Integration of systems for automatic mapping of medical reports into structured data

Leandro A. Ensina[1], Weber S. R. Takaki[12], Huei D. Lee[1], Newton Spolaôr[1], Narco A. R. Maciejewski[1], Claudio S. R. Coy[2], Feng C. Wu[12]

[1] Western Paraná State University
[2] University of Campinas

New technologies have supported diverse knowledge areas, including health sciences. Among these technologies, Internet and Artificial Intelligence stand out, for they facilitate applications such as telemedicine and the discovery of useful knowledge to help specialists from the domain area.

In this context, the use of ontologies enables the representation of knowledge and interoperability between systems [1]. In addition, the use of ontologies have also been proposed to convert natural language into structured data [2,4].

Therefore, data contained in medical reports such as descriptions of signs and symptoms, test results, treatment adopted, and the evolution of the patient can be extracted automatically utilizing computational techniques and ontologies, and later be submitted to pattern recognition and data mining methods [2].

The integration of two relevant systems [5]: the Telemedicine and Data Management Integrated System (SITGD) [3], and the Mapper Reports Ontologies (MRO) [4], both developed by the Bioinformatics Laboratory of the Western Paraná State University - LABI / Unioeste, aimed to expand the reach and the benefits of each system. Thus images and sequences of audio and video could be linked to structured data extracted from textual reports, and also professionals located in different sites could collaborate in real time.

SITGD is responsible for the real-time transmission of medical procedures and the management of patient data and exams, as well as for allowing doctors to take snapshot images, and the creation of textual reports. On the other hand, MRO has the purpose of mapping unstructured medical reports into a structured database with the support of an ontology, which contains mapping rules representing possible combinations of terms for the sentences of each report.

The mapping process is performed in two phases. In the first phase, the set of medical reports is processed to reduce the complexity of phrases and to identify characteristics that represent the knowledge contained in the reports. Intermediate files are then generated, such as the ontology and the standardization file, which are used for the mapping of new medical assessments. Therefore, this phase acts as a training step for the mapping process. In the second phase, the reports to be mapped are standardized using the intermediate files generated in the first phase. Afterward, the content of these reports is organized into a structured representation with table format.

The results obtained showed that the integration between the SITGD and the MRO systems is functional, allowing not only a single user authentication step, but also seamless navigation between both systems, as well as the automatic recognition, performed by the MRO, of the reports created within SITGD [5].

Thus, soon after the creation of a report, it can already be viewed in a structured representation, allowing health experts to identify relevant characteristics.

It is worth noting that the importance of structuring these reports lies in the fact that it allows for statistical analysis and pattern recognition, ultimately supporting medical research and the study of health conditions.
Future work includes the application of Knowledge Discovery in Databases (KDD) techniques in order to find relevant patterns in the structured data after the treatment performed by MRO, as well as to include features extracted from images captured during the exams. New information and knowledge can then be obtained, contributing with health experts in the understanding and treatment of diseases.

[1] O. Bodenreider and A. Burgun, Biomedical Ontologies*, Medical Informatics: knowledge management and Data Mining in biomedicine,* Springer, New York, pp. 211-236, 2005.

[2] H.D. Lee, M.C. Monard, D.F. Honorato, A.C. Lorena, C.A. Ferrero, A.G. Maletzke, W. Zalewski, C.S.R. Coy, J.J. Fagundes and F.C. Wu, Mapping unstructured data in digital and printed documents into attribute-value tables. In R.A.E. Andrade, J.M. Gomez and A.R. Valdes, editors, proceedings of the *Towards a Trans-disciplinary Technology for Business Intelligence, Gathering Knowledge Discovery, Knowledge Management and Decision Making,* Shaker Verlag, pp. 198-209, 2011.

[3] C.S.R. Coy, P.C. Marques Filho, M.R. Litchtnow, M. Fontana, L.A. Ensina, N. Spolaôr, W.S.R. Takaki, F.C. Wu, H.D. Lee, M. Fonteque Junior, J.J. Fagundes, R.F. Leal and M.L.S. Ayrizono, BR Patent INPI 512015001532-0, *SITGD*, 2015

[4] C.S.R. Coy, J.T. Oliva, H.D. Lee, F.C. Wu, J.J. Fagundes, R.B. Machado, N. Spolaôr, M. Fonteque Junior, R.F. Leal and M.L.S. Ayrizono, BR Patent INPI 512015000342-9, *MRO – Mapper Reports Ontologies*, 2015.

[5] L.A. Ensina, F.C. Wu, H.D. Lee, W.S.R. Takaki and J.T. Oliva, Construção de um módulo para mapeamento de laudos médicos para integração em um sistema computacional em telemedicina, *proceedings of 3º Encontro Anual de Iniciação Científica, Tecnológica e Inovação*, Cascavel, Brazil, 2017.