

# MAPEAMENTO AUTOMÁTICO DE LAUDOS MÉDICOS PARA EXTRAÇÃO DE CONHECIMENTO

Everton Alvares Cherman<sup>1</sup>, Newton Spolaôr<sup>2</sup>, Huei Diana Lee<sup>3</sup>, Daniel de Faveri Honorato<sup>4</sup>, Cláudio Saddy Rodrigues Coy<sup>5</sup>, João José Fagundes<sup>6</sup>, Juvenal Ricardo Navarro Góes<sup>7</sup>, Feng Chung Wu<sup>8</sup>

**Abstract** — *The Knowledge Discovery in Databases – KDD – process may help specialists on decision making. Information on the medical field is usually found on Medical Reports – MR – described on natural language. In order to apply KDD it is necessary to map this information on an attribute-value format. The Bioinformatics Laboratory/Unioeste, in partnership with the Computational Intelligence Laboratory/USP and the Service of Coloproctology/UNICAMP, is developing a MR automatic mapping method intending to reduce the necessary time of the process and to avoid possible subjectivity on the manual information mapping. In this method, a domain dictionary is build, supported by specialists, which helps the automatic MR mapping algorithm. This work presents two case studies about this method using information from 100 colonoscopy MR and on 609 upper digestive endoscopy MR, from which 82% and 100% were, respectively, automatically mapped, eliminating the subjectivity and reducing the necessary process time.*

**Index Terms** — *Automatic mapping, Knowledge extraction, Upper digestive endoscopy, Colonoscopy, Bioinformatics.*

## INTRODUÇÃO

A Descoberta de Conhecimento em Bases de Dados é, em geral, aplicada sobre dados no formato atributo-valor. Neste trabalho são apresentados dois estudos de caso para os quais foi realizado o mapeamento de Laudos Médicos – LM – de Endoscopia Digestiva Alta – EDA – e de colonoscopia para tabelas atributo-valor – TAV.

## MATERIAIS E MÉTODOS

Os exames de EDA e de colonoscopia contribuem no diagnóstico de enfermidades esofagogastroduodenais e colorretais, respectivamente [1]-[2]. Os LM descrevem as informações resultantes desses exames, as quais são mapeadas a partir de uma metodologia composta por duas

fases [3]. Na primeira fase são definidos, a partir de padrões textuais dos LM, os atributos da TAV e um dicionário do domínio. A segunda fase envolve a aplicação de um Algoritmo de Busca e Preenchimento – ABP –, o qual realiza o mapeamento dos LM para a TAV a partir do dicionário construído na primeira fase.

## RESULTADOS E DISCUSSÃO

A metodologia foi aplicada a 609 LM de EDA e a 100 LM de colonoscopia, dos quais foram gerados os dicionários. Após a aplicação do ABP foi realizada uma apuração dos valores não preenchidos na TAV por meio de um conjunto de termos não processados. Constatou-se que 82% e 100% dos valores foram preenchidos automaticamente para a TAV para os domínios da colonoscopia e EDA, respectivamente.

## CONCLUSÕES

A metodologia possibilitou o mapeamento automático dos LM, reduzindo o tempo necessário para o processo, e eliminou a possível subjetividade do mapeamento manual. A TAV e os dicionários construídos poderão ser utilizados para o mapeamento de outros conjuntos de LM.

## AGRADECIMENTOS

Ao Programa de Desenvolvimento Tecnológico Avançado – PDTA/FPTI – e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq – pelo auxílio por meio da linha de financiamento de bolsas.

## REFERÊNCIAS

- [1] CORDEIRO, F., “Endoscopia Digestiva”, 1994.
- [2] QUILICI, F. A., “Colonoscopia”, 2000.
- [3] HONORATO, D. D. F.; LEE, H. D.; MONARD, M. C.; WU, F. C.; MACHADO, R. B.; et al., “Uma metodologia para auxiliar no processo de construção de bases de dados estruturadas a partir de laudos médicos”, *Encontro Nacional de Inteligência Artificial*, 2005, 593-601.

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

<sup>2</sup> Newton Spolaôr, Laboratório de Bioinformática, Universidade Estadual do Oeste do Paraná, Foz do Iguaçu, PR, Brasil, newtonspolaor@hotmail.com

<sup>3</sup> Huei Diana Lee, Laboratório de Bioinformática, Universidade Estadual do Oeste do Paraná, Foz do Iguaçu, PR, Brasil, hueidianalee@gmail.com

<sup>4</sup> Daniel de Faveri Honorato, Laboratório de Inteligência Computacional, Universidade de São Paulo, São Carlos, SP, Brasil, dfaverih@gmail.com

<sup>5</sup> Cláudio Saddy Rodrigues Coy, Serviço de Coloproctologia da Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, SP, Brasil

<sup>6</sup> João José Fagundes, Serviço de Coloproctologia da Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, SP, Brasil

<sup>7</sup> Juvenal Ricardo Navarro Góes, Serviço de Coloproctologia da Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, SP, Brasil

<sup>8</sup> Feng Chung Wu, Laboratório de Bioinformática, Universidade Estadual do Oeste do Paraná, Foz do Iguaçu, PR, Brasil, Fundação Parque Tecnológico

Itaipu, Foz do Iguaçu, PR, Brasil, Serviço de Coloproctologia da Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, SP, Brasil,