Conference Paper

Leveraging Web Scraping and API Integration for Efficient Medical Device Data Management

Agapi Konstantina Liontou[1,*], Spilios Zisimopoulos[2] and Aris Dermitzakis[1]

[1] Biomedical Technology Unit, Department of Medical Physics, School of Medicine, University of Patras, Patras, Greece.
[2] Institute of Biomedical Technology, Patras, Greece.

[*] Corresponding Author Email: liontou.kon@gmail.com

ABSTRACT

Accurate identification and management of medical devices is of particular importance to ensure patient safety and regulatory compliance within healthcare systems. This paper presents a comprehensive exploration of medical device data retrieval, focusing on the integration of web scraping and Application Programming Interface (API) technologies. The utilization of Unique Device Identifiers (UDIs) and the Global Medical Device Nomenclature (GMDN) system is emphasized to enhance device authentication, attribute verification, and accurate categorization.

This paper introduces a state-of-the-art code implementation that combines web scraping techniques and API integration to address the challenges of retrieving and verifying device information. The code facilitates both access to data and healthcare professionals and stakeholders to make informed decisions based on reliable and up-to-date information. This is a significant and defining advance in the field, offering a powerful solution that is innovative as well as vital.

The paper concludes by discussing the potential impact of these developments on patient safety, regulatory compliance, and the overall advancement of healthcare technology. In addition, the importance of accurate device identification, the role of UDIs and GMDN, and the significance of the provided cutting-edge code are highlighted, providing valuable insights into the field of medical device data retrieval.

Keywords—*Web scraping, Medical device data management, API integration.*

INTRODUCTION

In today's data-driven landscape, the demand for effective data extraction techniques is critical. Patient safety, regulatory compliance, and informed decision-making hinge on the acquisition of accurate and up-to-date information about medical devices. However, obtaining this information can be a daunting task, given the disparate origins and formats of available data.

This paper examines the challenges associated with medical device data management and introduces a comprehensive methodology, proposing a combination of web scraping methods with the integration of Application Programming Interfaces (APIs), with a particular focus on two key elements: Unique Device Identifiers (UDIs) and Global Medical Device Nomenclature (GMDN) terms, pivotal for the identification and categorization of medical devices. Our objective is to present an integrated approach for gathering, validating, and employing medical device data from the AccessGUDID (Global Unique Device Identification Database), which acts as the authoritative source of device information, laying the foundation for our methodology. In addition, our methodology relies on the web-Praxis Medical Equipment Management Software (MEMS), developed by the Institute of Biomedical Technology (INBIT)[1], whose data can be accessed by appropriate healthcare unit users. A dataset of medical device information was provided for the purpose of this study. AccessGUDID database on the other hand is freely accessible.

The significance of swift medical equipment (ME) identification in today's fast-paced healthcare environment cannot be overstated. Particularly in the context of recalls and field safety notices (FSNs) issued by manufacturers, which contain vital information about affected device types, rapid identification plays a crucial role. Clinical engineers are entrusted with the immediate and appropriate response to these notices, ensuring the correct course of corrective actions or equipment withdrawals to prevent adverse events.

To address this pressing need, our developed software serves a dual purpose. It aids in the validation of ME within a hospital's inventory, ensuring precise matching with the corresponding UDI codes. UDI serves as a standardized system for identifying and tracing medical devices Additionally, it facilitates the accurate classification of ME into GMDN groups, a critical facet of the UDI system. GMDN serves as a standardized system for classifying medical devices worldwide. It is noteworthy that recalls and FSNs frequently include UDIs for affected devices, emphasizing the growing importance of UDI-based traceability. Moreover, according to MDR 2017/745[2], the traceability of devices by means of a UDI system should significantly enhance the effectiveness of the post-market safety-related activities for devices.

To address the complexities of gathering and verifying device-related data, this paper discloses an innovative code implementation. We combine web scraping techniques with AccessGUDID (Global Unique Device Identification Database) API integration, presenting a powerful solution. Our code adeptly extracts imperative device details from the AccessGUDID website and validates this information with data provided by web-Praxis. Additionally, our code integrates seamlessly with the AccessGUDID API, simplifying the retrieval of device-specific information dependent on UDIs, ensuring compliance with data usage regulations, and increasing the reliability and accuracy of the retrieved data.

