/~andi/finished\\_thesis.html](http://www.ifs.tuwien.ac.at/~andi/finished_thesis.html) (18/mar/2004).
- [Balasubramanian, 1994] Balasubramanian, V. (1994). State of the art review on hypermedia issues and applications. Relatório Técnico, Rutgers University, New Jersey, USA. Available at [http://www.e-papyrus.com/hypertext\\_review/index.html](http://www.e-papyrus.com/hypertext_review/index.html) (01/jul/2004).
- [Barros, 2003] Barros, P. G. (2003). Realidade virtual & multimídia. Disponível na WWW: [http://www.cin.ufpe.br/~if124/mult\\_conceitos.htm](http://www.cin.ufpe.br/~if124/mult_conceitos.htm) (19/jul/2004).
- [Baumeister et al., 1999] Baumeister, H., Koch, N., and Mandel, L. (1999). Towards a UML extension for hypermedia design. In *Proceedings of UML'99*, Fort Collins, USA. Disponível na WWW:<http://www.pst.informatik.uni-muenchen.de/projekte/forsoft/pubs/uml99.pdf>.

- [Besser, 2000] Besser, H. (2000). Digital longevity. In *Handbook for Digital Projects: a Management Tool for Preservation and Access*, pages 155–166, Andover MA: Northeast Document Conservation Center, Disponível na WWW: <http://www.nedcc.org/digital/dighome.htm> (18/mar/2004). Maxine K. Sitts.
- [Besser and Trant, 1995] Besser, H. and Trant, J. (1995). Introduction to imaging: Issues in constructing an image database. Relatório Técnico, Getty Information Institute.
- [Bimbo, 1999] Bimbo, A. D. (1999). *Visual Information Retrieval*. Hardcover.
- [Bishoff, 2003] Bishoff, L. (2003). Digital imaging best practices. Relatório Técnico, Western States Digital Standards Group and Digital Imaging Working Group, Disponível na WWW: [http://www.cdpheritage.org/resource/scanning/documents/WSDIBP\\_v1.pdf](http://www.cdpheritage.org/resource/scanning/documents/WSDIBP_v1.pdf) (31/mar/2004).
- [Bloemers, 2001] Bloemers, R. (2001). Electronic and digital signatures. *The eBusiness Group*.
- [Byers, 2003] Byers, F. R. (2003). Care and handling of cds and dvds: A guide for librarians and archivists. Relatório Técnico, Council on Library and Information Resources and National Institute of Standards and Technology, Disponível na WWW: <http://www.clir.org/pubs/abstract/pub121abst.html> (13/abr/2004).
- [California, 2001] California (2001). Digital image format standards. Relatório Técnico, California Digital Library, Disponível na WWW: <http://www.cdlib.org/news/pdf/CDLImageStd-2001.pdf> (31/mar/2004).
- [Carneiro et al., 1994] Carneiro, L. M. F., Coffin, M. H., Cowan, D. D., and Lucena, C. J. P. (1994). ADVcharts: A visual formalism for highly interactive systems. In *M. D. Harrison and C. Johnson editor, Software Engineering in Human-Computer Interaction*, Cambridge University Press.
- [Carvalho et al., 2004] Carvalho, S., Cardozo, A., Lopes, A., Zandonadi, V., Lazarini, O., Reggiani, L., Costa, E., and Grego, M. (2004). O mapa do vídeo. *Revista InfoExame*, (214):36–64.
- [Cecor, 2004] Cecor (2004). Centro de Conservação e Restauração de Bens Culturais Móveis - Cecor. Disponível na WWW: <http://cecor.eba.ufmg.br> (15/jan/2004).
- [Christodoulou et al., 1998] Christodoulou, S. P., Styliaras, G. D., and Papatheodorou, T. S. (1998). Evaluation of hypermedia application development and management systems. In *Hypertext 98*, Pittsburgh, USA.
- [Cinque et al., 2001] Cinque, L., Ciocca, G., Leviardi, S., Pellicano, A., and Schettini, R. (2001). Color-based image retrieval using spatial-chromatic histogram. *Image and Vision Computing*, 19(13):979–986.

- [Conklin, 1987] Conklin, J. (1987). *Computer*, chapter Hypertext: An Introduction and Survey, pages 17–41. IEEE Computer Society Press.
- [Conway, 2001] Conway, P. (2001). Preservação no universo digital. Relatório Técnico, Conservação Preventiva em Bibliotecas e Arquivos e CLIR - Council on Library and Information Resources, Disponível na WWW: [si-arq02.siarq.unicamp.br/cpba/pdf\\_cadtec/52.pdf](http://si-arq02.siarq.unicamp.br/cpba/pdf_cadtec/52.pdf) (17/mar/2004).
