SENIOR DESIGN PROJECT

AUTOMATION AND VISUALIZATION OF PATHOLOGICAL DATA FROM SUM HOSPITAL

Submitted by

AMIT ANAND
anandamit.ops@gmail.com

SUNIDHI SINGH
sunidhi.singh421@gmail.com

STIF SPEAR SUBBA
spearoff@gmail.com

SANGEET MOY DAS
dassangeet768@gmail.com

Guided by,

Dr. TRIPTI SWARNKAR
triptiswarnakar@soa.ac.in

START DATE : 04-01-2018
END DATE : 15-04-2018

TABLE OF CONTENT

ABSTRACT

SUM Hospital, a leading medical institution of India handling huge quantity of patient records on a day-to-day basis. We have noticed that this process is done manually, which is not time efficient, as analysis and visualization of the collected data take lot of man hour and has to be done frequently.

Our project focuses on creating a in-house application which can be used for automating this process of data visualization and analysis of pathological data collected from the patient. We plan to automate the complete process of data collection and with the use of certain Machine Learning algorithm plan to devise a model which can be used to make prediction based on patient history and antibiotics sensitivity pattern.

INTRODUCTION

2.1 PURPOSE:

The purpose of this project is to reduce the Time complexity and optimize the resources involved in analyzing the patient data collected daily in SUM Hospital using our automation tool created using Ruby on rails framework and prediction of medication using patient antibiotics sensitivity patterns using machine learning algorithms.

2.2 SCOPE:

Machine learning algorithms are applied to the large-scale, multidimensional datasets of the healthcare labeled data. The machine learning technique such as principle component analysis for dimensionality reduction is applied for creating the training models for identifying the false positive cases. Unsupervised methods can be applied with principle component analysis of dimensionality reduction for the identification of antibiotics sensitivity of a particular patient. Clustering analysis can be performed through supervised and unsupervised methods. In the supervised method, the machine algorithm clustering method finds the patterns and applies the technique for segmentation and builds the labeled data for predicting the outcome of the test data through classifiers. In case of unsupervised method such as K-means clustering, deep learning methods are applied that can identify the patterns in the clusters autonomously without the requirement of building the training labeled data.

SOLUTION

3.1 PROBLEM STATEMENT:

Most of the visualization were done manually which was not time efficient, our automation tool is responsible for taking patient data from the hospital system and using those data for analyzing patient history of antibiotic sensitivity pattern and based on the result provide output for better medication and overall decrease the redundant time process.

The original dataset used during the initial phase contains:

- 84 unique columns of different data.
- 1289 different patient data collected over the period of 1 month i.e; APRIL.

Major findings of each dataset:

1. Sensitivity to a particular antibiotic.
2. Result (positive/ negative) to a particular antibody.
3. Place of isolation of the organism like ICU, Cabins and NICU.
4. Isolates resistant to which antibiotics.

3.2 EXISTING SOLUTION:

The existing solution which has been previously implemented make use of the Microsoft access data entry tool for storing the patient medical records which was done manually and in the end of a calendar month calculations were made using the collected data for visualization using Bar Graph, Pie Chart etc.

Moreover World Health Organization has developed a Open Source desktop application which can be used by medical institution for analysis and visualization of medical data which can be further used for better health management for patient by providing them suitable medication based on their sensitivity test.

3.2.1 SHORTCOMINGS:

- The throughout process was manual which was tedious and not time efficient.
- The process was prone to human error of miscalculation and in case of medical data error can be fatal.
- The data was unstructured and not clean to undergo analysis process.

3.3 PROJECT DESIGN:

The purpose of this document is to outline the technical design of the automation application to analyse the pathological data and provide an overview for the implementation of the model. It's main purpose is to:

- Provide the link between the Functional Specification and the detailed Technical Design documents.
- Detail the functionality which will be provided by each component or group of components and show how the various components interact in the design.

- Provide a basis for the automation system detailed design and development.

As is true with any high level design, this document will be updated and refined based on changing requirements.

3.3.1 Description of the Problem:

The automation application is responsible for taking data stored in access DB using using the MS Office tool and then based on the functionalities defined in the modules, analysis and visualization of the data set will be done.
This procedure will tremendously reduce man hours and minimize error related to calculation which was the reason we discarded the first method of analysis.

