

## VIDEOTECA DIGITAL: A EXPERIÊNCIA DA VIDEOTECA MULTIMEIOS<sup>1</sup> DO IA/UNICAMP<sup>2</sup>

Maria Lúcia Figueiredo FAGUNDES<sup>3</sup>  
mafagu@iar.unicamp.br  
Gilberto dos Santos PRADO  
gtto@sol.com.br

### RESUMO

*Apresenta 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. Localiza a discussão conceitual no universo das bibliotecas digitais e propõe alterações na configuração atual de seu banco de dados.*

**Palavras-chave:** *Bibliotecas digitais, Internet, Metadados, Sistemas Multimídia, Sistemas de Recuperação da Informação, Videodigital.*

### 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 suggests transformations aiming a better functionality of the database.*

**Key Words:** *Digital Libraries, Digital Video, Internet, Metadata, Multimedia Systems, Retrieval Information Systems, Videolibraries.*

A coleção de vídeos do Departamento de Multimeios, IA/Unicamp, foi formada para oferecer apoio instrumental aos cursos de pós-graduação oferecidos pelo Instituto de Artes. Com mais de 1200 títulos abrangendo documentários, ficção e videoarte, é hoje considerada bastante

representativa da história do cinema. Seu acervo conta também com produções realizadas por professores e alunos do departamento de Multimeios.

Desde meados de 1993, este acervo tem sido organizado em um banco de dados.

<sup>(1)</sup> Videoteca Multimeios acessível no endereço online: <http://www.iar.unicamp.br/videotecalvideoteca.htm>

<sup>(2)</sup> Parte da Dissertação apresentada ao Curso de Mestrado em Artes do Instituto de Artes da UNICAMP, sob orientação do Prof. Dr. Gilberto dos Santos Prado, do Departamento de Multimeios.

<sup>(3)</sup> Coordenadora do Arquivo de Multimeios e da homepage do IA/UNICAMP.

A primeira organização da Videoteca foi feita em banco de dados não relacional (Microsoft Works para MS-DOS, versão 2.0). Embora esta metodologia tenha se mostrado limitada, consideramos seu uso de grande valia pois pudemos organizar e classificar nossa crescente coleção de vídeos ao longo dos anos.

Montamos a primeira estrutura do banco de dados nos espelhando na indicação do *Library of Congress Catalog: motion pictures and filmstrips* e visando a atender as necessidades de consulta demonstradas pelos usuários da Videoteca, em sua maioria, alunos e professores do Instituto de Artes da UNICAMP. Esta estrutura contemplou os seguintes campos de registro: título, diretor, produção, ano de produção, gênero, tempo de duração, bitola original, sistema/cromia, idioma, procedência, número de classificação e sinopse.

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 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 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 primeiro *software* usado: a limitação do tamanho dos campos, até 256 caracteres. Este problema poderia ser em parte minimizado se, na época, tivéssemos formatado um outro campo que contemplasse palavras-chave. Entretanto isso não foi feito devido à insegurança em estabelecer termos que não obedecessem a um vocabulário controlado, pois não tínhamos nenhum conhecimento de um tesouro, em português, específico para cinema e para arte em geral.

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

#### Recebendo uma interface Web

Com o estabelecimento da World Wide Web, foi iniciada em fins de 1996 a pesquisa para a disponibilização desses dados na rede Internet. Passamos a olhar o banco de dados da Videoteca Multimeios 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*.

Estabelecemos como objetivo o uso dos recursos oferecidos pelo sistema de redes remotas, 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 *hiperlinks*, imagens e sons, propusemos ampliar as possibilidades de consulta aos filmes da Videoteca Multimeios e avançar além do formato-padrão de apresentação e organização da informação que prioriza a linguagem escrita.

