

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE ARTES

Mestrado em Artes

**VIDEOTECA MULTIMEIOS: EM DIREÇÃO A UM
BANCO DE DADOS
MULTIMÍDIA**

Maria Lúcia Figueiredo Fagundes

Este exemplar é a redação final da tese
defendida pela Sra. Maria Lúcia Figueiredo
Fagundes e aprovada pela Comissão Julgadora
em 28/10/1999

Prof. Dr. Gilberto dos Santos Prado

Dissertação apresentada ao Curso de Mestrado
em Artes do Instituto de Artes da UNICAMP
como requisito parcial para a obtenção do grau
de Mestre em Artes sob orientação do Prof. Dr.
Gilberto dos Santos Prado

Campinas - 1999



20002653

UNIDADE	BC
N.º CHAMADA:	T/UNICAMP
	F139v
V.	Ex.
TOMBO BC/	40353
PROC.	278/2000
C	☐
D	☒
PREÇO	R\$ 11,00
DATA	12/02/2000
N.º CPD	

CM-00133197-1

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA CENTRAL DA UNICAMP

F139v

Fagundes, Maria Lucia Figueiredo

Videoteca multimeios : em direção a um banco de dados multimídia / Maria Lucia Figueiredo Fagundes. -- Campinas, SP : [s.n.], 1999.

Orientador : Gilberto dos Santos Prado.

Dissertação (mestrado) - Universidade Estadual de Campinas , Instituto de Artes.

1. Sistemas multimídia. 2. Sistemas de recuperação da Informação. 3. Bibliotecas digitais. 4. Videodigital. 5. Banco de dados relacionais. 6. World Wide Web (Sistema de recuperação da informação). I. Prado, Gilberto dos Santos. II. Universidade Estadual de Campinas. Instituto de Artes. III. Título.

AGRADECIMENTOS

a Ricardo Mayer de Aquino, companheiro e colaborador indescritível.

a Eveline Borges, amiga e maior incentivadora.

a Diogo, Bio, Tati e Gabri, pela paciência e alegria.

a Maria Alice Filadelfo, pelo apoio infra-estrutural caseiro.

a Miriam Manini, pelas discussões e revisão.

a Everaldo Luís Silva, parceiro de trabalho e introdutor do ciberespaço em meu mundo pessoal.

a Daniel Roseno, parceiro de trabalho e personagem central na interface *Web* do banco de dados Videoteca Multimeios.

a Adilson Ruiz, pelo incentivo.

a Comissão Fulbright, pelo financiamento de uma bolsa de estudos que propiciou o aprofundamento do tema biblioteca digital através do estágio de um ano no *Center for Advanced Technology* da *New York University*, entre setembro de 1998 e agosto de 1999, participando do projeto em andamento *Second Avenue Online: The Yiddish Theater Digital Archives*.

Resumo

Este trabalho discorre sobre a implantação de recursos multimídia e interface *Web* no banco de dados desenvolvido para a coleção de vídeos da Videoteca Multimeios, pertencente ao Departamento de Multimeios do Instituto de Artes da UNICAMP. Neste percurso, localiza a discussão conceitual no universo das bibliotecas digitais, comenta e descreve técnicas necessárias à concretização do projeto e propõe alterações na configuração atual do banco de dados Videoteca Multimeios.

Abstract

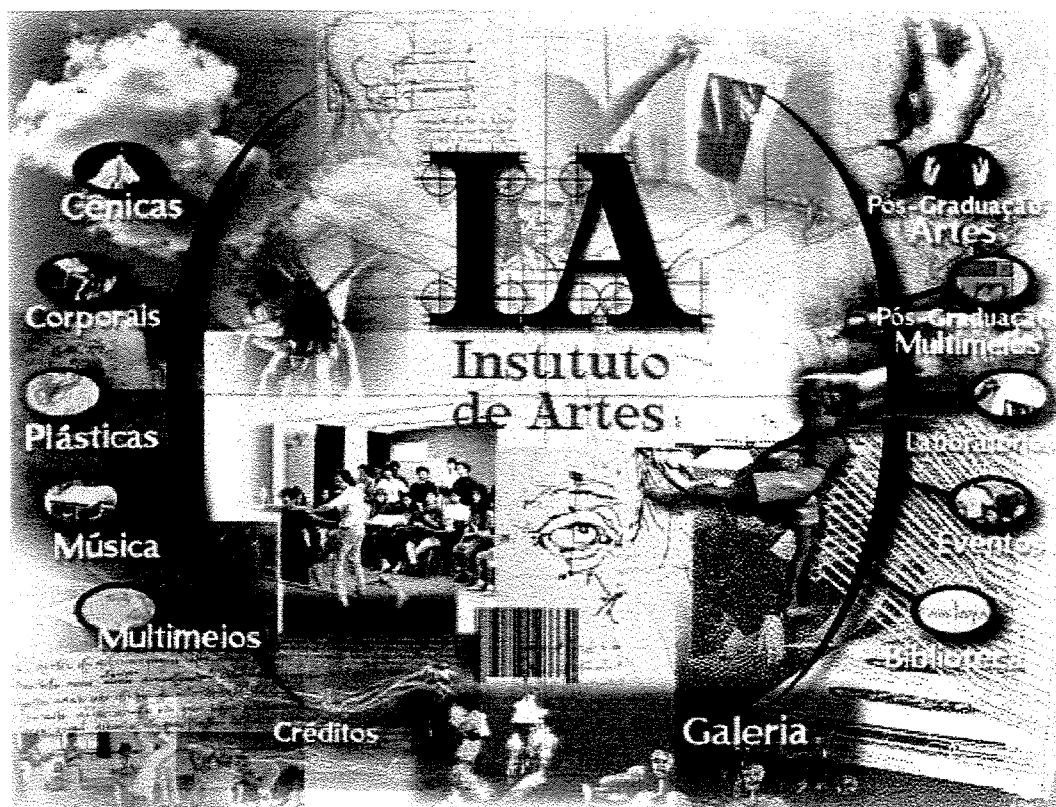
This work focus on the implementation of multimedia resources and a Web interface of the Videoteca Multimeios's video database. This videolibrary is located on the Multimedia Department of the Arts Institute of the State University of Campinas, UNICAMP. It spans issues concerning digital libraries, describes the path leading towards it's construction and finally suggests transformations aiming a better functionality of the database.

SUMÁRIO

Introdução	7
Objetivos	12
Capítulo 1	
Histórico da Videoteca Multimeios	15
Capítulo 2	
Videoteca Digital, Biblioteca Digital	24
2.1 Hipertexto, Multimídia, Hipermídia, Interatividade: novos paradigmas.....	24
2.2 Biblioteca Digital: definições e problemática	30
2.2.1. Arquitetura técnica.....	34
2.2.2. Construção de uma coleção digital	35
2.2.3. Digitalização	36
2.2.4. Metadado ou metadocumento	40
2.2.5. Nome e endereço.....	44
2.2.6. Direitos autorais	45
2.2.7. Preservação	45
2.3. Banco de Dados.....	45
Capítulo 3	
Videoteca Multimeios na Internet	53
3.1. Formatando a arquitetura.....	53
3.1.1. Estrutura do sistema : diagrama e descrição	53
3.1.2. Composição multimídia da base de dados	60
3.2. Implantando o projeto.....	60
3.2.1. Definição de um universo representativo para amostragem da pesquisa.....	60
3.2.2. Tecnologias usadas: apresentação e discussão.....	61
3.2.2.1 Banco de dados	61
3.2.2.2 Imagem fotográfica	66
3.2.2.3. Hemeroteca: tecnologia OCR	67
3.2.2.4. <i>Links</i>	67
3.2.2.5. Fragmentos de vídeos.....	67
3.3. Limitações/ Possibilidades para transformação.....	68
Considerações finais	73
Referências Bibliográficas	76

ILUSTRAÇÕES

<i>Home-page</i> do Instituto de Artes.....	6
Foto de Abel Gance	8
Foto do filme Napoleon	8
Foto de Luba Kadison Buloff.....	10
Foto de Annabelle Moore.....	11
Foto de Vannevar Bush	27
Desenho da máquina Memex	28
Diagrama do fluxo de dados do banco de dados Videoteca Multimeios.....	54
Página de rosto do <i>site</i> da Videoteca Multimeios.....	56
Ficha de busca do <i>site</i> da Videoteca Multimeios.....	57
Ficha com resultado resumido.....	58
Ficha completa de um registro em vídeo.....	59
Ficha ilustrativa da listagem de títulos no banco de dados Access.....	63
Ficha do procedimento de atualização de registro no banco de dados Access.....	64
Ficha de um registro no banco de dados Access.....	65
Foto do filme Corisco e Dadá e foto de Lampião e Maria Bonita.....	66
Ficha catalográfica de um <i>software</i> para museu.....	69
Ficha catalográfica de um <i>software</i> para museu.....	70
Ficha catalográfica de um <i>software</i> para museu.....	70
Ficha catalográfica de um <i>software</i> para museu.....	71
Ficha catalográfica de um <i>software</i> para museu.....	71



Home-page do Instituto de Artes

<http://www.ia.unicamp.br>

Rafael Peixoto Ferreira - Projeto Gráfico

Introdução

Há quatro anos, em janeiro de 1995, construir um *site* para disponibilizá-lo na então nascente *World Wide Web* era uma aventura considerável. Embora a face gráfica da Internet já estivesse bem conhecida, a construção de um *site* demandava o aprendizado da linguagem HTML (*Hypertext Mark-up Language*), usada para elaborar o hipertexto, e o conseqüente domínio de todos os seus comandos de formatação, ou *tags*, usados para definir o formato gráfico da página, das letras, inserir uma imagem, um arquivo sonoro ou um *link* para outro *site*. Árduos tempos. Nesta época, ao assumirmos a coordenação da *home-page* do Instituto de Artes da UNICAMP - e apesar das dificuldades iniciais - percebemos a potencialidade do novo meio referente à disponibilização, busca e recuperação da informação encarada como um todo. Com o passar dos meses, observamos que os sistemas hipermídia, presentes na *Web*, aprimoravam o grau de resolução, e diminuía gradativamente o tempo de espera para receber imagens. Passamos então a olhar o banco de dados da Videoteca Multimeios, que vínhamos organizando desde 1993, como um candidato a laboratório de experimentação em novas tecnologias de rede. Um banco de dados “estático”, contido em si mesmo, se assim podemos caracterizá-lo, com o recurso dos sistemas hipermídia, podia vislumbrar outros horizontes, expandir suas linhas de alcance, potencializadas pelas conexões oferecidas *online*.

Decididos a colocar na *Web*, não só a coleção de vídeos pertencente à Videoteca Multimeios como também inserir imagens e *links* de referência nas fichas catalográficas que acompanham seus registros, aprofundamo-nos no tema que resultou nesta pesquisa¹.

O ponto de partida para esse trabalho foi a memória: indagações a respeito do que tem sido escolhido como relevante pelas diversas culturas que mereça ser guardado, arquivado e colocado à disposição das comunidades. Hoje, a expansão da rede Internet possibilita essa estranha interconexão entre pessoas e lugares, dissolvendo nossa tão acomodada noção cultural de tempo e espaço. Temos através da *Web* um canal de comunicação com o enorme repositório de imagens fabricadas pela humanidade. Este imenso

¹ Pesquisa de mestrado interrompida em 1993 e agora retomada sob nova perspectiva.

arquivo visual nos chega como uma avalanche, não nos possibilitando processar devidamente tanta informação. Passamos rapidamente de uma imagem a outra, de um *site* a outro, de um texto a outro, como deglutidores visuais compulsivos, procurando ter um “idéia geral”, como se possível fosse, do conteúdo deste enorme acervo.

Na segunda década deste século, Abel Gance, teórico e realizador francês da área do cinema, anunciava para este meio de expressão, possibilidades que a técnica, então disponível, realizava com grandes limitações. Similar a tantas ocasiões visionárias em que a arte antecede em muito a tecnologia, Gance (*apud Agel, 1982:21*) diria, em 1927, no momento em que estava mergulhado no trabalho do qual resultaria seu épico *Napoléon*:



O diretor Abel Gance e o ator
Albert Dieudonné como Napoleão

“O tempo da imagem chegou... Todas as lendas, toda a mitologia e todos os mitos, todos os fundadores de religiões e as próprias religiões, todos os grandes vultos da História, todos os reflexos objetivos da imaginação dos povos desde milênios, todos, todos esperam sua ressurreição luminosa e os heróis se acotovelam em nossas portas para entrar. O tempo da imagem chegou!...”

Há algo de incrivelmente atual nas palavras de Abel Gance. Apesar de fazer referência exclusivamente ao cinema, é como se ele pudesse nos enxergar, hoje, conectados na rede mundial, descobrindo um manancial inesgotável de imagens, sendo testemunhas e compartilhando a maravilhosa produção visual da humanidade, deixando, literalmente, que personagens da História invadam nossas casas e nossas telas.

Foi em um desses momentos de verdadeira imersão imagética que decidimos sofisticar o banco de dados da Videoteca Multimeios. Por que não acrescentar imagens às informações da ficha técnica de um filme? Não é verdade que, ao olharmos uma cena fotografada de um filme que assistimos há tempos, ele, como por encanto, volta-nos à memória com detalhes e nuances? Além de imagens, por que não reunir reflexões criteriosas e abalizadas acerca de filmes, publicadas na imprensa e apresentá-las juntamente com o registro da informação de um filme? Como usar o recurso literário das “notas de rodapé”, que num livro remete a todo um conhecimento “paralelo”, acrescentando, enriquecendo o tema condutor, além de referenciar afirmações?

Em nossa realidade brasileira o acesso ao produto visual nunca foi fácil. Livros ilustrados são caros, freqüentar o cinema não é sinônimo de ter acesso a bons filmes, bibliotecas nem sempre são bem aparelhadas, arquivos e museus na eterna batalha por verbas para conservarem devidamente seus acervos...

A popularização do videocassete, por um lado, possibilitou maior acesso à filmes em geral. E, recentemente, temos a impressão de que entramos na dobra do tempo e milhares de personagens da História estão tendo sua “ressurreição luminosa”, como disse Gance, passando a fazer parte de nosso cotidiano.

Senão, como, por exemplo, conhecer a dança da borboleta de Annabelle Moore (embora digam ser uma imitação de Loie Fuller), filmada no estúdio de Thomas Alva Edison, o *Black Maria*, em West Orange, New Jersey, em 1895, e disponível no *site* da Biblioteca do Congresso? Ou como partilhar com Luba Kadison Buloff, velha dama do teatro Yiddish da Europa e

Estados Unidos, suas memórias e aventuras senão através de seu depoimento em vídeo e vasta coleção de fotos disponíveis na *Web*?²



Luba Kadison in *The Wich* by Anton Chekhov. Israel, 1967.

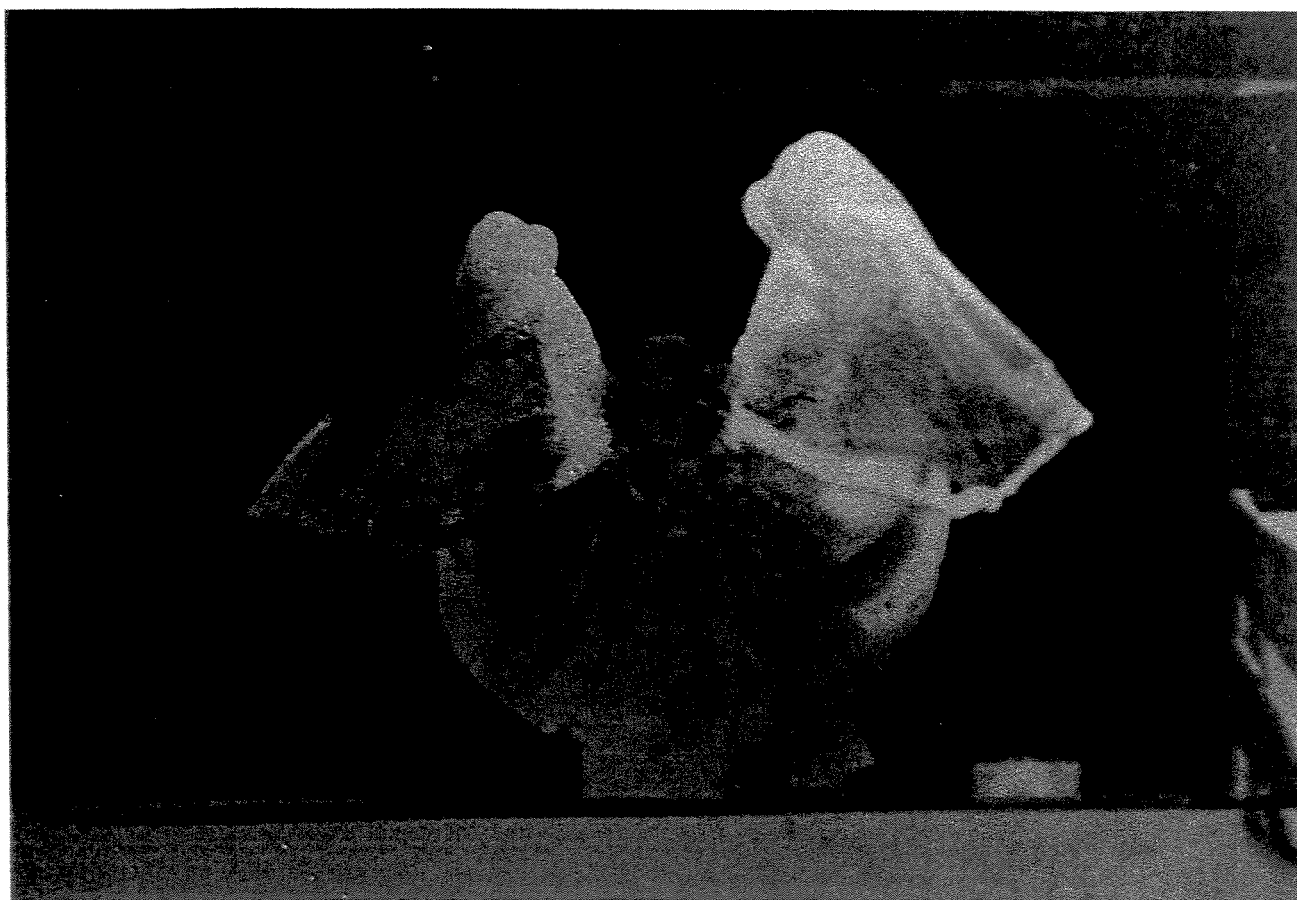
Parece precoce fazer uma avaliação do impacto desta revolução sobre nós e sobre como estamos adquirindo o conhecimento. Este tema, entretanto, não é matéria de nossas indagações neste trabalho.

Enquanto isso, tentemos contribuir para a disseminação de informações no espaço que nos cabe, neste caso, a Videoteca Multimeios, e, através de sua coleção de vídeos, agora ligados a documentos *Web*, providenciar para que nossos usuários possam mergulhar um pouco mais no mundo imagético, viajar através do tempo e compartilhar com atores e realizadores do cinema, suas visões de mundo, suas histórias locais, suas preocupações e conquistas.

² *Second Avenue Online: The Yiddish Theater Digital Archives Project*. Projeto desenvolvido no Center for Advanced Technology da New York University.

Nossa proposição é relatar como operacionalizamos a mudança de um banco de dados de registros catalográficos de vídeos para um *site* na *Web* com informações ampliadas.

Sabemos que as transformações na área digital têm ocorrido numa velocidade quase inimaginável e, por esta razão, uma apresentação como a que propomos agora corre o risco de se tornar obsoleta em pouco tempo. Entretanto, o produto concreto desta discussão, uma videoteca organizada em formato banco de dados em rede Internet, ainda não é coisa comum, pelo menos no Brasil.



Annabelle Serpentine Dance

1895. Edison Manufacturing Company. Distributed by Maguire and Baucus and their Continental Commerce Company. Photographed by W.K.L. Dickson and William Heise. With Annabelle Moore. Shot at Edison's Black Maria studio in West Orange, New Jersey, in early 1895. Print source: Museum of Modern Art. 27 sec.

