

# Medication/Adverse Drug Events Information Extraction in Discharge Summaries to Enrich Discharge Medication Instructions

Christine Kakalou, Pantelis Natsiavas, Theodore Dalamagas and Manolis Koubarakis

**Abstract**— Improved readability of Electronic Health Records (EHR) free-text could support better patient information provision and prevent confusion when EHR information are openly available. An informed domain-specific Information Extraction approach is introduced ultimately aiming to enrich discharge summaries with patient-friendly language. An evaluation of the initial step to extract medications and their respective Adverse Drug Events are presented.

## I. INTRODUCTION

Patients' access to their own Electronic Health Record (EHR) data has been recently growing. However, the complexity and highly domain-specific language of EHR narratives introduces challenges regarding patient engagement, e.g., stress and confusion [1]. Several works involving ML and NLP techniques have been proposed to assess and improve the readability of medical texts (e.g. [2]), tackling the low health literacy problem. In this work, we present the concept of a system that simplifies discharge medication instructions and enhances them with patient-friendly terms. We also provide preliminary results in terms of identifying potential mentions of Adverse Drug Events (ADEs) in the relevant discharge summaries.

## II. CONCEPT AND METHODOLOGY

We use free-text hospital discharge summaries from MIMIC-III [3]. Our approach proceeds as follows. First, we extract the medications prescribed during discharge and any ADEs mentioned, while we also retrieve the most prevalent ADEs from external data sources (e.g. DrugBank). Thus, (Medication, {ADE list}) pairs are created e.g. (*Lipitor*, {*tinnitus*, *sacroiliac joint stiffness*}). Then, using patient-friendly terminology resources (e.g., *ringing in ears* instead of *tinnitus*), we will assess the quality of the individual term suggestions. Then, a Knowledge Graph (KG), modelling the relevant information extracted from each discharge summary along with the enriched candidate friendly terms,

\* This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH –CREATE – INNOVATE (project code: T2EDK-00640).

C. K. is with the Institute of Applied Biosciences, Centre for Research & Technology Hellas, Thessaloniki, Greece, and the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens (phone: +306975852653; e-mail: ckakalou@certh.gr).

P. N. is with the Institute of Applied Biosciences, Centre for Research & Technology Hellas, Thessaloniki, Greece (e-mail: pnatsiavas@certh.gr).

T. D. is with the ATHENA Research Center, Information Management Systems Institute, Marousi, Greece (e-mail: dalamag@athenarc.gr).

M.K. is with the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens (e-mail: koubarak@di.uoa.gr)

will be created. Finally, we plan to develop a method that, through the KG traversal, facilitates the automatic generation of customized, patient-friendly text snippets to warn of probable ADEs, such as: "*Lipitor can give you ringing in your ears and stiffness in your lower back.*". The quality of the generated text snippets (in terms of their usefulness in dealing with low health literacy) will be evaluated both by domain experts and systematically using generic and domain-specific readability measures. We are currently completing the first step where we utilized the MedSpaCy clinical NLP library with the *Med7* NER model and the Spark NLP library with the pretrained *ner\_ade\_clinical* and *re\_ade\_clinical* models in a customized pretrained pipeline, and we report on preliminary results.

## III. RESULTS

To validate the effectiveness of the tools in our work's context, we used the gold annotated corpus from Track 2 of the 2018 National NLP Clinical Challenges (n2c2). The evaluation results are summarized in Table 1 below.

TABLE I. EVALUATION MEASURES FOR INFORMATION EXTRACTION

Extraction Task	Average Evaluation Measures		
	Precision	Recall	F1
Strict Medication NER	0.82	0.80	0.81
Lenient Medication NER	0.96	0.95	0.95
ADE-medication Relation Extraction	0.52	0.43	0.45

## IV. CONCLUSIONS

The initial evaluation of the proposed pipeline for the Information Extraction tasks required for the creation of medication-Adverse Drug Reaction pairs demonstrate that our approach is comparable to the State-of-the-Art [4][5]. Further work involves finetuning parameters and additional customization of the information extraction pipeline.

## REFERENCES

- [1] T. Davis Giardina and H. Singh, "Should Patients Get Direct Access to Their Laboratory Test Results?", *JAMA*, vol. 306, no. 22, 2011.
- [2] J. Zheng and H. Yu, "Assessing the Readability of Medical Documents: A Ranking Approach", *JMIR Medical Informatics*, vol. 6, no. 1, 2018.
- [3] A. Johnson et al., "MIMIC-III, a freely accessible critical care database", *Scientific Data*, vol. 3, no. 1, 2016.
- [4] A. Kormilitzin, N. Vaci, Q. Liu and A. Nevado-Holgado, "Med7: A transferable clinical natural language processing model for electronic health records", *Artificial Intelligence in Medicine*, vol. 118, 2021.
- [5] S. Henry, K. Buchan, M. Filannino, A. Stubbs and O. Uzuner, "2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records", *Journal of the American Medical Informatics Association*, vol. 27, no. 1, 2019.