Buscamos então:

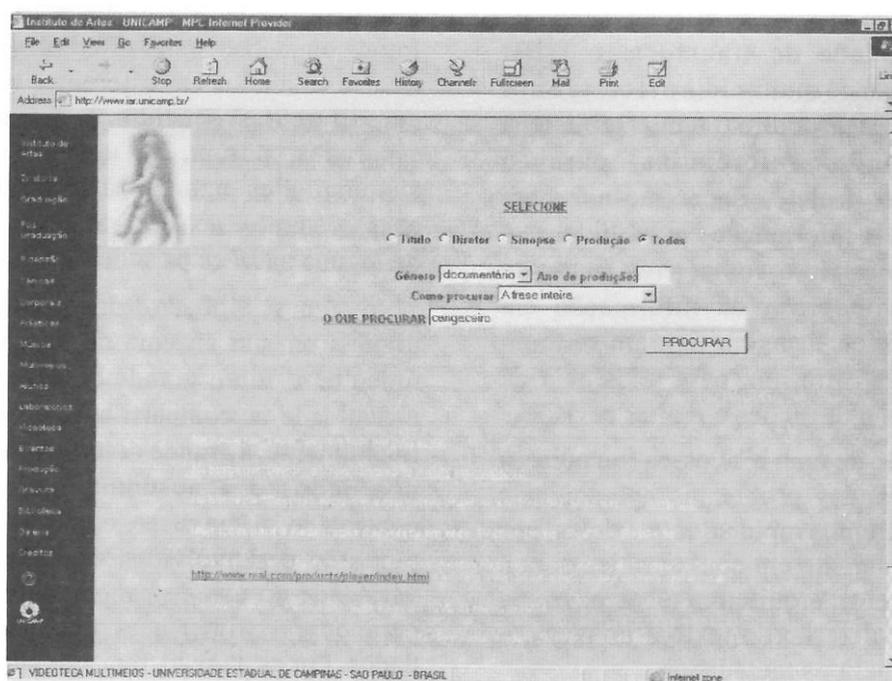
1. Transportar para a interface gráfica da Internet um banco de dados considerado tradicional na sua forma de apresentação da informação;
2. Inserir informações complementares à ficha técnica de um filme: fotos, fragmentos do vídeo, artigos de jornal ou revista, e *hiperlinks* remetendo a *sites* afins.

O transporte dos dados para um banco de dados relacional (Microsoft Access) viabilizou a interface Web e deixou em aberto a possibilidade da criação futura de outras tabelas para diferentes documentos, como, fotos e artigos de jornais do Arquivo Multimeios ou ainda coleções de vídeos de outros departamentos do Instituto de Artes. Naquele momento, não pensávamos ainda na possibilidade de construir um banco de dados com

conceito multimídia, apesar de estarmos usando recursos de imagem e de rede remota nos campos de registro. Assim sendo, ao invés de criarmos novas tabelas, decidimos pela ampliação de campos na tabela já existente, ficando sua estrutura acrescida dos campos: fotos, vídeos, críticas e *hiperlinks*. Estes campos de registros adicionais ficaram atados à ficha catalográfica sem possibilidade de serem recuperados *per si*. Para que o banco de dados da Videoteca Multimeios pudesse ser acessado via-Internet foi necessária a programação de inúmeros procedimentos (instruções SQL, CGI scripts e linguagem C).

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.



Escolhemos algumas produções para exemplificar o projeto-piloto e acoplamos à ficha catalográfica do filme informações provenientes de fotos, artigos de jornais e revistas, *links Web* e, em alguns casos, *clips* do filme (ou o filme todo). Construímos uma página dentro do *site* da Videoteca, chamada "Olho no Vídeo", onde divulgamos periodicamente produções escolhidas e/ou novas aquisições que receberam em suas fichas catalográficas informações adicionais provenientes de mídias diversas.

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. Com exceção do *site* do Cinemabrasil, ainda não encontramos nenhuma videoteca organizada em sistema de banco de dados usando recursos de hipertexto e multimídia. Entretanto, a forma em que sua tabela foi projetada não permite uma busca e recuperação da informação multimídia. No formato em que se apresenta, ainda é um banco de dados prioritariamente textual acrescido de informações visuais e hiperlinks. Sua adaptação para o formato multimídia faz parte do desenvolvimento do projeto.

### Biblioteca Digital, Videoteca Digital

A possibilidade de transformar o BD Videoteca em um banco multimídia é recente e nos aproximou da discussão sobre o tema "biblioteca digital". Buscando referências para o aprimoramento do projeto, deparamo-nos com pesquisas na área da informação em bibliotecas e museus digitais, visando o compartilhamento de acervos e busca integrada da informação em diferentes repositórios. Estas configuram-se como questões candentes e trazem, para a Ciência da Informação, novas indagações e muitas revisões.