Objetivos

Dividimos nosso trabalho em dois momentos: Na primeira fase da pesquisa não contávamos ainda em explorar o conceito de multimídia no banco de dados. Apesar da inserção de imagens e de estarmos rompendo com a predominância da linguagem escrita na forma de apresentação de informações em banco de dados, tínhamos consciência de não estar construindo um banco de dados multimídia, pois o sistema projetado não possibilita uma recuperação multimídia da informação.

No decorrer do trabalho, entretanto, avanços tecnológicos na área da Computação têm apontado para a integração de bases de dados, compartilhamento de acervos e busca integrada da informação em diferentes repositórios. Grandes mudanças na área de bibliotecas e museus digitais trazem à tona possibilidades antes não vislumbradas, fazendo-nos questionar o sistema por nós construído. Estamos, no momento, fazendo face à possibilidade de transformar nosso sistema, e, para isso, temos buscado referências em projetos em andamento, fato que tem colocado novos questionamentos e nos levado em direção ao conceito de biblioteca digital.

Coloquemos, então, distintamente, os objetivos de cada fase.

Na primeira fase, buscamos:

1. Transportar para a interface gráfica da Internet um banco de dados considerado tradicional na sua forma de apresentação da informação;
2. Inserir informações complementares à ficha técnica de um filme aproveitando as possibilidades técnicas oferecidas pelo sistema de redes remotas. Como resultado, ampliariamos as possibilidades de consulta e pesquisa do acervo de filmes da videoteca em questão.

Para operacionalizar estas metas tivemos que lidar com problemas de três ordens:

- conceitual:
 - na definição do conceito de informação;
 - na escolha de como representar a informação para que seja recuperável e que porção dela seria recuperável.

- técnica:
como seria feito o transporte para a interface *Web*;
como seriam usadas as técnicas de digitalização para a transformação de textos, fotos e vídeos em objetos digitais;
como seria construído um mecanismo de busca para a recuperação da informação.
- de pesquisa:
onde buscaríamos informação adicional a ser acrescentada nas sinopses dos vídeos.

Na segunda fase da pesquisa, temos lidado com a problemática da integração de diferentes bases de dados e padronizações necessárias para efetuar-las. Embora ainda não tenhamos efetuado a mudança do banco de dados para o formato multimídia - o que ocorrerá na sequência da apresentação deste trabalho - temos feito face às seguintes questões:

1. Como a área da Informação está respondendo aos problemas levantados pela aplicação da tecnologia de integração de bases?
2. Existem projetos em implantação?

No decorrer desta demonstração, buscaremos comentar todas estas questões.

Assim sendo, nosso tema se desenvolverá da seguinte maneira:

O capítulo 1 – Histórico da Videoteca Multimeios - traça a trajetória desta videoteca bem como a organização de sua informação. Aborda como pano de fundo a organização geral do arquivo Multimeios dentro da especificidade de seus documentos visuais. Em seguida, discorre sobre a organização das informações da videoteca em banco de dados, colocando os problemas encontrados e as soluções viáveis.

O capítulo 2 – Videoteca Digital, Biblioteca Digital – contextualiza e operacionaliza estes termos. É no ambiente da Internet que se desenvolve nossa pesquisa com o banco de dados da Videoteca Multimeios. Discutimos aqui a instauração de novos paradigmas através dos elementos que concretizam a comunicação por um sistema de redes remotas, quais sejam: o hipertexto, a multimídia, a hipermídia e a interatividade.

Neste percurso buscamos as raízes históricas das pesquisas que viabilizaram o que hoje conhecemos por Internet.

Comentamos o conceito de banco de dados levantando os novos elementos encontrados no contexto de redes remotas.

Trazemos a conceituação e discussão de uma biblioteca digital por ser este o formato mais aproximado de nosso objeto de pesquisa. Temos a oportunidade aí de discorrer sobre os grandes problemas operacionais gerais enfrentados pelas bibliotecas digitais, a maioria dos quais refletidos em nossa prática diária no percurso da pesquisa.

No capítulo 3 - Videoteca Multimeios na Internet – relatamos nossa experiência prática na formatação *Web* do banco de dados.

Baseados na colocação conceitual e teórica feita nos capítulos anteriores, aqui discorreremos sobre a operacionalização dos elementos que compõem a nossa videoteca digital.

Descrevemos a inserção do banco de dados da Videoteca Multimeios num sistema de redes remotas, fornecendo a especificação técnica e discorrendo sobre a forma de manutenção e atualização da base de dados.

Descrevemos também os registros provenientes de diferentes mídias que compõem o banco de dados da videoteca Multimeios e quais títulos em vídeo foram escolhidos para o projeto-piloto.

Na descrição das tecnologias implementadas, falamos primeiramente sobre o processo técnico de transposição de um banco de dados não relacional para um relacional e os ajustes sofridos para sua inserção num sistema de redes remotas. Na sequência, falamos sobre aquisição e tratamento da imagem, fotográfica e em movimento (vídeo), e sobre a aplicação de tecnologia OCR sobre textos.

Finalmente, dentro das possibilidades técnicas atuais, apresentamos as modificações a serem feitas na arquitetura do sistema para adaptá-lo ao formato multimídia.

Capítulo I

Histórico da Videoteca Multimeios

O Arquivo Multimeios, ligado ao Departamento de Multimeios do Instituto de Artes da UNICAMP, foi implantado no final de 1991. Na época contava com um acervo de 100 títulos em vídeo no formato VHS, 10 em Betacam, 7 em Umatic e 5 filmes em 35mm. Além disso, contava com um pequeno conjunto de fotografias. Mais tarde foi acrescido de roteiros para vídeo e TV e pequena hemeroteca sobre cinema. O Arquivo Multimeios tem como finalidade subsidiar e dar apoio didático às disciplinas dos cursos de pós-graduação em Multimeios e Artes, além de promover um atendimento à comunidade acadêmica, emprestando material e trocando informações com outras instituições e órgãos educacionais.

Sua videoteca abriga hoje cerca de 1200 (um mil e duzentos) títulos no formato VHS, cujos temas cobrem diferentes áreas de interesse dos cursos: documentário, ficção, videoarte, produções do departamento de Multimeios e de outros departamentos do Instituto de Artes. Este acervo, já considerado bastante representativo da cinematografia internacional, vem crescendo consideravelmente graças ao financiamento de projetos ligados a disciplinas teóricas da área de cinema dos cursos de pós-graduação do Instituto de Artes, bem como através de doações em forma de cópias em vídeo por parte de pessoas físicas e instituições culturais.

Na coleção fotográfica existem mais de 2000 negativos, cerca de 600 fotos e 180 diapositivos, acondicionados devidamente e parcialmente organizados.

Os roteiros para cinema, vídeo e TV são produzidos por alunos da disciplina Teoria e Prática em Roteirização do curso de mestrado em Multimeios. Estão disponíveis para consulta em banco de dados no software Works for Windows.

A pequena hemeroteca é formada por recortes dos jornais *O Estado de São Paulo* e *Folha de São Paulo*, dos anos 1994, 1995 e 1996, tendo o cinema como tema principal. Está parcialmente organizada.

Filme, vídeo e som são, para a área de documentação, considerados materiais “cegos”, ou seja, materiais que não mostram de imediato seu conteúdo. Necessitam da intermediação de uma máquina para terem seu conteúdo desvelado e a sua apresentação é feita em tempo real. É por esta razão que exigem por parte do indexador uma descrição mais detalhada ao serem processados documentalmente, principalmente no que se refere ao conteúdo. Oferecer ao pesquisador uma gama de informações bastante consistente minimiza o trabalho de percorrer em tempo real todo o material. Informações consistentes significam escolhas por parte do indexador, escolhas pessoais que devem ser baseadas em conceitos teóricos que definam e restrinjam a especificidade do material visual com relação à sua organização.

Dentro da classificação biblioteconômica, o material visual se diferencia do livro, tem muitas especificidades e apresenta dificuldades para sua organização. No Brasil, material visual não conta com um vocabulário controlado e, no geral, não existem normas universais standardizadas para sua organização. Recentemente a *Library of Congress* nos Estados Unidos, e a *British Library* na Inglaterra têm sido consideradas centros de referência internacionais para a reunião e discussão de convenções na documentação do material audiovisual. Até então, construíam-se vocabulários especializados para organizarem localmente coleções específicas, como é o caso do *National Film Archive* (USA), coleções de filmes da *Library of Congress* (USA) ou da *British Film Institute* (Inglaterra). Nos Estados Unidos, a catalogação da imagem tem sido feita usando-se tradicionalmente como padrão de vocabulário controlado o *The Library of Congress Subject Headings* (LCSH), desenvolvido para material textual. Com a sofisticação das bases de dados em imagem percebeu-se as limitações desta categorização, sendo desenvolvidos então dois tesouros específicos para materiais gráficos: o *Art and Architecture Thesaurus* (AAT) e o *Library of Congress Thesaurus for Graphic Materials* (LCTGM). Por esta razão, muitas linhas diretivas para o trato da informação visual nos chegam em inglês, o que dificulta a padronização de termos em português.

Em nosso caso específico, nortearmos a organização de nosso acervo (vídeos e fotos) baseados em consultas ao *Library of Congress Catalog: motion pictures and filmstrips*, no *Anglo American Cataloguing Rules*, nas Normas para Documentação da Associação Brasileira de Normas Técnicas e no Guia de Recomendações para a documentação fotográfica do fato cultural (FUNARTE/Centro de Documentação e Pesquisa). Além desse subsídio bibliográfico, contamos com a assessoria inicial dos técnicos do Centro de Memória e do Arquivo Edgar Leuenroth, ambos da UNICAMP.

Assumimos o conceito de informação tal como descrito no *Oxford English Dictionary*³ e discutido por Michael K. Buckland⁴ no texto “Information as Thing”: informação como representação de um conhecimento que tem forma tangível. Inclui dado (registros gravados eletronicamente), texto e documento (imagem, som) e objeto (descrição ou representação dele).

Inicialmente as informações da coleção de vídeos se resumiam ao título e diretor do trabalho, listadas alfabeticamente. O fluxo de alunos à procura de filmes com temas específicos levou-nos, durante o ano de 1993, a desenvolver um banco de dados que atendesse a essas necessidades. Era preciso, então, contemplar outras informações, principalmente os assuntos abordados e o ano de produção.

A escolha do *software Microsoft Works* para MS-DOS, versão 2.0, para dar forma a um banco de dados, mostrou-se, na época, eficiente para os objetivos propostos, além de ser compatível com o microcomputador PC XT existente no Arquivo. Hoje, esse sistema é considerado “acanhado” diante das possibilidades existentes, pois o banco de dados do *Works* é um banco de dados de arquivo não-relacional, em oposição a um banco de dados relacional. Como veremos mais adiante no capítulo 2, num banco de dados de arquivo não-relacional não se pode combinar ou comparar informações entre dois ou mais arquivos.

Montamos a estrutura do banco de dados Videoteca nos espelhando na indicação do catálogo de filmes da *Library of Congress* dos Estados Unidos

³ Oxford English Dictionary, 1989, vol.7, p.946 : “*Information-as-thing*: The term “information” is also used attributively for objects, such as data and documents, that are referred to as “information” because they are regarded as being informative, as “having the quality of imparting knowledge of communicating information”.

⁴ Buckland, Michael K., 1991

e contemplando os seguintes campos de informação: título, diretor, produção, ano de produção, gênero, tempo de duração, bitola original, sistema/cromia, idioma, procedência, número de classificação e sinopse.

Esta estrutura de campos foi montada com base nas necessidades de consulta aos vídeos demonstradas pelos usuários da Videoteca, em sua maioria alunos e professores do Instituto de Artes da UNICAMP. Naquele momento, priorizamos essas informações em detrimento do detalhamento da ficha técnica, omitindo informações tais como “elenco”, “diretor de fotografia”, “diretor de som” e outras. Lamentavelmente, esta omissão deixou uma lacuna no banco de dados Videoteca impedindo a recuperação de informações preciosas para um pesquisador de cinema, no que diz respeito a produções consideradas não-comerciais e que não constam dos grandes bancos de dados dos *sites* americanos (que são bastante completos na cinematografia mundial) e nem do *site* do Cinemabrasil⁵, *site* abrangente sobre longa-metragens brasileiros.

Uma boa parte do material do acervo da videoteca nos chega através de doações, em forma de cópia, sem nenhuma referência complementar. Para extrair dados deste material é necessário assistir à fita e analisá-la. A responsabilidade pelo conteúdo e pela confiabilidade da informação nos levaram a recorrer a pesquisas bibliográficas suplementares a cada dúvida ou inexistência de informações consideradas importantes.

As decisões tomadas no preenchimento de alguns dos elementos descritivos de um registro em vídeo foram da seguinte ordem:

1. Título: Filmes estrangeiros são catalogados com o título em português e no original. Produções sem um título específico, tais como registros de apresentação de palestras, seminários, teses, recebem como título o nome do apresentador ou do evento.
2. Diretor: este campo só é preenchido caso haja um crédito explícito para o responsável pela direção do trabalho.
3. Produção: este campo é preenchido mediante três condições: o crédito explícito, quando existente; o nome do Departamento de Multimeios e da UNICAMP para registros gerados no âmbito deste departamento, tais como palestras, seminários e eventos gravados por ele; e programas

⁵ Cinemabrasil.[Online]. Disponível na Internet via WWW.URL:<http://www.cinemabrasil.org.br>

gravados de transmissões televisivas, quando gerados pela emissora de TV, recebendo o nome da emissora.

4. Gênero: usamos a classificação adotada pelo cinema comercial, acrescido de categorias específicas. No todo, perfazem quinze gêneros: animação, comédia, dança, documentário, drama, entrevista, experimental, faroeste, ficção científica, *making of*, palestra, policial, suspense, videoarte e *videoclip* musical.
5. Bitola Original: informar sobre a captação original é importante na medida em que, conhecendo a matriz técnica que gerou o registro, o usuário saberá o que esperar da qualidade da imagem final que ele estará vendo na cópia em vídeo.
6. Procedência: cinco possibilidades são anotadas: se a fita é uma matriz (de um registro gerado na UNICAMP); uma cópia; uma fita selada (significando que foi adquirida mediante compra pelo Departamento de Multimeios); uma doação; ou uma cópia gravada de um programa de TV.
7. Sinopse: para preencher este campo, sempre tivemos grande cuidado em proporcionar uma boa representação da informação, escolhendo tanto termos gerais quanto específicos, buscando aumentar o poder de recuperação desta informação. Entretanto, esbarramos com um problema de estrutura do *software*: a limitação do tamanho dos campos, que comporta apenas 256 caracteres. Isto significa mais ou menos seis linhas datilografadas usando fonte 8, o que vem restringir o espaço para a descrição do conteúdo do trabalho em vídeo. Este problema poderia ser em parte minimizado se, na época, tivéssemos formatado um campo que contemplasse palavras-chave. Entretanto isso não foi feito devido à insegurança em estabelecer termos que não obedecessem a um vocabulário controlado, além de não termos nenhum conhecimento de um tesouro, em português, específico para cinema e para arte em geral.

Com o banco de dados Videoteca em progresso, e vislumbrando incrementar as possibilidades de recuperação e cruzamento de informações para ampliar o âmbito de pesquisas em cinema, foi iniciada uma hemeroteca de artigos sobre cinema visando a colocar estas informações em banco de dados.

Afinal, um dos objetivos de um profissional de informação é detectar as necessidades informacionais dos usuários de seus serviços e tentar prover a

infra-estrutura de informação para promover a consulta e a pesquisa. Apoiados nessa premissa, iniciamos o acompanhamento de artigos publicados nos jornais *Folha de São Paulo* e *O Estado de São Paulo* e durante pouco mais de dois anos (1994/96) preparamos este material para entrar em banco de dados quando fosse resolvido o problema de migração do *software Works* para um gerenciador de banco de dados relacional. Nossa idéia seria construir outra tabela no banco de dados contemplando os campos: título, autor, referência do jornal (ou da fonte usada), data da publicação do artigo e palavras-chave extraídas dos artigos dos jornais. Estas informações poderiam ser cruzadas com as informações da ficha videográfica de forma a oferecer ao pesquisador material suplementar à fita de vídeo. Os artigos de jornal em si estariam em pastas à disposição para consulta ou xerox no Arquivo Multimeios⁶.

O trabalho de preparação de uma hemeroteca é extremamente lento e artesanal. Além disso, exige a constituição de um vocabulário controlado que só começa a tomar forma depois de se ter extraído temas e se delineado o universo representado. Assim a catalogação por assunto começa a tomar forma. A parte mais artesanal de organização dos artigos em fichários e pastas só foi possível com a ajuda de bolsistas e tivemos que interromper o trabalho com a hemeroteca ao cessarem as bolsas. De qualquer forma, pudemos armazenar boa quantidade de artigos, principalmente durante o ano de 1995, ano de comemoração dos 100 anos do Cinema⁷.

Em uma avaliação posterior, examinando as condições práticas de trabalho que nos eram apresentadas, computamos um certo surrealismo àquele período que vivemos. Tínhamos, até fins de 1995, um único microcomputador no Departamento de Multimeios, um PC 286, localizado fisicamente na secretaria e servindo a todas as necessidades do Departamento. O microcomputador XT que serviu ao Arquivo Multimeios nos anos 1993/94 passou a apresentar inúmeros problemas e se tornou inadequado.

A alimentação do banco de dados, portanto, passou a ser feita no PC da secretaria, e, como consequência, os alunos passaram a não ter mais acesso

⁶ Naquela época nem imaginávamos a possibilidade de transformação digital dos artigos via *scanner* e OCR.

⁷ Muitos desses artigos fazem parte do banco de dados Videoteca na Internet.

direto àquelas informações. Para sanar minimamente o problema, eram impressas listagens contemplando título e/ou diretor, o que, convenhamos, era pouca informação comparada às possibilidades de pesquisa oferecidas por um banco de dados. Em vista disso, tanto a alimentação do banco de dados Videoteca como a preparação da hemeroteca demandavam a aquisição de um novo microcomputador para o Arquivo, o que não aconteceu.

Felizmente outra mudança se fez: em 1996, novos microcomputadores com maior capacidade de processamento foram adquiridos pelo departamento e a ligação do Instituto de Artes à rede Internet da UNICAMP foi completada. Este fato mudou completamente a direção do trabalho até então desenvolvido.

Acompanhando via-Internet as pesquisas de ponta na área de recuperação da informação, a equipe do Arquivo Multimeios passou a discutir a viabilidade de métodos e procedimentos compatíveis com modernas técnicas de disponibilização da informação, tanto visual como textual. Colocou-se o desafio de transformar um banco de dados “acanhado”, de filmes catalogados, em um *site* de pesquisa ampliado, disponibilizado na *home-page* do Instituto de Artes usando os recursos oferecidos pela tecnologia de rede.

Tínhamos, no início do trabalho, perguntas muito simples e objetivas que nos traçaram o caminho para as primeiras atividades:

Para quê fazer um *Website* do banco de dados da Videoteca?

Para quem direcionar? Quem são nossos usuários?

Que tipo de informação o usuário vem buscar?

Como prover esta informação e como ampliá-la?

Que desafios encontraremos e como solucioná-los?

Em que nos basearemos conceitualmente? Há experiências similares?

Havia de concreto o problema do acesso às informações da Videoteca Multimeios: não contávamos com um computador na sala da Videoteca, inviabilizando, portanto, a consulta ao banco de dados. Tentar solucionar esta lacuna transferindo a consulta de informações dos vídeos para um espaço virtual, sediado na *home-page* do IA, parecia ser uma atitude acertada.

Entre as metas propostas, além de continuar a atender seus usuários habituais, alunos e professores dos cursos de pós-graduação do Instituto de Artes, deveria ampliar a gama de interessados, pesquisadores de outras unidades e instituições. O desafio seria construir uma página que contemplasse todo o acervo de vídeos, sugerisse outras possibilidades de pesquisa e ampliasse quantitativa e qualitativamente a informação constante da ficha técnica de um filme.

