

WILLIAN ZALEWSKI ( UNIOESTE ) , HUEI DIANA LEE , WU FENG CHUNG , ANDRÉ  
GUSTAVO MALETZKE , CLÁUDIO SADY RODRIGUES COY , JOÃO JOSÉ FAGUNDES ,  
JUVENAL RICARDO NAVARRO GOÉS .

*willzal@gmail.com - UNIOESTE*

Com o crescente desenvolvimento tecnológico, as bases de dados estão tornando-se cada vez maiores. Para auxiliar na análise dessa grande quantidade de dados, processos apoiados pela área da computação, tal qual a de Descoberta de Conhecimento em Bases de Dados, tem sido desenvolvidos. Para que esse processo possa ser realizado, é necessário que os dados estejam representados em formatos apropriados como o atributo-valor. Neste trabalho é apresentado um sistema que tem como objetivo auxiliar a fase de pré-processamento no processo de Descoberta de Conhecimento em Bases de Dados permitindo a construção automática de formulários e de uma base de dados relacionados à doença de Chron. Esse sistema constitui parte de uma metodologia que foca o problema do mapeamento de dados contidos em formulários médicos para o formato atributo-valor.

repositório de dados;descoberta de conhecimento;doença de chron.

## Introdução

O avanço tecnológico tem permitido o armazenamento de um volume cada vez maior de informações em Base de Dados - BD - , tornando cada vez mais difícil a utilização de técnicas tradicionais de análise dos dados. Desse modo, torna-se necessária a aplicação de métodos de Inteligência Artificial para uma análise mais completa dessas informações. Um dos processos de apoio à análise de dados nessa área, é o processo de Descoberta de Conhecimento em Bases de Dados - DCBD - [1], o qual auxilia na extração e compreensão de conhecimento contido em BD. Nesse contexto, o processo de DCBD pode ser útil para auxiliar em pesquisas relacionadas à doença de Chron, a qual representa um desafio para médicos e pesquisadores, principalmente pelo aumento de sua incidência e por sua causa e cura serem ainda pouco conhecidos [2]. Porém, na área médica, grande parte dos dados está contida em laudos e formulários preenchidos manualmente, e para que o processo de DCBD possa ser aplicado, é necessário que as informações estejam mapeadas em BD, geralmente, no formato atributo-valor. São diversos os fatores que podem estar associados à ausência desses dados em BD, como a não existência de computadores em todos os ambulatórios médicos, a consideração por muitos de que a utilização de formulários manuscritos no relacionamento com o paciente o torna menos impessoal ou a necessidade por se manter registros impressos.

Freqüentemente, o processo de mapeamento de informações contidas em laudos e formulários médicos é realizado de modo manual. Essa tarefa, além de apresentar um alto custo de tempo, está sujeita a problemas de padronização no mapeamento das informações para a BD. Desse modo, torna-se importante a existência de métodos que realizem esse procedimento de maneira automática.

Este trabalho está inserido dentro do projeto de Análise Inteligente de Dados Aplicada ao Mapeamento de Dados Médicos - AIDMD - [3, 4] e constitui parte de uma metodologia para o mapeamento de dados relacionados à doença de Chron para uma BD no formato atributo-valor. O objetivo desse trabalho é apresentar um sistema que implementa uma metodologia para apoiar na construção automática de formulários médicos de múltipla escolha. Além disso, esse sistema auxilia na construção da BD que receberá as informações mapeadas a partir dos formulários preenchidos.

## Material e Métodos

Para a construção do sistema, foi utilizada a linguagem de programação Java (<http://www.sun.com>) [5], apoiando-se no paradigma orientado a objetos e no conceito de desenvolvimento baseado em três camadas [6]. Para auxiliar na construção do código Java utilizou-se o ambiente de desenvolvimento NetBeans IDE 5.0 (<http://www.netbeans.org>). Na camada de apresentação do sistema foi elaborada uma interface gráfica de fácil interação com o usuário, a qual possibilita a captação de informações na forma de perguntas e respostas, e a geração do formulário a partir desses dados. Para tanto, essas informações são codificadas na linguagem LaTeX (<http://www.latex-project.org>) e inseridas em um arquivo texto. Esse arquivo, após processado pelo aplicativo PDFLaTeX, presente no conjunto de ferramentas disponibilizadas pelo MiKTeX (<http://www.miktex.org>), permite gerar o formulário no formato PDF (*Portable Document Format*).

Em uma fase posterior, as informações preenchidas no formulário deverão ser mapeadas para a BD. Para tornar isso possível é construído um arquivo texto, denominado Arquivo de Interpretação, o qual permite relacionar cada informação preenchida no formulário ao respectivo campo na BD.

Como mencionado, utilizando as mesmas informações coletadas por meio da interface gráfica para a geração do formulário, é construída a BD. Essa BD irá comportar as informações mapeadas dos formulários e está sendo construída no Sistema de Gerenciamento de Banco de Dados - SGBD - [7], MySQL (<http://www.mysql.com>), o qual utiliza a linguagem SQL (*Structured Query Language*) para a manipulação dos dados e construção das tabelas. O processo de construção dos formulários e da BD é representado pela Figura 1.

