

# BIBLIOTECA DIGITAL DE TESES E DISSERTAÇÕES DO INSTITUTO DE FÍSICA GLEB WATAGHIN - UNICAMP : RELATO DE EXPERIÊNCIA

SPONCHIADO, R. A.<sup>1</sup> , VICENTE, V. S.<sup>2</sup>

Nos últimos anos pudemos ver várias iniciativas de Instituições nacionais e internacionais com relação ao trabalho de desenvolvimento e criação de Bibliotecas Digitais com a disponibilização e acesso do conteúdo integral de seus acervos.

Biblioteca Digital tem sido um tema cada vez mais discutido nas literaturas de ciência da informação e de informática.

Cunha (2000) analisa a evolução das bibliotecas em: Biblioteca tradicional moderna – Era I; Biblioteca Automatizada – Era II; Biblioteca Eletrônica – Era III; Biblioteca Digital e Virtual – Era IV.

De acordo com Kahn e Wilensky apud Cattelan (1997) Bibliotecas digitais são sistemas que armazenam, acessam, disseminam e gerenciam objetos digitais.

Segundo Machado (1999),

A biblioteca digital tem como característica uma coleção de documentos eminentemente digitais, independendo se foram criados na forma digital ou digitalizados a partir de documentos impressos, e permite, por meio de redes de computadores, compartilhar a informação instantânea e facilmente.

As teses e dissertações são importantes documentos que retratam o resultado de pesquisas geradas em Universidades cujos resultados podem alimentar novas pesquisas.

Em sua grande maioria, os exemplares ficam restritos às Instituições que geraram essas pesquisas, o que torna difícil a disseminação de seus conteúdos.

O objetivo principal de uma biblioteca digital de teses é a ampla divulgação do conhecimento gerado por uma Instituição, possibilitando acesso simultâneo aos documentos por usuários múltiplos, facilitando a integração de grupos de estudos com interesses comuns e retornando para a sociedade o resultado dos trabalhos científicos, garantindo os direitos intelectuais.

O Banco de Teses Digitais tem várias vantagens, das quais podemos citar: facilidade na disseminação das informações e na preservação dos documentos; inclusão de links para outros endereços eletrônicos; melhor

---

<sup>1</sup> UNICAMP - Instituto de Física Gleb Wataghin

<sup>2</sup> UNICAMP - Instituto de Física Gleb Wataghin

acesso público às pesquisas; disponibilidade de consultar o documento em qualquer momento e em qualquer local com acesso à Internet; menor necessidade de consultar o exemplar em papel, o que mantém o material no acervo em melhor estado de conservação; acesso simultâneo por múltiplos usuários; rapidez e democratização da informação.

Segundo Muller (2000) o crescimento das tecnologias da informação e comunicação tornará os usuários mais independentes das fontes de informação, fazendo com que a intermediação dos bibliotecários e suas visitas à biblioteca sejam menos freqüentes.

Mas, Haricombe apud Bezerra (2002) conclui afirmando que “essa tendência, ironicamente parece ter aumentado o nível de ajuda prestada pelos bibliotecários, na medida em que eles assumem novos papéis, tais como apoio técnico para navegação na Web e recuperação de informação”.

O Instituto de Física Gleb Wataghin criado em 1967, teve sua primeira tese de doutorado defendida em 1969 que foi orientada pelo seu primeiro Diretor o Prof. Marcello Damy de Sousa Santos.

O acervo de teses e dissertações da Biblioteca do Instituto de Física da Unicamp, conta com 1.128 documentos produzidos pelos alunos de pós-graduação de 1969 até hoje.

Através de iniciativa própria, o IFGW em concordância com a Biblioteca e a Coordenadoria de Pós-Graduação, determinou em 1999, um fluxo de procedimentos para facilitar o início de um projeto de digitalização das teses e dissertações produzidas por seus alunos.

Em 2000 foi iniciado o trabalho de digitalização de originais em papel, em 2001 foi desenvolvida em PHP a interface de busca e em 2002 o banco de dados foi disponibilizado com aproximadamente 180 documentos em texto integral.

Hoje a biblioteca digital está completa, com todas as 1120 Teses e Dissertações disponíveis em texto completo, sendo 658 de mestrado e 462 de doutorado.