Naquele momento, o *site* do Cinemabrasil era a única referência brasileira apresentando a temática cinema no formato de banco de dados. Atualmente com cerca de 500 títulos catalogados e projetando como objetivo catalogar um total de 3100 filmes brasileiros de longa-metragem, este *site* apresenta informações detalhadas na ficha técnica. Sua interface *Web* é um programa em inglês pré-formatado, direcionado para a indústria do cinema americano. Como apresenta inúmeros campos que não têm relação com a realidade da Videoteca Multimeios, optamos por criar um sistema que fizesse a interface entre o banco de dados já existente e a face gráfica da Internet⁸.

Até o presente momento, não encontramos nenhum outro *site* brasileiro que apresente uma videoteca *online* em formato de banco de dados; encontramos apenas alguns que apresentam sua videoteca na forma de índices ou listagens, sem o conceito de banco de dados, portanto, sem o conceito de procura ampliada⁹. Temos conhecimento de um projeto de uma Videoteca Digital, usando o conceito de banco de dados, que disponibilizará na *Web* fitas de vídeo para o ensino de graduação e pós-graduação da Faculdade de Biblioteconomia da PUC de Campinas, mas ainda não há resultados concretos que possamos referenciar.

⁸ Ver diagrama e descrição no capítulo 3.

⁹ Oficina Pedagógica de Multimídia, da Coordenadoria de Pesquisa e Extensão da Universidade Federal de Santa Catarina. [Online]: <http://www.ced.ufsc.br/videotec.html>

Encontramos muitas exibições *online*: exposições de trabalhos fotográficos, galerias de arte virtual apresentando *clips* de vídeos, porém, sem a organização dos dados em formato-banco.

Nos Estados Unidos encontramos atualmente vários *sites* sobre cinema organizados em banco de dados, porém, em 1996, nossa única referência era o *Internet Movie Database*¹⁰, que, apesar de priorizar a indústria norte-americana de filmes, consideramo-lo bastante abrangente na cinematografia internacional.

Antes de passarmos à descrição propriamente dita da pesquisa, faz-se necessário algumas considerações a respeito do conceito de videoteca digital bem como sua estrutura de apoio, o banco de dados, temas do capítulo seguinte.

¹⁰ Internet Movie Database [Online]: <http://www.moviedatabase.com>

Capítulo 2

Videoteca Digital, Biblioteca Digital

O surgimento da *World Wide Web* criou uma estrutura virtual para a informação, reunindo os mais variados e díspares assuntos ao alcance de um simples “click” de *mouse* no computador do usuário. Pouco a pouco, o desenvolvimento da tecnologia foi propiciando uma organização dessa informação. Com a ajuda de programas que utilizam ferramentas de busca não somente é possível organizar esta imensa coleção de documentos, classificando-a por assuntos, como é possível procurar (e possivelmente achar) qualquer tema a partir de uma palavra-chave. Hoje, o termo “surfar”, muito usado até há bem pouco tempo e que significa passar (ou passear) de um *site* a outro, a esmo, seguindo *hiperlinks*, tem sido substituído pelo termo “recuperar” (*retrieval*), objetivamente, uma informação específica. Há *home-pages* para todo tipo de assunto. E há quem diga que a Internet e sua face gráfica, o *World Wide Web*, é uma grande biblioteca digital.

Será?

Para muitos teóricos e pesquisadores, a *Web* não é uma biblioteca digital. Entretanto, antes de explorarmos o conceito de biblioteca digital há que se discutir um pouco os novos paradigmas onde se assentam as criações e publicações no formato *Web*.

2.1 Hipertexto, Multimídia, Hipermissão, Interatividade: novos paradigmas

Muito se tem discutido nos últimos anos sobre esta revolução trazida à área da informação por este novo suporte digital acompanhado da tecnologia de redes remotas¹¹. De fato, esta maneira de apresentar a informação é muito diferente do formato linear e hierárquico instaurado pelo advento da escrita

¹¹ Ambron, S. & Hooper, K., 1987. Laurel, B., 1990. Landow, G., 1992. Cotton, B. & Oliver, R., 1993. Guay, T., 1995.

e legitimado pela imprensa. Os elementos que fundamentam este sistema de gerenciamento da informação possibilitam ao indivíduo, durante o processo de conhecimento, anotar, relacionar, criar e compartilhar informações provenientes de várias mídias, desde texto a gráficos, som, fotografia, imagem em movimento e programas ao vivo.

O hipertexto é a espinha dorsal deste sistema. Um sistema hipertexto é composto de nós (conceitos) e ligações (relações). Um nó representa uma idéia ou conceito e pode conter texto, gráfico, animação, áudio, vídeo, imagem ou programa. Nós conectam-se a outros nós através de *links*, ligações. A estrutura resultante é denominada “teia”, o que deu origem ao nome *World Wide Web* dentro do projeto CERN¹².

Este sistema espelha a forma em que a mente humana processa, organiza e recupera informações. Ele cria um espaço informacional orgânico, oposto ao formato seqüencial imposto pelo paradigma da escrita. O primeiro a definir o termo hipertexto (*hypertext, non-linear text*) foi Ted Nelson, em 1965: “um corpo de material escrito ou pictórico interconectado de maneira complexa e dificilmente representável no papel, podendo conter sumários ou mapas de seu conteúdo e inter-relações além de anotações e notas de rodapé de quem o examine”¹³. Embora só recentemente tivéssemos conseguido operacionalizá-lo através de máquinas e programas que propiciaram sua concretização, este conceito já foi usado na literatura antiga. O *Talmud* usou muito o recurso do comentário sobre o comentário feito ao texto principal, bem como referências a trechos do *Talmud*, do *Torah* e *Tenach*. Épicos indianos como o *Ramayana* e o *Mahabharata*, com suas histórias originando outras histórias, também são considerados antigos protótipos de representação hipertextual¹⁴.

O espaço físico facilitador deste sistema, os computadores, e as redes de fibra ótica, transmissoras dos dados, têm apresentado desenvolvimento tecnológico rápido e constante, oferecendo à *Web* - espaço virtual onde ele

¹² Em 1989, Tim Berners-Lee, pesquisador do CERN, Laboratório Europeu de Partículas Físicas, Suíça, propôs um projeto que aglutinasse a Internet e os sistemas hipertextuais e multimídia com o fim de compartilhar informações com cientistas do mundo inteiro.

¹³ Nelson, Ted., 1965.: “*a body of written or pictorial material interconnected in a complex way that it could not be conveniently represented on paper. It may contain summaries or maps of its content and their interrelations; it may contain annotations, additions and footnotes from scholars who have examined it.*”

¹⁴ Balasubramanian, V. 1994.

se manifesta - a possibilidade de incorporar cada vez mais a multimídia, ou seja, formatos diferentes do textual para apresentar informação. Segundo Guay¹⁵, a multimídia é um novo e poderoso paradigma pois instaura a experiência multisensorial do conhecimento. Fundida ao hipertexto e à interatividade, dá origem à hipermídia. Cada um e todos inter-relacionados propiciam um ambiente extremamente rico e poderoso à informação e instauram novas formas de gerir e experienciar o conhecimento, podendo trazer radicais mudanças ao indivíduo e à sociedade. Para Pierre Lévy¹⁶, os documentos hipertexto são novas ferramentas cognitivas que abrem diferentes possibilidades cognitivas e intelectuais, vindo extrapolar àquelas oferecidas pelos documentos em papel, de leitura linear.

A *Web* é baseada em uma série de protocolos que tiveram seu desenvolvimento no *European Laboratory for Particle Physics* (CERN), na Suíça, em 1990.

O protocolo fundamental é o HTML - *hypertext mark-up language*. Este protocolo permite a publicação de páginas gráficas em forma de hipertexto. O usuário que vai percorrendo páginas ligadas por *hyperlinks* tem a impressão de estar folheando um único livro, quando na verdade cada página pode estar armazenada a quilômetros de distância uma da outra. Hipertexto possibilita “*links*” não somente entre diferentes documentos armazenados num mesmo computador, como entre documentos armazenados em qualquer computador da Internet. O protocolo que proporciona a organização dos documentos em sítios individualizados, dando-lhes um endereço, é o URL – *Uniform Resource Locator*. A comunicação entre computadores na Internet é feita através do protocolo TCP/IP - *Transmission Control Protocol/Internet Protocol*, que divide a informação em pacotes enviando-os por diversos caminhos e assegurando sua integridade no destino.

A possibilidade hoje concreta de acesso a uma gama inacreditável de assuntos numa fração de tempo e de forma incrivelmente fácil tem suas raízes históricas nos anos 40¹⁷, através de uma máquina denominada *Memex*

¹⁵ Guay, T., 1995.

¹⁶ Lévy, P., 1993

¹⁷ Embora a maior parte da literatura especializada aponte como marco o artigo de Bush, *As we may think*, de 1945, suas idéias já vinham sendo apresentadas numa série de artigos que datam do início dos anos 30,

por seu idealizador, Vannevar Bush¹⁸. Bush é considerado por muitos como o verdadeiro pai do hipertexto. Em 1945 ele escreveu um artigo, *As we may think*¹⁹, onde descreve como a Memex atuaria e como poderia estender a memória humana através da organização associativa da informação.



Vannevar Bush

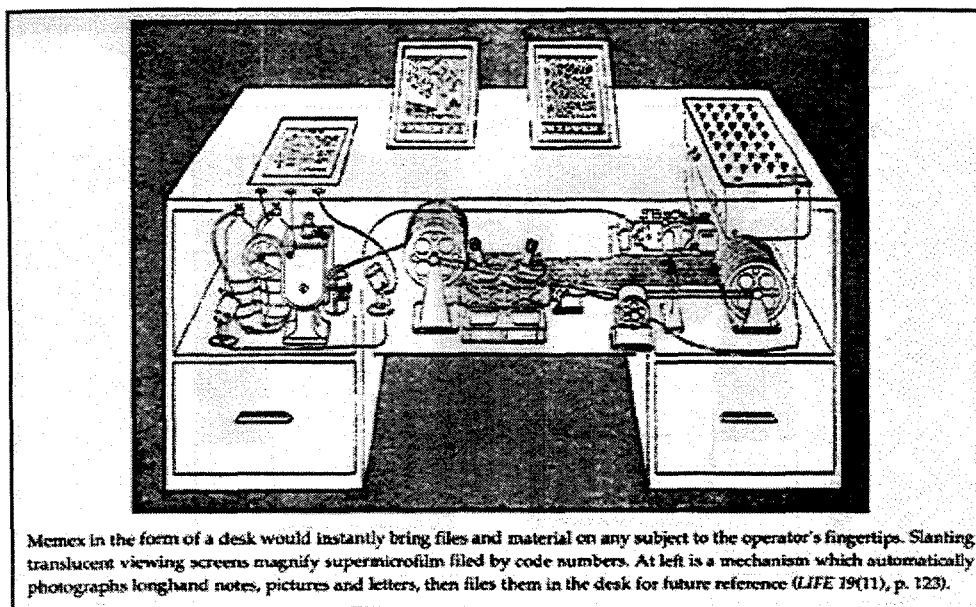
Bush compartilha a premissa teórica de que a mente humana trabalha por associações, criando uma teia de atalhos que conectam memórias e dados. Para ele, portanto, qualquer organização mecânica da informação deveria incorporar o mecanismo de associação. O que inicialmente motivou Vannevar Bush tem a ver com a explosão da literatura científica do pós-guerra e a dificuldade que pesquisadores enfrentavam para tomar conhecimento de todas as pesquisas publicadas em seus distintos campos de atuação. A Memex foi pensada como um sistema que facilitaria aos pesquisadores encontrarem informações de uma forma mais fácil do que lidando somente com documentos impressos.

segundo artigo apresentado por M. Nyce e Paul Kahn, 1992.

¹⁸ Vannevar Bush (1890-1974). Norte-americano, engenheiro elétrico, foi presidente do *National Defense Research Committee*, em 1940, e diretor do *Office of Scientific Research and Development*, em 1941. Desenvolveu o primeiro computador eletrônico analógico e o Analisador Diferencial. Ao término da Segunda Guerra Mundial, suas pesquisas focaram meios de reunir a imensa massa de trabalhos científicos desenvolvidos durante a guerra.

¹⁹ Bush, Vannevar, 1945.

Memex²⁰



A Memex guardaria dados em microfilmes que seriam projetados em uma mesa de trabalho através de várias telas, de modo a ser possível consultar vários microfilmes ao mesmo tempo, oferecendo ao usuário a possibilidade de costurar vários pedaços de informação montando veredas entre eles. Esta mesa comportaria também um *scanner* para o pesquisador inserir novas informações na trama já existente, bem como um espaço para anotações escritas a mão. Desta forma o pesquisador poderia construir um índice associativo. Apesar da Memex nunca ter sido construída, os conceitos nela embutidos, tais como o hipertexto e a interatividade, inspiraram outros trabalhos também visionários, como o de Ted Nelson.

Como citamos acima, o termo hipertexto²¹ foi cunhado por Theodor H. Nelson, nos anos 60. Naquela época, suas idéias soavam como utopia pois não só os conceitos embutidos nelas não eram familiares, como os computadores necessários para pô-las em prática ainda não existiam. Em 1974 ele publicou, como editor independente, um livro que era literalmente

²⁰ Foto e gráfico disponibilizados no endereço *Web*: <http://www.kerryr.net/pioneers/bush.htm>

²¹ Segundo definição de Owen Kelly (1996), hipertexto se refere a uma forma de armazenar, escolher e acessar dados. Parte da premissa de que os dados armazenados em um computador podem ser ordenados e acessados em qualquer ordem e de acordo com vários critérios.

dois-em-um: *Computer Lib* e *Dream Machines*. Esses livros foram concebidos de tal forma que um ia complementando o outro no decorrer da leitura.

Em *Dream Machines*, Nelson escreveu:

“ (...) apesar das desvantagens, o papel como “media” (mídia) tem sido no mínimo compatível; você pode guardar livros, revistas e anotações na mesma prateleira e as comparar usando uma mesma mesa de trabalho. Isto se torna impossível se você usa os novos meios eletrônicos, óticos ou magnéticos ou computadorizados. Cada qual necessita de ambientes separados. Você não pode fazer anotações em um videoteipe. O que nós precisamos é de um meio de apresentação e uma forma de arquivamento que seja tão padronizado quanto o papel e que consiga reunir a presente massa de informações espalhada pelos mais diversos suportes. A esperança talvez esteja numa “estrutura de dados compartilhados” (*shared-standard data structure*)²² .

A estrutura proposta por Nelson chamava-se Xanadu²³. Ele definiu este sistema como “somente uma coisa: uma nova forma de interconexão entre arquivos de computador - que corresponde à interconexão das idéias - que pode ser refinada e elaborada dentro de uma rede compartilhada.” Acrescentou ainda : “se você compreende esta forma de interconexão, então poderá avaliar seu potencial revolucionário”²⁴.

Para Guay (1995), as teorias de Marshall McLuhan²⁵ sobre aldeia global, mídias frias e quentes e mídias como extensões humanas, as de George Landow²⁶ abordando temas tais como não-linearidade, multilinearidade e metadocumento, e as desenvolvidas por Ted Nelson²⁷ acharam sua atualização na concretude virtual da *Web*.

²² Nelson, Ted. *Dream Machines*, p. 5. Microsoft Press, 1987.

²³ Este projeto ainda está em curso e sendo redefinido para a realidade dos dias de hoje, segundo diz o autor em sua *home-page*: <http://www.sfc.keio.ac.jp/~ted/XU/XuPageKeio.html>

²⁴ Op. cit.

²⁵ McLuhan, M., 1964, 1967, 1988, 1989.

²⁶ Landow, G., 1992.

²⁷ Nelson, T., 1965.

Tendo definido os elementos fundadores da rede *WWW*, podemos passar a discutir termos como biblioteca digital e videoteca digital, objeto de nosso trabalho.

2.2 Biblioteca Digital: definições e problemática

A princípio podemos pensar a videoteca digital como um braço, um ramo, uma especialidade da biblioteca digital, uma vez que usam os mesmos meios e se manifestam no mesmo ambiente. As discussões sobre o conceito de biblioteca digital (*digital libraries*) são anteriores às concernentes a *videolibraries*²⁸, esta bem recente e que diz respeito especificamente à indexação e recuperação da informação visual de forma automática. Há, por outro lado, uma boa documentação providenciada pela Unesco²⁹ referente a arquivos audiovisuais, nome tradicionalmente usado para setores específicos de arquivos ou bibliotecas tradicionais que contêm em seu acervo documentos armazenados em outro suporte que não o livro, tais como filmes, vídeos, fitas de áudio e fotografias, além de contarem com práticas diferentes para manipulação e conservação de seu material.

Como estamos considerando nosso objeto de pesquisa a Videoteca Multimeios em sua versão digital, não falaremos sobre conceitos de arquivos audiovisuais no sentido tradicional do termo. Também não aprofundaremos a discussão sobre *videolibraries* pois, como dissemos acima, a literatura existente neste tema se reporta aos métodos automatizados de indexação das informações e discorre sobre tecnologias tais como reconhecimento de som e de imagem.³⁰ Exploraremos, sim, o conceito de arquitetura de uma biblioteca digital, uma vez que a construção do formato digital da Videoteca Multimeios está baseada nos conceitos dessa primeira.

Deparamo-nos neste percurso com problematizações da Ciência da Informação. Não sendo nossa área de especialidade, tentamos apresentá-las sem discutí-las em profundidade, porém assumindo o desafio de aplicá-las de maneira correta em nosso projeto. Nossa pesquisa, concentrada no

²⁸ Não conhecemos um termo específico em português para o conceito de *videolibraries*.

²⁹ [Online]: <http://www.unesco.org>

³⁰ Christel, M., 1995, 1996; Smith & Christel, 1995; Smith & Chang, 1996; Christel & Pendyala, 1996.

uso técnico de mídias em rede remota, coteja, no entanto, a todo instante, a área da informação, notadamente a Ciência da Informação, especialmente a Biblioteconomia. Estas ciências nos fornecem categorias estandardizadas para a organização da informação e conceitos teóricos que definem a especificidade do material visual em relação à sua organização. Os avanços recentes na área de museus e bibliotecas digitais têm contribuído sobremaneira para o enriquecimento de nosso projeto.

Ao se fazer uma procura *online* pela palavra chave *digital library* tem-se como resultado uma enorme listagem compreendendo desde a Biblioteca do Congresso dos Estados Unidos até *sites* que catalogam dados sobre peças de automóveis e denominam suas informações com o termo biblioteca digital. Parece-nos que o termo biblioteca digital tem sido usado indiscriminadamente por setores muito diversos de sua atividade-fim original. Em vista de uma certa confusão que envolve o termo, tentemos caracterizá-lo.

A comunidade norte-americana dos bibliotecários e profissionais da informação tem usado ao longo dos anos diferentes termos para este conceito eletrônico de biblioteca: *eletronic library*, *virtual library*, *library without wall*, e, mais recentemente, *digital library*. Por outro lado, o conceito de biblioteca digital tem sido explorado e desenvolvido por outras comunidades que não os bibliotecários e tem apresentado significados específicos³¹:

- sob o ponto de vista da recuperação da informação, é uma grande base de dados;
- para os que trabalham na área de tecnologia de hipertexto, é uma aplicação particular dos métodos de hipertexto;
- para a área de gerenciamento de informação, é uma aplicação da *Web*;
- para a Ciência da Informação, é um degrau no contínuo processo de automação de bibliotecas começado há 25 anos.

O uso indiscriminado do termo *digital library* não é aprovado pela comunidade de bibliotecários, que não concorda com as seguintes acepções³²:

³¹ Nurnberg, et al, 1995 em Cleveland, G., 1998.

³² Op. cit.

- coleções de algoritmos computacionais ou programas de computador são considerados bibliotecas digitais pelos profissionais da Ciência da Computação e por desenvolvedores de *software*;
- para distribuidores e comerciantes de bases de dados, o que eles distribuem são bibliotecas digitais;
- para grandes corporações, uma biblioteca digital é o sistema de gerenciamento de informação que controla os documentos eletronicamente.

