

## CONSTRUÇÃO DE BASES DE DADOS APOIADA POR MÉTODOS COMPUTACIONAIS A PARTIR DE FORMULÁRIOS MÉDICOS: ESTUDO DE CASO EM DOENÇA DE CROHN

MALETZKE AG; LEE HD; ZALEWSKI W; MATSUBARA ET; VOLTOLINI RF; COY CSR; GÓES JRN; WU FC; FAGUNDES JJ

Lab. de Bioinformática/Univ. Est. do Oeste do Paraná-UNIOESTE/PTI e Serviço de Coloproctologia-UNICAMP

**OBJETIVOS:** Apresentar uma metodologia computacional que auxilia no mapeamento de formulários médicos para uma Base de Dados — BD — estruturada e um estudo de caso aplicado ao domínio médico relacionado à doença de Crohn. **MÉTODOS:** A metodologia proposta está dividida em três etapas: (1) Geração de Formulários e Construção da Base de Dados, (2) Construção de Padrões sobre Formulários e (3) Mapeamento de Formulários e Preenchimento da Base de Dados. A primeira etapa é responsável pela construção do formulário a partir de um conjunto de informações, pela construção de uma Base de Dados e de um Arquivo de Interpretação — AI, o qual é responsável por indicar o valor que deve ser armazenado na BD de acordo com as respostas que foram marcadas no formulário e o tipo de dado do atributo. O formulário construído está estruturado no formato de perguntas e respostas, sendo que uma determinada pergunta poderá conter várias respostas. Cada pergunta representa um atributo na BD e as respostas (campos ou marcas) um dos possíveis valores que serão conferidos a esse atributo. A partir da construção automática do formulário é possível realizar a padronização em relação à formatação e minimizar o custo de tempo na construção do formulário, do AI e da BD. Na segunda etapa é realizada a construção dos padrões de um formulário por meio da identificação e armazenamento de informações como tamanho, localização e formato dos campos do formulário, as quais serão consideradas na etapa seguinte. Para isso, é realizada a análise das estruturas horizontais e verticais de um documento digitalizado. Desse modo, esta etapa está dividida em três fases: Segmentação Horizontal, Segmentação Vertical e Construção de Padrões. Na Fase 1 é analisada a estrutura horizontal do formulário com intuito de construir um conjunto formado por Bandas de Texto — BT, a qual BT consiste em um conjunto de linhas contínuas, na qual cada linha possui uma frequência de pixels pretos maior ou igual a um determinado limiar. Após, para cada BT construída é analisada, na Fase 2, sua estrutura vertical de modo que seja construído um novo conjunto de BTs. Assim, cada BT gerada na Fase 2 representa possíveis objetos presentes no formulário como caracteres e marcas. Já na Fase 3 as BTs construídas na Fase 2 são classificadas com o intuito de identificar quais destas representam uma marca. As BTs classificadas como marcas terão suas coordenadas mapeadas para uma Base de Padrões — BP. Anteriormente à aplicação das Etapas 2 e 3, é necessário que os formulários sejam submetidos a uma fase de Pré-processamento, na qual esses são tratados para que possíveis ruídos e imperfeições indesejadas sejam removidos. Por último, considerando que as Etapas 2 e 3 já foram realizadas e considerando um formulário preenchido é realizada a terceira etapa, na qual esse formulário é mapeado para a BD por meio da análise da BP e do AI. A partir da metodologia proposta foi desenvolvido um Sistema Computacional que provê os recursos apresentados anteriormente e atua como uma ferramenta para aplicação desta metodologia. Este trabalho apresenta um estudo de caso aplicado à doença inflamatória de Crohn. A incidência dessa doença na população mundial tem aumentado nos últimos anos e suas causas e cura ainda são pouco conhecidas. Nesse contexto, o Laboratório de Bioinformática em parceria com o Serviço de Coloproctologia da Faculdade de Ciências Médicas da Universidade de Campinas elaboraram um protocolo contendo um conjunto variado de informações relacionadas à essa enfermidade, as quais foram consideradas relevantes para a análise. Dentre essas informações foi selecionado um subconjunto, conjuntamente com especialistas da área, para a aplicação da metodologia proposta. Esse subconjunto foi organizado em 17 questões por formulário, totalizando 50 campos a serem considerados para preenchimento. Após, 100 cópias do formulário modelo foram geradas e distribuídas a dez colaboradores, os quais preencheram os formulários recebidos de modo “*ad libitum*”, apenas fixando-se como critérios de preenchimento a realização da marcação próxima ao centro do campo e aplicando-se uma pressão habitual de escrita. Posteriormente, a Etapa 3 da metodologia foi aplicada a cada um dos cem formulários. **RESULTADOS:** Os resultados mostraram que 87 formulários apresentaram 100,00% de suas marcas preenchidas mapeadas corretamente para a BD, ou seja, todas as 17 alternativas de cada um desses formulários foram mapeadas corretamente para a BD. Outros oito formulários apresentaram um mapeamento correto de 94,12%, três formulários obtiveram 88,24% e dois formulários apresentaram 82,35% das marcas preenchidas mapeadas corretamente para a BD. Desse modo, obteve-se uma precisão de 98,82% com desvio padrão 3,45%.

**CONCLUSÕES:** A metodologia apresentada auxiliou no processo de mapeamento de informações, tornando-o uma tarefa menos custosa em relação ao tempo de mapeamento, e menos propensa a erros se comparada ao mapeamento manual. Além disso, torna possível a utilização de documentos impressos no registro das informações, característica considerada importante tanto para especialistas na área médica quanto para estabelecimentos que necessitam manter registros na forma de documentos em papel ou até mesmo pela não disponibilidade de computadores em todos os ambulatórios. Os especialistas avaliaram que a metodologia atendeu aos propósitos deste trabalho possibilitando a coleta de dados dispostos em documentos impressos para uma representação computacional. O sistema computacional desenvolvido a partir da metodologia será de grande utilidade em todas as etapas do mapeamento, pois é possível observar por meio dos resultados apresentados, que um dos fatores para obter-se um desempenho ainda mais eficiente é a possibilidade de realizarem-se ajustes nos parâmetros de cada etapa de acordo com as características de um determinado formulário.