3.3.2 System Architecture:

The system will be constructed from multiple distinct components:

- **Input Module**

  Initially, user_authentication takes place where he needs to provide the user_id and password. Its function includes:

  1. authenticate()

     Here the authenticity of the user will be verified, so that any unknown personnel do not get access to sensitive patient records.

2. readDatabase()

    This method will get access to all the data rows and columns using the DatabaseHandler class and then using the AutomationTool class the main functionalities will be implemented.

- **Processing Module**

    Initially with the help of the data collected in the input module automation will be done in this phase with the help of TestFilter class.

    The functions which the class invokes:

    1. getTestFilter()
    2. getTable()
    3. draw()

        This method is used to create the various graphs for data set visualization like bar graph, Pie charts, histogram etc.
    4. addFilter()
    5. createLineChart()
    6. createPieChart()
    7. init()
    8. addComponent()
    9. addFilter()
    10. fillTable()
    11. addElementToList()
    12. getList()
    13. ClickAction()
    14. etc...

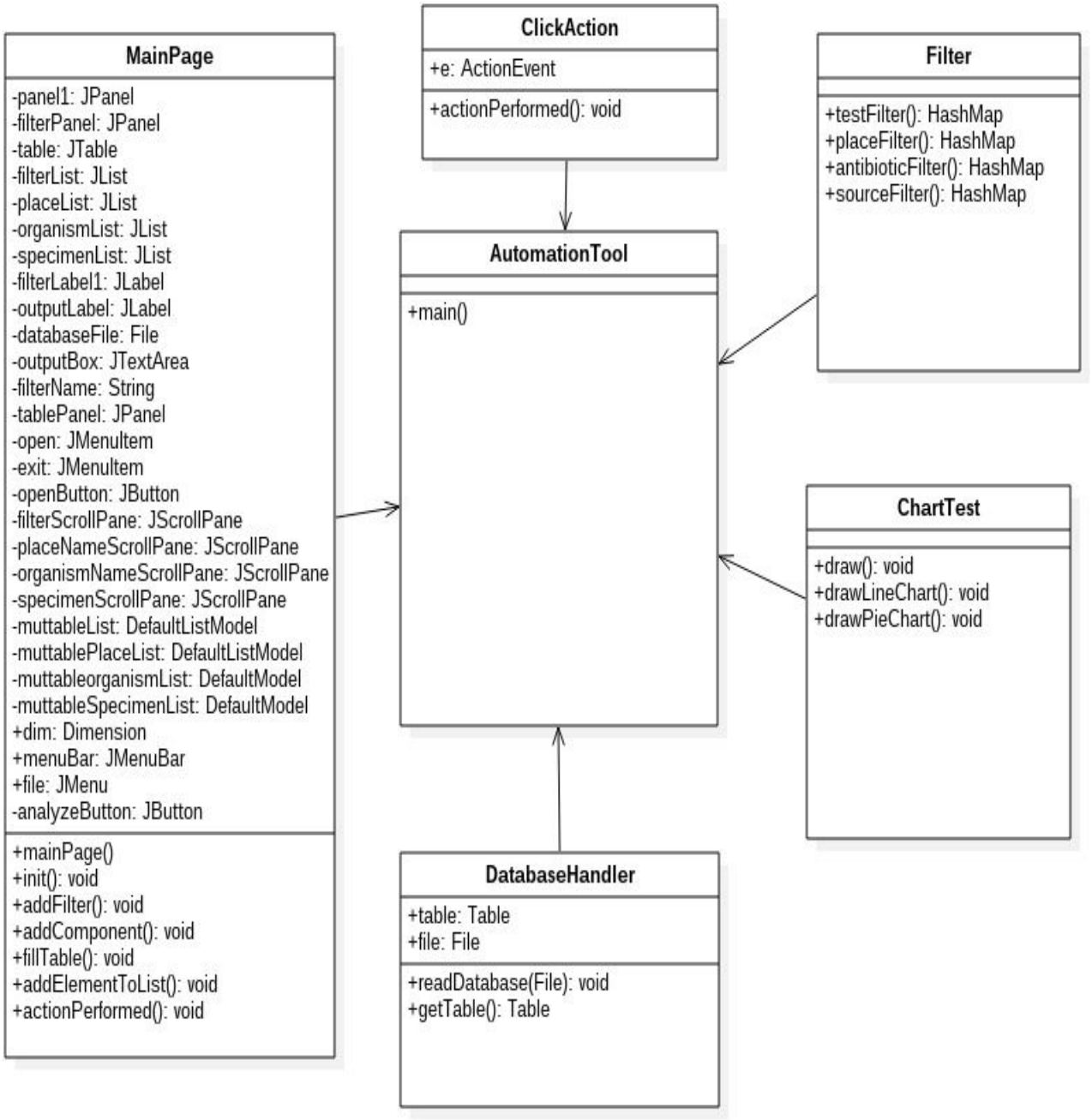
NOTE: Many more modules will be added as the development process proceeds.

