

Creating Medical Datasets using Robot Process Automation

Iulian ROMAN^{a,*}, Călin MUNTEAN^{b,c}, and Sebastian-Aurelian ȘTEFĂNIȚĂ^{a,c}

^a Department of Computer Science, Faculty of Mathematics and Computer Science, West University of Timișoara, Vasile Pârvan Bldv., no. 4, 300223 Timișoara, Romania.

^b “Victor Babeș” University of Medicine and Pharmacy, Eftimie Murgu Sq., no. 2, 300041 Timișoara, Romania.

^c Romanian Society of Medical Informatics, 300222 Timișoara, Romania.

E-mails: iulian.roman01@e-uvt.ro; cmuntean@umft.ro; sebastian.stefaniga@e-uvt.ro;

* Author to whom correspondence should be addressed

Abstract

A dataset is a collection of information, usually structured in a format that can be used in research or training various applications such as machine learning. Datasets are fundamental to many analytical processes because they provide raw data that can be used in different applications. While some datasets are readily available, others must be specially curated. Creating a dataset often involves automated data collection methods, with web scraping being a popular technique. Web scraping employs scripts to extract data from online sources, demanding robust programming skills and considerable maintenance effort. It is important to note that the method can break site policies because it may violate terms of service agreements. A proposed solution for gathering the required information is to create a Robot Process Automation (RPA) flow, which collects information from various sources as a normal user would do manually. A use case of RPA flow involves using it to compile a medical dataset that contains information about clinical trials for various drugs. The dataset will be utilized in researching the effects of placebo drugs on people. The information is available on the clinicaltrials.gov site and the results of different trials can be downloaded manually as files. Automating this process can be done by creating a flow in UiPath by linking predefined activity blocks, which simplifies the creation and maintenance of the workflow. Running the flow will follow the steps that a normal user would do for downloading the results of different trials from the site. This approach eases data collection, allowing medical staff to generate datasets, without the need for deep technical skills, that can be used in different investigations. In conclusion, an RPA flow is a straightforward way to gather information for a dataset, circumventing the complexities of traditional programming-based approaches like web scraping.

Keywords: Datasets; Robot Process Automation; UiPath; Data Collection; Data Extraction

