

Uso de Sistemas de Informação Multimídia em Acervos Permanentes

FLÁVIO HUMBERTO CABRAL NUNES¹
ARNALDO DE ALBUQUERQUE ARAÚJO¹
LUIZ ANTÔNIO CRUZ SOUZA²

¹UFMG - Universidade Federal de Minas Gerais
DCC - Departamento de Ciência da Computação
NPDI - Núcleo de Processamento Digital de Imagens
Caixa Postal 702 - CEP 30161-970 Belo Horizonte (MG)
(fhcununes,arnaldo)@dcc.ufmg.br

²UFMG - Universidade Federal de Minas Gerais
EBA - Escola de Belas Artes
CECOR - Centro de Conservação e Restauração de Bens Culturais Móveis
Av. Antônio Carlos, 6627 - CEP 31270-901 Belo Horizonte (MG)
luiz-souza@ufmg.br

Resumo. Este artigo discute o processo de digitalização de acervos, considerando a problemática da sua preservação digital. O uso de sistemas de informação para permitir um acesso ágil e simples também é tema de discussão. Como estudo de caso, é apresentado o processo de digitalização do Cecor. O Cecor produz documentos sobre o processo de restauração de obras que ele executa. Para gerenciar este acervo, foi criada uma aplicação multimídia que é discutida neste texto.

Palavras-Chave. Preservação digital, banco de dados, sistemas de informação multimídia, hipermídia, digitalização de acervos.

1 Introdução

As instituições mantenedoras de arquivos públicos ou outros acervos documentais enfrentam diversos problemas, decorrentes geralmente do grande acúmulo de documentos e de sua fragilidade, destacando-se os riscos decorrentes da degradação dos originais devido a sua manipulação direta e freqüente e da dificuldade de acesso às informações por parte dos pesquisadores e do público em geral [4].

A tecnologia digital surge como uma possibilidade de romper esse compromisso, permitindo dar amplo acesso às cópias digitais de alta qualidade dos documentos originais, ao mesmo tempo resguardando-os da manipulação desnecessária.

Do lado dos consulentes, o uso da tecnologia digital permite operações de busca muito mais rápidas e sofisticadas do que é possível com os instrumentos de pesquisa em papel, como inventários e índices.

As redes de computadores, em particular a *Internet*, adicionam a possibilidade da consulta remota, expandindo enormemente o universo de pesquisadores com acesso ao acervo [10].

Entretanto, documentos digitais que requerem retenção por longos períodos encontram problemas de acessibilidade por causa da obsolescência da tecnologia de *hardware* e *software*. É importante que, ao se adotar tecnologias mais novas e melhores, não haja a perda do acesso aos documentos históricos digitais.

Este trabalho discute sobre a implementação de sistemas de informação multimídia para disponibilização de acervos de bibliotecas e arquivos, considerando aspectos de preservação digital dos documentos digitais. A seção 2 apresenta alguns trabalhos de digitalização como motivação para este trabalho. A seção 3 aborda o processo de digitalização de acervos. Na seção 4, são discutidas questões de acesso e organização dos dados do sistema. O sistema implementado é descrito na seção 5. Por fim, a seção 6

apresenta algumas considerações finais sobre o trabalho.

2 Trabalhos Relacionados

Diversos trabalhos têm sido desenvolvidos no sentido de aplicar sistemas de informação multimídia em acervos históricos, arquivísticos e de bibliotecas ou de propor modelos e implementações para esses sistemas [10].

Um exemplo de instituição com projeto de digitalização de seu acervo é o Arquivo Público Mineiro (APM). Essa instituição centenária recolhe e conserva importantes documentos administrativos abrangendo os principais aspectos da história mineira, garantindo aos cidadãos total acesso aos mesmos. O APM está trabalhando para digitalizar todo o seu acervo e o Núcleo de Processamento Digital de Imagens (NPDI) tem um papel muito importante neste trabalho através de alunos de mestrado e iniciação científica. Inicialmente, foi feito um sistema que contemplava apenas a coleção do presidente Arthur Bernardes [9]. Posteriormente, esse protótipo foi expandido para outras coleções, levando em consideração aspectos de preservação digital e gestão documental e fluxo de trabalho [10].

O Projeto Portinari é também uma outra iniciativa de digitalização de acervos [15]. Este trabalho começou em 1979 com a catalogação de 4500 trabalhos e 25000 documentos sobre o trabalho de Cândido Portinari, o mais famoso pintor brasileiro. Em seguida, foi desenvolvida uma aplicação hipermídia contendo toda a informação resultante desta catalogação. Nesta aplicação, as obras de Portinari podem ser exploradas por tema, data e técnicas, assim como as relações entre as obras e os eventos em que foram exibidas [18].

