International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 07 July 2024
Page No: 2074- 2082
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0283

Parsing Unstructured Healthcare Data Using Regular Expressions

Dr. Naresh Patel K M [1], Dr. Azizkhan F Pathan[2], Mrs. Kotramma T S[3], Mrs. Usha K[4], Mrs. Bhuvaneshwari S B[5]

[1]Associate Professor, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India.

[2]Associate Professor, Department of Information Science and Engineering, Jain Institute of Technology, Davangere, Karnataka, India.

[3]Assistant Professor, Department of Information Science & Engineering, Jain Institute of Technology, Davangere, Karnataka, India.

[4,5]Assistant Professor, Department of Computer Science & Engineering, Jain Institute of Technology, Davangere, Karnataka, India.

Emails: *nareshpatela.is@gmail.com[1], apathan21@gmail.com[2], kotrammats@jitd.in[3], ushakrishnamurthy85@gmail.com[4], bhuvaneshwari.seacet@gmail.com[5]*

Abstract

Healthcare organizations produce a substantial volume of healthcare data on a daily basis, managing and monitoring of the healthcare data is tedious and results in data explosion. The healthcare data generated should be properly maintained in the data repository otherwise leads to privacy threat. Most of the healthcare data are unstructured in nature and needs to be organized in a better format. Extracting the meaningful insights from the unstructured healthcare data is complex and most of the time it results in meaningless data. Unless the unstructured healthcare data is transformed to structured healthcare data, proper insights cannot be drawn from the healthcare data. Parsing Unstructured Healthcare Data (PUHD) using the regular expressions helps in identifying the key value elements from the unstructured healthcare data which are the demographic data of the patient and analysis of the healthcare data from these key value elements becomes easier. So, there is a need to design a system that achieves automatic extraction of the key value elements from the unstructured healthcare data.

Keywords: *Demographic Data; Parsing; Regular Expressions; Structured Healthcare Data; Unstructured Healthcare Data.*

1. Introduction

Every day, healthcare organizations generate an enormous influx of data, spanning a wide spectrum of information types. This constant flow includes patient medical records, diagnostic test results, treatment plans, administrative documents, financial transactions and more. The sheer magnitude and diversity of this data create a complex landscape that demands careful management and vigilance. The process of overseeing and regulating healthcare data is no small feat. It involves numerous intricate tasks, such as data collection, storage, retrieval, analysis and sharing. Ensuring the accuracy, reliability and accessibility of this data is paramount to maintaining the quality of patient care, facilitating medical research and supporting informed decision-making [2]. However, the challenge lies in the potential for data proliferation. Without a well-structured approach to managing healthcare data, it can quickly multiply and spread across disparate systems, departments and locations. This can lead to redundancy, inconsistency and inefficiency. As data duplicates and diverges, it becomes increasingly difficult to maintain a unified view of patient information, medical histories and treatment outcomes. [1] This fragmentation can hinder collaboration among healthcare professionals and compromise the seamless delivery of care [3]. One of the critical reasons for effectively organizing and

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 07 July 2024
Page No: 2074- 2082
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0283