Para um grande número de pessoas, a *World Wide Web*, com sua imensa coleção de documentos, é, hoje, uma grande biblioteca digital. Muitos teóricos da informação não concordam com esta afirmativa. Para Clifford Lynch³³, a Internet e, mais notadamente, a *Web*, não foi concebida com as ferramentas adequadas à disponibilização e à recuperação organizada da informação. Desenvolveram-se na direção de um grande repositório caótico armazenando imensa quantidade de dados originalmente impressos, tornados digitais.

Em seu texto *Digital Libraries: definitions, issues and challenges*, Gary Cleveland³⁴ aponta-nos que, para operacionalizar o termo biblioteca digital, bibliotecários assumem como ponto de partida a premissa de que bibliotecas digitais são bibliotecas com os mesmos objetivos e funções de bibliotecas tradicionais, quais sejam: desenvolvimento e gerenciamento de coleções, análise de assuntos, criação de índices temáticos, provisão do acesso à informação e preservação do documento. Há todo um trabalho embasado em técnicas e procedimentos específicos para a organização dos dados visando a serem recuperados facilmente pelo usuário. O autor nos oferece também um resumo das características de uma biblioteca digital apoiado em várias discussões³⁵:

- bibliotecas digitais são a face digital de bibliotecas tradicionais e contemplam tanto coleções digitais quanto coleções no tradicional formato impresso;
- bibliotecas digitais também comportam material existente fora da fronteira física e administrativa de uma biblioteca digital;

³³ Lynch, C., 1997.

³⁴ Cleveland, G., 1998.

³⁵ Arms, 1995; Graham, 1995; Chepesuik, 1997; Lynch and Garcia-Molina, 1995; on Cleveland, G., 1998.

- bibliotecas digitais incluirão todos os procedimentos e serviços que fazem parte do eixo central de uma biblioteca tradicional. Entretanto, estes processos tradicionais que dão especificidade às bibliotecas tradicionais terão que ser revistos e readaptados para acomodarem as diferenças apresentadas pelos novos meios digitais;
- bibliotecas digitais devem idealmente proporcionar uma visão coerente de toda a informação contida dentro da biblioteca, não importa sua forma ou formato;
- bibliotecas digitais dirigir-se-ão à comunidades particulares ou constituídas, como bibliotecas tradicionais o fazem ainda hoje, embora estas comunidades possam estar dispersas por toda a rede Internet;
- bibliotecas digitais requerem os trabalhos especializados de bibliotecários e de cientistas da computação para serem viabilizadas.

Estas características expandem o conceito tradicional de biblioteca e enfatizam o trabalho científico do profissional da informação, enquanto integram novas tecnologias, novos métodos e novas mídias.

Ainda segundo Cleveland, o que não define uma biblioteca digital é a existência de um único sistema digital que ofereça acesso instantâneo a toda a informação constante nele, para todos os setores da sociedade, de qualquer lugar do mundo. Este conceito era propagado como viável, durante os primeiros tempos da Internet, quando não se tinha avaliado devidamente a complexidade da construção de bibliotecas digitais. O que temos hoje em dia na Internet é um amontoado de dados organizados em diferentes sistemas, dificilmente recuperáveis em toda amplitude e riqueza de suas informações, de maneira organizada. A possibilidade de intercâmbio e operacionalização desses dados está limitada a sistemas restritos, construídos com este propósito. Nesta direção caminham inúmeras pesquisas que buscam encontrar soluções para os imensos desafios colocados à Ciência da Informação pelas novas tecnologias.

A *Web*, analisada então sob este prisma, não é uma grande biblioteca digital; é considerada, sim, como um enorme repositório onde inúmeras coleções digitais organizadas coexistem com outra imensa massa de dados não organizados.

Como vimos, a polêmica sobre o termo biblioteca digital ainda acirra ânimos. Vamos nos apoiar nas definições apontadas pela comunidade dos bibliotecários e restringir o uso do termo aplicando-o ao contexto onde exista uma pré-organização da informação segundo certas normas pré-definidas visando a facilitar ao usuário a recuperação da informação.

A construção de uma biblioteca digital levanta, segundo alguns teóricos³⁶, grandes problemas operacionais que agrupamos em sete itens: arquitetura técnica, construção de uma coleção digital, digitalização, metadado ou metadocumento, nome e endereço, direitos autorais e preservação.

Nosso objetivo aqui é abordá-los em sua aplicabilidade ao nosso trabalho, uma vez que, *per si*, cada um destes itens mereceria um capítulo, ou mesmo uma tese.

2.2.1. Arquitetura técnica

Diferentemente da primeira forma de catalogação eletrônica usada por bibliotecas através do sistema monobloco OPAC (*Online Public Access Catalog*), hoje a construção de uma biblioteca digital se apoia numa coleção de diferentes sistemas que operam de forma integrada dentro de uma interface comum (a *Web* ou uma rede intranet), conectados através de uma rede eletrônica. Esta arquitetura depende dos seguintes componentes:

- redes locais de alta velocidade e conexão rápida via-Internet;
- banco de dados relacionais que acomodem formatos digitais variados;
- *softwares* de ferramentas de busca poderosos e acurados;
- vários servidores que acomodem serviços variados (arquivos de imagem digital, servidor *Web*, servidor para FTP – *File Transfer Protocol*);
- documentação eletrônica para integração dos diferentes sistemas.

Estes sistemas podem ser variados, dependendo do perfil da biblioteca digital: bases bibliográficas que apontem tanto para texto como para imagem; índices alfabéticos, temáticos, geográficos, cronológicos; *links* para outros endereços *Web* que tenham relação com a informação colocada; gráficos; jornais eletrônicos; fotos, som, vídeo. Estes diferentes materiais podem estar localizados em diferentes bancos de dados, em diferentes servidores em espaços físicos separados; porém, ao usuário que consulta o

³⁶ Kahn, R. & Wilensky, R., 1995; Arms, W.Y., 1995; Paepcke, A., 1996; Schatz, B. & Chen, H., 1996; Fleischhauer, C., 1996 e 1998; Arms, C.R., 1997; Cleveland, G., 1998.

sistema, eles aparecerão como um conjunto orgânico graças aos programas de integração e uso de alguns padrões para documentos.

A possibilidade de integrar várias bibliotecas digitais em um sistema de busca de informação único tem sido a grande questão do momento. A dificuldade em compartilhar informações deve-se, inicialmente, à grande diversidade de estruturação dos dados e à ausência de um vocabulário controlado. Além disso, como os projetos digitais têm sido desenvolvidos separadamente, há uma variedade no uso de bancos de dados, *softwares* para busca e interfaces que não compartilham dados. Por esta razão, tem sido quase impossível a integração de bibliotecas, arquivos e museus num sistema único de busca. Entretanto, estamos acompanhando a discussão de vários grupos com enfoque nas possibilidades de viabilizar, em parte, esta imensa tarefa, bem como temos notícias de projetos desenvolvidos em parceria que já disponibilizaram na *Web* vários repositórios integrados³⁷.

2.2.2. Construção de uma coleção digital

A construção de uma coleção digital envolve alguns ou todos os seguintes métodos:

- digitalização do material: convertendo papel ou outro meio (foto, som, vídeo) de uma coleção já existente para o formato digital;
- aquisição de um material originalmente digital (CD-ROMs, livros e/ou jornais eletrônicos)
- providenciando acesso digital a material localizado em outros repositórios que não o da coleção, através de *links Web* (neste caso podendo ser considerado como uma extensão da coleção).

³⁷ *The Research Libraries Group* (RLG) é uma corporação de universidades, arquivos, sociedades históricas, bibliotecas, museus e outras instituições culturais e de pesquisa, a maioria norte-americana, que provê informação integrada de suas bases de dados.

Videomuseum é uma associação de 43 museus e instituições culturais francesas que abrigam coleções em arte moderna e contemporânea.

Electronic Library Image Service for Europe (ELISE) é um projeto liderado pelo *Victoria and Albert Museum* e congrega 10 instituições culturais (compreendendo Inglaterra, Irlanda, Bélgica e Holanda).

Advanced Information Hub in Gifu abriga um sistema integrado onde participam 10 museus do Japão.

International Web Collaboration for Medieval Studies Archives, liderado pela *Princeton University*, USA, integra várias bases de dados sobre arte cristã.

Dance Heritage Coalition é uma aliança entre instituições com o objetivo de centralizar acervos de dança. Fazem parte: *Dance Collection of the New York Public Library for the Performing Arts*; *The Harvard Theatre Collection*; *Harvard University*; *Library of Congress*; *Ohio State University*; *San Francisco Performing Arts Library and Museum*; *American Dance Festival*; *Jacob's Pillow Dance Festival*.

A primeira opção envolve critérios de escolha sobre o que disponibilizar em formato digital, e custos altos. Temos notado a crescente tendência de instituições desenvolverem projetos em parceria numa tentativa de minimizar custos de produção além de evitar o problema de replicar dados. Entretanto, como apontamos acima, a integração de bases e sistemas diferentes não é um problema de resolução simples, não só pelo problema técnico de uso de diferentes *softwares* para base de dados como também, e principalmente, pela dificuldade de trocar e compartilhar dados indexados sem uma padronização.

2.2.3. Digitalização

Este termo de uso recente significa converter para o formato eletrônico um dado que esteja armazenado em um sistema analógico ou um suporte fixo (livro, jornal, foto, pintura, filme, vídeo, áudio), usando tecnologias variadas. Significa também selecionar formatos para arquivar dados e os disponibilizar em rede, levando em conta fatores tais quais a qualidade da reprodução, a conveniência do acesso, a longevidade do formato, o custo da produção e a opção por formatos standardizados ou já adotados na prática.

Para textos, os formatos padronizados são os que empregam a *Standard Generalized Markup Language* (SGML – ISO8879) e o *Hypertext Markup Language* (HTML - RFC 1866)³⁸. Estas linguagens viabilizam a codificação de textos para serem vistos na *Web* e hoje em dia estão bastante popularizadas.

SGML é uma linguagem aceita internacionalmente e compõe-se de uma série de regras para definir e expressar a estrutura lógica e o conteúdo de diferentes tipos de documentos eletrônicos. Podemos chamá-la de “linguagem-mãe”. É usada para descrever documentos tão variados e heterogêneos como uma descrição de manuscritos antigos ou os registros médicos de pacientes de uma clínica. Entretanto, é limitada a documentos-texto pela impossibilidade de executar procedimentos que impliquem a geração de eventos tais como a execução de áudio, vídeo, gráficos e imagens.

A linguagem SGML faz uso de *Document Type Definitions* (DTDs) que são regras que definem marcas, *tags*, para especificar documentos de variados

³⁸ Farta documentação online: <http://www.oasis-open.org/cover/sgml-xml.html>

tipos. Por exemplo, a linguagem HTML é uma das aplicações da SGML, ou um caso particular onde DTDs parametrizam *tags*, indicações específicas que viabilizam a aparência de um documento e sua ligação com outros na *Web*.

EAD DTD (*Encoded Archival Description by Document Type Definition*)³⁹, ou *Archival Finding Aids* são outra aplicação da SGML, usada para codificar informações sobre coleções de um arquivo considerada padrão. Um documento codificado em EAD DTD (por exemplo, uma coleção de um determinado grupo de dança com materiais diversos como fotos, figurinos, desenhos de cenários, vídeos, programas, etc, pertencente a um determinado arquivo), consiste de dois segmentos: o primeiro informa sobre ele, documento (título, quem compilou, quando) e o segundo segmento informa, hierarquicamente, sobre o conteúdo (da coleção, de um grupo, de uma série).

Está em desenvolvimento a versão abreviada e simplificada da SGML, a XML - *Extensible Markup Language*⁴⁰. Ela tenta resolver problemas colocados pela rigidez da HTML (que tem formato fixo). Uma das vantagens da XML será a viabilização do uso da linguagem SGML na *Web*. Atualmente, os documentos escritos em SGML necessitam da intermediação de uma aplicação (como o Panorama, por exemplo) para serem vistos na *Web* em formato HTML. Com a implementação da XML, transmitir e compartilhar documentos via-*Web* se tornará mais fácil, tanto em SGML quanto em HTML. A XML conserva as habilidades estruturais da SGML redefinindo parâmetros e removendo um grande número de procedimentos complexos e nem sempre utilizados, facilitando sobremaneira a definição de documentos.

Falamos sobre formatos padronizados para documentos-texto. Passemos agora para o universo das imagens e sons.

As imagens e sons podem ser armazenados e comprimidos em diferentes formatos, de acordo com seu objetivo-fim, em alta, média ou baixa resolução, ou ainda as três. O ideal seria armazenar um arquivo na mais

³⁹ O site com informações sobre EAD pode ser consultado no endereço online:

<http://www.loc.gov/ead/eadback.html>

⁴⁰ Informações detalhadas sobre XML são encontradas no endereço online: <http://www.ucc.ie/xml/>

alta resolução, sem compressão, como “original digital”, e outros em média e baixa resolução e comprimidos, para pesquisa e modelo rápido em *sites Web*. A tecnologia digital para a imagem fixa está melhor resolvida e é diferente da usada para imagem em movimento e som.

Atualmente, as bibliotecas digitais com acervos de imagem fixa têm feito uso de dois formatos de apresentação na *Web*:

- Imagem-ícone (*thumbnail image*): pequena imagem apresentada geralmente junto com o registro bibliográfico, ou o título do trabalho. Oferece ao pesquisador a possibilidade de decidir se quer ou não ver o formato maior e de melhor qualidade. O formato técnico dessa imagem é em GIF⁴¹ com compressão GIF, não passando de 30 kb em tamanho de arquivo.
- Imagem de acesso: imagem maior e de melhor qualidade, porém ainda comprimida para atender a um rápido acesso por parte do usuário. Normalmente tem formato técnico JPEG⁴², com compressão JPEG e tamanho de arquivo até 150 kb.

A imagem digital considerada “matriz” deve ser armazenada preferencialmente em CD-ROMs (pois é possível que cada imagem seja um arquivo de mais de 1 MB em tamanho), sem compressão, em formato TIFF⁴³. Este recurso preserva a imagem para um futuro em que possa haver mudanças nos padrões de compressão, bem como reserva uma matriz de alta qualidade para processar reproduções.

Como apontamos acima, a tecnologia para áudio e imagem em movimento na *Web* é bastante recente e ainda não bem resolvida tecnicamente, ainda que *sites* que ofereçam arquivos de áudio estejam cada vez mais populares. A digitalização de um vídeo envolve o armazenamento e operações com um volume tão grande de dados que dificilmente um computador pessoal conseguiria processar, não fosse o recurso da compressão do arquivo de imagem/som.

⁴¹ GIF (*Graphics Interchange Format*). Formato desenvolvido pela CompuServe para transmissão de imagens por linhas discadas. Este formato usa codificação LZW (*Lempel-Ziv and Welsh*) algoritmo para comprimir imagens e reduzir tempos de transmissão.

⁴² JPEG (*Joint Photographers Expert Group*). Formato que permite a transferência de arquivos para várias plataformas. JPEG é baseado em codificação da imagem por transformação matemática, o que oferece altas taxas de compressão embora com perda de informações. É possível comprimir imagens em até um quinto (1/5) do tamanho original sem perda de qualidade perceptível.

⁴³ TIFF (*Tagged Image File Format*)

Owen Kelly⁴⁴ chama a atenção para a dimensão do problema da imagem em movimento num processo de transmissão de dados. Para termos a noção do volume de dados necessário para se transmitir um filme através da *Web*, façamos o seguinte exercício: suponhamos que a resolução da tela de seu computador esteja ajustada para 640 x 480 *pixels*⁴⁵. Suponhamos ainda que o filme será apresentado em 256 cores e com taxa de repetição de 24 quadros por segundo. A quantidade de dados transmitida, neste caso, é de 58.982.400 bits por segundo. Se considerarmos que a taxa de transmissão dos modems mais recentes é de 53.600 bits por segundo, concluimos que a taxa de transferência de um filme não é ainda possível sem que haja uma grande compressão dos dados.

Por esta razão, no momento da captura digital de um vídeo ou de uma fonte sonora, já devem ser estabelecidos os parâmetros de compressão. Uma vez capturado, o arquivo é editado (ou não) e exportado recebendo nova compressão. O resultado de tanta compressão afeta claramente a qualidade da imagem e da informação, principalmente se seu destino é a *Web*.

Temos encontrado disponível em *sites Web* arquivos de vídeo em 3 formatos diferentes:

- MPEG-1, de alta resolução e grande tamanho, conseqüentemente demandando um longo tempo para o *download* no computador do cliente;
- QuickTime, de média resolução e menor em tamanho, porém o fato de necessitar ser carregado inteiramente no computador do usuário (*downloading*) demanda também uma certa demora.
- RealVideo, uma tecnologia que possibilita ao usuário ver/ouvir o arquivo vídeo ou áudio (através do RealAudio) no momento em que suficiente conteúdo (o que demora apenas alguns segundos) tenha chegado, não necessitando ser carregado por completo. Esta solução é chamada de *streaming media*.

Ao clicar em um item de *streaming media*, o *download* do arquivo começa. Mas, ao invés de ter que carregar o arquivo inteiro no computador para que ele possa ser visto (ou ouvido), ele pode começar a ser rodado no momento em que suficiente conteúdo (comumente chamado *buffer*) tenha chegado e assegurado um *broadcasting* contínuo, e isto demora apenas alguns

⁴⁴ Kelly, O. 1996 (notes, 8)

⁴⁵ A tela do computador é formada por uma matriz de pontos chamados de *pixels* ou *dots*.

segundos. Esta solução veio substituir muitos dos arquivos de vídeo e áudio apresentados na *Web* nos dois primeiros formatos listados acima, porém, com óbvia perda de qualidade e informação.

2.2.4. Metadado ou metadocumento

Metadados são dados que descrevem o conteúdo e os atributos de um determinado item de uma biblioteca digital. Na biblioteca tradicional, são registros bibliográficos ou catalográficos que descrevem um documento. Enquanto as bibliotecas contam com normas para o registro de metadados do material bibliográfico, arquivos e museus nunca contaram com padrões internacionais, ou mesmo nacionais, para a estruturação da informação, desenvolvendo individualmente métodos e práticas desconectadas umas das outras.

Os arquivos norte-americanos com acervo artístico introduziram a catalogação eletrônica no final dos anos 70. Nos anos iniciais, o único modelo existente era o que estava sendo aplicado aos livros, e que praticamente replicava eletronicamente as fichas catalográficas manuais. Manteve-se esse modelo, considerando-se que era importante assegurar a estrutura como vinha sendo aplicada manualmente nestes arquivos desde o final do século passado. Entretanto, diferentemente do registro bibliográfico, onde a informação sobre cada livro é limitada e consistentemente estruturada, em arquivos e museus isso nunca aconteceu devido à inexistência de convenções para descreverem seus acervos. A maioria dos arquivos e museus desenvolveu estruturas próprias para classificar elementos tais como data, nome do artista, estilo, proprietário, dados da mobilidade do objeto, dados descritivos da peça/objeto/foto, etc. O resultado disso é uma grande heterogeneidade na forma de documentar o que hoje chamamos objeto visual, fruto da conversão digital de uma foto, ou da foto de um objeto.

No momento em que alguns projetos em parceria foram alavancados, constatou-se a diversidade de maneiras no trato da informação visual. Por outro lado, o desenvolvimento de *hardware* e *softwares* que possibilitam a integração de dados localizados em bases diversas tem exigido medidas urgentes na busca de padronização na catalogação. Hoje em dia, muitos seminários, encontros e congressos entre a comunidade de bibliotecários, profissionais de informação e acadêmicos da área da computação debatem a

necessidade de diretrizes e padrões para catalogação eletrônica de textos e imagens⁴⁶.