3 Digitalização do Acervo

Poucas tecnologias têm oferecido um grande potencial para pesquisa e ensino como a tecnologia de imagens digitais. Ela possibilita examinar objetos raros e únicos fora dos ambientes controlados de museus e bibliotecas. Ela também permite exibir e ligar coleções ao redor do mundo, quebrando as barreiras físicas de acesso. Outra vantagem é ampliar o universo de usuários sem distinção de classes sociais ou econômicas [8]. O uso de imagens digitais também contribui para a preservação dos documentos, pois estes deixam de ser manipulados diretamente.

Entretanto, os computadores não oferecem soluções simples para a preservação documental. A informação digital é mais sujeita à adulteração, vandalismo e acidentes do que a analógica; seus suportes são frágeis e perecíveis. Devido ao fato de os dados de computador dependerem de complexos sistemas cooperantes de *hardware* e *software*, a rápida obsolescência torna-se uma ameaça constante de perda de acesso aos acervos.

A primeira consideração a fazer na determinação dos parâmetros de qualidade das imagens digitais é o seu propósito. Essa questão afeta todas as decisões subsequentes, desde a resolução até o formato de armazenamento [10].

Há um consenso crescente dentro da comunidade de preservação que um número de arquivos de imagem devem ser criados para cada documento para atender um conjunto de usos. Uma imagem de arquivo ou mestre deve ser criada, representando a cópia de mais alta qualidade e sendo bastante fiel ao documento original [12]. Esta imagem é arquivada em mídias CD-ROM ou DVD. A partir desta cópia, vários arquivos de acesso podem ser produzidos para atender usos específicos.

No processo de conversão dos documentos da representação analógica para digital, deve-se considerar alguns aspectos. A resolução do documento deve capturar todos os detalhes significantes do documento original. Quanto maior a resolução, maior é a quantidade de detalhes capturados, mas o tamanho do arquivo também é aumentado consideravelmente. O aumento da profundidade do *bit*, número de *bits* usados para representar cada *pixel*, permite capturar mais tons de cinza ou de cor. Esse aumento afeta os requisitos da resolução, o tamanho do arquivo e o método de compressão utilizado [5]. Isto é válido para as imagens de arquivo. Já a resolução das cópias de acesso dependerá do seu uso.

Processos de tratamento de imagens permitem melhorar a qualidade da imagem digitalizada, mas alteram as características do arquivo. Para efeitos de preservação não é adequado que as cópias mestre passem por esses tratamentos sob penalidade de diminuir sua fidelidade em relação ao seu original.

O formato de arquivo a ser adotado para as imagens mestre deve suportar a resolução, a profundidade de *bit*, a informação de cor e os metadados necessários. O formato também pode ser armazenado compactado ou não. Os tipos de compressão podem ser com perdas ou sem perdas de informação, sendo que as imagens mestre não deve ser compactadas em um formato com perda de informação.

4 Acesso e Armazenamento da Informação

4.1 Informação Multimídia

Multimídia e digitalização de imagens são fortes tendências em computação. Como a multimídia realça a qualidade e a quantidade da informação, muitas áreas da tecnologia da informação estão abraçando as plataformas, tipos de dados e ambientes multimídia.

Algumas pessoas definem multimídia em termos de suporte a vídeo digital e interatividade; outros são mais liberais e incluem imagens, áudio, gráficos 2D/3D e mesmo texto. A multimídia está trazendo melhoramentos para a interação homem-máquina. A multimídia fornece uma forma mais expressiva e rica de representação, exibição e interação com a informação [14].

4.2 Gerência de Dados Multimídia

O tamanho dos dados multimídia é frequentemente grande e o armazenamento de grandes quantidades de dados é uma tarefa normalmente alocada a sistemas de gerenciamento de banco de dados (SGBD) [7].

Para ADJEROH [6], o SGBD multimídia é o cerne de um sistema de informações multimídia e pode ser entendido como o conjunto de programas e rotinas usados para definir, criar, armazenar, indexar, gerenciar e pesquisar bancos de dados multimídia. Nesse contexto, um SGBD multimídia deve prover suporte para os dados multimídia da mesma forma que um SGBD tradicional suporta dados alfanuméricos simples. Na verdade os diferentes tipos de dados complexos envolvidos em um banco de dados multimídia requerem métodos especiais para gerenciamento, armazenamento, acesso, indexação e recuperação dos dados, tornando o SGBDMM uma ferramenta mais sofisticada que o SGBD tradicional [6, 16].