- [Cornell, 2000] Cornell (2000). Moving theory into practice: Digital imaging tutorial. Relatório Técnico, Cornell University Library.
- [David, 1997] David, M. M. (1997). Multimedia databases. In *Database Programming & Design*, volume 10, pages 26–35.
- [Eakins and Graham, 1999] Eakins, J. P. and Graham, M. E. (1999). Content-based image retrieval. Relatório Técnico, JISC Technology Applications Programmer.
- [Elmasri and Navathe, 2000] Elmasri, R. and Navathe, S. B. (2000). *Fundamentals of Database Systems*. Addison - Wesley, 3a. Edição.
- [Garzotto et al., 1995] Garzotto, F., Mainetti, L., and Paolini, P. (1995). Hypermedia design analysis. In *Communications of the ACM*, volume 8, pages 74–86.
- [Garzotto et al., 1993] Garzotto, F., Paolini, P., and Schwabe, D. (1993). HDM - a model-based approach to hypertext application design. In *ACM Transactions on Information Systems*, volume 11, pages 1–26.
- [Ghafoor, 1995] Ghafoor, A. (1995). Multimedia database management systems. In *ACM Computing Surveys*, volume 27, pages 593–598.
- [GIFF, 2004] GIFF (2004). Graphic Interchange Format (GIF) - Homepage. Disponível na WWW: <http://www.serve.com/jb/gif.e.htm> (31/mar/2004).
- [Gonzalez and Woods, 1992] Gonzalez, R. C. and Woods, R. E. (1992). *Digital Image Processing*. Addison-Wesley Publishing Company.
- [Grosky, 1997] Grosky, W. I. (1997). Managing multimedia information in database systems. In *Communications of the ACM*, volume 40, pages 72–780.
- [Heminger and Robertson, 2000] Heminger, A. R. and Robertson, S. (2000). The digital rosetta stone: A model for maintaining long-term access to static digital documents. In *Commnications of the AIS*, volume 3, Atlanta, USA. Association for Information Systems.
- [Isakowitz et al., 1995] Isakowitz, T., Stohr, E., and Balasubramanian, P. (1995). RMM: A methodology for the design of structured hypermedia applications. In *Communications of the ACM*, volume 8, pages 34–44.

- [Jacobson et al., 1999] Jacobson, I., Booch, G., and Rumbaugh, J. (1999). *The Unified Software Development Process*. Addison-Wesley.
- [Jacobson et al., 1997] Jacobson, I., Christerson, M., Jonsson, P., and Övergaard, G. (1997). *Object-Oriented Software Engineering*. Addison-Wesley.
- [JavaScript, 2004] JavaScript (2004). Javascript.com(TM) - The definitive javascript resource: Javascript tutorials, free java scripts, source code and other scripting resources. Disponível na WWW: <http://www.w3schools.com/js/default.asp> (03/dez/2004).
- [JPEG, 2004] JPEG (2004). Joint Photographic Experts Group (JPEG) - Homepage. Disponível na WWW: <http://www.jpeg.org> (31/mar/2004).
- [Katzenbeisser, 2001] Katzenbeisser, S. (2001). On the design of copyright protection protocols for multimedia distribution using symmetric and public-key watermarking. In *Workshop Notes of the 12th International Workshop on Database and Expert Systems Applications, Fifth International Query Processing and Multimedia Issues in Distributed Systems Workshop (QPMIDS'2001)*, pages 815–819, IEEE Computer Press.
- [Khoshafian and Baker, 1996] Khoshafian, S. and Baker, A. B. (1996). *Multimedia and Imaging Databases*. Morgan Kaufmann Publishers, Inc.
- [Kyaw and Boldyreff, 1998] Kyaw, P. and Boldyreff, C. (1998). A survey of hypermedia design methods in the context of world wide web. Relatório Técnico, University of Durham, Disponível na WWW: <http://www.dur.ac.uk/liz.burd/tr.03-98.ps>.
- [Lahanier et al., 2002] Lahanier, C., Aitken, G., Shindo, J., Pillay, R., Martinez, K., and Lewis, P. (2002). Eros: An open source, multilingual research system for image content retrieval dedicated to conservation-restoration exchange between cultural institutions. In *Proceedings of the ICOM 13<sup>th</sup> Triennial Meeting*, pages 287–294, Rio de Janeiro, Brasil.