overseeing healthcare data is to safeguard patient privacy and data security. Medical records contain sensitive and confidential information about individuals, including personal identifiers, medical conditions, medications and procedures. Mishandling or unauthorized access to this data can have severe consequences, not only in terms of legal and regulatory repercussions but also in eroding patient trust. To address these challenges, healthcare organizations must implement robust data management strategies. This includes adopting standardized data formats, implementing secure data storage practices, establishing data governance frameworks and utilizing advanced technologies for data analytics and insights. By doing so, healthcare organizations can ensure that data remains accurate, easily accessible and protected against potential breaches. Unstructured data is essentially information that does not adhere to a predefined data model or follow a specific organizational structure. Unlike structured data that fits neatly into rows and columns within a database, unstructured data lacks a consistent format or schema. This type of data doesn't conform to rigid rules or predefined patterns, making it more challenging to categorize, process, and analyze compared to structured data [4]. Unstructured data can take various forms, such as text documents, emails, social media posts, images, audio recordings, video files and more. Due to its lack of standardization, unstructured data often requires specialized techniques and tools for extraction, transformation and interpretation. Because of its diverse and irregular nature, unstructured data presents unique challenges and opportunities. Advanced technologies like natural language processing (NLP), machine learning and data mining are employed to extract insights from unstructured data sources [8]. By transforming unstructured data into a more structured or semi-structured format, organizations can unlock valuable information hidden within these seemingly disorganized sources. This process can lead to improved decision-making, enhanced data-driven strategies and a deeper understanding of complex phenomena. Parsing Unstructured Healthcare Data (PUHD) through the utilization of regular expressions is a powerful technique that plays a pivotal role in unraveling valuable insights from the complex landscape of unstructured healthcare data. Regular expressions, often abbreviated as regex, are a tool for pattern matching and text manipulation. [5] When applied to PUHD, they act as a digital scalpel, skillfully extracting relevant pieces of information that are essential for patient care, medical research, and decision-making. These "key value elements" encapsulate vital aspects such as patient demographics, diagnoses, treatments, medications, dates, and more. Healthcare data varies in format across different organizations, making it challenging to process directly. To process such data, we present the Parsing Unstructured Healthcare Data (PUHD) algorithm that contains regular expressions for identifying the key elements from the textual data that are the necessary details of the person and also helps in making analysis. Data frame functions in Python transform the identified key elements into a structured format, forming relational tables. [6] figure 1 shows the proposed Architecture. By employing regular expressions to perform PUHD algorithm, healthcare professionals and data analysts can achieve several key benefits. [7]

1.1.Precise Data Extraction

Regular expressions allow for highly specific pattern matching, ensuring that only the desired data elements are extracted. This precision minimizes the risk of errors or omissions in the extracted information.

1.2.Efficiency in Data Handling

Manually sifting through large volumes of unstructured data is time-consuming and error-prone. Regular expressions automate the extraction process, saving valuable time and reducing the chances of human oversight.

1.3.Enhanced Data Analysis

Extracting key value elements facilitates efficient data analysis. The organized, structured data can be easily subjected to various analytical techniques, aiding in trend identification, anomaly detection, and predictive modeling.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 07 July 2024
Page No: 2074- 2082
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0283

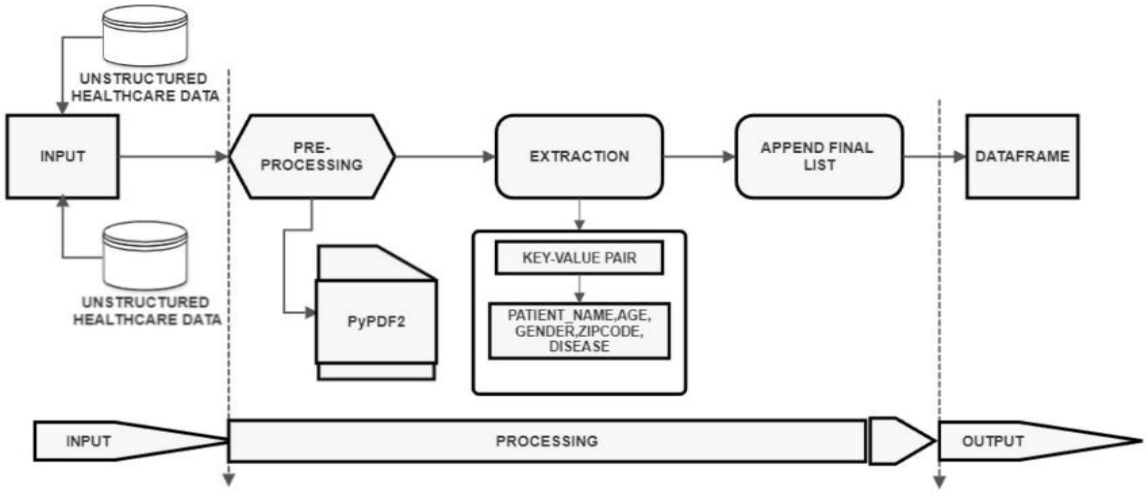


Figure 1 Proposed Architecture for Transforming Unstructured Healthcare Data to Structured Healthcare Data