4.3 Tipos de Consultas

Os tipos de consultas envolvendo documentos multimídia estende os tipos de consulta existentes em SGBD's tradicionais. As consultas podem ser feitas sobre o conjunto de metadados dos documentos. O termo metadado significa dado acerca do dado e é utilizado para denotar os atributos associados a um documento digital. Os metadados adicionam valor ao documento, fornecendo informações importantes

acerca do seu conteúdo, formato e história administrativa [13].

Um vocabulário controlado pode ser criado para cada documento. O problema dessa abordagem é que ela depende da subjetividade humana. Em uma fotografia de uma construção, uma pessoa pode descrevê-lo como um prédio e outra como um edifício.

Outro tipo de consulta é a consulta baseada no conteúdo do documento. Um exemplo seria recuperar todas as imagens que tenham um carro vermelho. Esses métodos determinam a resposta das consultas através de níveis de similaridade entre o documento de pesquisa e os aqueles existentes no banco de dados. As pesquisas podem ser interativas, permitindo ao usuário refinar sua consulta pelo resultado de uma consulta anterior [14].

5 O Sistema Restaura

5.1 O Cecor e seu Acervo

O Cecor, Centro de Conservação e Restauração de Bens Culturais Móveis, desenvolve trabalhos na área de conservação e restauração de pinturas, esculturas e obras sobre papel [1]. Durante o processo de restauração, é executado um minucioso processo de documentação e arquivamento de todos os dados levantados das peças em restauração. Desde 1978, a instituição reúne um representativo acervo, composto em sua maioria de dossiers, fotografias e slides das várias técnicas fotográficas utilizadas para o estudo de obras de arte. A Figura 1 apresenta uma pintura e algumas análises físicas as quais ela foi submetida. A Figura 2 ilustra alguns tipos de documentos que compõem o acervo do Cecor.

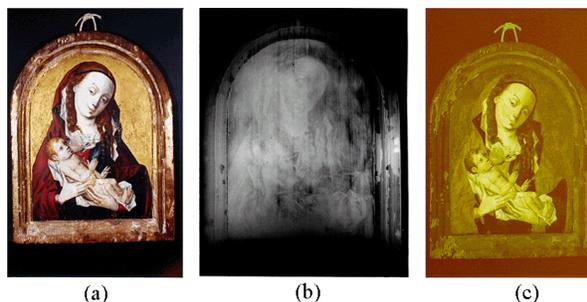


Figura 1: (a) Pintura de uma Madona. (b) Raio-x da figura (a). (c) Ultravioleta da figura (a).

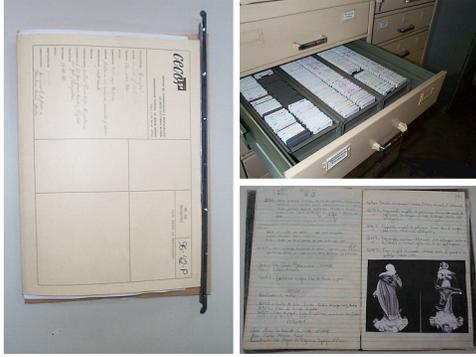


Figura 2: Alguns tipos de documentos existentes no acervo do Cecor: fichários, slides e cadernos de análises.

5.2 Protótipo

O modelo de gerenciamento do acervo que era utilizado possui vários problemas. O registro das obras era feito manualmente. Além de ser um trabalho cansativo e monótono, o restaurador perdia muito tempo com esse processo que poderia ser usado em outras tarefas mais importantes. Outro problema é a preservação dos documentos originais. Como a busca era feita manualmente, muitos documentos acabavam sendo manipulados sem necessidade. Com o tempo, começavam a aparecer nos documentos degradações originadas desta manipulação excessiva e desnecessária. Também havia o desejo de tornar as informações deste acervo acessíveis a pesquisadores externos ao Cecor e ao público em geral. No antigo modelo somente os funcionários e professores do Cecor tinham acesso a essas informações. Qualquer pessoa que quisesse ter esse acesso, tinha que se dirigir até a instituição e pedir às pessoas autorizadas pelas informações de que necessitava.

