

# UM SISTEMA PARA AUXILIAR NO PROCESSO DE MAPEAMENTO DE FORMULÁRIOS MÉDICOS PARA BASES DE DADOS ESTRUTURADAS

Willian Zalewski<sup>1</sup>, Hwei Diana Lee<sup>2</sup>, André Gustavo Maletzke<sup>3</sup>, Cláudio Saddy Rodrigues Coy<sup>4</sup>, João José Fagundes<sup>5</sup>, Juvenal Ricardo Navarro Góes<sup>6</sup>, Feng Chung Wu<sup>7</sup>

**Abstract** — Focused on helping researches related to Crohn disease, the Bioinformatics Laboratory/UNIOESTE/PTI in partnership with the Coloproctology Service/FCM-UNICAMP developed a protocol of information, considered relevant by specialists, regarding to this disease. Because of the great quantity of attributes defined in this protocol, a methodology and a computational system were developed to automatically map data contained in medical forms manually filled. These forms were structured in a question and answers layout, in which the form's multiple choice and numeric fields are filled. An experimental evaluation of this methodology using fifty forms filled by ten collaborators was held in this work. The mapping performed by the system presented a 99.60% precision in multiple choice fields and 96.23% in numeric fields. The use of this methodology supported by this computational system provided time and subjectivity elimination or minimization during the mapping process.

**Index Terms** — Data Mining, Artificial Intelligence, Bioinformatics, Crohn Disease.

## INTRODUÇÃO

Na área de medicina, pesquisas relacionadas a determinadas enfermidades como a doença de Crohn [1] necessitam da coleta de uma grande quantidade de dados. No entanto, freqüentemente, os dados estão dispostos em laudos e formulários impressos. Nesse contexto, processos computacionais poderiam ser úteis como ferramentas de apoio para pesquisas relacionadas a essas doenças. Desse modo, neste trabalho é apresentado um método eficiente e menos dispendioso para realizar a coleta desses dados.

## MATERIAIS E MÉTODO

A partir da metodologia desenvolvida pelo Laboratório de Bioinformática/UNIOESTE/PTI em conjunto com o Serviço de Coloproctologia/FCM/UNICAMP e o Laboratório de Inteligência Computacional/USP [2] foi construído um sistema computacional para realizar o

mapeamento automático de formulários médicos impressos preenchidos manualmente. Por meio da utilização desse sistema foi gerado um formulário modelo, a partir do qual foram realizadas 50 cópias. Esses formulários foram preenchidos por dez colaboradores, de modo que cada um preencheu cinco formulários. Posteriormente, os 50 formulários foram digitalizados e os dados preenchidos foram mapeados, com o auxílio de uma rede neuronal, para uma base de dados com a utilização do sistema.

## RESULTADOS E DISCUSSÃO

Considerando todos os formulários, o mapeamento 250 dos campos de múltipla escolha obteve 99,60% de precisão. Em relação ao reconhecimento dos 504 dígitos preenchidos nos campos numéricos foi constatada uma precisão de 96,23%. Com base nesses resultados, foi possível observar que o custo de tempo e a subjetividade foram eliminados ou minimizados durante o mapeamento.

## CONCLUSÃO E TRABALHOS FUTUROS

Os resultados obtidos neste trabalho foram considerados satisfatórios pelos especialistas do domínio. Trabalhos futuros incluem a aplicação de métodos computacionais para a extração de padrões que possam estar contidos na base de dados construída por meio do mapeamento dos formulários.

## AGRADECIMENTOS

Ao Programa de Desenvolvimento Tecnológico Avançado – PDTA/FPTI – pelo auxílio por meio da linha de financiamento de bolsas.

## REFERÊNCIAS

- [1] SANTOS, J. C. M., "Doença de Crohn: Aspectos Clínicos e Diagnósticos", *Revista Brasileira de Coloproctologia*, Vol. 19., 1999, 276-285.
- [2] MALETZKE, A. G.; LEE, H. D.; ZALEWSKI, W.; MATSUBARA, E. T.; VOLTOLINI, R. F.; et al., "Mapeamento de Informações Médicas descritas em Formulários para Bases de Dados Estruturadas", *VII Workshop de Informática Médica*, 2007.

<sup>1</sup> Willian Zalewski, Laboratório de Bioinformática, Universidade Estadual do Oeste do Paraná, Av. Presidente Tancredo Neves, 6731, 85866-900, Foz do Iguaçu, PR, Brasil, willzal@gmail.com

<sup>2</sup> Hwei Diana Lee, Laboratório de Bioinformática, Universidade Estadual do Oeste do Paraná, Foz do Iguaçu, PR, Brasil, hueidianalee@gmail.com

<sup>3</sup> André Gustavo Maletzke, Laboratório de Inteligência Computacional, Universidade de São Paulo, São Carlos, SP, Brasil, andregustavom@gmail.com

<sup>4</sup> Cláudio Saddy Rodrigues Coy, Serviço de Coloproctologia da Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, SP, Brasil

<sup>5</sup> João José Fagundes, Serviço de Coloproctologia da Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, SP, Brasil,

<sup>6</sup> Juvenal Ricardo Navarro Góes, Serviço de Coloproctologia da Faculdade de Ciências Médicas, Campinas, SP, Brasil

<sup>7</sup> Feng Chung Wu, Laboratório de Bioinformática, Universidade Estadual do Oeste do Paraná, Fundação Parque Tecnológico de Itaipu, Foz do Iguaçu, PR, Brasil, Serviço de Coloproctologia da Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, SP, Brasil, wufc@pti.org.br