

Saberes e práticas contemporâneas em gestão e inovação na Educação Profissional e em Sistemas Produtivos

Armazenamento de Arquivos em modo Binários em Banco de Dados MySQL: Estudo de Caso Apoiado em um Sistema com Suporte a Catalogação de Trabalhos de Graduação

Jader G.T.L. Pedro¹, Kesley R.G. Alves²; Marcio J. Romero³; Gustavo C. Bruschi⁴; Luis A. da Silva⁵

Resumo - Este artigo tem como objetivo apresentar um modelo de dados relacional simplificado para implantação de um sistema com suporte a catalogação de trabalhos de graduação, facilitando o acesso aos documentos de forma remota a um baixo custo de armazenamento, apresentando como diferencial a implantação ao suporte de campos binários para armazenamento dos arquivos diretamente no banco de dados. Com a aplicação do modelo proposto foi possível provar que é viável a implementação de um banco de dados para gestão de documentos através do uso de campos binários para armazenamento de arquivos.

Palavras-chave: Banco de Dados, Objetos Binários, Gestão de Arquivos.

Abstract - This paper aims to present a simplified relational data model for the implementation system with support for the cataloging of undergraduate work, thus facilitating access to documents remotely at a low cost of storage, introducing as differential the deployment to the support of the use of binary fields to store the files directly in the database. With the application of the proposed model it was possible to prove that it is feasible to implement a database for document management using binary fields for file storage.

Keywords: Database, Binary Objects, File Management.

1. Introdução

Devido ao crescimento da demanda de material acadêmico produzido por alunos das instituições de ensino e a complexidade de se fornecer acesso a estes documentos sob demanda, enxergou-se a dificuldade de manter estes conteúdos alocados apenas na forma física, no qual se tem um alto custo de

¹ Faculdade de Tecnologia de Bauru, jader.gabriel@gmail.com

² Faculdade de Tecnologia de Bauru, kesley_10@hotmail.com

³ Faculdade de Tecnologia de Bauru, marcinho.r@hotmail.com

⁴ Faculdade de Tecnologia de Bauru, gustavo.bruschi@fatec.sp.gov.br

⁵ Faculdade de Tecnologia de Bauru, luis.silva51@fatec.sp.gov.br

Saberes e práticas contemporâneas em gestão e inovação na Educação Profissional e em Sistemas Produtivos

armazenamento e uma disponibilidade limitada aos leitores de acordo com Santos Neto e Santos (2013). A cada semestre são produzidos uma gama de novos artigos e trabalhos de graduação, aumentando significativamente o acervo acadêmico das instituições de ensino do país, no período de 2002 a 2007 observou-se um crescimento de 49% na produção de artigos científicos. Atualmente, os materiais de pesquisas produzidos ficam disponibilizados em bibliotecas físicas para consultas, gerando gastos para mantê-los em bom estado de conservação e disponíveis aos leitores (MUCCIOLI et al., 2007).

Este artigo tem como objetivo apresentar um modelo relacional para banco de dados acadêmicos dando suporte a um sistema para catalogar todos os trabalhos dos alunos de uma instituição de ensino qualquer, facilitando assim, o acesso a estes documentos de forma remota a um baixo custo de armazenamento; e ainda, ampliando a acessibilidade através de plataformas heterogêneas, tais como: dispositivos móveis, *notebooks*, *desktops*, entre outros. Este artigo apresenta os seguintes diferenciais: i) a apresentação de um modelo simplificado, mas consistente que atende a necessidade básica de um modelo acadêmico baseado em artigos de trabalho de graduação e ii) a implantação ao suporte do uso de campos binários para armazenamento dos arquivos diretamente no banco de dados.

O presente trabalho está dividido como segue: na Seção 2 são abordados os referenciais teóricos importantes para o desenvolvimento dos experimentos. Na Seção 3 são descritos os materiais e métodos que foram utilizados na condução dos experimentos. Na Seção 4 é apresentado o modelo de dados proposto, a fim de comprovar a efetividade da proposta, juntamente com os resultados dos experimentos. E por fim, na Seção 5 são apresentadas as considerações finais.