Há muitos obstáculos a enfrentar para superar estes grandes e novos desafios colocados à área da Informação, fruto da revolução digital. Muitas perguntas têm sido feitas: no âmbito de uma mesma instituição, como integrar diferentes mecanismos de acesso ao dado que foram desenvolvidos eletronicamente ao longo dos anos? Como criar um sistema coerente de acesso a reproduções digitais provenientes de materiais de diferentes classes? E, num passo além, como lidar com a interoperabilidade entre instituições, de forma a integrar arquivos digitais de assuntos similares, porém de diferentes instituições, em um sistema integrado de pesquisa?

Sabemos que há uma variedade de fontes de informação que são organizadas por instituições culturais, museus e bibliotecas e que, tomando um formato eletrônico, desembocam na rede *WWW*:

- Registros de fotos e gravuras;
- Descrição de objetos e artefatos de artes;
- Livros, jornais e outras publicações;
- Imagens digitais das fontes citadas acima;
- Catálogos de bibliotecas;
- Texto de documentos diversos;
- Instrumentos de pesquisa (*Finding aids*) de coleções de arquivos;
- Descrição de coleções de arquivos;
- Descrição de fontes de pesquisa;
- Base de dados com vários campos de registro;
- Páginas *Web* apresentando itens de coleções e exposições;
- Páginas *Web* sobre instituições culturais.

Algumas dessas informações estão editadas em páginas *Web*, outras (como a maioria das bases de dados) não foram originalmente escritas em formato

⁴⁶ Dos Estados Unidos, citamos três importantes projetos: *The Dublin Core Metadata Workshop*, o *Vision Project* e o *MESL – Museum Educational Site Licensing Project*. Na comunidade européia, *CATRIONA (Cataloguing and Retrieval of Information Over Networks Applications)*, e os encontros da *EVA – Electronic Imaging and the Visual Arts*. No Brasil, os Seminários Nacionais de Bibliotecas Universitárias e a iniciativa da Fundação Getúlio Vargas.

Web, mas apenas chegam ao usuário em formato *Web, on-the-fly*⁴⁷ mediante *scripts* previamente definidos.

Dirigido inicialmente (e exclusivamente) para páginas *Web* efetivamente construídas, apontamos o *Dublin Core Metadata Element Set*⁴⁸ como uma das tentativas de padronizar a entrada de informação. Desenvolvido pela comunidade acadêmica internacional de profissionais ligados à informação mediante uma série de encontros internacionais⁴⁹, muita polêmica e volumosos relatórios, um núcleo de elementos descritivos foi acordado, tendo como objetivos facilitar a recuperação da informação através da descoberta de fontes de pesquisa e promover a interoperabilidade semântica entre metadados hospedados em sistemas heterogêneos. O *Dublin Core* é um conjunto de quinze elementos e sua fácil manipulação atende também a não-profissionais da área da informação, podendo ser utilizado por qualquer criador de página *Web*. Apresenta a vantagem de oferecer uma maneira de apresentar “informação sobre a informação” que seja facilmente rastreada pelos inúmeros mecanismos de busca encontrados na rede Internet.

Segue uma lista dos quinze elementos⁵⁰:

Title (Título): o nome dado ao trabalho/objeto digital pelo autor ou editor;
Author or Creator (Criador): a(s) pessoa(s) ou organização primeiramente responsável pela criação do conteúdo intelectual do trabalho. Por exemplo, autor, no caso de documentos escritos; artista, fotógrafo ou ilustrador, no caso de documentos visuais;
Subject and Keywords (Assunto): o tema do trabalho, ou palavras-chave que descrevam o conteúdo. O uso de vocabulário controlado e esquemas formais de classificação é encorajado;
Description (Descrição): descrição textual do conteúdo do trabalho. Resumos, no caso de documentos textuais; ou descrição do conteúdo, no caso de objetos visuais;

⁴⁷ Página *Web* formatada segundo parâmetros pré-definidos por programação.

⁴⁸ A página principal do *Dublin Core Metadata Initiative* pode ser encontrada no endereço online: <http://purl.oclc.org/dc/>

⁴⁹ Os relatórios de todos os encontros pode ser consultado nos arquivos da revista *D-Lib Magazine*, no endereço online: <http://www.dlib.org/dlib.htm>

⁵⁰ Esta lista está disponível no endereço online: http://purl.oclc.org/dc/about/element_set.htm

Publisher (Editor): a entidade responsável por disponibilizar o trabalho em sua presente forma, tal como uma editora, uma universidade, uma corporação;

Contributors (Outros Colaboradores) : pessoa(s) ou organização não especificada no elemento Criador que fez significativa contribuição intelectual ao trabalho, cuja contribuição, porém, é secundária com relação à de pessoas ou organizações especificadas no elemento Criador (por exemplo, editores, tradutores, ilustradores);

Date (Data): a data em que o trabalho foi disponibilizado em sua presente forma. Recomenda-se o uso de um número de oito dígitos no formato AAAA/MM/DD – (ano/mês/dia)⁵¹;

Resource Type (Tipo de Recurso): o gênero do objeto, tal como uma *home-page*, uma novela, um poema, um relatório, um ensaio, um dicionário. Para fins de interoperabilidade semântica, sugere-se a escolha do gênero dentre uma lista que está sendo desenvolvida⁵²;

Format (Formato): o formato eletrônico que representa o trabalho/objeto digital, tal como, um texto em HTML, ASCII, um arquivo *Postscript*, um arquivo executável *Windows*, etc. Esta informação indica a possível necessidade de *software* ou *hardware* específico para disponibilizar o objeto. Para fins de melhor interoperabilidade semântica, o registro do formato deve ser escolhido dentre uma lista que está sendo desenvolvida;

Resource Identifier (Identificação): uma cadeia de caracteres (*string*) ou um número usado para identificar exclusivamente o objeto, como URLs, ISBN, etc;

Source (Fonte): matriz geradora do objeto digital. Por exemplo, o ISBN e outras referências da versão impressa de um livro de poemas que gerou uma página HTML do mesmo;

Language (Idioma): em que língua está o conteúdo do trabalho/objeto⁵³;

Relation (Relação): relacionamento com outros objetos digitais;

Coverage (Cobertura): características espaciais e temporais do objeto. O uso deste elemento está ainda em fase experimental;

Rights Management (Direito Autoral): sugere-se o uso de um *link* para uma notificação sobre normas acordadas entre as partes.

⁵¹ Estas normas encontram-se no endereço *online*: <http://www.w3.org/TR/NOTE-datetime> . Este item tem sido objeto de polêmicas devido à rigidez com que indica a data a ser colocada.

⁵² No momento em que este documento foi referenciado a lista de gêneros ainda não estava disponível. As discussões em curso podem ser encontradas no endereço *online*: <http://sunsite.berkeley.edu/Metadata/types.html>

⁵³ Ver normas para preenchimento no endereço *online*: <http://ds.internic.net/rfc/rfc1766.txt>

Todos estes campos, após preenchidos, serão codificados em HTML e farão parte do cabeçalho da página *Web* sem, no entanto, serem visíveis ao usuário. Entretanto, sua localização na página *Web* permite que as informações ali contidas sejam facilmente indexadas ou encontradas pelos diferentes mecanismos de busca que operam na rede Internet. Vê-se, portanto, que a adoção deste sistema padronizado de organização de metadados poderá facilitar e ampliar a gama de informação recuperável. Entretanto, como notamos, o *Dublin Core* atua primordialmente em páginas estáticas *Web*, não sendo possível codificar informação em páginas geradas *on-the-fly*, ou seja, páginas que são geradas automaticamente segundo um programa pré-determinado. Por ter se mostrado um efetivo indexador de metadados, esta lacuna que apresenta tem sido trabalhada no sentido de buscar formas que estendam o uso do *Dublin Core* a documentos não gerados em páginas *Web* manualmente construídas⁵⁴.

Outra iniciativa que poderíamos apontar na busca de padrões para a disponibilização de informação digital é o *Vision Project* (*Visual Resources Sharing Information Online Network*)⁵⁵. Uma colaboração entre o *Research Libraries Group*, o *Visual Resources Association* e o *Getty Information Institute*, este projeto busca desenvolver padrões para descrever documentos visuais. Um modelo classificatório para documentos visuais baseado em um sistema padronizador desenvolvido pela *Visual Resources Association*, chamado *Core Categories for Visual Resources* está sendo testado e avaliado pela comunidade de arquivos e museus norte-americanos e canadenses. Este modelo incorpora o *Art & Architecture Thesaurus*⁵⁶ e o *Union List of Artist Names*⁵⁷, desenvolvidos pelo *Getty Information Institute*.

2.2.5. Nome e endereço

No contexto da biblioteca digital, a questão do nome de um objeto digital está relacionada também à sua localização virtual. É o nome que dá ao objeto digital sua particularidade, sua existência e o torna passível de ser

⁵⁴ Discussões sobre o tema no endereço online: <http://www.rlg.org/metawg.html>

⁵⁵ *Vision Project* pode ser consultado online: <http://www.oberlin.edu/~art/vra/vision.html>

⁵⁶ É possível consultar o *Art & Architecture Thesaurus* online: http://shiva.pub.getty.edu/aat_browser bem como as *Categories for the Description of Works of Art*: <http://www.getty.edu/gri/standard/cdwa/HOMEPAGE.HTM>

⁵⁷ Consulte online: http://shiva.pub.getty.edu/ulan_browser/

encontrado; corresponde ao ISBN de uma biblioteca tradicional. Porém, da maneira como as páginas *Web* estão sendo estruturadas, os nomes estão sempre vinculados ao endereço URL (*Uniform Resource Locator*), o atual método de identificar objetos na Internet. Os URLs apresentam, em uma linha, vários nomes separados por pontos e barras significando itens que deveriam estar separados. São eles: o método pelo qual o documento é acessado (http, ou ftp, por exemplo), o nome da máquina, o país que a hospeda, a localização do documento na máquina, o nome do arquivo onde está o documento (este último não é necessariamente um nome original, como por exemplo, “projetos”, ou “bio”, ou “index”, muito usados genericamente). Instala-se, então, um grande problema, pois qualquer dos itens que sofra uma modificação ou mudança de endereço pode ocasionar o “desaparecimento” do item conhecido.

Tentando minimizar estes problemas, grupos de estudo buscam soluções através de esquemas globais de identificadores únicos que persistam além do tempo de duração da organização que originou a informação.

2.2.6. Direitos autorais

O simples fato da informação digital ser facilmente copiada e acessada remotamente por muitos usuários dispara um problema de difícil solução, ou seja, o controle sobre os direitos autorais.

2.2.7. Preservação

Preservação no mundo digital diz respeito à durabilidade da informação enquanto suporte manipulável e à veloz obsolescência desta tecnologia. Embora poucos anos tenham se passado desde o início da era digital, sabemos existir grande quantidade de trabalho já perdido por estar em um suporte que não tem mais o meio físico para decodificá-lo.

2.3. Banco de Dados

Na qualidade de realizadores de *sites* para a *Web* e necessitando criar banco de dados na Internet sem, no entanto, sermos profissionais da área de computação, deparamo-nos com a urgência de compreender alguns conceitos que lhe são inerentes para conseguir operacionalizá-lo e tirar partido de suas potencialidades.

A definição mais simplificada de um banco de dados poderia ser um conjunto organizado de informações. Dentro desta concepção, uma agenda telefônica é um banco de dados não computadorizado. A diferença está na forma como registramos os dados. Numa agenda telefônica, as informações estão gravadas no papel. Num banco de dados informatizado, os dados são gravados no disco rígido do computador, de acordo com uma certa estrutura pré-determinada por um programa específico. O programa que permite criar e manter o banco de dados é chamado de Sistema de Gerenciamento de Banco de Dados, ou SGBD (*Database Management System*, ou *DBMS*). Dados, no universo da computação, são vários tipos de informação que podem ser representados em formato digital, incluindo textos, fotos, som, programas em vídeo, ilustrações, desenhos, base de dados, programas de computador e combinações deles todos. Os bancos de dados em formato digital mantêm os registros automaticamente organizados e permitem a procura ou consulta de registros específicos, baseada em algum(ns) campo(s) do registro.

Tomando como exemplo a videoteca Multimeios, cada registro de filme é constituído de vários campos: título, diretor, gênero, número de classificação, sinopse, etc. Podemos fazer uma consulta sobre a filmografia de um diretor específico, abrangendo determinados anos e somente os seus filmes do gênero “comédia”, por exemplo. Ou procurar, na base de dados inteira, quais os filmes que contêm a palavra “índio” na sinopse. Um banco de dados classifica automaticamente as informações resultantes de uma consulta.

Os primeiros bancos de dados surgiram no final dos anos 70. Eram processados em grandes equipamentos e respondiam às necessidades informacionais de grandes empresas e corporações. Bastante complexos, necessitavam de programação e pessoal especializado para operá-los.

Com o lançamento dos PCs - *personal computers* - tornou-se possível o processamento de informações em pequenos aparelhos de mesa, embora o volume dessas informações fosse ainda bem pequeno, comparado com o que é possível nos dias atuais. Para se ter uma idéia, nos anos 80, o *software* dBase, líder dos bancos de dados relacionais para microcomputadores, operava em plataforma DOS e sua versão III Plus ocupava dois disquetes de 360 KB e podia ser rodada a partir de unidades

de disquete. A versão IV, lançada em 1988, saltou para 3,8 MB. Acontece que, naquela época, apenas cerca de 30% dos microcomputadores – modelo XT - eram equipados com disco rígido, que, na sua maioria, tinham apenas 20 MB. Portanto, a nova versão do banco de dados dBase teve que esperar o mercado se equipar com PCs mais potentes.

Um artigo publicado em 1994 na Revista Informática Exame⁵⁸ nos dá a dimensão de um passado recente que, pelos parâmetros atuais, parece remoto....:

“Há dez anos um banco de dados era utilizado de forma monousuária para atender no máximo às necessidades de um departamento. Um controle de estoque ou sistema de faturamento eram manuseados em seus departamentos por apenas um usuário de cada vez. Aos poucos a tecnologia de redes locais se estabeleceu e passou a exigir um novo patamar de características dos bancos de dados. Os sistemas passaram a ser multiusuários para atender às necessidades corporativas.(...) O *software* ACCESS, lançado no final de 1992, primeiro produto especializado em bancos de dados produzido pela Microsoft, aportou no mercado com a missão de se tornar uma espécie de dBase dos anos 90, ou o dBase da plataforma Windows. (...) A facilidade promovida pelo ambiente gráfico faz com que os bancos de dados dBase e Clipper, que detinham, sob o DOS, a maior base instalada junto a usuários e programadores, não tenham perfil coerente com o banco de dados da década de 90. A interface gráfica permite ao usuário o acesso a todos os elementos do sistema através da utilização de mouse, ícones e caixas de diálogo.(...) Idealmente, os novos bancos de dados devem oferecer acesso a qualquer tipo de base de dados e em diferentes plataformas. Esse ponto é muito importante para empresas que tiveram o desenvolvimento da informática ocorrendo de forma independente em máquinas de grande porte e em micros e que têm agora de integrar ou compartilhar dados.”

Podemos falar em alguns tipos de bancos de dados:

⁵⁸ Revista Informática Exame, ano 9 – n. 98 maio/1994 p.46-56

1. Banco de dados hierárquico: composto de tabelas, seus registros estão relacionados de uma maneira imutável, ou seja, não é possível estabelecer relacionamentos novos. Neste caso os relacionamentos são sempre do tipo um-para-um ou um-para-muitos;
2. Banco de dados relacional: os dados são armazenados em mais de uma tabela e é possível estabelecer relacionamentos entre as tabelas, baseando-se nas informações em comum, e gerando novos registros. Este tipo de banco de dados é apropriado para o relacionamento muitos-para-muitos. Os bancos de dados relacionais empregam uma linguagem de consulta padronizada para acesso a banco de dados;
3. Banco de dados direcionado a objeto: manipula dados, ou conjunto de dados, como “objetos” que apresentam atributos.

Um banco de dados é, na verdade, a representação de uma realidade complexa. Se o aplicarmos num contexto amplo, ele deve simular uma situação em que vários elementos se relacionam para produzir um resultado. Para ser representativa de todas as situações bem como das interconexões desejadas, a arquitetura de um sistema de banco de dados necessita ser bem definida desde o início. Ao se construir um sistema, é necessário, além do conhecimento técnico das ferramentas de banco de dados, um profundo entendimento da realidade que este sistema irá representar.

De forma bem simplificada, tomemos a seguinte situação como exemplo: ao informatizar o acervo da videoteca Multimeios, gostaríamos que ele ficasse interligado a uma hemeroteca formada por artigos sobre cinema publicados em jornais e revistas. Neste caso, ao escolher um *software* que gerenciasse tanto a base de dados da Videoteca Multimeios como a base de dados formada pelos textos da hemeroteca, teríamos que escolher um programa de banco de dados relacional, para estabelecer relações entre um filme e uma crítica sobre ele.

Como veremos no Capítulo 3, o fato de não termos escolhido um gerenciador de banco de dados mais robusto no início da automatização da Videoteca, premidos por limitações de várias ordens (computador PC XT incapaz de rodar um programa de banco de dados mais potente), limitou

suas possibilidades de expansão e relacionamentos e criou uma série de problemas no momento da migração para um banco de dados relacional e operacionalização para a plataforma *Web*.

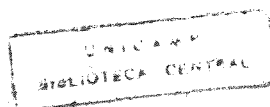
O conceito de rede em informática veio trazer maior complexidade ao já complexo universo dos bancos de dados.

Podemos falar em dois tipos de redes: as redes locais, que interligam computadores relativamente próximos uns dos outros, situados na mesma sala ou, no máximo, no mesmo prédio; e as redes remotas, projetadas para conectar computadores distantes uns dos outros, em prédios diferentes, em cidades e mesmo países diferentes.

A comunicação entre computadores, que permite receber ou transmitir dados ou arquivos, pode ser feita através de linhas telefônicas ou através de cabos ligando diretamente um computador a outros. Para que computadores troquem informações, é necessário tanto um meio de transmissão físico como também devem executar um programa que permita a comunicação entre eles. Embora seja possível que os computadores façam intercâmbio de informações usando ondas de rádio, é mais comum o uso de um simples cabo que liga computadores próximos, ou uma rede local que liga vários computadores num único escritório ou prédio, ou ainda os canais de fibra ótica ou de microondas do sistema telefônico que ligam computadores em longas distâncias. A conexão por sistema telefônico é atualmente usada porque os fios já estão instalados praticamente em todos os lugares onde colocamos um computador.

Embora qualquer um destes meios físicos possa estabelecer um caminho de dados entre dois computadores, é necessário um programa de comunicação - um sistema operacional de rede - para estabelecer o método específico de comunicação e para selecionar e enviar dados ou receber dados e armazená-los no próprio disco. Além do sistema operacional, a comunicação via-rede é objetivada pela intermediação de um modem:

“A maioria dos sistemas telefônicos é projetada para carregar sinais de frequência de voz que variam continuamente. Como transmitir bits de dados computacionais através de linhas telefônicas? É aí que entra o papel do modem (o nome “modem” é uma abreviatura de MODulador/DEModulador).



Os computadores transmitem e recebem seus fluxos de bits através de um modem. Este dispositivo permite que se comuniquem imprimindo (ou modulando) dados computacionais num sinal transportador de frequências de voz. O modem na extremidade receptora reverte este processo e restaura os dados do computador para sua forma original. As velocidades de transmissão de modem são medidas em bits por segundo (bps). Todos os modems realizam as mesmas funções básicas de modulação/demodulação, mas diferem na velocidade de transmissão de dados, nos conjuntos de comandos de computador que entendem (tais como ajustar a velocidade de transmissão, discar um número, desligar), nos recursos extras que possuem e no fato de residirem dentro ou fora do computador.”⁵⁹

A operacionalização de banco de dados em rede está ligada a um desenvolvimento da indústria da computação que envolve o conceito de cliente/servidor.

