

Mapeamento de Informações Médicas descritas em Formulários para Bases de Dados Estruturadas*

André Gustavo Maletzke^{1,2}, Huei Diana Lee^{1,2}, Willian Zalewski¹, Edson Takashi Matsubara², Richardson Floriani Voltolini¹, Cláudio Sady Rodrigues Coy³, João José Fagundes³, Juvenal Ricardo Navarro Góes³, Wu Feng Chung^{1,2,3}

¹Centro de Engenharias e Ciências Exatas – Universidade Estadual do Oeste do Paraná
Laboratório de Bioinformática – LABI
Parque Tecnológico Itaipu – PTI
Caixa Postal 39, 85856-970 – Foz do Iguaçu, PR, Brasil

²Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo
Laboratório de Inteligência Computacional – LABIC
Caixa Postal 668, 13560-970 – São Carlos, SP, Brasil

³Faculdade de Ciências Médicas – Universidade Estadual de Campinas
Serviço de Coloproctologia
Caixa Postal 6111, 13083-970 – Campinas, SP, Brasil

{andregustavom, hueidianalee, willzal}@gmail.com

Abstract. *Computational processes are increasingly being used to support analysis and extraction of patterns embedded in great quantities of data. Nevertheless, some areas, such as medicine, frequently register information through traditional methods using printed documents, disabling the direct application of such processes. This work presents a methodology to assist the semi-automation of medical information mapping, from printed forms, to structured databases. It is also presented a case study in which the methodology is applied to construct a database from information related to Crohn disease.*

Resumo. *Processos computacionais são cada vez mais utilizados para auxiliar na análise e extração de padrões contidos em grandes volumes de dados. Todavia, algumas áreas como a de medicina registram frequentemente as informações por meio de métodos tradicionais utilizando-se de documentos impressos, impossibilitando a aplicação desses processos. Nesse contexto, este trabalho apresenta uma metodologia para a semi-automatização do processo de coleta de informações médicas, por meio da utilização de formulários impressos, para uma Base de Dados estruturada. Também é apresentado um estudo de caso no qual essa metodologia é aplicada na construção de uma Base de Dados a partir de informações referentes à doença de Crohn.*

1. Introdução

Atualmente, a quantidade de dados armazenada em meios digitais aumenta com uma velocidade considerável, tornando a análise desses dados uma tarefa crescentemente complexa por meio de métodos manuais. Desse modo, cada vez mais busca-se desenvolver e

*Trabalho desenvolvido com o apoio da Fundação Parque Tecnológico de Itaipu – PTI

estudar métodos e processos que possam auxiliar na análise de dados por meio de ferramentas computacionais [1]. Um dos processos de apoio à análise de dados é o processo de Descoberta de Conhecimento em Bases de Dados — DCBD [2]. Esse processo pode ser aplicado para construir modelos que representem o conhecimento contido em grandes volumes de dados de modo que possam auxiliar especialistas no processo de tomada de decisão. O DCBD é um processo iterativo e interativo, dividido em três fases: Pré-processamento, Mineração de Dados e Pós-processamento. A fase de Pré-processamento é a que apresenta maior custo de tempo e tem por objetivo conhecer a aplicação e os dados e prepará-los para a fase seguinte. A fase de Mineração de Dados apresenta como finalidade a aplicação de métodos computacionais que permitam a extração e a representação do conhecimento contido nos dados [3]. Já na fase de Pós-processamento é realizada a avaliação e a validação dos modelos construídos com o auxílio de especialistas [1].

Na área de medicina, os dados encontram-se, freqüentemente, em formatos desestruturados ou semi-estruturados dificultando a análise por meio de métodos computacionais. Esses dados podem estar representados, por exemplo, em laudos médicos ou formulários impressos contendo diversas informações como sintomatologia e histórico de pacientes. Para que esses dados possam ser analisados por meio da aplicação de métodos computacionais, como o processo de DCBD, é necessário que estejam representados em um formato adequado, como exemplo, o formato atributo-valor [4]. Diversos fatores estão relacionados à indisponibilidade desses dados em uma representação adequada ao DCBD, como a não existência de computadores em ambulatórios médicos, a consideração por muitos desses profissionais de que a utilização de documentos impressos torna o relacionamento com o paciente menos impessoal ou a necessidade de manter-se um registro impresso [5].