- [Lanzelotte et al., 1993] Lanzelotte, R. S. G., Marques, M. P., Penna, M. C. G., Portinari, J. C., Ruiz, I. D., and Schwabe, D. (1993). The portinari project: Science and art team up together to help cultural projects. In *Proceedings of the 2nd International Conference on Hypermedia and Interactivity in Museums (ICHIM'93)*, Cambridge, UK.
- [Laudon and Laudon, 1999] Laudon, K. C. and Laudon, J. P. (1999). *Sistemas de Informação com Internet*. LTC - Livros Técnicos e Científicos S.A., 4a. Edição.
- [Lopes et al., 2003] Lopes, A., Zandonadi, V., Machado, C., and Reggiani, L. (2003). Um superguia para você tirar de letra como guardar seus dados. *Revista InfoExame*, (205):37–63.
- [Maniatis et al., 2001] Maniatis, P., Giuli, T. J., and Baker, M. (2001). Enabling the long-term archival of signed documents through time stamping. Relatório Técnico, Computer Science Department, Stanford University, California, USA, Disponível na WWW: <http://www.arxiv.org/abs/cs.DC/0106058> (21/jun/2004).

- [Melo, 1988] Melo, R. N. (1988). *Bancos de Dados Não Convencionais: A Tecnologia do BD e suas Novas Áreas de Aplicação*. VI Escola de Computação, Campinas.
- [MySQL, 2004] MySQL (2004). Mysql: The world's most popular open source database. Disponível na WWW: <http://www.mysql.com> (05/fev/2004).
- [Nara, 1998] Nara (1998). Nara guidelines for digitizing archival materials for electronic access. Relatório Técnico, National Archives and Records Administration, Disponível na WWW: [http://www.archives.gov/research\\_room/arc/arc\\_info/guidelines\\_for\\_digitizing\\_archival\\_materials.pdf](http://www.archives.gov/research_room/arc/arc_info/guidelines_for_digitizing_archival_materials.pdf).
- [Niblack, 1986] Niblack, W. (1986). *An Introduction to Digital Image Processing*. Prentice-Hall International.
- [Nikon D100, 2004] Nikon D100 (2004). Câmera digital da Nikon. Disponível na WWW: <http://www.dpreview.com/reviews/nikond100/> (20/dez/2004).
- [Nunes et al., 2002] Nunes, F. H. C., Souza, L. A. C., Araújo, A. A., and Correa, M. A. (2002). Disponibilizando e preservando o acervo sobre conservação e restauração de bens culturais móveis do cecor. *Revista Eletrônica de Iniciação Científica (REIC)*, Disponível na WWW: <http://www.sbc.org.br/reic>, 2(3).
- [Pansanato, 1999] Pansanato, L. T. E. (1999). EHDM: Um método para o projeto de aplicações hipermídia para ensino. Dissertação de Mestrado, Universidade de São Paulo, São Carlos, Brasil. Disponível na WWW: [www.cp.cefetpr.br/luciano/public/dissertacao\\_mestrado.pdf](http://www.cp.cefetpr.br/luciano/public/dissertacao_mestrado.pdf) (05/mai/2004).
- [Paula Filho, 2000] Paula Filho, W. P. (2000). *Multimídia: Conceitos e Aplicações*. LTC - Livros Técnicos e Científicos Editora S.A.
- [Paula Filho, 2001] Paula Filho, W. P. (2001). *Engenharia de Software: Fundamentos, Métodos e Padrões*. LTC - Livros Técnicos e Científicos Editora S.A., 2a. Edição.
- [PDF, 2004] PDF (2004). Portable Document Format (PDF) - Homepage. Disponível na WWW: <http://partners.adobe.com/asn/tech/pdf/specifications.jsp> (31/mar/2004).
- [Pereira, 2004] Pereira, M. A. C. (2004). Introdução à internet. Homepage. Disponível na WWW: <http://www.marco.eng.br/historia.htm> (21/jun/2004).
- [PHP, 2004] PHP (2004). PHP: Hypertext Preprocessor. Disponível na WWW: <http://www.php.net> (03/dez/2004).
- [PNG, 2004] PNG (2004). Portable Network Graphics (PNG) - Homepage. Disponível na WWW: <http://www.libpng.org/pub/png/> (31/mar/2004).



- [Portinari, ] Portinari. Projeto Portinari - Homepage. Disponível na WWW: <http://www.portinari.org.br> (02/abr/2004).
- [Prabhakaran, 1997] Prabhakaran, B. (1997). *Multimedia Database Management Systems*. Kluwer Academic Publishers.
- [Rossi, 1996] Rossi, G. H. (1996). *Um Método Orientado a Objetos para o Projeto de Aplicações Hiperímida*. Tese, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brasil. Disponível na WWW: <http://www-lifia.info.unlp.edu.ar/~fer/oohdm/> (05/fev/2004).