In summary, this paper delves into the intricate realm of medical device data management, leveraging advanced techniques for data extraction and integration. Our integrated approach not only promises to enhance patient safety and regulatory compliance but also equips healthcare professionals with the tools needed for informed decision-making in an increasingly dynamic healthcare landscape.

BACKGROUND

In the modern era, accurate and timely management of medical device data is essential for various stakeholders in the healthcare industry. In this section, we talk about the background of the key elements of this paper's methodology: Unique Device Identifiers (UDIs) and Global Medical Device Nomenclature (GMDN), in addition to an overview of web scraping and Application Programming Interface (API) integration.

Unique Device Identifiers (UDIs)

Unique Device Identifiers or UDIs, are alphanumeric codes for medical devices, offering a standardized global means of identification. They enable precise tracking from manufacturing to post-market surveillance. UDIs encode key device information like manufacturer, model, and production date. Their implementation has significantly advanced healthcare, enhancing patient safety, regulatory compliance, and supply chain management. UDIs empower healthcare professionals to quickly access device data, identify recalls, and respond to adverse events effectively. [3]

Global Medical Device Nomenclature (GMDN)

The Global Medical Device Nomenclature (GMDN) is an internationally recognized system for naming and categorizing medical devices. It offers a structured classification system that facilitates global communication in healthcare. GMDN codes categorize devices based on their purpose, structure, and operation, enabling precise comparisons. GMDN adoption has streamlined regulations, research, and product development in the medical device industry. It promotes consistency in terminology and categorization, as well as seamless compatibility and cooperation among healthcare stakeholders for sharing accurate device information.[4]

Web Scraping and API Integration

In the contemporary era, extracting data from various web sources has become essential for various domains, including healthcare. In this section, a foundational understanding of the main data extraction methods: web scraping and Application Programming Interfaces (APIs), is provided. Web scraping, also known as web extraction or harvesting, involves the automated extraction of data from websites and their subsequent storage for analysis or retrieval. This method, widely recognized for its efficiency and accuracy, has evolved significantly over the years. Modern web scraping tools have become versatile, capable of parsing markup languages, integrating with computer vision and natural language processing techniques, and simulating human browsing behavior.[5,6]

APIs, on the other hand, serve as intermediaries between software applications, enabling seamless communication and data exchange.

Web Scraping vs. API

Web scraping, with its capability to access data from multiple web pages and repositories, excels in collecting large volumes of heterogeneous data efficiently. It offers flexibility in data collection and analysis, making it a valuable tool in domains such as computer vision and natural language processing.

APIs, on the other hand, provide a structured and controlled means of accessing specific data from applications or software. While they offer advantages such as standardized interfaces and faster data extraction, they also come with limitations in terms of functionality and access to a single website or predefined functions.[5, 7, 8]

METHODS AND MATERIALS

Python for Web Scraping

Python is a highly favored choice for web scraping due to several key reasons. Firstly, Python is a very popular programming language because of its simplicity and learnability, facts that make it accessible even to those with little programming experience. Its extensive community support ensures readily available assistance for overcoming coding challenges.

Furthermore, Python excels in web scraping because of the readability of the code. Python code is designed for easy understanding, promoting clear and concise programming practices. This readability not only enhances productivity during development but also facilitates code understanding, even when revisiting it after some time has passed. This attribute promotes more efficient code maintenance and facilitates code reuse.[8] For these reasons, Python is used in all the software we developed.

Our Code

We present an integrated code solution that combines web scraping via Beautiful Soup (Version 4, Python Library for HTML and XML Parsing, Crummy, Cambridge, MA, USA) with AccessGUDID API integration. Tailored

for healthcare professionals, this tool simplifies access, validation, and analysis of device data based on UDIs and GMDN names while ensuring data compliance.