Desse modo, torna-se necessário o desenvolvimento de métodos e ferramentas que auxiliem no processo de estruturação da informação de modo semi-automático, buscando-se, principalmente, a padronização e o ganho de tempo. Este trabalho é parte do projeto de Análise Inteligente de Dados aplicado ao Mapeamento de Dados Médicos — AIDMD [6], e tem como objetivo apresentar uma metodologia que auxilia no processo de mapeamento de formulários médicos para Bases de Dados — BD — estruturadas e um estudo de caso aplicado ao mapeamento de informações referentes à doença de Crohn. O AIDMD está sendo desenvolvido pelo Laboratório de Bioinformática — LABI — da Universidade Estadual do Oeste do Paraná — UNIOESTE — em parceria com o Laboratório de Inteligência Computacional — LABIC — da Universidade de São Paulo / São Carlos — USP — e o Serviço de Coloproctologia da Faculdade Ciências Médicas da Universidade Estadual de Campinas — UNICAMP.

Este trabalho está estruturado da seguinte maneira: na Seção 2 é descrita a metodologia proposta, na Seção 3 é apresentado um estudo de caso aplicado à doença de Crohn e na Seção 4 são apresentadas as conclusões deste trabalho e trabalhos futuros.

2. Metodologia Proposta

Métodos de Inteligência Artificial vêm sendo cada vez mais utilizados para auxiliar na análise e extração de padrões contidos em grandes volumes de dados. Nesse contexto, foi proposta uma metodologia para auxiliar na coleta e no armazenamento de dados, de maneira estruturada em uma Base de Dados, a partir de formulários médicos de múltipla

escolha. A metodologia desenvolvida foi dividida em três etapas:

1. Geração de Formulários e Construção da Base de Dados;
2. Construção de Padrões sobre Formulários;
3. Mapeamento de Formulários e Preenchimento da Base de Dados.

Na Figura 1 são apresentadas as etapas da metodologia proposta. A primeira etapa

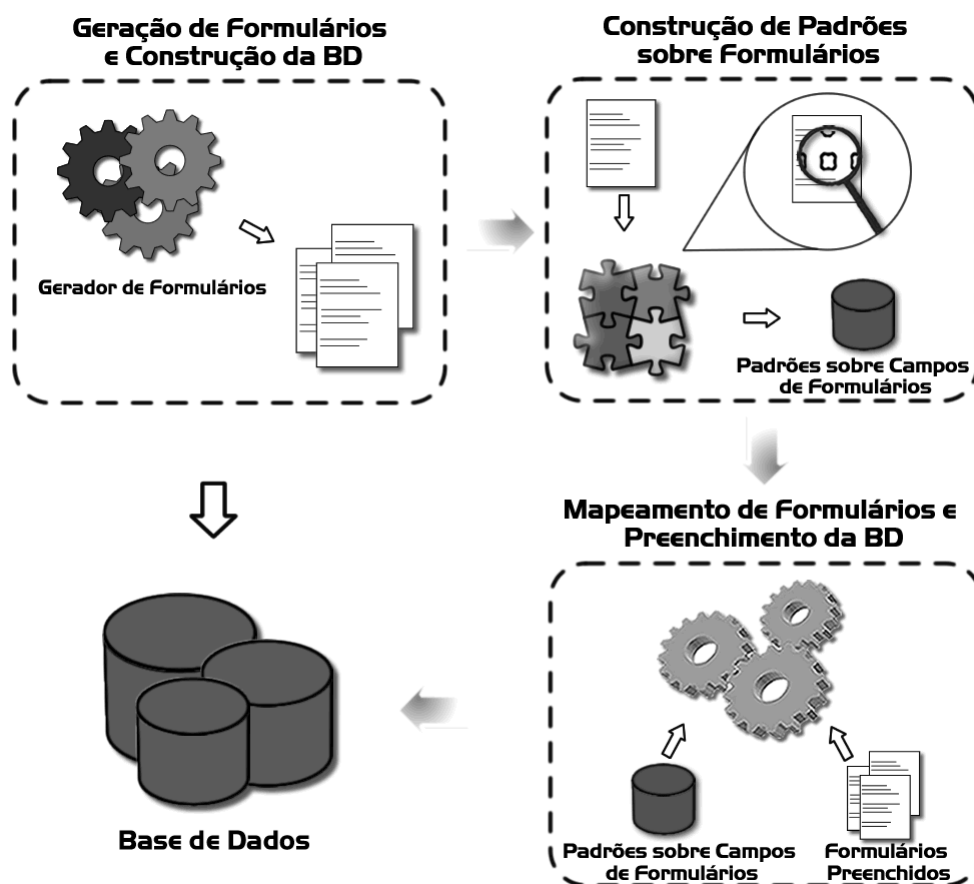


Figura 1. Etapas da metodologia proposta.