2. Referencial Teórico

Base de dados científica foram criadas com a intenção de disponibilizar em um único acervo centenas de revistas científicas, fazendo com que o pesquisador tenha um ganho de tempo ao procurar o assunto desejado. De um modo mais simplificado, pode-se dizer que essas bases são um local onde encontramos centenas de artigos e revistas, sem a necessidade de se ficar navegando por diversos locais. É importante ressaltar que todas as publicações disponíveis em uma base de dados estão respaldadas por qualidade e originalidade, devido aos criteriosos processos de seleção (TEIXEIRA, 2011).

Existem dois principais tipos de bases, as que oferecem o texto de forma integral e as bases que apenas referenciam, ou seja, que apenas indicam a existência do artigo, ficando a cargo do leitor procurá-lo no site específico de origem. Segundo Meneghini (2003) o site Scielo, mantenedor de uma das maiores bases científicas nacionais, teve como propósito em sua criação, permitir que as produções científicas brasileiras se tornassem cada vez mais visíveis em âmbito internacional e constituir uma base de dados sólida que

Saberes e práticas contemporâneas em gestão e inovação na Educação Profissional e em Sistemas Produtivos

fornece ferramentas que pudessem avaliar a produção nacional de conhecimento.

De acordo com Guimarães et al. (2011) o MySQL é um Sistema Gerenciador de Banco de Dados (SGBD) relacional, utilizado comumente para gerir bases de aplicações livres, geralmente voltadas para a *Web*, utiliza como linguagem de manipulação de dados a Linguagem de Consulta Estruturada ou *Structure Query Language* (SQL). Apresenta fácil integração com a linguagem de programação PHP e também possui característica multiplataforma de Sistemas Operacionais (SO), ainda conta com desempenho de alto nível e uma boa estabilidade na gestão de seus processos.

Atualmente, vem se tornando um dos SGBD's mais utilizados em conjunto com aplicações *Web* no mundo, inclusive por corporações famosas como: Facebook, Suzuki e Google, entre outros. Um recurso ainda pouco explorado nesta plataforma são os armazenamentos de arquivos diretamente nos SGBD's, com o uso do *Binary Large Object* (BLOB) realizando a conversão de arquivos de textos, áudios, filmes, entre outros; em códigos binários permitindo seu armazenamento e gerenciamento diretamente nas colunas das tabelas do banco de dados.

Para Dionísio (2015), o MySQL é preparado para lidar com grande quantidade de dados, o que torna viável a sua utilização na proposta de realização da construção de uma biblioteca acadêmica, garantindo assim a integridade dos arquivos e um acesso rápido aos conteúdos sem prejudicar a performance do banco de dados.

Em sua manipulação é possível trabalhar com tipo de dado BLOB que realiza o armazenamento de arquivos convertidos em dados binários, podendo estes serem imagens, PDF, *Word*, entre outras extensões de arquivos. Dentre as vantagens na utilização deste tipo de dado se destacam: integridade referencial, baixa quantidade de conflito de nomes de arquivos, possibilidade de replicação em vários servidores, além de, o tipo de dado BLOB pode ser utilizado como uma alternativa à necessidade de um repositório de arquivos dedicado para armazenamento, pois, com a utilização deste recurso é possível armazenar no próprio servidor MySQL os arquivos de diversas mídias, eliminando assim a necessidade de um servidor específico de arquivos, para o armazenamentos dos mesmos.

Um campo de uma tabela formatado para este tipo de dado é na realidade uma coleção de dados binários armazenados como uma única entidade em um SGBD. A capacidade do BLOB para armazenamento equivale a 16 *bits*. Já o tipo MEDIUMBLOB, possui a capacidade de armazenamento equivalente à 24 *bits*.

Como qualquer recurso de banco de dados, os campos do tipo BLOB possuem suas particularidades, dentre as quais podemos destacar, não podem ser atribuídos como chaves primárias, além do fato de não podermos utilizar os comandos *GROUP* e *SORT* na realização de consultas SQL.