Mais uma vez de forma bem simplificada:

“um servidor de banco de dados é um computador que roda um Sistema de Gerenciamento de Banco de Dados em uma linguagem estruturada de perguntas, por exemplo, em SQL (*Structured Query Language*). Os computadores-clientes enviam pedidos em instruções SQL ao computador-servidor de banco de dados, que por sua vez acessa o banco de dados armazenado para processar e retornar os resultados ao computador cliente”.⁶⁰

Quando esta ação se dá no espaço virtual da *Web*, o computador-cliente roda uma aplicação para interagir com o banco de dados através de programas específicos para a *Web*, como o *Netscape* ou o *Internet Explorer*.

Observe no diagrama do Capítulo 3 como esta operação é realizada no caso específico da Videoteca Multimeios.

⁵⁹ Rubin, Charles, *Usando Microsoft Works*. Tradução: Insight Serviços de Informática, 1991, Editora Campus Ltda.

⁶⁰ Lowe, D., 1995.

O conceito de rede trouxe mudanças inimagináveis às funções já desempenhadas pelos primeiros bancos de dados, com a possibilidade de acessar arquivos em repositórios distantes e relacionar dados hospedados em diferentes bases de dados.

A recuperação de informações no ambiente *Web* é efetivada através dos chamados mecanismos de busca (*search engines*):

“As ferramentas de busca descendem do “grande pai” dos utilitários de busca em bancos de dados, o WAIS (*Wide Area Information Server*), baseado no protocolo Z39.50, um padrão para o formato de banco de dados e utilitários de busca concebido há mais ou menos uma década. Os utilitários de busca trabalham todos de modo semelhante e são constituídos por duas partes principais: o indexador e o mecanismo de busca. O indexador examina todo o seu banco de dados - isto é, todos os documentos HTML, ASCII (*American Standard Code for Information Interchange*) ou outros no seu *site WEB* - e cria um arquivo (geralmente chamado de fonte) que atua como um índice de conteúdo para todos os dados. Este arquivo fonte anotarà as palavras que existem no arquivo de dados original, onde estão, quantas vezes aparecem e onde encontrá-las. Quando um usuário pede que localize um certo termo, o mecanismo de busca procura o termo na fonte, o que é muito mais eficiente do que pesquisar em todo o *site*.”⁶¹

Uma ferramenta de busca que varre automaticamente a WWW através dos endereços URL (*Uniform Resource Locator*) é chamada *Robot*. Existem diferentes tipos de *robots* que executam diferentes tarefas. Um *robot* programado para medir a extensão e crescimento da *Web* é conhecido pelo nome de Wanderer. Um *robot* programado para obter referências sobre documentos específicos (de áreas específicas, por exemplo, medicina, artes, computação, etc.) é conhecido pelo nome Knowbot. Um *robot* que cria automaticamente um banco de dados a partir de links *Web* é chamado Spider.

⁶¹ Revista Internet World, número 9, Maio 1996, p. 60-62

A função do *robot* é coletar informações sobre os documentos hospedados nos *sites* que ele visita; estas informações são indexadas e colocadas em uma base-de-dados que pode ser consultada usando-se um mecanismo de busca com interface para o usuário.

Nos dias atuais notamos que a maioria dos desenvolvedores e comerciantes de programas de bancos de dados têm se preocupado em construir versões *Web* para seus produtos, ou seja, têm desenvolvido ferramentas de busca compatíveis com os protocolos *Web*. Isto, todavia, não elimina a necessidade de programação suplementar para a conexão entre o ambiente *Web* e um banco de dados específico. Examinaremos, no capítulo em seqüência, como isso ocorre no caso de nosso projeto.

Capítulo 3

Videoteca Multimeios na Internet

3.1. Formatando a arquitetura

3.1.1. Estrutura do sistema : diagrama e descrição

A rede Ethernet da Universidade Estadual de Campinas, UNICAMP, é formada por um conjunto de sub-redes independentes que servem a diferentes unidades. Quatro tecnologias compõem o sistema de redes: ATM, FDDI, Ethernet 10mbps (*megabits per second*) e FastEthernet 100mbps.

A sub-rede do Instituto de Artes é uma rede Ethernet com 10 mbps de taxa de transmissão na parte interna e conexão com o *backbone* da UNICAMP em Fast Ethernet de 100 mbps.

A especificação do sistema usado pelo banco de dados Videoteca no Instituto de Artes é a seguinte:

O servidor IAR é um Pentium 350, Disco Rígido de 9.2 Gigabytes;

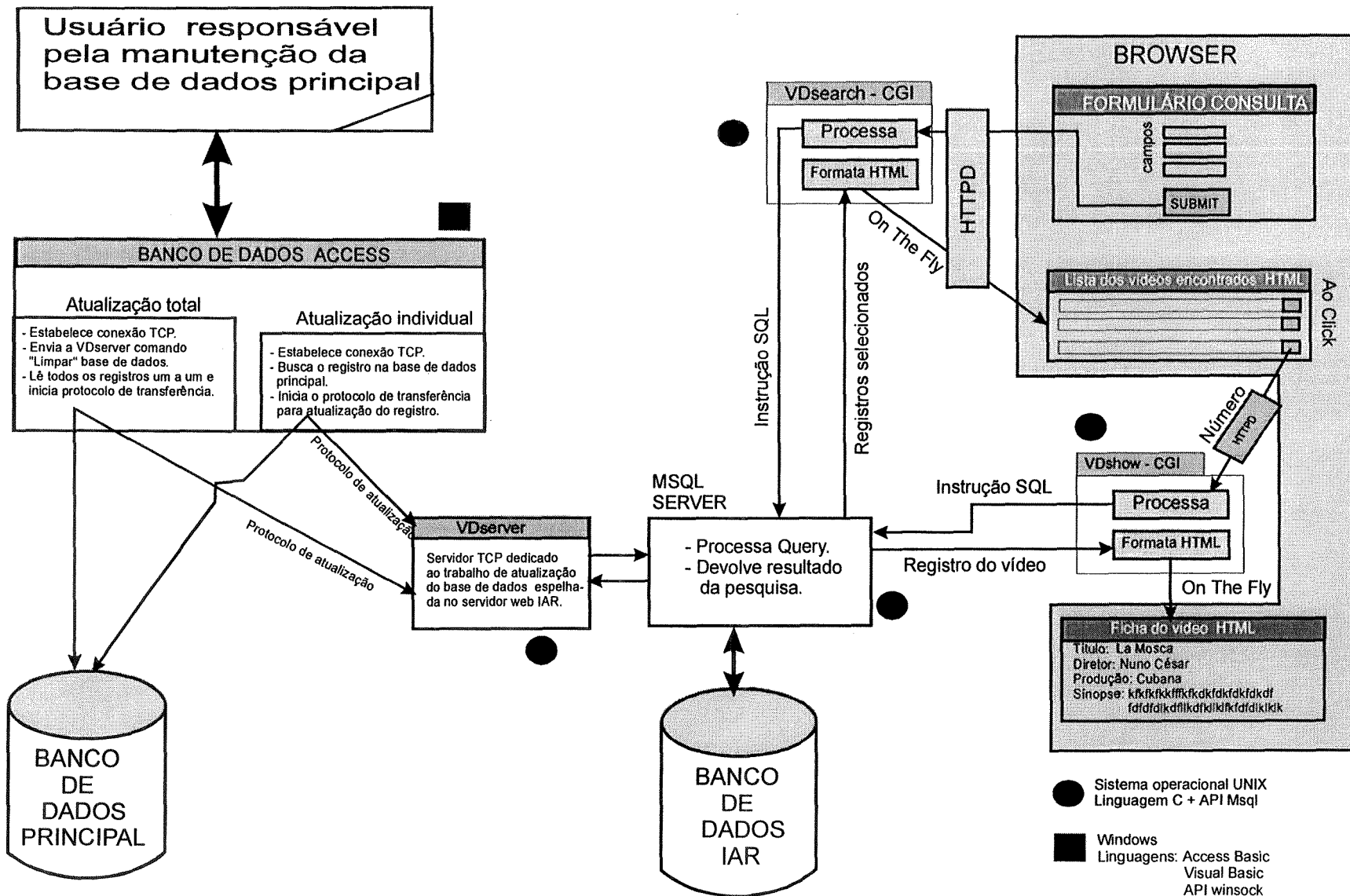
O servidor *Web* Apache roda Unix como sistema operacional;

O servidor SQL utilizado para gerenciar o banco de dados na *Web* é um Mini-SQL.

Todos os procedimentos para a implantação e operacionalização do sistema foram escritos pelo administrador de redes do Instituto de Artes.

Vejamos como funciona o sistema do banco de dados da Videoteca Multimeios.

DFD - Diagrama do Fluxo de Dados
Banco de dados da videoteca do Departamento de Multimeios



O lado esquerdo do gráfico mostra a atuação de quem alimenta e atualiza o banco de dados da Videoteca. O banco de dados Principal, localizado fisicamente no servidor do Laboratório Midiatec, Departamento de Multimeios, usa um programa construído em Microsoft Access. Ali é feita a alimentação e atualização dos registros em vídeo no formato ficha técnica, bem como são inseridos os registros numéricos das fotos, dos arquivos-texto de críticas ao vídeo e os endereços *Web* dos *hyperlinks* e dos vídeos. Esta base (ou os registros atualizados) é transferida para a base de dados *Web* (espelho da base de dados Principal) no servidor banco de dados IAR, localizado fisicamente em sala própria do Instituto de Artes.

Esta transferência foi, numa primeira fase, feita via-ftp (*File Transfer Protocol*). Mais tarde, ao programa em Access foram acrescentados procedimentos utilizando protocolo TCP/IP para estabelecer conexão com outro programa, escrito em linguagem C, no ambiente Unix onde fica a base de dados *Web*. O objetivo destes *scripts* é enviar remotamente qualquer registro atualizado no banco de dados Principal para o servidor SQL, que é o sistema gerenciador na base de dados *Web* (também chamado banco de dados IAR). Este procedimento simplificou bastante a tarefa de atualização do banco de dados⁶².

Olhando no lado direito do gráfico, vemos como se processa a interação entre o usuário, em consulta ao *site Web* da Videoteca, e o sistema.

A interface *Web* opera no modo cliente-servidor manipulando hipertextos codificados em linguagem HTML e acessível através dos programas-clientes (*browser*, como Netscape ou Internet Explorer). Imaginemos uma situação onde o usuário faz uma consulta *online*, por exemplo, digita “Hitchcock”, clica em “Diretor” e coloca o ano de 1963. Sua pergunta é enviada ao servidor *Web* (HTTPD) que, através de um programa CGI⁶³ desenvolvido exclusivamente para esta tarefa – VDSearch -, processa a pergunta em instruções SQL e a envia ao servidor Mini-SQL. Este, por sua vez, processa a pergunta restringindo a busca no banco de dados IAR apenas a filmes de 1963 que tenham Hitchcock como diretor; devolve

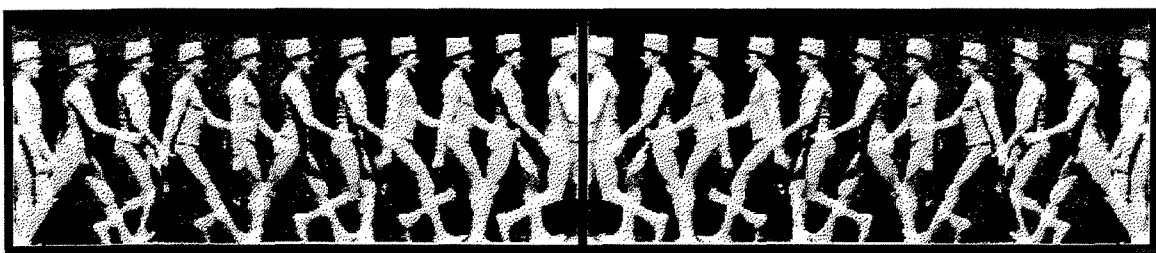
⁶² ver ilustração na página 64.

⁶³ CGI (Common Gateway Interface): padrão internacional para executar processos externos. CGI scripts são programas que rodam no servidor desempenhando tarefas específicas, principalmente para estender as possibilidades da *Web*.

então o resultado da pesquisa ao VDSearch, que formata em HTML o resultado com a lista do(s) vídeo(s) encontrado(s).

Esta lista não mostra a ficha completa dos registros. Caso o usuário queira ver a ficha completa, novos procedimentos entram em ação:

o servidor *Web* (HTTPD) é novamente chamado e aciona um programa escrito em linguagem C – VDSHOW – que, através de uma instrução SQL, aciona o servidor MSQl que, por sua vez, vai buscar no banco de dados IAR a ficha completa do registro em vídeo solicitada. Esta ficha é formatada *on-the-fly* pelo programa VDSHOW que a apresenta em HTML no *browser* do usuário.



VIDEOTECA MULTIMEIOS

[projeto](#) | [busca](#) | [utilização](#) | [vídeos](#) | [links](#)

Há quatro possibilidades de recuperação da informação no BD Videoteca:

- qualquer palavra, ou combinações de palavra (utilizando lógica booleana) nos campos título, diretor, produção e sinopse;
- pesquisa por gênero, exclusivamente, obtendo um panorama geral de todos os títulos no gênero escolhido;
- pesquisa por ano de produção, exclusivamente, obtendo um panorama geral de todos os títulos realizados no ano escolhido;
- pesquisa combinando os campos gênero e/ou ano de produção com somente um dos seguintes campos: título, diretor, produção ou sinopse.

**SELECIONE**

☐ Título ☒ Diretor ☐ Sinopse ☐ Produção ☐ Todos

Gênero ▼ Ano de produção:
Como procurar A frase inteira ▼

O QUE PROCURAR**PROCURAR**

Na busca por "Título", escreva tudo em maiúsculo.

Na busca por "Diretor", escreva somente as iniciais em maiúsculo.

Na busca por "Sinopse", escreva qualquer palavra-chave, ou combinações de palavra, em minúsculo.

Instruções para a visualização dos vídeos em rede: RealNetworking :
RealVideo/RealAudio

Para visualizar os clips de vídeos, instale gratuitamenteo RealPlayer, última versão, da Progressive Networks: <http://www.real.com/products/player/index.html> e escolha a versão adequada ao seu sistema operacional.

Hardware mínimo: Pentium 90 multimídia com 16Mb de memória RAM

Conexão mínima: modem 28.8 Kbps ou rede local

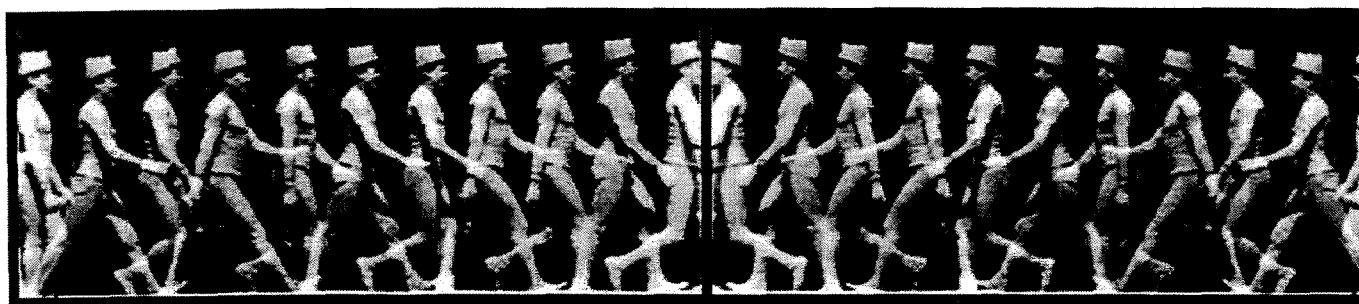


Visitante número

000799

O resultado da pesquisa pode ser apresentado de duas formas:

- em formato resumido, com o título, nome do diretor, ano de produção e um ícone para acessar a ficha técnica completa;



RELAÇÃO DE VÍDEOS ENCONTRADOS

TÍTULO	DIRETOR	ANO	
AMPEXO	Adilson Ruiz	1990	
CARLOTA AMOROSIDADE	Adilson Ruiz	1987	
CEM OSWALD ANOS - VIDEOVIDA DE UM POETA	Adilson Ruiz	1990	
EGO OU SÓ MESMO VENDO COMO É QUE DÓI	Adilson Ruiz	1990	
ENTREVISTA COM ROGÉRIO CERQUEIRA LEITE	Adilson Ruiz	1991	
ESPANHA DE MARIA	Adilson Ruiz	1992	
INFINITA TROPICÁLIA	Adilson Ruiz	1982	
MAKING OF RAPSÓDIA PARA CINEMA E ORQUESTRA	Flailda Siqueira/Adilson Ruiz	1991	
MULTIMEIOS	Adilson Ruiz/ Haydée Dourado	1984	
MÁRIO - UM HOMEM DESINFELIZ - Série Asp. Cultura Bra	Adilson Ruiz	1993	
RAPSÓDIA PARA CINEMA E ORQUESTRA	Adilson Ruiz	1992	
UNIVERSIDADE NA ERA DOS MULTIMEIOS, A	Adilson Ruiz	1986	
UZE Brioloco	Adilson Ruiz	1990	
VAGALUME	Alice Ruiz	1996	



- detalhado, mostrando as informações constantes em todos os campos existentes na ficha técnica, além de ícones para fotos, fragmentos de vídeo e *links* que remetem a críticas sobre o vídeo e/ou outros *sites* na Internet.

INSTITUTO DE ARTES VIDEOTECA

Acervo:MultiMeios

Título:UZEBRIOLOCO

Diretor:Adilson Ruiz

Produção:Studio AR Cinema Vídeo

Local:BR/SP

Duração:9'

Ano de produção:1990

Bitola:35 mm

Sistema:PalM/cor **Idioma:**português

Gênero:experimental **Classificação:**009 **Procedência:**Telec.C.Com.Unicamp

Informações Técnicas

Sinopse

Filme-cantata que se remete ao "Ébrio", de Vicente Celestino, interpretado pelo dramaturgo José Celso Martinez Correa. Uma espécie de biografia metafórica do diretor José Celso, nos anos 80.



80k



3.1.2. Composição multimídia da base de dados

Fazem parte do banco de dados Videoteca Multimeios:

- as fichas catalográficas com informações sobre cada título em vídeo;
- fotos significativas referentes aos filmes, pesquisadas em livros e escolhidas pela equipe da Videoteca;
- fragmentos de vídeo (imagem e som) ou, em certos casos, o vídeo todo;
- artigos de jornais (e/ou revistas) em hipertexto;
- *links Web* remetendo a *sites* afins.

3.2. Implantando o projeto

3.2.1. Definição de um universo representativo para amostragem da pesquisa

A escolha dos filmes em vídeo para o projeto-piloto obedeceram ao critério de primeiramente dar visibilidade às produções do Instituto de Artes, em seguida chamar a atenção para aquisições exclusivas do acervo e, por fim, divulgar os grandes mestres da cinematografia.

Nesta ordem, foram escolhidos os títulos:

Morayngava, de Regina Muller (professora no Departamento de Artes Corporais), e Virgínia Valadão;

Rapadura, Taiada, Melado, produção conjunta de alunos do Departamento de Multimeios com coordenação de Marcelo Costa Souza;

Uzebrioloco, de Adilson Ruiz (professor no Departamento de Multimeios)

Corisco e Dadá, de Rosemberg Cariri;

Cangaceiro, de Lima Barreto;

Ti Oluwa Ni Ile, de Tunde Kelani (produção do cinema nigeriano⁶³);

Almonds & Raisins, a History of the Yiddish Cinema, de Russ Karel;

Le Golen, de Julien Duvivier;

The Dybuk, de Michal Waszynski;

The Golden Age of Second Avenue, a Loving Chronicle of Yiddish Theatre in America, de Morton Silverstein;

⁶³ O Departamento de Multimeios promoveu, em 1995, um seminário sobre cinema nigeriano com a presença do professor Onookome Okomo.

Yidl Mitn Fidl, de Joseph Green;⁶⁴

The Lodger, de Alfred Hitchcock;
Os Pássaros, de Alfred Hitchcock.⁶⁵

O *site* da videoteca na *home-page* do Instituto de Artes passou a contar com uma chamada para a divulgação mensal de produções escolhidas e/ou novas aquisições, com espaço para comentários sobre os filmes escolhidos.