tem por objetivo gerar um formulário modelo, a partir de um conjunto de informações, relacionadas a um determinado domínio, e construir uma Base de Dados, na qual essas informações deverão ser mapeadas, juntamente com um Arquivo de Interpretação — AI. A segunda etapa é responsável pela construção dos padrões a partir de um formulário modelo não preenchido. Esses padrões são referentes, por exemplo, à localização, ao formato e ao tamanho das marcas ou dos campos do formulário. É importante notar que a proposição das Etapas 1 e 2 tem como objetivo auxiliar na agilização do processo de mapeamento de informações de formulários médicos para BD estruturadas, possibilitando realizar o mapeamento de formulários cuja estrutura e organização das informações sejam desconhecidas. Desse modo, o processo de mapeamento torna-se uma tarefa menos custosa e mais eficiente, se comparada a métodos em que todas as regiões de interesse devem ser mapeadas manualmente. Já a terceira etapa, na qual são considerados formulários preenchidos, é responsável pelo mapeamento desses formulários, de maneira que a BD construída na Etapa 1 seja preenchida. No entanto, antes das Etapas 2 e 3

serem realizadas, é necessário que os formulários sejam submetidos a uma fase de Pré-processamento. Nessa fase, os formulários são tratados com o objetivo, por exemplo, de remover possíveis ruídos e corrigir imperfeições indesejadas. Portanto, nessa fase podem ser aplicados diversos métodos existentes na área de Processamento Digital de Imagens, como filtros lineares e morfológicos [7].

Pré-processamento

Como mencionado, o objetivo desta fase é preparar os formulários para que possam ser analisados pelas Etapas 2 e 3 da metodologia proposta. Nesta fase são realizadas as seguintes operações: binarização, correção da escala e da inclinação e segmentação do formulário. A binarização da imagem de um formulário, consiste em transformar uma imagem, representada por diversos níveis de cinza, em outra contendo somente duas cores, preto (1) ou branco (0) [7]. Posteriormente, é realizada a correção no ângulo de inclinação do formulário, caso seja necessário, por meio da aplicação de um método de interpolação. Para isso, é realizada a leitura da Marca de Referência — MR — do formulário, representada por uma reta situada no cabeçalho do formulário e que se estende por toda sua largura, a qual é ilustrada na Figura 2. É realizada também a correção na escala da imagem, caso o formulário tenha sido digitalizado em uma resolução superior ou inferior a definida como padrão. O método de interpolação *Nearest Neighbor* foi escolhido como padrão a ser utilizado, o qual foi determinado por meio da análise dos experimentos desenvolvidos em [5]. Por último, é realizado o processo de segmentação, que consiste em dividir uma imagem em outra menor ou em um grupo de imagens menores [8]. Para realizar a segmentação, são utilizadas informações da MR como localização e comprimento. Portanto, após a segmentação, o novo formulário gerado é formado pela área retangular, delimitada pelas coordenadas (x_0, y_0, x_f, y_f) , onde x_0 e y_0 são referentes às coordenadas iniciais da MR e x_f e y_f são referentes aos comprimentos da MR e do formulário, respectivamente. Com o término desta fase, o formulário estará no formato apropriado para ser processado pela Etapa 2 e, posteriormente, pela Etapa 3.