3.3.3 System Architectural Design:

We have chosen Object Oriented Designing for the application for automating the analysis and visualization process. Object-oriented design is the process of planning a system of interacting objects for the purpose of solving a software problem. It is one approach to software design. An object contains encapsulated data and procedures grouped together to represent an entity. The 'object interface', how the object can be interacted with, is also defined. An object-oriented program is described by the interaction of these objects. Object-oriented design is the discipline of defining the objects and their interactions to solve a problem that was identified and documented during object-oriented analysis. We have chosen this architecture because of the following reasons:

- Easier maintenance.
- Objects may be understood as stand-alone entities.
- Objects are potentially reusable components.
- For some systems, there may be an obvious mapping from real world entities to system objects.

2.3.4 Class Diagram:



SYSTEM REQUIREMENTS

4.1 OVERALL DESCRIPTION:

- PROJECT ABSTRACT

  Project Name: Automation and visualization of pathological data from SUM Hospital.

  Project Member: Amit Anand
  Sunidhi Singh
  Stif Spear Subba
  Sangeet Moy Das

  Date: 9th February 2018

- SYSTEM CONSTRAINTS

  Hardware Constraints:

  CLIENT: Medical Personnel, Hospitals.

  MACHINE: Processor- Pentium 4, RAM-1GB (minimum)

  Software Constraints:

  CLIENT: The client should have browser.

  DEVELOPER: Ruby on rails

  Design Standard Compliance:

  The system shall be implemented in Ruby on rails on the front end with postgre SQL as the prefered DataBase.

Software Environment:

TYPE	**NAME**
OS	Windows, Linux, MacOS
Language	Ruby, Python
Library	Ruby on Rails, Postgre SQL