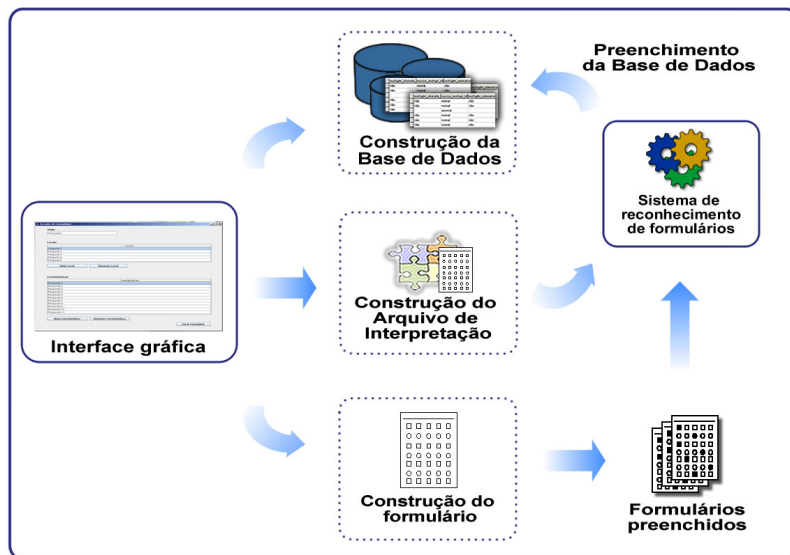


Figura 1 - Representação do processo de construção do formulário e da BD.

## Resultados e Discussão

O sistema desenvolvido neste trabalho, disponibiliza uma interface gráfica, a qual permite ao usuário manipular todos os recursos oferecidos pelo sistema. Por meio dessa interface, a organização das informações no formulário torna-se uma tarefa do sistema, permitindo ao usuário preocupar-se somente com o conteúdo dos dados que irão compor o formulário.

Para a geração dos formulários, utilizou-se a ferramenta MiKTeX, a qual é uma implementação do sistema LaTeX para a plataforma Windows. O LaTeX constitui um pacote de macros que permite por meio do TeX processar e imprimir textos com alta qualidade tipográfica. O TeX é um programa usado para o processamento eletrônico de textos e fórmulas matemáticas. Tanto o TeX como o LaTeX são programas bastante estáveis há muitos anos, tornado-os sistemas confiáveis. Os arquivos que compõem um documento LaTeX possuem formato ASCII (*American Standard Code for Information Interchange*). Essa propriedade possibilita a manipulação do código contido nesses arquivos por ferramentas tradicionais que trabalham sobre este tipo de formato. No caso de formatos binários, aplicações com essas características não seriam viáveis. O formulário gerado segue a formatação PDF, a qual facilita a troca de arquivos entre diferentes plataformas operacionais e possibilita a geração do formulário com um tamanho reduzido.

Com a utilização desse sistema é possível padronizar os formulários quanto aos tipos de dados que serão armazenados na BD. Posteriormente essa BD deverá ser preenchida por meio de um sistema que realiza a extração de informações em formulários preenchidos. Para tornar isso possível, o Arquivo de Interpretação gerado pelo sistema contém informações que possibilitam relacionar perguntas e respostas do formulário a um campo na BD.

### **Conclusões**

Neste trabalho foi apresentado um sistema, o qual implementa uma metodologia para a automatização do processo de construção de formulários de múltipla escolha e de uma BD referentes à doença de Chron. A utilização desse sistema, possibilitará a redução no custo de tempo na construção dos formulários e a minimização na ocorrência de erros de padronização no mapeamento das informações para a BD. Esse sistema está sendo utilizado como componente de um sistema para a extração de informações contidas em formulários médicos relacionados à doença de Chron, preenchidos de forma manual.

### **Agradecimentos**

Ao Programa de Desenvolvimento Tecnológico Avançado - PDTA/FPTI-BR - pela concessão de bolsa de iniciação científica.

### **Referências Bibliográficas**

1. U. Fayyad; G. Piatetsky-Shapiro; P. Smyth in Second International Conference on Knowledge Discovery and Data Mining, Menlo Park, CA, 1996, pages 82-88.
2. F. Cordeiro; J. S. M. Filho; J. C. Prolla. *Endoscopia Digestiva*. Ed.; MEDSI, Rio de Janeiro, 1994.
3. D. F. Honorato; H. D. Lee; M. C. Monard; F. C. Wu; R. B. Machado; A. P. Neto; C. A. Ferrero in Anais do V Encontro Anual de Inteligência Artificial, XXV Congresso da Sociedade Brasileira de Computação, Porto Alegre, 2005, pages 593-601.
4. H. D. Lee, Tese de Doutorado, ICMC-USP, 2005.
5. P. J. Deitel; H. M. Deitel. *Java: Como Programar*. Ed.; Bookman, Porto Alegre, 2003.
6. R. S. Pressman. *Engenharia de Software*. Ed.; McGraw-Hill, Rio de Janeiro, 2002.
7. H. F. Korth; A. Silberschatz. *Sistemas de Banco de Dados*. Ed.; Makron Books, São Paulo, 1995.