1.4.Clinical Decision Support

Access to accurate patient details through key value elements empowers healthcare providers to make well-informed clinical decisions. They can quickly identify relevant medical history, treatment plans, and risk factors, leading to improved patient outcomes.

1.5.Research Insights

Researchers can leverage PUHD to efficiently gather data for studies and investigations. Regular expressions can help identify patients with specific conditions, enabling researchers to analyze treatment outcomes and disease progression. [9]

1.6.Data Integration

Extracted key value elements can be integrated into electronic health records (EHR) systems or other structured databases. This integration enhances data continuity, interoperability and the overall quality of healthcare data. Parsing Unstructured Healthcare Data (PUHD) algorithm using regular expressions is a transformative approach that allows for the efficient extraction of key value elements from unstructured healthcare data. This process greatly simplifies the analysis and utilization of healthcare data, leading to improved patient care, research insights and informed decision-making within the healthcare domain.

2. Literature Survey

Mohamed Mehfoud Bouh, Forhad Hossain and Ashir Ahmed “A Machine Learning Approach to Digitize Medical History and Archive in a Standard Format”. In: 9th International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2023). In this paper the authors present extracting structured data from unstructured scanned documents using technologies such as OCR and NLP, and ML models to extract useful information standardizing it in a common format like Fast Healthcare Interoperability (FHIR) or Open Electronic Health Record (EHR), and using methods like ML and BERT models to generate predictions is a relatively new field. Healthcare organizations look for ways to use the enormous quantity of data included in scanned medical documents to enhance patient outcomes and reduce costs, this approach has gained increasing attention in recent years. Kye Hwa Lee et. al. “ANNO: A General Annotation Tool for Bilingual Clinical Note Information Extraction”, In: The Korean Society of Medical Informatics (2022). In this paper the authors proposed ANNO is a Docker-based web application that users can freely use without worrying about dependency issues. Human annotators can share their annotation markups as regular expression sets in a dictionary structure and cross-check their annotated corpora with one another. The main features of ANNO include dictionary-based regular expression sharing, cross-check functionality for each annotator, and standardized input (Microsoft Excel) and output

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 07 July 2024
Page No: 2074- 2082
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0283

(extensible markup language [XML]) formats. Dipali Baviskar, et. al." Efficient Automated Processing of the Unstructured Documents Using Artificial Intelligence". In: IEEE Access (2021). In this paper, the authors proposed AI-based approaches have a strong potential to extract useful information from unstructured documents automatically due to advancements in natural language processing (NLP) and machine learning. These technologies enable the analysis of text data to identify patterns, entities, and relationships that are not immediately apparent. NLP techniques like tokenization, named entity recognition (NER), and sentiment analysis help in breaking down and understanding complex texts. Machine learning models, particularly deep learning, can be trained on vast datasets to improve their accuracy in tasks like classification and information retrieval. Additionally, AI can handle large volumes of data efficiently, making it possible to process documents at a scale that would be impractical for humans. This automation not only saves time but also enhances the consistency and reliability of the extracted information, providing valuable insights for various applications such as data mining, business intelligence, and content management. Agostino Forestiero, et. al. "Natural language processing approach for distributed health data management". In: Euromicro International Conference on Parallel, Distributed and Network- Based Processing (2020). In this paper, the authors proposed Natural Language Processing (NLP) approach for distributed health data management. The increasing use of digital health data, like electronic health records (EHRs), has led to store an unprecedented amount of information. Managing this large amount of data can often introduce issues of information overload, with potential negative consequences on clinical work, such as errors of omission, delays, and overall patient safety. Health data are represented with vectors obtained through the Doc2Vec model. The effectiveness of the approach was proved performing a set of preliminary experiments exploiting a tailored implemented simulator. Jaina Bansal, Amarnath Poddar, et. al. "Identifying a Medical Department Based on Unstructured Data: A Big Data Application in Healthcare". In: International Journal for Research in Engineering Technology (IJRET) (2019). In this paper, they proposed a system that will scan prescriptions, referral letters and medical diagnostic reports of a patient, process the input using OCR (Optical Character Recognition) engines, coupled with image processing tools, to direct the patient to the most relevant department. They have implemented and tested parts of this system wherein a patient enters his symptoms and/or provisional diagnosis, the system suggests a department based on this user input. Veena G, et. al. "Relation Extraction in Clinical Text using NLP Based Regular Expressions". In: 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT) (2019). In this paper, the authors proposed a relation extraction system for medical field related data. The major aim of this work is to retrieve different medical related data and to find out the relation between extracted medical data. Medical data usually contains a lot of unstructured or semi-structured data; by implementing methods like labeling and path similarity analysis we are able to convert it into a structured or classified form. Other methods that we use in our work are web scrapping, regular expressions, and part-of-speech tagging. All these methods are implemented in python. Using "Relation extraction in clinical texts" algorithm.