4.2 Screenshots:

Figure 1: Application Home

Figure 2: Showing patient data base in the browser

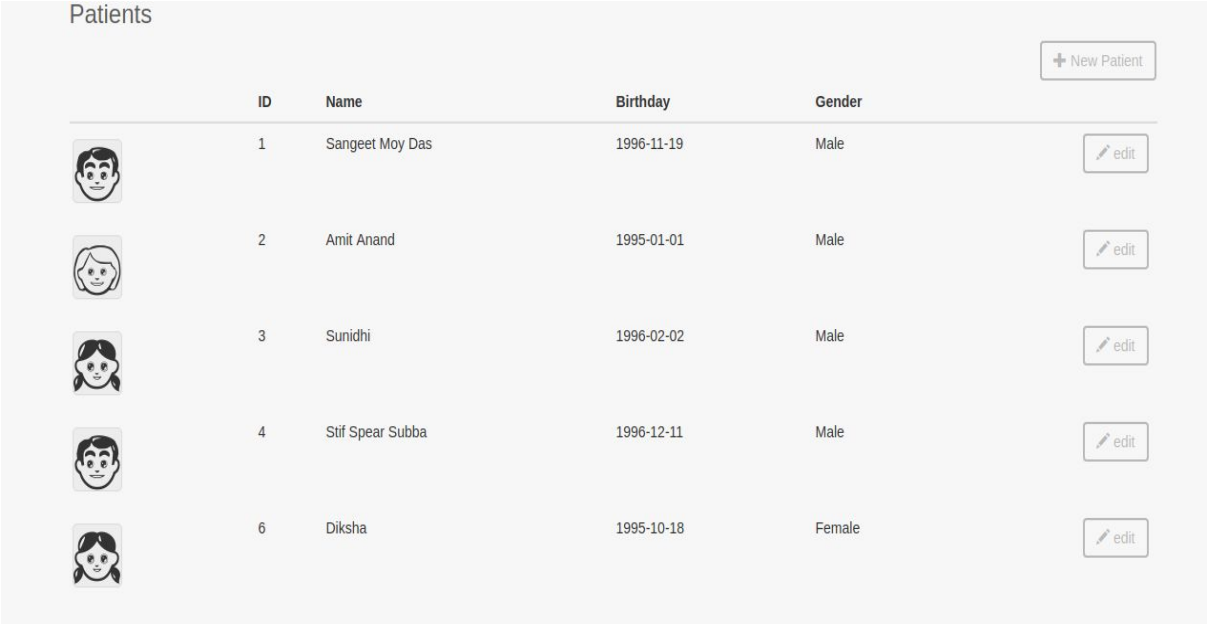


Figure 3: Editing patient database via browser.

4.3 SYSTEM EVENTS AND DATA FLOW:

- SYSTEM ACTIVITIES:

  System will take data from the patient medical records and then compute on the following factors:

  - Places from where sample Isolated
  - Organisms
  - Antibiotics sensitivity
  - Collection source

  The output after analyzing the Data Set:

  1. Total number of cultures received.
  2. Total positive cases.
  3. Total number of samples from different location like Cabin Ward, ICUs, NICUs.
  4. Organism isolated from different locations.
  5. Interpretation of result from antibiotic sensitivity test.

- USER CHARACTERISTICS:

  1. Computer Proficiency: Beginner
  2. Should be able to make sense of the analysed medical data.

EXPECTED RESULT

The automation tool which was developed for analysis and visualization of the patient records collected at SUM Hospital should be able to produce the following expected results:

- Data should be filtered using the requirement of the user.
- Analysis should be time efficient.
- Error free visualization should be done.
- Prediction of antibiotic given the patient medical history.
- The machine learning model which will be trained using our data set should acquire at least 80% efficiency.
- The automation tool should function in a bug free manner.
- It should be simple to use for user with beginner level proficiency in using such application.
- User authentication function should work with minimum breach possibility.

PROJECT TIMELINES

DELIVERABLES REGISTER:-

Deliverable Description	Planned Start Date	Planned End Date	Status/ Comments
Creating the UI of the automation tools which will be used for analysis and visualization of the patient data.	7-09-17	28-09-17	Delivered
Setting up the database for storing the patient record in our in-house automation application.	29-09-17	18-10-17	Delivered
Setting up the backend which was primarily written using JAVA with use of libraries like awt, swing, jackcess, jfreechart for added feature for filtering the data for visualization.	19-10-17	04-12-17	Delivered
Setting up the front end of the web app which would act as the UI for the database of the app via which data will be entered.	29-01-18	19-02-18	--todo--
Creating a MySQL database to store data, to be input directly from the automation web app and merging it with the front end.	20-02-18	07-03-18	--todo--
Adding more filters based on the query given by the client.	08-03-18	15-03-18	--todo--
Testing phase before final delivery and deployment.	16-03-18	26-03-18	--todo--
Deployment	27-03-18	31-03-18	--todo--

*dates are subject to change based on client and developer feedback exchange.

Siksha 'O' Anusandhan (Deemed to be University)
Established U/S –3 of UGC Act - 1956
INSTITUTE OF TECHNICAL EDUCATION AND RESEARCH
Jagamohan Nagar, Bhubaneswar-30
Department of Computer Science and Engineering

Sample Survey Questions

It includes 10 questions by the following group members:

Sangeet Moy Das
Stif Spear Subba
Amit Anand
Sunidhi Singh

Questionnaire for the app: Automation and Visualization of hospital Pathological Data.