3.2.2. Tecnologias usadas: apresentação e discussão

3.2.2.1 Banco de dados

A base de dados da Videoteca Multimeios, iniciada no final de 1992 no *software* Microsoft Works para MS-DOS, versão 2.0, teve que passar por modificações para receber uma interface WWW. Esta primeira versão do Works foi transportada para o Works for Windows sem maiores problemas. Entretanto, por armazenar somente informações textuais e ser um banco de dados não relacional, deixou de ser adequado aos novos objetivos propostos, quais sejam, disponibilizar na Internet sua base de dados acrescida de imagens, vídeos, textos de jornais e *hiperlinks*. Dessa forma, foi efetivada a conversão para o *software* Microsoft Access, um *software* para banco de dados relacional. Esta operação foi bastante trabalhosa.

Salvamos a base de dados em *arquivo.dbf* dBase III. O campo sinopse foi cortado em parte pois, tendo cerca de 300 caracteres no Works, foi exportado com apenas 51. Além disso, devido à existência de alguns campos sinopse sem registro, o transporte automático foi feito sequencialmente tirando, portanto, da ordem e preenchendo os campos que estavam vazios (sem sinopse) em ordem crescente. Como resultado, nenhuma sinopse mesclou acertadamente com seu devido registro de filme. Após várias tentativas de acertar as sinopses manualmente, optamos por criar um campo numérico para cada registro que servisse de referência para a nova migração.

⁶⁴ Estes títulos em vídeo foram doados pela coordenação do projeto *Second Avenue Online: The Yiddish Theater Digital Archives*, New York University, Center for Advanced Technology, New York.

⁶⁵ A Videoteca Multimeios conta com a filmografia quase completa de Alfred Hitchcock.

Na segunda tentativa usamos outro procedimento: salvamos em 2 partes: a) o banco inteiro sem o campo sinopse, como arquivo.dbf , e, b) somente a sinopse com o campo numérico (que criamos para mesclar na junção), como arquivo.txt, salvando como “ texto e tabulações – DOS”.

Foi feito um programa em Clipper para ler arquivo.txt e gravá-lo em arquivo.dbf, quebrando em 4 campos de 100 caracteres cada . Foi feita a vinculação do arquivo.dbf no Microsoft Access através de um programa em Access Basic, que faz a união dos 4 campos da sinopse em um só.

Finalmente resolvido, foram criados no banco de dados Access os campos fotos, vídeos, críticas e links. Naquele momento não pensamos em criar distintos arquivos de dados para os diferentes registros – fotos e críticas – pois o conceito e, principalmente, a operacionalização de bancos multimídia relacionais era ainda desconhecido para nós. Apenas um ano depois é que tomamos contato pela primeira vez com projetos e discussões sobre o tema.

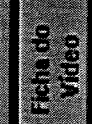
VIDEOTECA

Departamento de MultiMeios

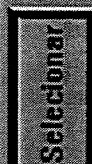
Título	Diretor	Class	Gênero	Ano	Número
MORANGOS SILVESTRES (MILD STRAWBERRIES)	Ingmar Bergman	177/451	drama	1957	454
MORAYNGAVA	Regina Müller/Virginia Valadão	515	documentário	1997	789
MORDENDO LÁPIS	Maria Leandra Bizello	071	empimental	1990	455
MORONETA	Etienn Samain/Murilo J. Machado	038	documentário	1984	456
MORONETA - (Reportagem)		038	entrevista		457
MORTE CANSADA, A	Fritz Lang	743	drama	1921	1055
MOUR DUBEN: MA VIE DE TSIGANE MANOUCHE	Yasuhiro Oimori	340	drama	1974	871
MULHER 90	Daniel Filho/Roberto Talma	089	entrevista	1990	458
MULHER DO ATRADOR DE FACAS, A (Curtas Gramado 89)	Nilson Villas Boas	338	comédia	1989	176
MULHER DO FAZENDEIRO, A (THE FARMER'S WIFE)	Alfred Hitchcock	762	drama	1928	1081
MULTIMEOS	Adilson Ruiz/Haydée Dourado	028	documentário	1984	459
MULTIPLICAÇÃO DE MARIA, A	Fernando de Tacca	268	documentário	1995	460
MUNDO NO AR, O (V Fest. Fotóptica)	Olhar Eletrônico	062	documentário	1987	461
MUSEU DA IMAGINAÇÃO	Milton Jesus/Geraldo Porto	091	documentário	1989	462
MÚSICO E O CAVALO, O	Telmo Carvalho	797	animação	1986	1125
NA DÚVIDA, SONHE COMIGO	Alzira Pimenta	070	ficção	1990	463
NÁGUA	Má Fagundes/Silvia Laurentiz	096	video arte	1992	464
NANA	Jean Renoir	375/437		1926	465
NANOOK, O ESQUIMÓ (NANOOK OF THE NORTH)	Robert Flaherty	286/SV004	documentário	1922	466
NÃO SOMOS COMPUTADORES		119	documentário	1988	467
NAPOLEON	Abel Gance	060	drama épico	1927	468
NAPOLEON	Abel Gance	201	drama épico	1927	469
NAPOLEON (2 Épocas)	Sacha Guitry	356/411	drama épico	1955	470



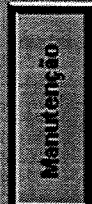
Imprimir Fiche



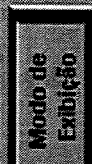
Fiche do Video



Selecionar



Manutenção



Modo de Exibição



Transporte para WEB



Ordenar por:

Departamento de MultiMeios

Título	Diretor	Class	Gênero	Ano	Número
--------	---------	-------	--------	-----	--------

BANCO DE DADOS DA VIDEOTECA - Atualização ON-LINE

TÍTULO:	MORAYNGAVA	Número	789
DIRETOR:	Regina Müller/Virginia Valadão		
PRODUÇÃO:	UNICAMP/Centro de Trabalho Indigenista		
ANO:	1997	ESTÉTICA	
GÊNERO:	documentário		
LOCAL:	BR		
DURAÇÃO:	16'	SINOPSE	
BITOLA:	Betacam	Morayngava, o "desenho das coisas", Yngiru, a "caixa das almas". Os filmes, sonhos de pajé. Assim os índios Asurini definem o vídeo, recém chegado em sua aldeia. Ao descobrirem que é possível guardar suas imagens, os velhos lamentam não ter gravado seus antepassados, mas resolvem registrar a iniciação de	
SISTEMA:	NTSC/COR		
IDIOMA:	asurini/portugu		
CLASSIFICAÇÃO:	515		
PROCEDÊNCIA:	doação diretoras		
ACERVO:	Multimeios		

Microsoft Windows Sockets Version
Running on Windows 95.
versão=1 Maior versao=1
Número de sockets=256
Tamanho do datagrama=65467

143.106.55.66

Enviar registro

Abortar

FOTOS	789.jpg	CRÍTICAS	
		VÍDEO80k	http://www.canalweb.unicamp.br/
		VÍDEO20k	

LINKS

NAPOLEON (2 Épocas)

Sacha Guitry

356/411

drama épico

1955

470



Ficha do Vídeo

Selecionar

Manutenção

Modo de Exibição



Imprimir Ficha

Refazer base de dados na WEB

Imprimir Listagem

Transporte para WEB

Ordenar por:

VIDEOTECA

Banco de dados da Videoteca

videoteca

Número: 129

TÍTULO: MORAYNGAVA

DIRETOR: Regina Müller/Virgínia Valadão

PRODUÇÃO: UNICAMP/Centro de Trabalho Indigenista

ANO: 1997

GENERO: documentário

LOCAL: BR

DURAÇÃO: 16'

BITOLA: Betacam

SISTEMA: NTSC/COR

IDIOMA: asurini/portugu

CLASSIFICAÇÃO: 515

PROCEDÊNCIA: doação diretores

Acervo: Multimídios

FOTO1

FOTO

VIDEO 20K

VIDEO 80K
http://www.canalweb.unicamp.br/nr

LINKS

CRITICAS

ESTÉTICA

SINOPSE
Morayngava, o "desenho das coisas", Yngiru, a "caixa das almas". Os filmes, sonhos de palé. Assim os índios Asurini definem o vídeo, recém chegado em sua aldeia. Ao descobrirem que é possível guardar suas imagens, os velhos lamentam não ter gravado seus antepassados, mas resolvem registrar a iniciação de

MOURDUBEN, MA VE DE TSOANE MANOUCHE Yasuhiro Umori 1974 871

MULHER 90 Daniel Filho/Roberto Talma 1990 458

MULHER DO ATRADOR DE FACAS, A (Curias Gramado 89) Nilson Vilas Boas 1989 176

MULHER DO FAZENDERO, A (THE FARMER'S WIFE) Alfred Hitchcock 1928 1081

Imprimir Ficha **Refazer base de dados na WEB** **Imprimir Listagem** **Transporte para WEB** **Ordenar por:**

Associar Críticas **Ficha do Vídeo** **Selecionar** **Manutenção** **Modo de Exibição**

3.2.2.2 Imagem fotográfica

Após a escolha de fotos representativas de cenas de filmes ou contextos onde a informação do registro do filme se insere, faz-se o escaneamento da imagem com resolução de 150 dpi (*dots per inch*) e tratamento através do *software* Adobe Photoshop para melhoria do contraste, luminosidade e finalização com moldura e/ou montagem. A imagem recebe então compressão de 50% no formato jpg (*Joint Photographic Experts Group*). Cada arquivo de foto deve ter no máximo 150K (kilobytes) para não demandar tempo de espera ao ser carregada no computador cliente. Cada foto que faz parte da base de dados recebe um número correspondente ao número do registro que a referencia e é esse número que integra o campo foto no banco de dados. O arquivo.jpg é transferido via-ftp para um sub-diretório de fotos no diretório Videoteca, dentro da raiz do servidor *Web HTTP*.



Marcos Antônio e Benedito
posam para foto de Benjamin Abrão



Dina Paes e Chico Diaz: "Corisco e Dadá", de Rosemberg Cariry

3.2.2.3. Hemeroteca: tecnologia OCR

Como parte do projeto-piloto, alguns dos textos da hemeroteca foram escolhidos dentro do universo representativo do projeto. Foram escaneados com resolução de 300 dpi (*dots per inch*) com aplicação da tecnologia OCR (*Optical Character Recognition*). Dois *softwares* para OCR foram testados: TextBridge Classic e OmniPage. Como esta conversão ainda não está bem solucionada, o texto resultante aparece com muito “lixo” de caracteres não reconhecidos devidamente, sendo necessárias correções manuais no processador de texto Word. Isto feito, podem então ser transformados em hipertexto, salvando o arquivo.rtf em arquivo.html, podendo receber *hiperlinks*. Repetimos então o mesmo procedimento descrito anteriormente para as fotos: cada um desses arquivos recebe um número correspondente ao registro na base de dados a que ele referencia e este número vai integrar o campo “críticas” no banco de dados. O arquivo.html é transferido via-ftp para um sub-diretório de críticas no diretório Videoteca, dentro da raiz do servidor HTTP.

Para aprender sobre esta tecnologia consultamos o projeto *The Rosenberg Papers - OCR for Scanned Images of Legal Documents*.⁶⁶

3.2.2.4. Links

Foram feitas algumas pesquisas na rede usando vários mecanismos de buscas e encontramos referências adicionais aos filmes de Hitchcock, aos filmes nacionais apontados (Corisco e Dadá e O Cangaceiro), e a algumas produções do cinema Yiddish. Estes *hiperlinks* têm seu endereço http registrado na base de dados dentro do campo “links” e são transferidos ao servidor IAR no procedimento de atualização da base de dados. Muitos outros *hiperlinks* estão sendo colocados porém, como não se referem ao universo de filmes aqui delimitado, não nos reportaremos a eles.

3.2.2.5. Fragmentos de vídeos

Até o presente momento, a digitalização de vídeos ou de fragmentos de vídeos tem sido feita em parceria com o Centro de Computação e Centro de Comunicação da UNICAMP. Os arquivos estão armazenados no servidor

⁶⁶ *The Rosenberg Papers – OCR for Scanned Images of Legal Documents*. [Online].
<http://www.columbia.edu/acis/cria/rosenberg/>

Canal *Web* e recebem um endereço *http* que é colocado no campo vídeo do banco de dados.

Como apresentado no Capítulo 2, a tecnologia para vídeo na Internet ainda faz face a muitos problemas. Tamanho excessivo do arquivo gerado (pedindo, portanto, um servidor robusto para armazenamento), necessidade de alta compressão com conseqüente perda de informação e qualidade de imagem são algumas das dificuldades concretas encontradas. No momento, não contamos com a possibilidade de armazenamento nos servidores do Instituto de Artes e a atual política de disponibilização de vídeos em rede da UNICAMP ainda não contempla grande quantidade de vídeos a serem colocados na *Web*.

3.3. Limitações/ Possibilidades para transformação

Como já dissemos no início deste trabalho, a maneira como foi estruturado o banco de dados, com uma só tabela contendo todos os campos para registros de formatos diferentes (ficha catalográfica, fotos, críticas, vídeo e hiperlinks), impôs limites à sua expansão, amarrou à ficha catalográfica do vídeo as informações suplementares provenientes dos artigos de jornal e das fotos, não permitindo uma busca e recuperação da informação multimídia. Toda e qualquer informação constante de artigos de jornal que está na base de dados, repousa desconhecida até o momento em que alguém consulte a ficha catalográfica do filme onde ela está aprisionada. O mesmo dizemos para as fotos referentes a filmes. Também comentamos que na época não tínhamos nenhum modelo onde nos espelhar para criar um verdadeiro sistema multimídia no formato banco de dados. Estamos já direcionando nossos esforços visando à sua reformulação imediata.

Uma das possibilidades de modificá-lo será a criação de duas novas tabelas dentro do Access, para cada um dos formatos: fotos e hipertexto de críticas. Estas tabelas serão estruturadas em campos de registro e receberão registros completos sobre cada item, nos moldes usados pela Biblioteca do Congresso dos Estados Unidos, de forma a deixá-las preparadas para serem integradas em quaisquer sistemas. Isto fará com que cada um desses campos se torne recuperável por um mecanismo de busca, além de ter sua informação cruzada com outros campos de outros registros de tabelas diferentes. Para cumprir essa reformulação, todas as programações feitas

em C, SQL e CGI, apontadas na descrição do diagrama acima, terão que ser refeitas, dando novos parâmetros de busca, recuperação e interface *Web*.

Sabemos da existência de *softwares* planejados para a integração de bases de dados e que já se encontram disponíveis no mercado comercial. Muitos museus nos Estados Unidos têm usado estes sistemas prontos, ou semi-prontos. Eles já vêm formatados para atender a grande parte das necessidades de tratamento da informação e variam na forma de apresentação, ênfase a aspectos, sistema aberto ou fechado (que determina a possibilidade de adaptação às especificidades do cliente) e robustez do banco de dados. Entretanto, não faz parte de nossa realidade optar por um sistema desses, pois não temos a complexidade de acervo de um museu, além do preço destes produtos os tornarem inviáveis para a nossa realidade imediata. Ainda que não optemos por eles, é de grande valia conhecermos sua estrutura e abrangência. Notemos que o sistema contempla o registro de livros, caixas de documentos em arquivo, objetos e coleção fotográfica, como se pode notar pelos formatos de fichas de banco de dados usadas em um destes sistemas⁶⁷:

The screenshot shows the 'Photo Catalog' application window. At the top, there is a menu bar with icons for Prior, Next, Add, Browse, Delete, Edit, Find, Print, Source, and Exit. Below the menu bar, the main area is divided into several sections. On the left, there is a 'Collection' section with the title 'HCHS Photograph Collection'. Below this, there are fields for 'Source' (Child, Elmer A.), 'Accession#' (1977.007), 'Object ID' (1977.007.0019), 'Title' (Kailua Village 1908), 'Place' (Kailua/North Kona/Hawaii Is.), 'Photographer' (Child, L.C.), and 'StorageLoc' (Photo File). In the center, there are fields for 'Objname' (Print, Photographic) and 'Date' (1908). On the right, there are fields for 'Catalog Date' (04/21/1984), 'Status' (OK), 'Neg#' (00748;Dup), 'Frame#' (empty), 'Print Size' (9'x 12 1/2'), 'Film Size' (empty), 'Slide#' (empty), and 'Other#' (6320). Below these fields, there is a 'Description' section with a text area containing a detailed description of the photograph. To the right of the description, there is an 'Image Management' section with a small thumbnail image of the photograph and a file name 'On 1\19770070019.TIF'. At the bottom right, there is a timestamp '03/23/1998 08:48 AM'.

Collection	HCHS Photograph Collection		Catalog Date	04/21/1984	by	rlh
Source	Child, Elmer A.		Status	OK		
Accession#	1977.007	Objname	Print, Photographic	Neg#	00748;Dup	
Object ID	1977.007.0019	Date	1908	Frame#		
Title	Kailua Village 1908		Print Size	9'x 12 1/2'		
Place	Kailua/North Kona/Hawaii Is.		Film Size			
Photographer	Child, L.C.	Photographer info		Slide#		
StorageLoc	Photo File		Other#	6320		

Description | People 3 | Copyright | Notes | Classification | Content | Condition | Custom

Kailua village as seen from the second story of the Hackfeld building near wharf. Kaiakeakua beach is fitted out for cattle shipping with a temporary cattle holding pen. Elmer Child says that there was a larger holding pen up around the corner of Palani & Kuakini junction where the ranchers held cattle before moving them to this smaller pen. Approximately 60 cattle would be held there and approximately 10 would be moved at a time to await shipping. Three women in street. Pen was also used for sumo wrestling on boat days

Image Management

On 1\19770070019.TIF

03/23/1998 08:48 AM

⁶⁷ PastPerfect Museum Software. [Online]: <http://www.pastime2000.com/>

Library Catalog				27 Books Sorted by Title						
Prior	Next	Add	Browse	Delete	Edit	Find	Source	Print	Exit	
Collection	Hawaii County Historical Society Library					Cat Date	06/03/1997 by MLP			
Source	AltaMira Press					LCCN	78-15994			
Phys Desc	Bibliography: p.155, Includes index					ISBN	0-910050-37-6			
Accessn#	1997.001	Objname	Book			ISSN				
Object ID	1997.001.0010	Call#	AM139.R44			Status	On-Loan			
Storage Loc	Main Library					Condition	Good			
Author	Reibel, Daniel B.					Copy#	1 02/27/1998 11:02 PM			
Title	Registration Methods for the Small Museum									
2nd Title										
Subjects Publisher Summary Series - Edition Status Notes Custom Museum registration methods American Association of State and Local History Accessioning, Cataloguing 										

Archives Catalog				14 Archives Sorted by Object ID						
Prior	Next	Add	Browse	Delete	Edit	Find	Source	Print	Image	Exit
Collection	Records Kobayashi Sodaworks, Ltd.					Catalog Date	02/27/1998 by MLP			
Source	Kobayashi, Chiyono					Status	OK			
Accessn#	1984.013	Objname	Records			Condition	Excellent			
Object ID	1984.013.0001	Container	MSS Box 15-22			02/28/1998 08:51 AM				
Identity Statement Context Container List Structure Conditions of Access & Use Applied Materials Custom Title Records of the Kobayashi Sodaworks, Ltd. Dates of creation 1898-1941 Level of description Fonds Multilevel Linking Extent of the unit of description 4 linear feet contained in 8 document cases Storage Location HCHS Archives Room 6 Shelf 5 Scope and Content / Abstract Company records of the Kobayashi Sodaworks of Hilo Hawaii, 1918-1941. The files contain, correspondence relating to business affairs, personal correspondence, immigration documents, shipping and receiving invoices. Files also contain information relating to the U.S. Government shutdown of Japanese-owned businesses in Hawaii.										