3. Existing System

The healthcare industry currently prioritizes structured data, which is neatly organized into predefined models, facilitating straightforward information retrieval and analysis. In contrast, unstructured healthcare data, comprising clinical notes, medical images and patient records lacks a standardized format, presenting significant challenges for direct use. Extracting valuable insights from this unstructured data necessitates extensive preprocessing and transformation. By converting unstructured data into structured formats, healthcare professionals can leverage advanced analytical tools more effectively. This transformation not only streamlines data analysis but also enhances patient care by providing more

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 07 July 2024
Page No: 2074- 2082
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0283

accurate and timely information. Ultimately, structured data supports better decision-making processes, driving improvements in healthcare outcomes and operational efficiencies.

4. Problem Statement

The healthcare industry faces significant challenges in utilizing unstructured data such as clinical notes, medical images, and patient records due to its lack of standardized format. Unlike structured data, which is easily organized and analyzed, unstructured data requires extensive preprocessing and transformation to extract valuable insights. This inefficiency hampers the ability of healthcare professionals to fully leverage advanced analytical tools, impacting the quality of patient care and decision-making processes. Therefore, there is a critical need to develop effective methods for transforming unstructured healthcare data into structured formats to enhance data analysis, improve patient outcomes, and optimize operational efficiencies in the healthcare sector.

5. Proposed System

To address the challenges posed by unstructured healthcare data, a comprehensive solution involves developing and implementing advanced data preprocessing and transformation techniques. This includes leveraging Parsing Unstructured Healthcare Data (PUHD) algorithm to analyze and structure discharge summary of the patients that are in Pdf. By creating robust pipelines that automate these processes, healthcare data can be effectively converted into structured formats. Additionally, integrating these solutions with existing healthcare information systems will facilitate seamless data flow and accessibility. This approach will enable healthcare professionals to harness the full potential of advanced analytical tools, leading to improved patient care, more accurate decision-making and enhanced operational efficiencies across the healthcare sector.