A key aspect of our methodology involves cross-referencing data derived from web scraping with information from Praxis, our primary data source. This information is structured in an Excel file and has been collected by humans, checked, and entered into the Praxis database. This quality control step identifies disparities in device attributes, improving data accuracy. For example, discrepancies in GMDN names trigger further investigation to address updates or errors.

Our comprehensive validation process involves verifying GMDN names, selecting the most prevalent name when multiple names exist, and aligning device attributes across sources, ultimately enhancing the reliability of medical device data.

Data Import

To begin data retrieval, the code begins by importing data from an Excel file provided by Web Praxis. This file includes critical information about medical devices and the specific columns of interest include General Group, Specific Group, Manufacturer, Model, and Comments, where Unique Device Identifiers (UDIs) are often found. The initial Excel file includes 279 records, representing distinct medical devices.

UDI Extraction

Next, the code extracts the 14-digit UDIs from the appropriate field (if any because not all records have a registered UDI). This step is crucial for subsequent operations as UDIs serve as the primary key for accessing device information.

Web Scraping and API Integration

The heart of the methodology lies in its ability to combine web scraping and API integration for comprehensive data collection. Web scraping extracts data based on UDIs, while API calls are made to the AccessGUDID database to retrieve detailed device information. This hybrid approach ensures that even devices without readily available UDIs can be identified and analyzed.

Data Validation and Presentation

As data is retrieved, the code simultaneously validates and cross-references it to ensure accuracy and reliability. The data are then organized and presented in a structured format for further analysis and reporting.

Data Cross-Referencing

For enhancing data reliability, the code employs cross-referencing. It verifies device attributes like GMDN terms by comparing data from Praxis with data obtained through web scraping or API integration. This process ensures consistency across multiple sources and is a vital quality control step.

The aforementioned accomplishments were realized through the utilization of the following libraries: *requests* for making HTTP requests to interact with web services and retrieve data from web servers, *pandas* for data manipulation, *tkinter* for creating a GUI, *re* for regular expressions, *time* for time-related operations, and Beautiful Soup for web scraping purposes to pull the data out of HTML and XML files, as it helps in navigating, searching, and modifying the parse tree. Beautiful Soup creates a parse tree from page source code that can be used to extract data easily.

RESULTS

The results were quite satisfactory. Out of the 279 records initially provided, we found UDI matches for 193 records (69% success rate). By implementing a complex search based on the company name and brand name (this was done because there were cases that we did not have a UDI—device type match a priori), we were able to identify 14 more records which means 207 records in total (74% success rate).

The output generated by the code is presented in a structured Excel format (Figure 1), providing a comprehensive overview of matched UDIs, associated company names, brand names, GMDN names, cross references, and definitions.

Where:

1st column: UDI, UDIs are listed. It includes both the UDIs that were initially matched and the "N/A" symbol for equipment that had no UDI in the initial data and required an advanced search.

2nd column: Company Name as registered on the website.

3rd column: Brand Name as registered on the website.

4th column: GMDN Name as registered on the website.

5th column: GMDN Cross Reference. The term "same" is displayed for records that have the same GMDN Name on both the website and the initial Excel data. For records that are not the same, a different description is provided. This description was obtained from the initial data.

6th column: GMDN Definition as registered on the website for the devices that were "identified" by the UDI. For the devices that we did not know the UDI beforehand, and we made an advanced search, we had many results. So, the GMDN definition was obtained by examining the first page of search results on the website generated by the advanced search. We identified all the GMDN names and applied a sorting algorithm to select the names that appeared most frequently.