Existem apenas 8 teses que ainda não foram disponibilizadas no Banco, porque a pedido dos autores, estão aguardando a publicação de artigos em revistas científicas, cujos resultados foram citados nessas teses.

A Biblioteca digital de Teses do IFGW, baseada em modelos pesquisados, tem características próprias de apresentação e conteúdo.

O formato utilizado para disponilizar os arquivos em texto completo foi o PDF (Portable Document Format), que mantém o “lay out” original do documento, garante a integridade do texto e permite transmissão rápida via rede.

Os recursos de hardware utilizados foram:

- Scaners:

2 Scaner Microtek 900 dpi

1 Scaner HP 6100 2400 dpi

1 Scaner ScanJet 5370C 1200 dpi

- 1 Scaner HP 4C 2400 dpi
- 1 Scaner HP ScanJet 5470CXI 1200 dpi
- Microcomputadores:
  - 1 Pentium 64 Mb – 233 MHz
  - 1 Pentium 64 Mb – 200 MHz
  - 1 AMD DURON 128 Mb – 1300 MHz
  - 3 Pentium 32 MB – 166 Mhz
- Servidores
  - Repositório: Pentium III – 256 Mb – 550 MHZ – 3 discos de 18 Gb, sendo 1 disco de 18 Gb para as teses. (4,8 Gb é o espaço ocupado para armazenar as 1120 teses).
  - Espelho : Pentium III – 256Mb – 450 MHZ – 3 discos de 30 Gb.

O software utilizado foi o Adobe Acrobat e os editores de texto utilizados foram o MS Word e Front Page.

Os recursos humanos: 9 pessoas estiveram diretamente envolvidas, sendo: 1 bibliotecária coordenadora do projeto, 1 bibliotecária , 1 analista de sistemas, 1 estagiário de informática, 1 estagiário de biblioteconomia e 4 bolsistas.

O sistema de busca foi desenvolvido localmente em PHP, os metadados utilizados foram os campos do formato MARC do registro bibliográfico - <http://www.ifi.unicamp.br/ccjdr/teses/>

Cada Tese/Dissertação é cadastrada no Sistema através dos dados: autor, título, orientador, co-orientador, tipo de tese (Mestrado ou Doutorado), ano de publicação e palavras-chave. A interface de busca permite a pesquisa por qualquer palavra contida em qualquer destes campos.

O Resumo é apresentado separadamente do documento completo, em português e inglês, no formato HTML (Hypertext Markup Language) cujo processo inclui captura, revisão e correção do texto.

Procedimento 1 – acervo posterior a 1999, recebido em mídia eletrônica (PS)

As Teses/Dissertações são entregues pela pós-graduação, em mídia eletrônica (disquete, CD ou Zip-drive) no formato Post Script, acompanhada de autorização do autor para disponibilização na internet.

Após conferência dos arquivos, o texto completo é transformado em PDF através de conversão utilizando-se o Adobe Distiller.

Para cada T/D são gerados 5 arquivos:

- 1 PDF do texto completo
- 1 PDF dos resumo
- 1 HTML do resumo
- 1 PDF do resumo em inglês
- 1 HTML do resumo em inglês

- backup dos 5 arquivos

Procedimento 2 – acervo anterior a 1999 em papel

As Teses/Dissertações são digitalizadas a partir do texto original em papel, encadernado.

Os padrões variam de acordo com a qualidade do documento original.

A resolução média adotada é de 200 dpi, o formato a ser salvo o arquivo é PDF.

A cor que oferece melhor resultado qualitativo é “escala de cinzas”, porém para originais de boa qualidade usa-se o “branco e preto” pois o arquivo gerado fica menor.

Procedimentos alternativos:

A Tese/Dissertação desencadernada oferece melhores condições de trabalho.

Caso o original esteja muito claro é aconselhável fotocopiar a T/D e escanear as cópias em “branco e preto”.

Originais datilografados devem ser escaneados com baixa resolução (75dpi) para não refletir o verso da página escaneada.

Para páginas com imagens utiliza-se resolução de 100 dpi, “escala de cinzas”.

Para páginas coloridas, 200 dpi, cor original.

Todas as alterações de padrões influenciam no tamanho do arquivo final que varia entre 198 Kb e 10.2 Mb.

Foram discutidos neste trabalho alguns dos aspectos mais importantes da implementação da Biblioteca Digital de Teses e Dissertações do IFGW – UNICAMP.