6. PUHD Algorithm

Algorithm: Parsing Unstructured Healthcare Data (PUHD)
Input: Unstructured Healthcare Data in PDF format
Output: List of all unstructured data in structured format as Data frame
Begin initializing the buffer lists
bufferlist ← ””
Instantiate the PDFReader from the PyPDF2 library
reader = PdfReader(file_path)
retrieve
pdf_text = page.extract_text ()
Split the string into a list at each newline:
myvalue =
mystr.splitlines()
Iterate through each line in listitem
do
Assign key-value pairs to each list line, separated by ‘:’
Append
finallist←create
key_valuepair()
{demographic data of the patient}
if line!={key_value} format then
Apply regex to perform a pattern search on the keyword
Transfer the remaining patterns into a List
Map the list to a dictionary, assigning keys 0 through n
Based on the entries
Append finalist=finalist.append
endif
end while
Transform the finalist with key_value pairs as dataframe
Data frame= pandas.dataframe({Key_value})
Return the updated data frame as return data frame
The proposed algorithm is used to draw the useful insights from the unstructured healthcare data using regular expressions. The PUHD algorithm accepts the unstructured healthcare data in pdf format which in turn transformed to text file using PyPDF2 library. The PyPDF2 package provides the features like cropping, splitting, merging, transforming of the pdf files. The demographic data of the patients are extracted using regular expressions. Iterate through each line, splitting list items to identify strings that match regular expressions, and store them in a Python dictionary. Create key-value pairs for each new line, which will then be appended to the final list. The key-value pairs now contain the

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 07 July 2024
Page No: 2074- 2082
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0283

demographic data of the patient. Regular expressions are effective tools used to match strings and text processing. The recognized demographic data of the patient are then appended to the list. A data frame function is now used to create the data frame which is size mutable depending on the demographic data recognized. Now the structured healthcare data is available in the form of relational table. The data analyst and researchers can now make a significant analysis or contribute to the society by drawing the useful insights.

7. Dataset Description

In this paper, patient unstructured healthcare data is meticulously gathered from local hospitals, ensuring a rich and diverse collection of clinical narratives. The data collection process involves direct collaboration with hospital administrative and IT departments to access electronic health records (EHRs) containing discharge summaries. These summaries, authored by healthcare providers, include detailed accounts of patient history, diagnostic findings, treatment regimens, and follow-up instructions. The unstructured nature of these documents allows for the capture of nuanced clinical insights that structured data often overlooks. The physical collection process is conducted with stringent adherence to ethical guidelines and patient confidentiality protocols. All identifiable patient information is anonymized to protect privacy. The collected data undergoes a secure transfer to our research facility, where it is stored in a controlled environment, ensuring data integrity and security. This dataset provides a robust foundation for exploring natural language processing (NLP) techniques, enabling the extraction of valuable clinical insights and fostering advancements in healthcare informatics. By leveraging this rich dataset, we aim to enhance predictive analytics, support clinical decision-making, and ultimately contribute to the improvement of patient care and outcomes.