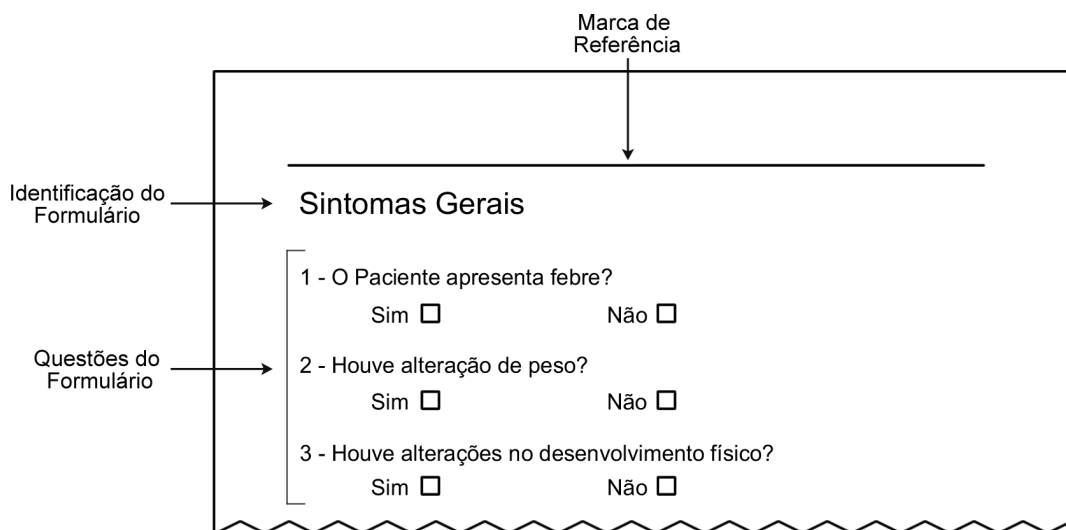


Figura 2. Representação parcial de um formulário.

2.1. Etapa 1: Geração de Formulários e Construção da Base de Dados

Nessa etapa são construídos os formulários a partir de um conjunto de atributos e/ou características, uma BD para a qual esses atributos deverão ser mapeados pelo processo de mapeamento e um Arquivo de Interpretação, o qual contém as regras de preenchimento da BD para cada tipo de formulário construído e está representado na linguagem *Extensible Markup Language* — XML¹. O AI é utilizado na Etapa 3 da metodologia e 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.

Os formulários construídos estão estruturados no formato de perguntas e respostas, sendo que uma determinada pergunta poderá conter várias respostas, as quais são representadas pelo seu significado seguido de uma marca ou campo. Neste trabalho esses campos ou marcas são representados pela figura geométrica de um quadrado, identificando a região que deverá ser preenchida. Cada pergunta representa um atributo na BD e as respostas um dos possíveis valores que serão conferidos a esse atributo.

Após a definição das perguntas e suas respectivas respostas, a construção dos formulários é realizada de maneira automática por meio da geração de um arquivo na linguagem \LaTeX^2 , o qual é processado pelo aplicativo \PDF\LaTeX , presente no conjunto de ferramentas disponibilizadas pelo \MikTeX^3 , gerando o formulário no formato *Portable Document Format* — PDF. Desse modo, é possível realizar a padronização em relação à formatação e minimizar o custo de tempo na construção dos formulários, do AI e da BD. Na Figura 2 é ilustrada uma representação parcial de um formulário construído com informações referentes à sintomatologia de paciente.

2.2. Etapa 2: Construção de Padrões sobre Formulários

Nesta etapa é realizado o reconhecimento das áreas de interesse, ou seja, dos campos de um formulário modelo não preenchido. Desse modo, informações como a localização de cada campo de um formulário são considerados padrões de um determinado formulário. Esses padrões serão utilizados na Etapa 3 para realizar o mapeamento dos formulários preenchidos para a Base de Dados. O processo de construção desses padrões baseia-se no algoritmo descrito em [9], que explora o fato de que grande parte dos documentos digitalizados possui uma estrutura vertical e horizontal. Esta etapa é dividida em três fases:

1. Fase de segmentação horizontal;
2. Fase de segmentação vertical;
3. Fase de construção de padrões.

Fase de segmentação horizontal

Nesta fase é analisada a estrutura horizontal dos formulários, por meio da geração de um Histograma de Projeção Horizontal — HH [9]. O HH consiste em contabilizar para cada linha horizontal do formulário a quantidade de pixels pretos presentes nessa linha, considerando a espessura de uma linha equivalente a um pixel. Assim, esse histograma é

¹<http://www.w3.org/XML>

²<http://www.latex-project.org>

³<http://www.miktex.org>

representado por um vetor de inteiros, em que cada posição do vetor armazena a quantidade de pixels pretos correspondente à linha analisada. Posteriormente, a partir da análise do HH, o formulário é segmentado por meio da determinação de um Limiar de Segmentação — LS — (normalmente 0 ou 1) em Banda de Texto — BT — e Banda Branca — BB. A BT consiste em um conjunto de linhas contínuas que apresentam quantidade de pixels pretos maior ou igual ao LS. Já uma BB consiste em um conjunto de linhas contínuas que possuem menor quantidade de pixels pretos em relação ao LS. Portanto, na segunda fase somente o conjunto de BTs será considerado, pois o conjunto de BBs representa áreas do formulário que não contém nenhuma informação a ser analisada.