- [Rothenberg, 1998] Rothenberg, J. (1998). Avoiding technological quicksand: Finding a viable technical foundation for digital preservation. Relatório Técnico, Council on Library and Information Resources, Disponível na WWW: <http://www.clir.org/pubs/reports/rothenberg/contents.html> (18/jun/2004).
- [Rui and Huang, 1999] Rui, Y. and Huang, T. S. (1999). Image retrieval: Current techniques, promising directions and open issues. *Journal of Visual Communication and Image Representation*, pages 39–62.
- [Rumbaugh et al., 1999] Rumbaugh, J., Jacobson, I., and Booch, G. (1999). *The Unified Modeling Language: Reference Manual*. Addison-Wesley.
- [Russell, 1999] Russell, K. (1999). Digital preservation: Ensuring access to digital materials into the future. Relatório Técnico, Disponível na WWW: <http://www.leeds.ac.uk/cedars/Chapter.htm> (18/jun/2004).
- [ScanMaker, 2004] ScanMaker (2004). *ScanMaker 9800XL: Scanner* de alta resolução da *Microtek*. Disponível na WWW: <http://www.microtekusa.com/sm9800xl.html> (20/dez/2004).
- [Siberschatz et al., 1999] Siberschatz, A., Korth, H. F., and Sudarshan, S. (1999). *Sistema de Banco de Dados*. São Paulo: Makron, 3a. Edição.
- [Silva, 1995] Silva, E. C. (1995). Sistemas de gerenciamento de documentos para centros de documentação e informação. Dissertação de Mestrado, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.
- [Smith, 1999] Smith, A. (1999). Why digitize? Relatório Técnico, Council on Library and Information Resources, Disponível na WWW: <http://www.clir.org/pubs/reports/pub80-smith/pub80.html>.
- [Sun, 2002] Sun (2002). Digital library technology trends. Relatório Técnico, Sun Microsystems, Disponível na WWW: [http://www.sun.com/products-n-solutions/edu/whitepapers/pdf/digital\\_library\\_trends.pdf](http://www.sun.com/products-n-solutions/edu/whitepapers/pdf/digital_library_trends.pdf).

- [Swetland, 2000] Swetland, A. J. G. (2000). Introduction to metadata: Setting the stage. Relatório Técnico, Getty Research Institute, Disponível na WWW: <http://www.getty.edu/research/institute/standards/intrometada> (15/jan/2004).
- [Tanenbaum, 1996] Tanenbaum, A. S. (1996). *Computer Networks*. Prentice-Hall, Inc., 3a. Edição.
- [Smarty, 2004] Smarty (2004). Smarty: Template engine. Disponível na WWW: <http://smarty.php.net/> (03/dez/2004).
- [TIFF, 1992] TIFF (1992). TIFF - revision 6.0. Relatório Técnico, Adobe Developers Association, Disponível na WWW: <http://partners.adobe.com/asn/developer/pdfs/tn/TIFF6.pdf> (31/mar/2004).
- [Valle Jr., 2003] Valle Jr., E. A. (2003). Sistemas de informação multimídia na preservação de acervos permanentes. Dissertação de Mestrado, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil. Disponível na WWW: <http://www.npdi.dcc.ufmg.br> (05/mai/2004).
- [Van Bogart, 1995] Van Bogart, J. W. C. (1995). Magnetic tapes storage and handling: A guide for libraries and archives. Relatório Técnico, National Media Laboratory, Disponível na WWW: [http://www.clir.org/pubs/reports/pub54/2what\\_wrong.html](http://www.clir.org/pubs/reports/pub54/2what_wrong.html) (14/abr/2004).
- [Waters and Garret, 1996] Waters, D. and Garret, J. (1996). Preserving digital information - task force on archiving of digital information. Relatório Técnico, The Commission on Preservation and Access and The Research Libraries Group, Disponível na WWW: <http://www.rlg.org/ArchTF/> (17/mar/2004).
- [Waugh et al., 2000] Waugh, A., Wilkinson, R., Hills, B., and Dell'oro, J. (2000). Preserving digital information forever. In *Proceedings of the Fifth ACM Conference on Digital Libraries*, pages 175–184, San Antonio, USA.
- [Wielinga et al., 2001] Wielinga, B. J., Schreiber, A. T., Wilemaker, J., and Sandberg, J. A. C. (2001). From thesaurus to ontology. In *K-CAP'01*, pages 194–201, Victoria, British Columbia, Canada. Disponível na WWW: <http://www.cs.vu.nl/~guus/papers/Wielinga01a.pdf>.