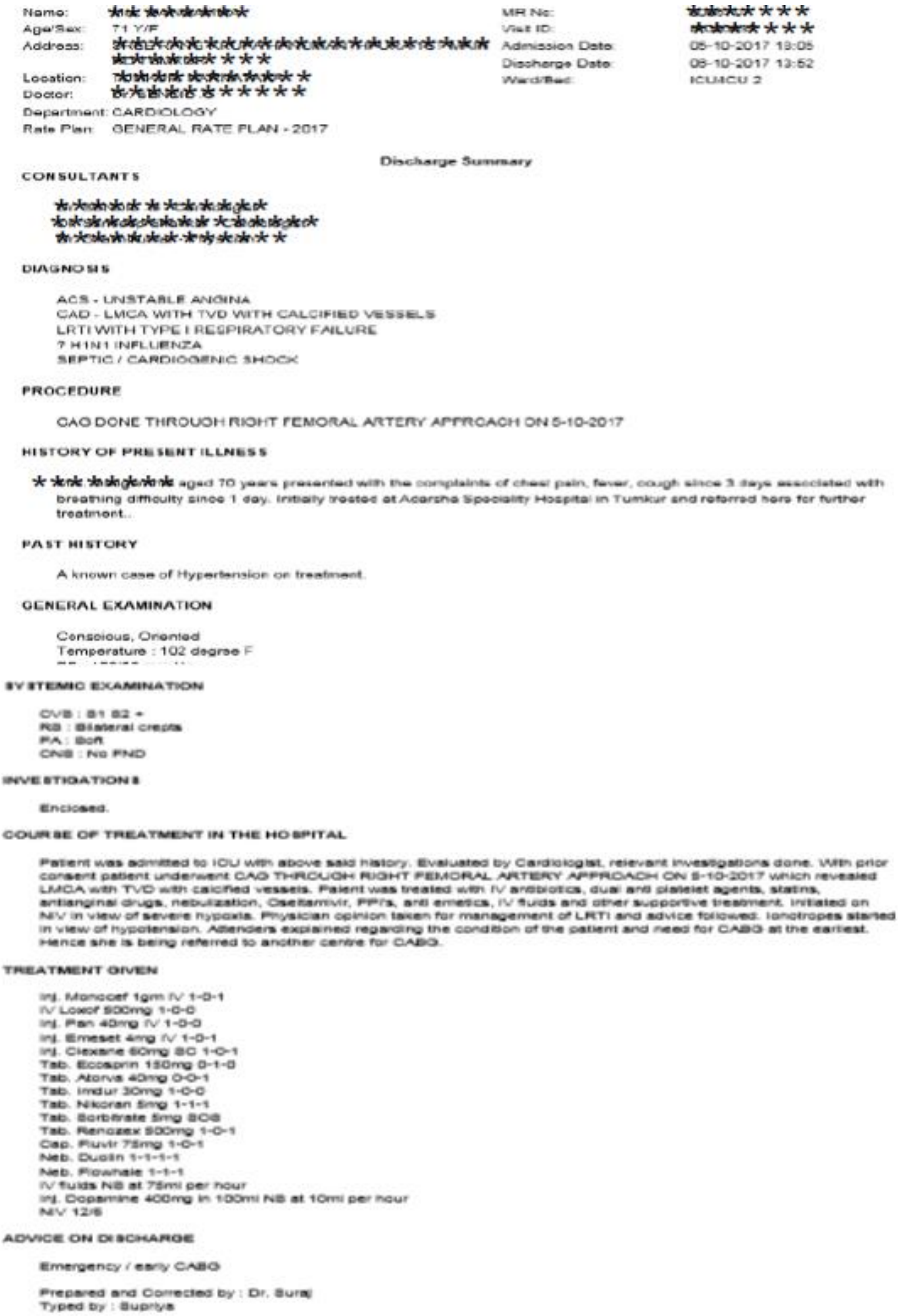
8. Results And Discussion

In this section, the execution time of the proposed technique is evaluated through practical experiments. These experiments are carried out on a system equipped with an Intel i5 processor and 4GB of RAM. The computational environment is set up using the Anaconda distribution, which facilitates the transformation of unstructured healthcare data into a structured format. During the experiments, the execution time required to process each individual record is measured separately to provide detailed performance insights. The model receives unstructured healthcare data as input, specifically consisting of discharge summary reports collected from local hospitals. The format of discharge summary reports differs across hospitals, as each institution follows its own unique system for storing patient data. This variation in formatting poses a challenge for standardizing and processing the unstructured healthcare data effectively. Initially the considered data will be in PDF format, the data is converted to text using PyPDF. By parsing these text files, we extract essential demographic data. Lastly the structured healthcare data is organized into a relational table using a data frame. The execution time depends on the length of the discharge summary reports. It is shorter for reports with fewer pages and increases as the number of pages in the reports grows. The unstructured healthcare data consists of discharge summary reports, which contain comprehensive information about the patient. An example of unstructured healthcare data is shown in Fig. 2. In larger cases, demographic data suffice for decision-making and data analysis. The collected unstructured healthcare data is parsed to obtain structured healthcare data, as depicted in Table 1.

Table 1 Sample Unstructured Healthcare Data Transformed to Structured Data.