Fase de segmentação vertical

Nesta fase são analisadas as BTs por meio da geração de um Histograma de Projeção Vertical — HV — que consiste em contabilizar a quantidade de pixels pretos presentes em cada coluna de uma determinada BT, considerando a espessura de uma coluna equivalente a um pixel. Então, cada HV é particionado por meio do LS, em um conjunto de HV menores, os quais irão compor o Conjunto de Busca — CB. Desse modo, os elementos do CB, denominados de HVB, representam cada elemento presente no formulário, como marcas, caracteres e possíveis ruídos que não foram tratados na fase de Pré-processamento.

Ainda nesta fase é realizada a construção da Base de Exemplos — BE, para a qual deve-se selecionar um subconjunto de marcas ou campos presentes no formulário, as quais serão utilizadas como exemplos para rotular os elementos do CB. Um exemplo é representado pelo Histograma de Projeção Vertical, denominado HVE, de uma marca de interesse. Assim, para cada marca selecionada é armazenado o seu HVE juntamente com suas respectivas coordenadas na BE.

Fase de construção de padrões

A fase de construção de padrões é responsável pela construção da Base de Padrões — BP — na qual todas as marcas do formulário são mapeadas e suas respectivas coordenadas são armazenadas. A BP será utilizada na Etapa 3 da metodologia para identificar quais marcas de um determinado formulário foram preenchidas. A construção da BP consiste em classificar todos os elementos presentes no CB, utilizando-se para isso os elementos da BE. Desse modo, foi implementado um classificador baseado no algoritmo *Nearest Neighbor* [10]. Esse método classifica um exemplo não rotulado por meio da atribuição da classe do seu vizinho mais próximo, de acordo com a medida de similaridade utilizada. O algoritmo implementado compara cada elemento do CB com os elementos presentes na BE utilizando uma técnica baseada na comparação de séries temporais, denominada *Dynamic Time Warping* — DTW [11]. Caso seja encontrada a similaridade entre o HVB em análise com algum HVE, a comparação é finalizada. Caso contrário, a comparação é realizada novamente, porém entre as frequências H_x do HVE e as frequências H_{x+t} do HVB, com x variando de 0 até o comprimento do HVE, t variando de 1 até n em intervalos de uma unidade e n sendo um valor inteiro menor ou igual à diferença de tamanho existente entre o HVE e o HVB. Para determinar a similaridade entre dois histogramas é necessário estabelecer uma medida de distância, que indique quão similares são dois histogramas. Neste trabalho foi utilizada a medida de distância *Chi-square* — χ^2 — para determinar o grau de similaridade entre um elemento do CB e um elemento da BE [12, 13].

2.3. Etapa 3: Mapeamento de Formulários e Preenchimento da Base de Dados

A Etapa 3 tem por objetivo identificar, a partir de um formulário preenchido, quais marcas foram assinaladas e, posteriormente, preencher a BD. Para isso, são utilizados os padrões construídos para identificar as marcas na etapa anterior e o Arquivo de Interpretação para realizar o preenchimento da BD. Os formulários preenchidos são submetidos, primeiramente, à fase de Pré-processamento, na qual possíveis imperfeições são tratadas e a área de interesse é segmentada. Após, os formulários são processados para a identificação das marcas que foram preenchidas. Os seguintes passos são realizados:

- Segmentação de cada marca do formulário por meio das coordenadas contidas na Base de Padrões;
- Contabilização da percentagem de pixels pretos que cada marca segmentada ocupa;
- Identificação das alternativas marcadas por meio da determinação de um Limiar de Preenchimento — LP.

Em seguida, é realizada a análise do AI que irá indicar, de acordo com a marca assinalada no formulário, qual será o atributo da BD que deverá ser preenchido, o valor e o tipo de dado desse atributo. Na Figura 3 são apresentados todos os passos realizados desde a utilização do gerador de formulários para a construção do formulário padrão até o mapeamento das informações contidas nos formulários preenchidos para a BD.