A Figura 1 demonstra um esquema representativo do processo de conversão de um arquivo PDF para binário. Primeiramente o arquivo original passa por um pré-processamento que pode ocorrer em uma linguagem de programação. O arquivo passa pela conversão de seu conteúdo original em

Saberes e práticas contemporâneas em gestão e inovação na Educação Profissional e em Sistemas Produtivos

binário, após concluído esse processo é gerado o objeto binário que é armazenado no banco de dados.

Importante destacar que os bancos de dados MySQL não armazenam os arquivos originais, apenas seus objetos binários, necessitando assim um processamento inverso para converter o objeto binário no arquivo original na sua recuperação, para posterior visualização.

Figura 1 – Processo de conversão campo BLOB.



Fonte: Os autores.

3. Método

Para a realização dos experimentos foi utilizado um ambiente virtual configurado através do *software VirtualBox*, composta por uma máquina virtual com 4 GB de RAM e contendo um espaço em disco dedicado de 25 GB, a qual recebeu a instalação do banco de dados MySQL, sobre o SO *Windows Server-2008*.

Com os estudos realizados chegou-se a definição das seguintes entidades para um modelo ideal para a proposta apresentada, composta por: *Autores*, *Cursos*, *Eventos*, *Trabalhos*, *Palavras_Chaves*, *Apresentacao*, *Areas*, *Temas*, *Autores_Trabalhos* e *Trabalhos_PalavrasChaves*, onde a entidade *Trabalhos* é a principal do modelo, pois é responsável por armazenar os trabalhos recebidos e referenciar a busca dos conteúdos.

O Modelo de Entidade Relacionamento (MER) foi construído para atender os requisitos de uma biblioteca acadêmica de acordo com a proposta apresentada, conforme é possível observar na Figura 2.

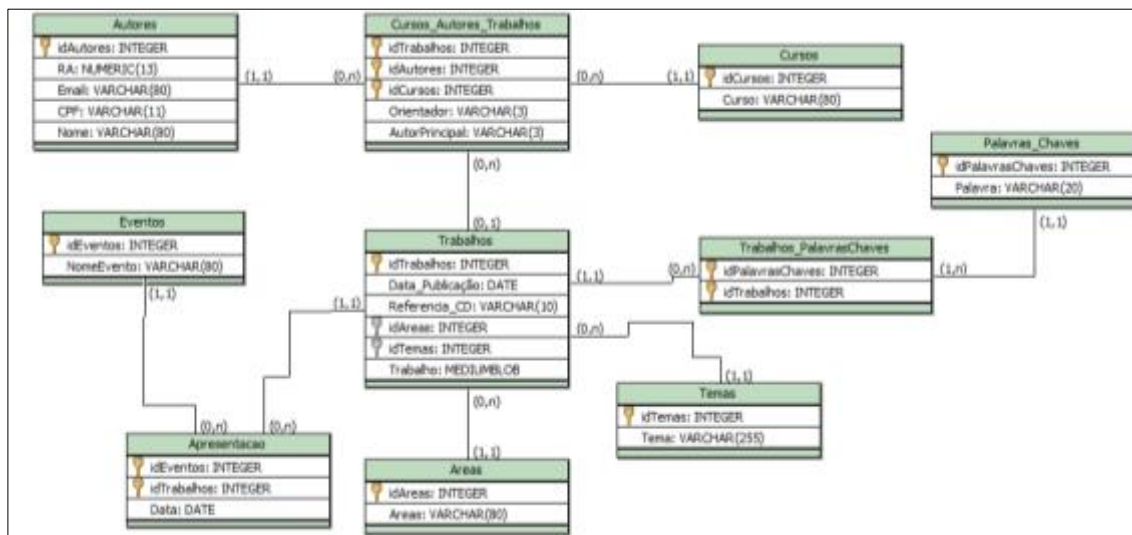
O modelo foi preparado para receber os trabalhos em formato PDF e armazena-los como objetos binários.