UDI	Company Name	Brand Name	GMDN Name	GMDN Cross Reference	GMDN Definition
5037048915	ADVANCED STERILIZATION PRODUCTS SERVICES INC.	STERRAD	Biological sterilization indicator reader	Same	A mains electricity (AC-p
0554103765	Medistim ASA	Medistim MiraQ Ultimate System	Cardiovascular ultrasound imaging system	Same	An assembly of mains el
0018405802	DRTECH Corporation	Exprimer	Indirect flat panel x-ray detector	Same	An electrically-powered
0018405802	DRTECH Corporation	Exprimer	Indirect flat panel x-ray detector	Same	An electrically-powered
8258308578	BAYER MEDICAL CARE INC.	MEDRAD® Avanta Fluid Management	Angiography contrast medium injection syste	Same	An assembly of devices
0590563547	Beckman Coulter, Inc.	TQ-Prep Workstation	Specimen processing instrument IVD	Same	An electrically-powered
2903389629	BECTON, DICKINSON AND COMPANY	FACSCanto	Flow cytometry analyser IVD	Same	An electrically-powered
4695022588	ONE LAMBDA, INC.	LABScan3D™	HLA class I & II antigen tissue typing IVD, kit,	Same	A collection of reagents
0414007903	Siemens Healthcare Diagnostics Products GmbH	RN II System	Nephelometry immunoassay analyser IVD	Same	An electrically-powered
4470804122	Fujirebio Europe	TENDIGO	Blot processor IVD	Same	A mains electricity (AC-p
5630929764	Roche Diagnostics GmbH	cobas e 602 Immunoassay Analyzer	Chemiluminescent immunoassay analyser IVI	Same	An electrically-powered
5630946341	Roche Diagnostics GmbH	cobas 8000 cobas ISE module	Multiple clinical chemistry analyser IVD, labo	Same	An electrically-powered
6950784074	INSTRUMENTATION LABORATORY COMPANY	ACL TOP 750	Coagulation analyser IVD, laboratory	Same	An electrically-powered
5630036950	Roche Diagnostics GmbH	cobas b 123 <i> POC system	Metabolic profile clinical chemistry analyser	Same	A stationary mains elect
5219000126	Biocartis NV	Idylla™ System	Thermal cycler nucleic acid amplification ana	Same	An electrically-powered
5710001318	Maquet Critical Care AB	FLOW-i C40	Anaesthesia workstation, general-purpose	47770 Μονάδα αναισθησί	An assembly of electron
3739000678	LIGHTLAB IMAGING, INC.	ILUMIEN™	Coronary optical coherence tomography syst	Same	An assembly of optical a
0103197006	Edwards Lifesciences LLC	HemoSphere	General purpose multi parameter bedside m	47485 Ενδαγγειακό σύστη	An electrically powered
4729890010	BOSTON SCIENTIFIC CORPORATION	FFR Link	Cardiovascular ultrasound imaging system	47485 Ενδαγγειακό σύστη	An assembly of mains el
5479115142	BIOTRONIK SE & Co. KG	Renamic	Cardiac pulse generator programmer	Same	An external device inten
4734210713	ST. JUDE MEDICAL, INC.	EnSite™	Cardiac mapping system	Same	An assembly of devices
4729948322	BOSTON SCIENTIFIC CORPORATION	ΠΥΤΗΙΑ HDx™	Cardiac mapping system	Same	An assembly of devices
4729938712	BOSTON SCIENTIFIC CORPORATION	RHYTHMIA HDx™	Transducer signal amplifier	Same	An electronic device tha
5036013952	ETHICON ENDO-SURGERY, LLC	N/A	Electrosurgical/ultrasonic surgical system ge	Same	A mains electricity-powe
4261076838	OLYMPUS Winter & Ibe GmbH	Olympus	Electrosurgical system generator	Same	An electrically-powered

FIGURE 1. Format of the output Excel File.

DISCUSSION

The high success rates in both types of searches indicated that the data provided by web-Praxis, although challenging to manage due to its unstructured format, demonstrate a fairly accurate recording of medical equipment, with correct UDIs, GMDN names, and definitions.

Matching all 279 medical records cannot be possible due to inaccuracies inherent in the databases. This creates issues for our model, often leading to many closely related outcomes. Even for a human observer, distinguishing between these outcomes is extremely challenging. So, reaching a perfect 100% match rate is impossible and would demand a lot of human effort from an experienced eye, potentially even reviewing unrelated results. Hence, a 74% match rate is considered satisfactory under these circumstances.

Overall, the output provides a comprehensive overview of the matched UDIs, associated company names, brand names, GMDN names, cross-references, and definitions, allowing for further analysis and verification of the recorded medical equipment data.