Object Catalog - History

16 Objects Sorted by Object Name

Collection: **Wai'ahukini Archaeology Collection**
 Catalog Date: 08/28/1997 by MLP

Source: University of Kailua
 Status: OK

Accession#: 1997.002 Objname: Pounder, Poi
 Disp Value: Excellent

Object ID: 1997.002.0001 Other Name: Pohaku ku'i 'ai
 Early Date: A.D 500 +- 60

Material: Stone
 Late Date:

Storage Loc: Room 14 Cabinet 7 Drawer 3
 Temp Loc:

Site Name: Wai'ahukini, Kealahakua, Hawaii
 Site#: HA-B22-140
 Other Info: Op A - 01546

Prov.	Description	Dimensions & Notes	Condition	Maintenance	Related 2	Custom
Collector	Smith, Duncan PhD, PI Wai'ahukini Archaeology Proj					
Found	Op. A Grid Square 6 Strata B					
Made	Unknown					
Used	Food preparation					
Owned						
Pl of Origin	Hawaii					

Image Management

On 801/poipounder.tif

01/04/1996 04:54 PM

Key Word Search

Object Collection

☐ Photos
 ☒ Objects
 ☐ Archives
 ☐ Library

Search field: Material/Medium for: WOOD
 ☒ and
 ☐ or
 Search field: Accession# for: 1995

Objectid	Object Name	Material / Medium	Store Loc
1995.003.0001	Table	Wood/Varnish/Metal	Farm Storage Facility
1995.003.0011	Board, Ironing	Wood/Metal	Farm Storage Facility
1995.003.0019	Stool	Wood/Nails	Farm Storage Facility Shel
1995.003.0028	Barrel	Wood/Bamboo	Farm Storage Facility Shel
1995.005.0004	Machine, Sewing	Wood/Leather/Metal/Varnish	Farm Storage Facility
1995.005.0020	Brazier	Cast iron/Wood	Farm Storage Facility Shel
1995.005.0040	Scoop	Wood/Metal	Farm Storage Facility Shel
1995.035.0001	Safe, Food	Wood/Metal/Glass	Farm Storage Facility

Records searched: 16 Found: 8

Retomemos, por fim, os sete itens relacionados no capítulo 2, referentes a bibliotecas digitais e comentemos cada um deles segundo sua aplicabilidade em nosso projeto:

- **Arquitetura Técnica:** temos o ambiente de rede adequado, banco de dados relacional, softwares e servidores necessários. Falta-nos um servidor exclusivo para vídeos digitalizados se quisermos aumentar o número de fragmentos em vídeo a serem disponibilizados no banco de dados. Falta-nos também reescrever os procedimentos de integração para tornar nosso sistema multimídia.
- **Construção de uma coleção digital:** consideramos já ter uma coleção digital.
- **Digitalização:** temos usado formatos padronizados para a digitalização de documentos.
- **Metadado ou Metadocumento:** adotamos a convenção do *Dublin Core* no *site* da Videoteca bem como nos outros *sites* que compõem a *home-page* do Instituto de Artes. Continuaremos acompanhando discussões e pesquisas no exterior e no Brasil sobre o tema da padronização de metadocumentos.
- **Nome e endereço:** com a nova formatação de diferentes bases de dados para fotos e críticas de jornal, haverá a necessidade de novos endereçamentos destes documentos.
- **Direitos autorais:** acompanharemos as discussões sobre este tema polêmico bem como a política adotada na UNICAMP para a questão dos direitos autorais.
- **Preservação:** como estamos apenas replicando documentos existentes em outros suportes que não o digital, nossas preocupações neste campo se dirigem mais à conservação física dos vídeos e fotos que fazem parte de nosso acervo.

Considerações finais

Como objetivo inicial, esta pesquisa vislumbrou tirar partido dos recursos oferecidos pelo sistema de redes remotas, ampliando as possibilidades de consulta aos filmes da Videoteca Multimeios, disponibilizando em um único local, qual seja, o *site* da Videoteca na *home-page* do Instituto de Artes, informações complementares à ficha técnica de um filme. Usando a flexibilidade do dado digital e inserindo imagens e sons, buscamos ir além do formato-padrão de apresentação e organização da informação que prioriza a linguagem escrita.

Neste processo podemos dizer que o banco de dados da Videoteca conseguiu avançar em alguns aspectos quando comparado aos *sites* de informação sobre cinema, principalmente os brasileiros, ao fazer uso de recursos de hipertexto e multimídia oferecidos pela tecnologia de redes remotas.

Partimos da organização da Videoteca em banco de dados não relacional. Embora esta metodologia tenha se mostrado limitada, consideramos que seu uso foi de grande valia pois pudemos organizar e classificar nossa crescente coleção de vídeos ao longo dos anos.

No momento em que nos deparamos com condições favoráveis, pudemos alterar a configuração do banco de dados. O programa escolhido – Microsoft Access – tem se mostrado adequado aos objetivos propostos, além de ser compatível com outros programas de banco de dados, caso queiramos fazer nova migração.

A escolha de um banco de dados relacional procurou atender a vários objetivos: propiciar condições para uma interface *Web*, possibilitar a ampliação de campos na base de dados existente e, como meta futura, possibilitar a criação de outras bases para diferentes documentos, como, artigos de jornais da hemeroteca e fotos.

Decidimos pela ampliação de campos na ficha já existente de registro dos vídeos numa tentativa de oferecer mais informações ao nosso usuário.

Escolhemos algumas produções para exemplificar o projeto-piloto e acoplamos à ficha catalográfica do filme outras informações provenientes de fotos, artigos de jornais e revistas, *links Web* e, em alguns casos, *clips* do filme (ou o filme todo).

Para que o banco de dados da Videoteca Multimeios pudesse ser acessado via-Internet foi necessário a programação de inúmeros procedimentos, o que foi efetivado em colaboração com o administrador de redes remotas do Instituto de Artes.

Fizemos um *site* para a Videoteca na *home-page* do Instituto de Artes. Temos notado que a página da Videoteca vem sendo acessada com crescente frequência, como nos mostra o marcador automático de acessos. Com a recente modificação feita neste *site*, implantando os parâmetros do Dublin Core, por um lado, e anunciando as implementações feitas – inserção de fotos, vídeos, etc. - , esperamos atingir um maior número de usuários em pouco tempo.

Durante a implantação do projeto defrontamo-nos com questionamentos que nos levaram à pesquisa de referências para nossa prática. Encontramos trabalhos extremamente interessantes e discussões muito proveitosas para nosso amadurecimento profissional. Vimos deficiências em nosso trabalho e colocamos metas de aprimoramento. Percebemos que a formatação escolhida por nós para apresentar informação adicional advinda de outras fontes – as fotos, os artigos de jornal, os *links Web* e os vídeos – também está limitada. Aprendemos ser possível ampliar sobremaneira a recuperação da informação existente neste material suplementar.

Refletir sobre este percurso e apresentar o caminho trilhado coloca-nos abertos à críticas e sugestões. O aprimoramento do trabalho é um de nossos objetivos.

Temos duas metas para um futuro desenvolvimento do trabalho: a primeira, imediata, explicitada acima, centra-se na busca integrada dos documentos multimídia que já compõem nossa base de dados, recuperando exclusivamente informações textuais. No caso das imagens, significa recuperar o metadado, o que foi escrito sobre o conteúdo da imagem.

A segunda é mais ambiciosa e diz respeito ao uso de tecnologias para reconhecimento de imagem e som na tentativa de recuperar a imagem pelo conteúdo.

Estamos acompanhando o andamento de projetos e aplicações das tecnologias que integram *speech recognition*, *language understanding* e *image understanding* para digitalizar, indexar e recuperar automaticamente dados em áudio e vídeo.⁶⁸

Gostaríamos de encarar os desafios que se colocam diante dessas veredas, tais como, a representação do conteúdo de um vídeo, numa perspectiva de sua manipulação e recuperação.

⁶⁸ *Visor Project*, desenvolvido no *Institute for Image Data*, Inglaterra; *Informedia Digital Video Library Project*, em desenvolvimento na Carnegie Mellon University, Pittsburgh, PA; *Visual Web Search Tools*, desenvolvido na Columbia University, Nova Iorque

Referências Bibliográficas

AGEL, H.(1982) Estética do Cinema. Cultrix, SP.

AMBRON, S. & HOOPER, K., Eds.(1987) Multimedia in Education. Cupertino: Apple.

ARISTARCO, G e T. (1990) Novo Mundo das Imagens Eletrônicas. Edições 70, LDA.

ARMS, Caroline R. (1997) Access Aids and Interoperability. National Digital Library Program, Library of Congress. [Online]. Disponível na Internet via WWW.URL: <http://memory.loc.gov/ammem/award/docs/interop.html>

ARMS, W.Y.(1995) Key concepts in the architecture of the digital library. *D-lib Magazine*. July, 1995 [Online}. Disponível na Internet via WWW.URL: <http://www.dlib.org/dlib/July95/07arms.html>

BALASUBRAMANIAN, V. (1994) State of the Art Review on Hypermedia Issues And Applications. Graduate School of Management, Rutgers University, Newark, New Jersey.

BALAZ, B. (1952) Theory of the Film. Dennis Dobson, Londres.

BAZIN, Andre. (1991) O Cinema - Ensaios. Brasiliense, SP.

BERNADET, J.C.& RAMOS, A F. (1988) Cinema e história do Brasil. Contexto, SP.

BUCKLAND, Michael K. Information as Thing. in *Journal of the American Society for Information Science*. 42(5):351-360, 1991;.

BUSH, Vannnevar. *The Atlantic Monthly*; July, 1945; As We May Think; volume 176, No.1; pages 101-108 [Online]. Disponível na Internet via WWW.URL: <http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm>

CANEVACCI, M. (1984) Antropologia do Cinema. Brasiliense, SP.

CHEPESUIK, R. (1997) The future is here: America's libraries go to digital. *American Libraries*, 2(1), 47-49.

CLEVELAND, Gary. (1998) Digital Libraries: Definitions, Issues and Challenges. IFLA Universal Dataflow and Telecommunications Core Programme, Occasional Paper, 8., p.2-3.

CONY, C.H. (1967) Charles Chaplin. Civilização Brasileira, RJ.

SMITH, J. R. & CHANG.F. VisualSEEK: a Fully Automated Content-Based Image Query System, Proceedings, ACM Multimedia '96 Conference, Boston, MA, November 1996. [Online] Disponível na Internet via WWW.URL: <ftp://ftp.ee.columbia.edu/pub/CTR-Research/advent/public/papers/smith96e.pdf>

CHRISTEL, M., (1995) Addressing the Contents of Video in a Digital Library. Carnegie Melon University, USA. [Online] Disponível na Internet via WWW.URL: <http://www.cs.tufts.edu/~isabel/christel/christel.html>

CHRISTEL, M., & PENDYALA, K., (1996) Informedia Goes to School: Early Findings from the Digital Video Library Project. D-Lib Magazine, September 1996. [Online]. Disponível na Internet via WWW.URL: <http://www.dlib.org/dlib/september96/informedia/09christel.html>

COTTON, B. & OLIVER, R..(1993) Understanding Hypermedia. London: Phaidon.

EISENSTEIN, S.(1969) Reflexões de um Cineasta. Zahar, RJ.
 _____ (1986) El Sentido del Cine. Ed. Siglo Veintiuno.
 _____ (1986) La Forma del Cine. Ed. Siglo Veintiuno.

FELLINI, F. (1986) Entrevista sobre o Cinema; realizada por Giovanni Grazzini. Civilização Brasileira, RJ.

FLEISCHHAUER, Carl. (1996). Digital Historical Collections: Types, Elements, and Construction. National Digital Library Program, Library of Congress. [Online]. Disponível na Internet via WWW.URL: <http://lcWeb2.loc.gov/ammem/elements.html>
 _____ (1998). Digital Formats for Content Reproductions. National Digital Library Program, Library of Congress. [Online]. Disponível na Internet via WWW.URL: <http://memory.loc.gov/ammem/formats.html>

GRAHAM, P.S. (1995). Requirements for the digital research library. [Online]. Disponível na Internet via WWW.URL: <http://aultnis.rutgers.edu/texts/DRC.html>

GRANGER, S., HARMSSEN, L. & HEMSLEY, J.R. (1996). Magnets. Museum and Galleries New Technology Study.Vasari Project. [Online] Disponível na Internet via WWW.URL: <http://www.vasari.co.uk/magnets/>

GUAY, Tim.(1995). Web Publishing Paradigms. Simon Fraser University. [Online]. Disponível na Internet via WWW.URL: <http://hoshi.cic.sfu.ca/~guay/Paradigm/Paradigm.html>

ISMAEL, J.C. Cinema e Circunstância. Coleção Buriti, 6. São Paulo Editora S.A, 1965

KAHN, Robert & WILENSKY, Robert. (1995) A Framework for Distributed Digital Object Services. [Online]. Disponível na Internet via WWW.URL: <http://www.cnri.reston.va.us/home/ctr/arch/k-w.html>

KELLY, Owell.(1996). Digital Creativity. Calouste Gulbenkian Foundation, pág. 36-39 e notas de rodapé).

LANDOW, G.(1992). Hypertext. Baltimore: John Hopkins University Press.

_____. (1992). From Memex to Hypertext: Vannevar Bush and the Mind's Machine. Ed. James M. Nyce and Paul Kahn. Boston: Academic Press, Pp.xi + 367. [Online] Disponível na Internet via WWW.URL: http://www.stg.brown.edu/projects/hypertext/landow/cv/Reviews/Nyce_977.html

LAUREL, B. Ed. (1990). The Art of Human-Computer Interface Design. Reading: Addison Wesley.

LEGRAND, Catherine (director) (1995) Chronicle of the Cinema. Dorling Kindersley Publishing Inc., New York

LÉVY, Pierre. (1993). As Tecnologias de Inteligência: o futuro do pensamento na era da informática. Rio de Janeiro, Ed 34, p.208.

LOWE, D.(1995). Cliente Servidor Para Leigos. tradução Andreia Guimarães dos Santos. - São Paulo: Berkeley Brasil.

LYNCH, Clifford (1997). Searching the Internet. Scientific American, March 1997. [Online]. Disponível na Internet via WWW.URL: <http://www.sciam.com/0397issue/0397lynch.html>

LYNCH, C. A & GARCIA-MOLINA, H. (1995). Interoperability, scaling and the digital libraries research agenda: a report on the May 18-19, 1995 *II TA Digital Libraries Workshop*. [Online]. Disponível na Internet via WWW.URL: <http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/main.html>

McLUHAN, M.. & POWERS, B.(1989). The Global Village. Oxford: Oxford.

_____. McLuhan, E. (1988). Laws of Media The New Science. Toronto: Toronto.

_____. (1967). The Medium is the Message. New York: Bantam.

_____. (1964). Understanding Media: The Extension of Man. New York: New American Library.

METZ, C. (1972). A Significação no cinema. Perspectiva, SP.

MORIN, E. (1970). O Cinema ou o homem imaginário. Tradução António-Pedro Vasconcelos. Moraes Editores.

NELSON, Ted. (1965). A File Structure for the Complex, The Changing and The Indeterminate, ACM 20th National Conference.

NYCE, M. & KAHN, P. From Memex to Hypertext: Vannevar Bush and the Mind's Machine. Ed. James, Boston: Academic Press, 1992. [Online]. Disponível na Internet via WWW.URL: http://www.stg.brown.edu/projects/hypertext/landow/cv/Reviews/Nyce_977.html

PAEPCKE, Andreas. (1996). Digital Libraries: Searching is not Enough. D-Lib Magazine, May 1996. [Online]. Disponível na Internet via WWW.URL: <http://www.dlib.org/dlib/may96/stanford/05paepcke.html>

PAGLIA, Camille. (1998). The Birds. British Film Institute, London.

SADOUL, G. (1983) História do Cinema Mundial. Vol I, II. Livros Horizonte, Lisboa.

SCHATZ, Bruce & CHEN, Hsinchun. (1996). Building Large-Scale Digital Libraries. Computer, May 1996. [Online]. Disponível na Internet via WWW.URL: <http://www.computer.org/pubs/computer/dli/>

SMITH, J. R. & CHANG, S.-F. Searching for Images and Videos on the World-Wide Web, CU/CTR Technical Report 459-96-25, 1996 [Online] Disponível na Internet via WWW.URL: <ftp://ftp.ctr.columbia.edu/CTR-Research/advent/public/papers/96/smith96e.ps>

SMITH, M. & CHRISTEL, M. (1995). Automating the Creation of a Digital Video Library. Proc. ACM Multimedia '95 (November 5-9, 1995), pp. 357-358. [Online] Disponível na Internet via WWW.URL: <http://www.cs.cmu.edu/People/msmith/mm-95-msmith.html>

Projetos Citados:

Advanced Information Hub in Gifu. [Online] Disponível na Internet via WWW.URL: <http://digital-museum.gr.jp/dmc>

American Memory: Historical Collection for the National Digital Library. [Online] Disponível na Internet via WWW.URL: <http://memory.loc.gov/ammem/ammemhome.html>

CATRIONA: [Online] Disponível na Internet via WWW.URL: <http://wp269.lib.strath.ac.uk:5050/Cat2/index.html>

Dance Heritage Coalition. [Online] Disponível na Internet via WWW.URL: <http://www.danceheritage.org>

ELISE – Electronic Library Image Service for Europe. [Online] Disponível na Internet via WWW.URL: <http://severn.dmu.ac.uk/elise>

Informedia Digital Video Library Project. [Online] Disponível na Internet via WWW.URL : <http://www.informedia.cs.cmu.edu/>

International *Web* Collaboration for Medieval Studies Archives. [Online] Disponível na Internet via WWW.URL: <http://www.princeton.edu/~ica/>

MESL Museum Educational Site Licensing Project. [Online] Disponível na Internet via WWW.URL: <http://www.gii.getty.edu/mesl>

Second Avenue *Online*: The Yiddish Theater Digital Archives Project. [Online] Disponível na Internet via WWW.URL: <http://www.yap.cat.nyu.edu>. O projeto em questão trata da recuperação da história do teatro Yiddish em Nova Iorque. Através de um *site Web*, apresenta, em sistema multimídia de banco de dados integrado, um riquíssimo material proveniente de coleções pessoais, tais como, fotos, cartazes, cartas, partituras, discos e gravações em vídeo com depoimentos de atores e familiares; reúne também referências a centros de documentação que tratam desta temática além de integrar as informações bibliográficas e arquivísticas pertinentes ao tema da *Bobst Library*, biblioteca universitária da *New York University*.

The 4th Dublin Core Metadata Workshop Report. [Online] Disponível na Internet via WWW.URL: <http://www.dlib.org/dlib/june97/metadata/06weibel.html>

The Research Libraries Group (RLG). [Online] Disponível na Internet via WWW.URL: <http://www.rlg.org>

VASARI Project [Online] Disponível na Internet via WWW.URL : <http://www.vasari.co.uk>

VIDEOMUSEUM. [Online] Disponível na Internet via WWW.URL: <http://www.videomuseum.fr>

VISOR Project. [Online] Disponível na Internet via WWW.URL : <http://www.unn.ac.uk/iidr/visor.html>

Visual *Web* Search Tools. [Online] Disponível na Internet via WWW.URL: <http://www.cnmtc.columbia.edu/search.html>

Catálogos:

CLÁSSICOS DO CINEMA, Catálogo. Mostra organizada pela Fundação Cinemateca Brasileira e Museu da Imagem e do Som, Campinas, 1979.

FILMOTECA, Catálogo. Empresa Brasileira de Filmes (Embrafilme). Ministério da Cultura.

USANDO MICROSOFT WORKS - Charles Rubin - Tradução: Insight Serviços de Informática, 1991, Editora Campus Ltda.

Revistas:

Revista INFORMÁTICA EXAME ano 9 - n.98 maio/1994 pág.46-56

Revista Internet World, número 9, Maio 1996, “Busca personalizada na sua home-page”, pág.60-62.

Revista SP-Cultura. Secretaria de Estado da Cultura, ano 1, número 1, agosto de 1982:

Revista Jornal do Brasil Programa. Ano 6, n. 781, p.4. (foto de Albert Dieudonné como Napoleão, no filme Napoleon, de Abel Gance)