Por esses problemas, esta instituição se mostrou como ambiente ideal para implementar um protótipo de sistema de informação multimídia, chamado Restare, que pudesse validar o estudo feito até então. O processo de desenvolvimento do Restare foi o PRAXIS que é um processo de *software* que visa o desenvolvimento de de aplicações interativas extensíveis. Ele fornece o grau de formalidade e controle para um projeto como este [11]. Praxis utiliza a *Unified Modeling Language* (UML) como linguagem de modelagem. A abordagem orientada a objeto parece ser mais natural para a maioria das aplicações de bancos de dados multimídia [17]. O sistema de banco de dados adotado é o *MySQL* [2], devido ao fato de ser um sistema livre com baixo custo e também por possuir uma vasta

documentação.

O Praxis possui as seguintes fases de desenvolvimento: concepção, elaboração, construção e transição. Na concepção, a implementação de um sistema é justificada. Na Elaboração, a especificação do produto é detalhada o suficiente para modelar conceitualmente o domínio do problema. Nesta fase, foram identificados três ambientes que representam áreas distintas presentes dentro do Cecor:

- **Restauração** - Processo de restauração propriamente dito.
- **Orçamento** - Compreende informações sobre os custos dos trabalhos de restauração.
- **Laboratório** - Informações sobre análises físicas e físico-químicas a que são submetidas as amostras retiradas obras.

Na construção, é desenvolvida (desenhada, implementada e testada) uma versão completamente operacional do produto. Nesta fase, foram criados pacotes que encapsularam as funções de cada um dos ambientes citados acima. A Figura 3 apresenta um modelo UML simplificado do sistema e a Figura 4 ilustra uma tela do sistema.

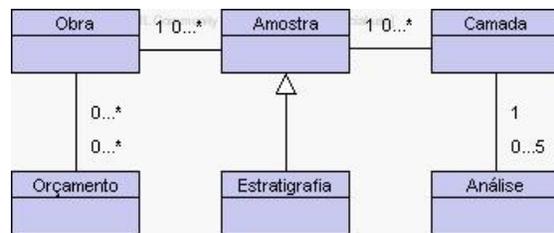


Figura 3: Modelo do sistema (simplificado).

Finalmente, a fase de transição coloca o produto à disposição dos usuários. Para essa fase, foi feito um treinamento do pessoal do Cecor, os usuários potenciais do sistema, para que pudessem utilizar a aplicação. Esse treinamento foi necessário devido ao precário conhecimento de computadores da comunidade de usuários que iria utilizar o Restare.

Os documentos digitalizados pertencem ao acervo de fotografias e slides da instituição. As imagens mestre foram digitalizadas no formato TIFF, formato que preenche a maioria dos requisitos de preservação, a uma resolução de 600 dpi (*dot per inch*). Mas essas imagens têm o inconveniente de serem muito grandes, sendo inviáveis para o tráfego



Figura 4: Tela de identificação da obra.

em rede. Assim, não servem como imagens de acesso, sendo necessário a utilização de formatos que utilizem compressão. O formato JPEG foi o escolhido para as imagens de acesso.

6 Discussão e Conclusão

Este artigo discute sobre a utilização de sistemas de informação multimídia para gerenciar acervos permanentes. Simplesmente digitalizar os documentos não resolve o problema. É necessário estabelecer os tipos de usos para os documentos digitais e aspectos, como a longevidade das cópias digitais, precisam ser considerados. Por isso, políticas de preservação digital devem ser seguidas para que todo o trabalho de digitalização não seja perdido com a evolução tecnológica.

O uso da multimídia permite uma interação homem-máquina mais próxima e de maior qualidade. A tecnologia digital possibilita livrar o homem de tarefas monótonas e deixar que ele se ocupe de tarefas que agreguem valor. Tarefas demoradas de consulta podem ser feitas rapidamente com o uso de SGBD's e com uma gama muito maior de consultas possíveis. Isso mostra como é interessante o uso de sistemas computacionais no gerenciamento de acervos onde é necessário permitir que seus usuários tenham condições de acessar as informações que necessitam.

Para ilustrar todas as vantagens, um sistema de informação (Restaura) foi desenvolvido para gerenciar o acervo do Cecor. O Restaura ainda está uma fase inicial, possuindo apenas interfaces de cadastro e atualização de dados e um conjunto limitado de consultas. O amplo acesso ao acervo ainda não foi possível, sendo restrito ao Cecor. A criação de

um módulo de acesso com um universo mais amplo de consultas faz parte dos planos desse projeto. Sistemas, como o Narcisse [3] e o APM [10, 9], possuem esse tipo de interface para o usuário. Mas o Restaura conseguiu agilizar o trabalho de pesquisa das informações no acervo. Por enquanto, as buscas nos sistemas são baseadas em texto, mas trabalhos estão sendo feitos para implementar mecanismos de buscas baseadas no conteúdo dos documentos.