In general, the implemented code combining web scraping techniques and integration with the AccessGUDID API has proven to be an essential tool for retrieving, validating, and analyzing medical device data based on UDIs and Global Medical Device Nomenclature (GMDN) names.

The need for web scraping arises from the vast amount of device information available on websites like AccessGUDID. Web scraping enables the efficient extraction of specific data elements, such as company names, brand names, GMDN names, cross-references, and definitions, from complex HTML structures. By automating the data retrieval process, this method provides instant access to the most up-to-date device information, ensuring accuracy and timeliness.

Integrating an API (in this case, the AccessGUDID API) further enhances the code's functionality and reliability. By leveraging the API, the code establishes a secure connection to the comprehensive device database provided by AccessGUDID, ensuring compliance with data usage policies, facilitating seamless data retrieval, and enhancing the reliability and accuracy of the obtained device information.

Moreover, the use of UDIs plays a crucial role in device identification and traceability. UDIs provide a standardized system for the unique identification of medical devices, ensuring many benefits for both patient and equipment safety. The code's ability to cross-check device data with

UDIs and GMDN names ensures the authenticity of devices, verifies their attributes, and allows for accurate categorization. This cross-checking process allows healthcare professionals to identify and address potential discrepancies, ultimately ensuring patient safety and improving informed decision-making.

CONCLUSION

In conclusion, the combination of web scraping, API integration, and the use of UDIs addresses the challenges of accessing, validating, and analyzing medical device data. The implemented code streamlines these processes, enabling users to efficiently retrieve reliable device information, verify device characteristics, and eventually make informed decisions. The code can be periodically used to ensure the fidelity of UDI codes and GMDN group classification of ME in a medical equipment inventory, to address the dynamic nature of the above systems, and to update the information for new ME types added to the inventory. As technology evolves, more and more advancements in web scraping and API integration will contribute to even more efficient and accurate device data management and analysis in the healthcare industry.

Many suggestions can be taken into account for even greater effectiveness of the solution we proposed. Firstly, validating and cleaning the input data is significant for ensuring data quality in general. Additionally, automating code execution as well as batch processing capabilities can improve efficiency, especially when dealing with large volumes of data. Moreover, integration with healthcare or inventory management systems is significant, in order to synchronize data and improve decision-making capabilities.

Collaborations and partnerships with regulatory bodies, healthcare institutions, or manufacturers can facilitate data sharing and drive industry-wide improvements in device identification and data management practices. Moreover, continuous data monitoring is recommended both to periodically retrieve and update device information from reliable sources so as to maintain data accuracy and relevance. Finally, comprehensive documentation and user support materials is necessary to be provided, as they would help users to use the code effectively and maximize its potential.

ACKNOWLEDGMENTS

We would like to acknowledge the invaluable contribution of the web-Praxis software, as a primary data source for this research.

REFERENCES

1. INBIT. Available online: https://www.inbit.gr/en/.
2. Official Journal of the European Union. The European Parliament and The Council of the European Union; 2024. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0745.
3. Unique Device Identification System (UDI System), FDA. Available online: https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance/unique-device-identification-system-udi-system.
4. GMDN. Available online: https://www.gmdnagency.org/.
5. Sathya, A.C., Rao, P. Mastan, Babu, S. Development of a Covid-19 Information Dashboard to Access the Health Care Resources and Requirements Online. In 2022 3rd International Conference on Computing, Analytics and Networks (ICAN), Rajpura, Punjab, India; IEEE; 2022; pp. 1–8. https://doi.org/10.1109/ICAN56228.2022.10007167.
6. Zhao, B. Web Scraping. In Encyclopedia of Big Data. Laurie A. S., Connie L. M. Springer: Cham, Switzerland; 2017;pp. 1–3.
7. Web Scraping vs API: What's the Difference and Similarity. 2023 Available online: https://www.scrapingdog.com/blog/web-scraping-vs-api/.
8. Khder, M.A. Web scraping or web crawling: State of art, techniques, approaches and application. *Int J Adv Soft Computing Appl*. 2021;13(3):144–168. https://doi.org/10.15849/IJASCA.211128.11.