Após identificado todos os atributos das entidades, foram definidos como as mesmas estão relacionadas. A entidade *Trabalhos* se relaciona com a entidade *Temas* recebendo como chave estrangeira o campo *idTemas* de cardinalidade *n* para *1*. Da mesma forma é possível visualizar que entidade *Areas* é relacionada com *Trabalhos* utilizando o campo *idAreas* como chave estrangeira de cardinalidade *n* para *1*. Nota-se que entidade *Trabalhos* se relaciona com a entidade *Eventos* através da entidade associativa *Apresentacao*, que possui como atributos as chaves primárias *idEventos* e

Saberes e práticas contemporâneas em gestão e inovação na Educação Profissional e em Sistemas Produtivos

idTrabalhos se tornando uma chave primária composta. A necessidade de se criar uma entidade associativa deve-se ao fato do relacionamento ser de *n* para *n*.

Figura 2 – Modelo Entidade Relacionamento.



Fonte: Os autores.

Para as entidades *Trabalhos* e a entidade *Palavras_Chaves* foi necessária a criação de mais uma entidade associativa com o nome de *Trabalhos_PalavrasChaves*, que tem como atributos as chaves primárias *idPalavrasChaves* e *idTrabalhos* se tornando uma chave primária composta, novamente a necessidade de se criar uma entidade associativa deve-se ao fato do relacionamento de ambas serem de *n* para *n*.

A entidade *Trabalhos* relaciona-se com a entidade *Cursos* e *Autores* através uma entidade associativa *Autores_Trabalhos* no qual as chaves compostas para este relacionamento são: *idTrabalhos*, *idAutores* e *IdCursos*, ambas as entidades com cardinalidade definida de *n* para *n*.

4. Resultados e Discussão

Diante da ideia proposta, foram aplicadas regras no banco para que o armazenamento de arquivos receba documentos do tamanho de até 16 MB com a utilização de colunas *MEDIUMBLOB*. A problemática encontrada a princípio foi ajustar a configuração do servidor *Apache* no arquivo *PHP.ini*, que limitava o *upload* de arquivos em 2 MB, sendo necessária alterar o parâmetro *upload_max_filesize*, que é responsável por delimitar o tamanho máximo de arquivos enviados de 2 para 16 MB: `upload_max_filesize=16M`.

Conforme descrito na Figura 3, foi criada uma classe na linguagem de programação PHP para automatizar o processo de conversão dos arquivos originais em binário. A variável `$_FILES` é responsável por selecionar os

Saberes e práticas contemporâneas em gestão e inovação na Educação Profissional e em Sistemas Produtivos

arquivos em formato PDF com nome temporário e em seguida recuperar o seu tamanho. A função `fopen()` é utilizada para abrir o arquivo, por se tratar de um arquivo binário, é necessário utilizar a opção “`rb`” onde a letra “`r`” solicita a leitura do arquivo e a letra “`b`” força a abertura do arquivo em modo binário. Após esta etapa, o comando `fread()` é responsável pela leitura dos dados do arquivo aberto. O arquivo é então aberto na variável `$conteudo` onde seu tamanho é medido em bytes e encapsulado.

Figura 3 – Código em linguagem PHP de conversão do arquivo PDF para binário.

```
$arquivo = $_FILES['pdf']['tmp_name'];  
$tamanhoArquivo = $_FILES['pdf']['size'];  
$fp = fopen($arquivo, "rb");  
$conteudo = fread($fp, $tamanho);  
$conteudo = addslashes($conteudo);  
fclose($fp);
```

Fonte: Os autores.

Após a conversão dos arquivos originais em binários notou-se um aumento na ordem de 30% no tamanho dos arquivos gerados, em relação ao tamanho original.

Para comprovar a funcionalidade do banco na recuperação dos arquivos, foi criado um sistema de busca conforme ilustrado pela Figura 4, onde este relaciona todos os trabalhos conforme a pesquisa feita pelo usuário e recupera as informações.

Figura 4 – Query para selecionar os trabalhos filtrando de acordo com o tema e classificando em ordem alfabética.

```
SELECT tema, idTemas FROM temas  
WHERE tema LIKE '%" . $tema . "%' ORDER BY tema ASC;
```

Fonte: Os autores.