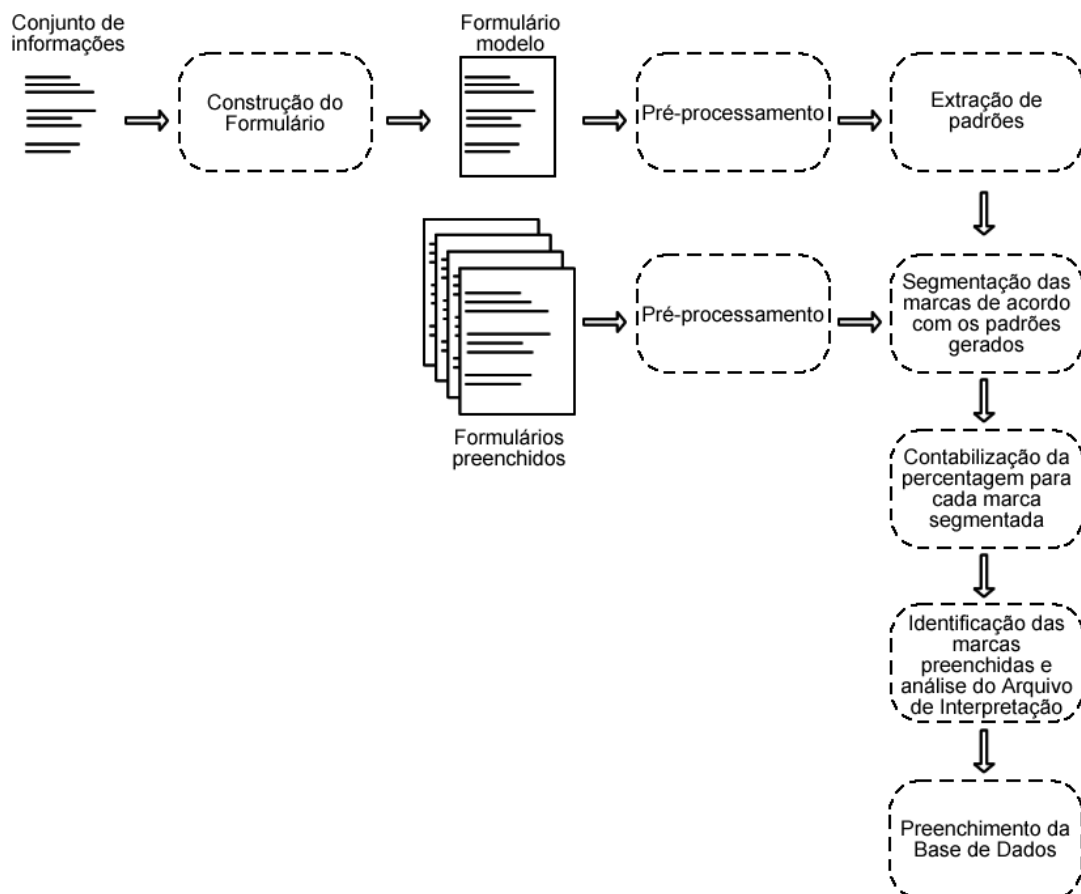


Figura 3. Detalhamento das Etapas 1, 2 e 3 da metodologia proposta.

Com o objetivo de prover suporte para a aplicação da metodologia proposta, foi desenvolvido um Sistema Computacional constituído por três módulos, os quais representam cada uma das etapas dessa metodologia. Esse sistema foi implementado na linguagem de programação JAVA⁴, utilizando o Sistema de Gerenciamento de Banco de Dados *MySQL*⁵ e a API *Java Advanced Imaging* — JAI — para a manipulação dos formulários digitalizados [14].