Pretendeu-se mostrar as etapas da geração da informação, como essa informação está organizada, quais as etapas da sua implementação e as soluções encontradas para alguns problemas.

Mostrou-se os recursos tecnológicos utilizados, o fluxo de informação, a geração e organização dessa informação no Banco de Teses Digitais.

A Biblioteca de Teses e Dissertações do IFGW da UNICAMP 100% digital foi o principal resultado deste projeto.

Sendo a primeira biblioteca específica do Brasil com seu acervo de 1128 Teses e Dissertações totalmente digitalizado, num total de 123.763 páginas, desenvolveu um trabalho pioneiro de digitalização de documentos e gerou como resultado a definição de alguns padrões de escaneamento de originais em papel, parte destes já citados anteriormente nos métodos/procedimentos e outros como:

- Média do número de páginas das teses IFGW/UNICAMP = 110 Páginas
- Média do tamanho dos arquivos PDF: 6 Mb
- Média de tempo para escanear/Digitar um Resumo/Abstract de tese = de 30 minutos a 60 minutos

- Média de tempo para escanear um Texto Integral = de 2 horas e 57 minutos até 7 horas e 30 minutos

Outro resultado importante é a integração de sistemas, pois a Biblioteca Digital de Teses e Dissertações permite uma interação com a Base Acervus/Virtua, onde foi criada uma metodologia para catalogar o registro das novas teses usando os dados da ficha catalográfica e disponibilizando esse registro na Base Acervus/Virtua. Informou-se que o documento estava disponível apenas em formato digital e fez-se o link no parágrafo 856 (localização e acesso eletrônico) do formato MARC. Dessa forma esse registro pôde ser disponibilizado também na Biblioteca Digital da UNICAMP. Assim, as teses mais recentes foram inseridas na Biblioteca Digital antes mesmo que o exemplar em papel chegasse ao acervo.

Por outro lado, destacou-se o desenvolvimento de um banco de dados local e de uma interface de busca própria, porém com características técnicas (metadados) suficientes para propiciar a integração com outros sistemas como a Biblioteca Digital da UNICAMP e a Biblioteca Digital Brasileira (IBICT).

#### **Resumo:**

Este artigo mostra o desenvolvimento do banco de dados de teses e dissertações digitais do IFGW – Instituto de Física Gleb Wataghin - UNICAMP. É apresentada a metodologia de criação de acervo digital com base na produção científica atual, gerada em mídia eletrônica desde o final de 1999 e na retrospectiva em papel, totalmente digitalizada.

#### **Palavras-chave:**

Biblioteca Digital ; Digitalização ; Teses e Dissertações ; Conversão de Documentos em Papel para Formato Digital.

#### **Referências**

BEZERRA, E. P. ; ARAÚJO, E. A. de ; BEZERRA, E. P. A análise da construção de uma ferramenta tecnológica : Biblioteca Digital Paulo Freire. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 20., 2002, Fortaleza. **Anais...** Fortaleza: UFC, 2002. 1CD.

CATTELAN, P. Bibliotecas digitais : alternativa viável para gerenciar o caos na Internet. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO, 18., 1997, São Luiz.. **Anais...** São Luiz: Collecta, 1997. (versão eletrônica: disquete)

CUNHA, M. B. da Construindo o futuro : a biblioteca universitária brasileira em 2010. **Ciência da Informação**, v.29, n.1, p. 71-89, jan./abr. 2000. Disponível em: <<http://www.ibict.br/cionline/artigos/>>

HARICOMBE, L.J. Introducion : service to remote users. **Library Trends** , 47, n.1, Summer 1998.

KAHN, R. ; WILENSY, R. **A framework for distributed digital object services**, 1995. Disponível em: <<http://www.cnri.reston.va.us/home/cstr/arch/k-w.html>>

MACHADO, R. N. Biblioteca do futuro na percepção de profissionais da informação. **Transinformação**, v.11, n.3, p. 215-222, set./dez. 1999.

MUELLER, S. P. M. Universidade e informação : a biblioteca universitária e os programas de educação à distância – uma questão ainda não resolvida. **DataGramZero – Revista de Ciência da Informação**, v.1, n.4, ago. 2000. Disponível em: <[http://www.dgz.org.br/ago00/Art\\_01.htm](http://www.dgz.org.br/ago00/Art_01.htm)>