	Name	Age	Gender	State
0	Mr. NAGARAJ	81	M	BANGALORE, KARNATAKA
1	Mrs. GANGAMMA	71	F	TUMKUR, KARNATAKA
2	Mr. SATHYANARAYAN	83	M	BANGALORE, KARNATAKA
3	Master DARSHAN	12	M	BANGALORE, KARNATAKA
4	Mr. MUNIRAJU M V	51	M	BANGALORE, KARNATAKA
5	Mr. GANGAPPA	91	M	BANGALORE, KARNATAKA

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 07 July 2024
Page No: 2074- 2082
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0283



Name: ************ MR No: **********
Age/Sex: 71 Y/F Visit ID: **********
Address: ************************************ Admission Date: 05-10-2017 19:05
************ Discharge Date: 06-10-2017 13:52
Location: ************* Ward/Bed: ICU/ICU 2
Doctor: ************
Department: CARDIOLOGY
Rate Plan: GENERAL RATE PLAN - 2017

Discharge Summary

CONSULTANTS

****
****
****

DIAGNOSIS

ACS - UNSTABLE ANGINA
CAD - LMCA WITH TVD WITH CALCIFIED VESSELS
LRTI WITH TYPE I RESPIRATORY FAILURE
? H1N1 INFLUENZA
SEPTIC / CARDIOGENIC SHOCK

PROCEDURE

CAG DONE THROUGH RIGHT FEMORAL ARTERY APPROACH ON 5-10-2017

HISTORY OF PRESENT ILLNESS

********** aged 70 years presented with the complaints of chest pain, fever, cough since 3 days associated with breathing difficulty since 1 day. Initially treated at Adarsha Speciality Hospital in Tumkur and referred here for further treatment.

PAST HISTORY

A known case of Hypertension on treatment.

GENERAL EXAMINATION

Conscious, Oriented
Temperature : 102 degree F

SYSTEMIC EXAMINATION

CVS : S1 S2 +
RS : Bilateral crepts
PA : Soft
CNS : No FND

INVESTIGATIONS

Enclosed.

COURSE OF TREATMENT IN THE HOSPITAL

Patient was admitted to ICU with above said history. Evaluated by Cardiologist, relevant investigations done. With prior consent patient underwent CAG THROUGH RIGHT FEMORAL ARTERY APPROACH ON 5-10-2017 which revealed LMCA with TVD with calcified vessels. Patient was treated with IV antibiotics, dual anti platelet agents, statins, antianginal drugs, nebulization, Oseltamivir, PPI's, anti emetics, IV fluids and other supportive treatment. Initiated on NIV in view of severe hypoxia. Physician opinion taken for management of LRTI and advice followed. Ionotropes started in view of hypotension. Attenders explained regarding the condition of the patient and need for CABG at the earliest. Hence she is being referred to another centre for CABG.

TREATMENT GIVEN

Inj. Monocef 1gm IV 1-0-1
IV Loxof 500mg 1-0-0
Inj. Pan 40mg IV 1-0-0
Inj. Emeset 4mg IV 1-0-1
Inj. Clexane 60mg SC 1-0-1
Tab. Ecosprin 150mg 0-1-0
Tab. Atorva 40mg 0-0-1
Tab. Imdur 30mg 1-0-0
Tab. Nikoran 5mg 1-1-1
Tab. Sorbitrate 5mg SOS
Tab. Ranozex 500mg 1-0-1
Cap. Fluvir 75mg 1-0-1
Neb. Duolin 1-1-1-1
Neb. Flowhale 1-1-1
IV fluids NS at 75ml per hour
Inj. Dopamine 400mg in 100ml NS at 10ml per hour
NIV 12/6

ADVICE ON DISCHARGE

Emergency / early CABG

Prepared and Corrected by : Dr. Suraj
Typed by : Supriya

Figure 2 **Sample Unstructured Healthcare Data**

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 07 July 2024
Page No: 2074- 2082
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0283

The execution time of processing the discharge summary report varies based on its length. The system's analysis indicates that an increase in the number of pages leads to longer processing times for the model to transform the data into structured healthcare data, as illustrated in the accompanying display. Figure 3 shows the execution Time graph.