3. Aplicação da Metodologia Proposta à Doença de Crohn

A doença de Crohn é uma enfermidade inflamatória com componente imunológica que causa distúrbios em seu funcionamento e acarreta o surgimento de sinais e sintomas. Foi descrita pelo médico Burril B. Crohn em 1932 e até os dias de hoje a etiologia e a fisiopatologia dessa doença são pouco conhecidas [15, 16]. Neste estudo de caso é apresentada a aplicação da metodologia proposta à doença inflamatória de Crohn. Para tanto, o LABI em parceria com o Serviço de Coloproctologia da Faculdade de Ciências Médicas da UNICAMP elaboraram um protocolo contendo um conjunto de informações, consideradas relevantes pelos especialistas envolvidos de ambas as instituições, relacionadas à doença de Crohn como resultados de exames clínicos, sintomatologia, histórico alimentar do paciente entre outros. Neste estudo de caso será abordado somente um subconjunto de informações relacionado à sintomatologia de pacientes para a construção do formulário, conforme apresentado na Figura 4. A partir dessas informações e por meio da aplicação da Etapa 1 da metodologia, foram construídos o formulário modelo, a Base de Dados e o Arquivo de Interpretação. Esse formulário foi impresso e, posteriormente, digitalizado e processado pela fase de Pré-processamento. Após, o formulário foi submetido à Etapa 2 da metodologia proposta para que os padrões presentes fossem identificados e armazenados — Figura 3. Como mencionado, os padrões referem-se à localização das marcas ou dos campos presentes nos formulários, os quais representam as respostas de cada questão do formulário. Uma vez que os padrões tenham sido identificados e armazenados, cópias do formulário poderão ser disponibilizadas para que sejam utilizadas nos ambulatórios.

- Febre: (Sim/Não)
- Alterações do peso: (Sim/Não)
- Alterações do desenvolvimento físico: (Sim/Não)
- Alterações do revestimento cutâneo: (Sim/Não)
- Anemia: (Sim/Não)
- Feridas na boca: (Sim/Não)
- Presença de dor na parede abdominal: (Sim/Não)
- Mal estar pós-prandial: (Sim/Não)
- Diarréia: (Sim/Não)
- Hemorragia digestiva: (Sim/Não)

Figura 4. Conjunto de informações.

Com o intuito de avaliar a metodologia proposta, foram realizadas quarenta cópias desse formulário, as quais foram distribuídas a dez colaboradores para que fossem completadas. O preenchimento dos formulários foi realizado de modo “ad libitum”, apenas fixando-se como critério a realização do preenchimento próximo ao centro de cada marca, aplicando-se uma pressão normal de escrita. Após, os quarenta formulários foram submetidos aos passos apresentados na Figura 3.

Neste estudo de caso foi utilizado um LP igual a 10%, ou seja, se os campos de cada formulário analisado apresentarem uma porcentagem de pixels pretos maior ou igual

⁴<http://www.sun.com/>

⁵<http://www.mysql.com/>

a 10% em relação à quantidade total de pixels presentes no campo analisado, essa marca será considerada como preenchida. Após a identificação de todos os campos assinalados foram realizadas a leitura e a análise do Arquivo de Interpretação para identificar quais atributos da BD deveriam ser preenchidos, o valor e o tipo de dado desse atributo.

Uma análise dos resultados mostrou que do total de quarenta formulários, 24 obtiveram 100% dos campos mapeados para a BD, 13 obtiveram entre 80% e 99% e 3 formulários obtiveram menos de 80% das marcas preenchidas mapeadas para a Base de Dados. É relevante ressaltar que os formulários que obtiveram percentagem de mapeamento inferior a 80% foram preenchidos de modo menos legível e preciso e, observou-se que os campos desses formulários foram assinalados com um marcador fino e o preenchimento não foi realizado próximo ao centro da marca. Desse modo, observa-se que 92,75% da BD foi preenchida com uma precisão de 82,50% e um desvio padrão de 17,08%. Esse resultado, ainda poderá ser melhorado por meio de ajustes nos parâmetros dos métodos e dos algoritmos aplicados durante todo o processo, desde a fase de Pré-processamento até a etapa de Mapeamento dos Formulários e Preenchimento da BD. Outras características, não referentes aos métodos e aos algoritmos implementados, como espessura e cor do marcador utilizado poderão ser considerados para obter uma melhoria nos resultados.

4. Conclusão

Neste trabalho foi apresentada uma metodologia para a semi-automatização do processo de coleta de informações, por meio da utilização de formulários de múltipla escolha, para uma Base de Dados estruturada. Apresentou-se também um estudo de caso aplicando a metodologia proposta a um subconjunto de informações referentes à sintomatologia de paciente relacionadas à doença inflamatória de Crohn. Essa doença tem despertado grande interesse de médicos e pesquisadores, e trata-se de uma enfermidade cujas causas e cura ainda são pouco conhecidas.

Esta metodologia proporcionou uma diminuição no custo de tempo e uma maior padronização no processo de mapeamento de informações para uma Base de Dados. Ainda, essa metodologia provê suporte para que especialistas da área médica possam continuar utilizando documentos impressos durante o relacionamento com os pacientes, o que é considerado, por muitos especialistas, uma maneira mais pessoal de relacionar-se com os pacientes, além da não disponibilidade de computadores em todos os ambulatórios, entre outros fatores.