O universo de usos futuros do acervo digital criado também é promissor. Simulações do processo de degradação podem ser feitas por computador, permitindo determinar como era a peça em seu estado original. Uma outra possível aplicação seria determinar o resultado de uma intervenção, auxiliando no processo de decisão. Estas aplicações são apenas alguns exemplos de usos futuros do acervo digital produzido. Por isso, a adoção de medidas de preservação permite que os documentos possam ser utilizados mesmo que seus usos não tenham sido ainda descobertos.

Referências

- [1] Centro de Conservação e Restauração de Bens Culturais Móveis - Cecor. Disponível na WWW: <<http://cecor.eba.ufmg.br>>, 15/jan/2004.
- [2] MySQL - Homepage. Disponível na WWW: <<http://www.mysql.com>>, 05/fev/2004.
- [3] Narcisse. Disponível na WWW: <<http://www.culture.gouv.fr/documentation/lrmf/pres.htm>>, 15/jan/2004.
- [4] Arquivo público mineiro: Projeto de digitalização de imagens. Relatório Técnico, Prodemge, Belo Horizonte, 1995.
- [5] Moving theory into practice: Digital imaging tutorial. Relatório Técnico, Cornell University Library, Disponível na WWW: <<http://www.library.cornell.edu/preservation/tutorial/contents.html>>, 14/03/2004 2000.
- [6] D. A. Adjero and K. C. Nwosu. Multimedia database management - requirements and issues. In *IEEE Multimedia*, volume 4, pages 24-33, Julho-Setembro 1997.
- [7] P. M. G. Apers, H. M. Blanken, and M. A. W. Houtsma. *Multimedia Databases in Perspective*. Springer, 1997.

- [8] H. Besser and J. Trant. Introduction to imaging: Issues in constructing an image database. Relatório Técnico, Getty Information Institute, Disponível na WWW: <http://www.getty.edu/research/conducting_research/standards/introimages/>, 14/03/2004 1995.
- [9] N. S. de Andrade. Sistemas de informação multimídia. Dissertação de Mestrado, Fundação João Pinheiro e Universidade Federal de Minas Gerais, Brasil, Novembro 1998.
- [10] E. A. do Valle Jr. Sistemas de informação multimídia na preservação de acervos permanentes. Dissertação de Mestrado, Universidade Federal de Minas Gerais, Brasil, Fevereiro 2003.
- [11] W. P. P. Filho. *Engenharia de Software: Fundamentos, Métodos e Padrões*. LTC - Livros Técnicos e Científicos Editora S.A., segunda edição, 2001.
- [12] F. S. Frey and J. M. Reilly. Digital imaging for photographic collections: Foundations for technical standards. Relatório Técnico, Image Permanence Institute, Rochester Institute of Technology, Disponível na WWW: <<http://www.rit.edu/~661www1/subpages/digibook.pdf>>, 14/03/2004 1999.
- [13] A. Gilliland-Swetland, T. Gill, M. Woodley, and M. Baca. Introduction to metadata: Pathways to digital information. Relatório Técnico, Getty Research Institute, Disponível na WWW: <<http://www.getty.edu/research/institute/standards/intrometada>>15/jan/2004.
- [14] S. Khoshafian and A. B. Baker. *Multimedia and Imaging Databases*. Morgan Kaufmann Publishers, Inc., 1996.
- [15] R. S. G. Lanzelotte, M. P. Marques, M. C. G. Penna, J. C. Portinari, I. D. Ruiz, and D. Schwabe. The portinari project: Science and art team up together to help cultural projects. In *Proceedings of the 2nd International Conference on Hypermedia and Interactivity in Museums (ICHIM'93)*, Cambridge, UK, September 1993.
- [16] S. Marcus. Foundations of multimedia database systems. *Journal of ACM*, 43(3):474–523, Maio 1996.
- [17] B. Prabhakaran. *Multimedia Database Management Systems*. Kluwer Academic Publishers, 1997.
- [18] G. H. Rossi. *Um Método Orientado a Objetos para o Projeto de Aplicações Hipermídia*. Tese, Pontifícia Universidade Católica do Rio de Janeiro, Disponível na WWW: <<http://www.lifia.info.unlp.edu.ar/~fer/oohdm/>>, 05/fev/2004, Brasil, Julho 1996.