A construção de uma biblioteca (ou museu) digital levanta grandes problemas operacionais dos quais comentaremos rapidamente 3 deles, por se refletirem diretamente em nossa prática com a Videoteca Multimeios: arquitetura técnica, uso de padrões para digitalização de texto e imagens, e busca de padrões para os metadados.

### 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 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 o grande desafio 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, já temos notícias de projetos desenvolvidos em parceria que já disponibilizaram na *WEB* vários repositórios integrados. Poderíamos citar os seguintes:

*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 8 instituições com o objetivo de centralizar acervos de dança.

Consideramos a experiência *web* da Videoteca Multimeios como um protótipo para o desenvolvimento de uma arquitetura aberta e ampla que abrigue outras coleções de vídeos do Instituto de Artes, bem como acervos fotográficos como o do Arquivo Multimeios e o acervo de obras da Galeria de Artes. Neste sentido estamos já projetando formatações para estas novas bases de dados com possibilidades de compartilhamento de acervos e cruzamento de informações.

## 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 estandardizados 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). Estas linguagens viabilizam a codificação de textos para serem vistos na Web e hoje em dia estão bastante popularizadas.

Está em desenvolvimento a versão abreviada e simplificada da SGML, a XML - *Extensible Markup Language*. Ela tenta resolver problemas colocados pela rigidez da HTML (que tem formato fixo). 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.

No caso da Videoteca Multimeios, os textos que compõem o campo "críticas" são arquivos.html e ficam armazenados em sub-diretório do diretório Videoteca, dentro da raiz do servidor *Web* do IA. Já as telas *web* com os resultados das buscas efetuadas na base de dados da videoteca são formatadas em HTML "on-the-fly".

No armazenamento de imagens fixas usamos formatos padronizados como JPEG e GIF.

Sabemos que a tecnologia para imagem em movimento (como também para som) ainda não está bem resolvida devido ao grande volume de dados transmitidos. Por este motivo, ao digitalizarmos nossos vídeos, optamos pela formatação em RealVideo, solução de *streaming media* de alta compressão. Este formato apresenta resolução de imagem muito inferior aos arquivos de vídeo em QuickTime ou ainda em MPEG-I, porém com a vantagem de poderem ser rapidamente visualizados pelo usuário.

## 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. 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 imagens. Citamos alguns deles:

Nos Estados Unidos, 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.

No *site* da Videoteca Multimeios adotamos o Dublin Core, conjunto de 15 elementos codificados em HTML que apresentam "informação sobre a informação" de uma página *web*. Sua localização na página *web* permite que as informações ali contidas sejam facilmente encontradas e indexadas pelos diferentes mecanismos de busca que operam na rede Internet.

Entretanto, 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 (que é o caso das páginas que trazem o resultado das buscas efetuadas na base de dados da Videoteca Multimeios).

De qualquer forma, o Dublin Core mostra-se um excelente recurso para a padronização de informações de páginas *web*, principalmente para aquelas cujos dados possam usar algum repertório de tesouros já intitucionalizados, como, por exemplo, o *Art & Architecture Thesaurus*, ou o *Union List of Artist Names*. Infelizmente, em nosso caso, não podemos tirar partido de toda a potencialidade do Dublin Core por não contarmos ainda com tesouros aceitos nacionalmente, em português, específico para artes.

#### Novos desafios

A possibilidade de oferecer vídeos digitalizados ao nosso usuário descortina um novo panorama de pesquisa, qual seja, a recuperação da imagem pelo conteúdo. Temos acompanhando o andamento de projetos e aplicações de tecnologias que integram *speech recognition*, *language understanding* e *image understanding* para digitalizar, indexar e recuperar automaticamente dados em áudio e vídeo. Algumas destas pesquisas tem sido aplicadas à educação, como a do Informedia Digital Video Library, em desenvolvimento na Carnegie Mellon University, PA, EUA, abrindo um campo ilimitado para a recuperação da informação.

#### PROJETOS CITADOS

Advanced Information Hub in Gifu. [Online] Disponível na Internet via WWW.URL:

<http://digital-museum.gr.jp/dmc>

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.URI: <http://www.princeton.edu/~ica/>

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

The 4<sup>th</sup> 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.URI: <http://www.vasari.co.uk>

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