Os especialistas do domínio avaliaram que a metodologia proposta atendeu ao objetivo deste trabalho possibilitando que os formulários fossem mapeados para uma BD estruturada. O sistema implementado a partir da metodologia proposta será de grande utilidade durante todas as etapas dessa metodologia, pois possibilita que os distintos parâmetros dos algoritmos aplicados ao mapeamento sejam ajustados para tipos diferentes de formulários, buscando-se melhores resultados para cada caso.

Trabalhos futuros, incluem a aplicação de processos apoiados por métodos computacionais, como o processo de DCBD, para a extração de padrões contidos na Base de Dados construída por meio desta metodologia. Outro trabalho inclui o desenvolvimento de um módulo para o mapeamento de caracteres manuscritos, possibilitando um maior nível de detalhamento dos dados mapeados e um enriquecimento dos padrões construídos a partir da BD gerada. Atualmente, estão também sendo realizadas novas avaliações

considerando amostras de dados mais amplas.

Agradecimentos: Ao Programa de Desenvolvimento Tecnológico Avançado — PDTA/FPTI-BR — pelo auxílio por meio da linha de financiamento de bolsas.

5. Referências

- [1] S. O. Rezende, *Sistemas Inteligentes: Fundamentos e Aplicações*, Manole, Barueri, Brasil, 2003.
- [2] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “Knowledge discovery and data mining: Towards a unifying framework”, In *Second International Conference on Knowledge Discovery and Data Mining*, pages 82–88, Menlo Park, USA, 1996.
- [3] I. H. Witten and E. Frank, *Data mining: practical machine learning tools and techniques*, Elsevier, San Francisco, USA, 2 edition, 2005.
- [4] D. F. Honorato, H. D. Lee, M. C. Monard, F. C. Wu, R. B. Machado, A. P. Neto, and C. A. Ferrero, Uma metodologia para auxiliar no processo de construção de base de dados estruturadas a partir de laudos médicos, In *V Encontro Internacional de Inteligência Artificial*, Porto Alegre, Brasil, 2005.
- [5] A. G. Maletzke, H. D. Lee, F. C. Wu, E. T. Matsubara, C. S. R. Coy, J. S. Fagundes, and J.R.N. Góes, Uma metodologia para auxiliar no processo de mapeamento de formulários médicos para bases de dados estruturadas, In *X Congresso Brasileiro de Informática em Saúde*, Florianópolis, Brasil, 2006.
- [6] H. D. Lee, Seleção de atributos importantes para a extração de conhecimento de bases de dados, Tese de Doutorado, ICMC - USP, São Carlos, Brasil, 2005.
- [7] R. C. Gonzalez and R. E. Woods, *Digital image processing*, Prentice Hall, New Jersey, USA, 2 edition, 2002.
- [8] Y. J. Zhang, “Evaluation and comparison of different segmentation algorithms”, In *Pattern Recognition Letters*, volume 18, pages 963–974, New York, USA, 1997, Elsevier Science Inc.
- [9] T. H. Minh and H. Bunke, Analysis and understanding of giro check forms, Relatório técnico, University of Berne, Switzerland, 1992.
- [10] T. M. Mitchell, *Machine learning*, McGraw-Hill, Boston, USA, 1997.
- [11] P. A. Morettin and C. M. Toloí, *Análise de séries temporais*, Edgard Blücher, São Paulo, Brasil, 2 edition, 2006.
- [12] Ulysses Doria, *Introdução à Bioestatística: para simples mortais*, Elsevier, São Paulo, Brasil, 1999.
- [13] S. Belongie, J. Malik, and J. Puzicha, “Matching shapes”, In *Eighth IEEE International Conference on Computer Vision*, pages 454–461, Vancouver, Canada, 2001.
- [14] L. H. Rodrigues, *Building imaging applications with java technology*, Addison Wesley, Boston, USA, 2001.
- [15] F. Cordeiro, J. S. M. Filho, and J. C. Prolla, *Endoscopia Digestiva*, MEDSI, Rio de Janeiro, Brasil, 1994.
- [16] R. S. Contran, V. Kumar, and T. Collins, *Robbins Patologia Estrutural e Funcional*, Guanabara Koogan, 6 edition, 2000.