A construção do *script* para leitura dos dados relaciona os itens da entidade *Temas* associados ao curso desejado como mostrado na Figura 5. O sistema ainda possibilita a busca de trabalhos através dos campos Curso e Áreas relacionadas.

Figura 5 – Query para selecionar os trabalhos filtrando de acordo com o Curso.

```
SELECT t.tema, t.idtemas, c.idcursos FROM cursos_autores_trabalhos cat  
INNER JOIN cursos c ON cat.idcursos= c.idcursos  
INNER JOIN trabalhos tr ON cat.idtrabalhos= tr.idtrabalhos  
INNER JOIN temas t ON t.idtemas = tr.idtemas  
WHERE c.curso = '" . $_GET["curso"] . "'  
ORDER BY tema ASC;
```

Fonte: Os autores.

Saberes e práticas contemporâneas em gestão e inovação na Educação Profissional e em Sistemas Produtivos**5. Considerações Finais**

Foi possível provar com esta proposta que é viável a implementação de um banco de dados para gestão de documentos através do uso de campos BLOB para armazenar arquivos em formato binário diretamente no banco de dados. Visto que possibilita maior segurança e integridade, pois todos os dados, inclusive os arquivos, são administrados pelo SGBD, eliminando assim a necessidade de um servidor de arquivos externo.

Outro item importante a ser destacado, é a relação do tempo para converter e desconverter o arquivo original em objeto binário e objeto binário no arquivo original, que não foi significativa, assim não comprometendo a recuperação ou a exibição dos arquivos.

Como ponto desfavorável, o fato dos arquivos serem armazenados em colunas binárias no banco de dados traz um aumento significativo no espaço de armazenamento em disco utilizado pelo SGBD's, frente ao armazenamento tradicional de um servidor de arquivos comum. Dentro dos experimentos realizados houve um aumento na ordem de 30%, mas isso pode variar de acordo com a quantidade e tipos de arquivos que serão armazenados.

Como trabalhos futuros pode-se indicar a realização de mais pesquisas relacionadas ao armazenamento de arquivos utilizando a forma binária, tanto em SGBD's relacionais, quanto nos não-relacionais, com aplicação de casos práticos.

Referências

- DIONÍSIO, E. J. *Armazenando Imagens no MySQL*. 2015. Disponível em: <<http://www.devmedia.com.br/armazenando-imagens-no-mysql/32104>>. Acesso em: 22 maio 2017.
- GUIMARÃES, M.C.S.; SILVA, C.H.; SANTANA, R.A.L. *Métricas em saúde coletiva: bases quantitativas e qualitativas para a criação de um índice de citação da literatura nacional em Saúde Coletiva*. Florianópolis, 2011. Relatório de pesquisa para o Projeto CNPq - Processo 403522/2008-0. 2011.
- MENEGHINI, R. *O projeto Scielo (Scientific Electronic Library onLine) e a visibilidade da literatura científica Periférica*. 2003. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-40422003000200001>. Acesso em: 18 jun. 2016.
- MUCCIOLI, C.; Campos, M.; GOLDCHMIT M.; DANTAS, P.E.C.; BECHARA, S.J.; COSTA, V.P. *A produção científica no Brasil*. 2007. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0004-27492007000400001>. Acesso em: 14 nov. 2016.
- SANTOS NETO, J.A.; SANTOS, R.P. *A gestão de documentos nos arquivos acadêmicos e a portaria MEC n°. 1.224/2013*. 2015. Disponível em:

**Saberes e práticas contemporâneas em gestão e inovação na Educação Profissional e em Sistemas
Produtivos**

<<http://www.uel.br/revistas/uel/index.php/infoprof/article/view/23398>>. Acesso em: 09 nov. 2016.

TEIXEIRA, M.V. *Você sabe o que são as bases de dados científicas?* Disponível em: <<https://bibliotecaucs.wordpress.com/2011/10/27/voce-sabe-o-que-sao-as-bases-de-dados-cientificas/>>. Acesso em: 02 nov. 2016.