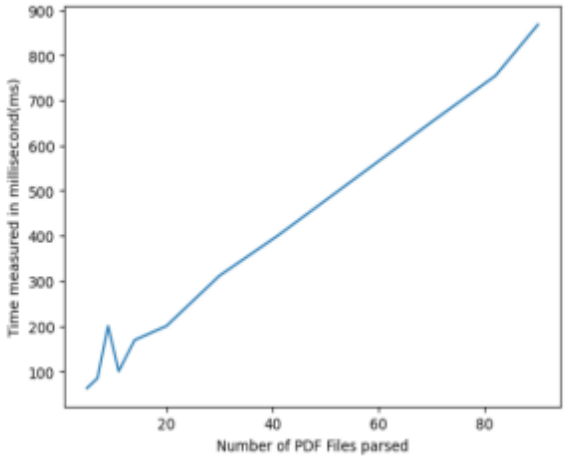


Figure 3 Execution Time for Parsing Unstructured Healthcare Data

Conclusion

Analyzing healthcare data and making informed decisions is increasingly crucial in current trends. Extracting insights directly from healthcare data can be challenging or inconclusive without proper formatting into a relational table structure. Formatting the data in this way facilitates clearer conclusions and streamlines the analytical process. Structuring healthcare data into a relational table is a crucial initial step for its further processing. This organized format lays the foundation for efficient data analysis, decision-making, and other critical tasks in healthcare management and research. The PUHD algorithm plays a crucial role in transforming healthcare data into a structured format which now contains the demographic data of the patient. Once the data is structured, researchers and decision-makers can effectively analyze it and contribute to society through informed insights and actions. The PUHD algorithm investigated has less execution time for the reports having less number of pages and execution time gradually increases with number of pages increases.

References

[1]. Mohamed Mehfoud Bouh, Forhad Hossain and Ashir Ahmed, “A Machine Learning Approach to Digitize Medical History and Archive in a Standard Format”, In: 9th International Conference on Information and Communication, Technologies for Ageing Well and e-Health (ICT4AWE), 2023. doi: 10.5220/0011986400003476.

[2]. Kye Hwa Lee et. al. “ANNO: A General Annotation Tool for Bilingual Clinical Note Information Extraction”, In: The Korean Society of Medical Informatics, 2022. doi: 10.4258/hir.2022.28.1.89.

[3]. Dipali Baviskar, et. al., “Efficient Automated Processing of the Unstructured Documents Using Artificial Intelligence”. In: IEEE Access (2021). doi:10.1109/ACCESS.2021.3072900.

[4]. Agostino Forestiero, Giuseppe Papuzzo, “Natural language processing approach for distributed health data management”, 28th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), 2020. doi: 10.1109/PDP50117.2020.00061.

[5]. Jaina Bansal, Amarnath Poddar, R. G_Roy, “Identifying a Medical Department Based on Unstructured Data: A Big Data Application in Healthcare”, International Journal for Research in Engineering Technology (IJRAET), 2019. doi: 10.22214/ijraet.2019.36972.

[6]. Veena G, R. Hemanth, Hareesh, Jithin, “Relation Extraction in Clinical Text using NLP Based Regular Expressions”, 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), 2019. doi: 10.21275/24422134802

[7]. Wencheng Sun, Zhiping Cai, Yangyang Li, Fang Liu, Shengqun Fang, Guoyan Wang, “Data Processing and Text Mining Technologies on Electronic Medical Records”, Journal of Healthcare Engineering, 2018. doi: 10.1155/2018/4302425

[8]. Tomoya Matsumoto, Wataru Sunayama, Yuji

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 07 July 2024
Page No: 2074- 2082
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0283

Hatanaka, Kazunori Ogohara, “Data Analysis Support by Combining Data Mining and Text Mining”, 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), 2017. doi: 10.1109/IIAI-AAI.2017.165

[9]. Jagruti Jangal Wagh, Jidnyasa Dharmik Gongane, Ashvini Tulshiram Dukare, “Unstructured Data Mining and Its Application”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), 2016.