Qn No.	Description
1.	**Type of data you deal with?** ❑ Mixed ❑ Images ❑ Tables ❑ User details ❑ Others ……...
2.	**Total number of expected users of the application?** ❑ 1 to 5 ❑ 6 to 10 ❑ 11 to 20 ❑ Others …….
3.	**How often do you analyze your data?** ❑ Daily ❑ Weekly ❑ Monthly ❑ Quarterly ❑ Others ……..
4.	**Do you require a tutorial for your application?**

	❑ Yes ❑ No ❑ Maybe
5.	**How would you like the output of data to be?** ❑ Numbered tabular format ❑ Graph analysis ❑ Both ❑ Other …….
6.	**How would you like the authentication system to be?** ❑ Yes, Password protected ❑ No ❑ Not required
7.	**Do you want your application to provide prediction based on patient sensitivity history?** ❑ Yes ❑ No ❑ Maybe
8.	**Do you want your application to provide support to analyse image based patient data?** ❑ Yes ❑ No ❑ Maybe
9.	**Do you want a version of the application which patients can use to enter their personal data and get prediction based on them?** ❑ Yes ❑ No ❑ Maybe
10.	**In your previous method of analysis which was the most difficult phase you encountered?** ❑ Data Entry ❑ Calculation ❑ Analysis ❑ Visualization

SRS DOCUMENT

Prepared by:-
Stif Spear Subba
Sangeet Moy Das
Amit Anand
Sunidhi Singh

1. INTRODUCTION

The following subsections of the software Requirements Specifications (SRS) document provide an overview of the entire SRS.

1.1 PURPOSE:

- The SRS document will provide a detailed description for the analysis and visualization of the medical data received from the different:
    - Departments of SUM hospital. The designed application will accurately calculate the positivity test for a given culture.
    - It will be used by the HOD of the microbiology department and other staff members of the department.

1.2 SCOPE:

- The software product to be produced is the automation of calculation and analysis of microbiological data provided by SUM Hospital.

- The first subsystem will do the automation and visualization and the second subsystem will accurately calculate the positivity test of the given culture.

1.3 DEFINITIONS & ABBREVIATIONS:

- SRS- Software Requirements and Specification
- End Users- The people who will be using the System.

2. THE OVERALL DESCRIPTION

Describes the general factors that affect the product and its requirements. This section does not state specific requirements. Instead it provides a background for those requirements, which are defined in the next section, and makes it easier to understand.

2.1 PRODUCT PERSPECTIVE:

The visualization and analysis application for SUM is an independent stand-alone system.It is totally self contained.

2.1.1 HARDWARE INTERFACES:

The automation module will be placed in the System of all the Pathological Lab.

2.1.2 SOFTWARE INTERFACES:

All databases for the application will be configured using postgre SQL. These database contain 84 data fields used for classifying the data, which can be modified by the end user.

2.2 PRODUCT FUNCTIONS:

- Allows automation of calculation of resistivity and sensitivity of each organisms for different antibiotics.
  -> Read the data from the database.
- Calculation of positivity and negativity test of each organisms based on different antibiotics.
- Percentage calculation of each type of organisms from the given sample.
- Print all the calculations in tabular form.

2.3 USER CHARACTERISTICS:

- Experience of Software: Minimal
- Technical Expertise : Web Browser

2.4 ASSUMPTIONS AND DEPENDENCIES:

- 'Blank spaces' in file denotes "no test is done for the sample".
- 'S' denotes the "sensitive" organisms to given antibiotics.

- 'SS' denotes "slightly sensitive" organisms to the given antibiotics.
- 'R' denotes "Resistant" organisms to the given antibiotics.

3. SPECIFIC REQUIREMENTS

This section contains all the Software requirements at a level of detail, that is sufficient to enable designers to design a system to satisfy those requirements and to testers to test that the system satisfies those requirements.

3.1 EXTERNAL INTERFACES:

The Microbiological department will use the standard input/output for a personal computer or mobile phones .This includes the following:

- Keyboard
- Mouse
- Monitor
- Web browser

3.1.1 SOFTWARE INTERFACES:

The System make use of a Postgresql database and ruby on rails for the UI.

3.1.2 HARDWARE INTERFACES:

The System shall run on a browser file supporting system.

3.2 FUNCTIONAL REQUIREMENTS:

Functional requirements define the fundamental actions that the system must perform.

1)reading MS access database.

2)Counts the Interpretations of results.

- 2.1)Sensitive.
- 2.2)Intermediate.
- 2.3)Slightly Sensitive.
- 2.4)Resistant

3)Count the total number.

- 3.1)Culture Sample received.
- 3.2)Positive cases.

4)Counts Total number of samples

- 4.1)from ward
- 4.2) total positive cases.

5)Counts Organisms isolated from wards.

6)Reading infection data from OPD and

7)Count the total number of samples and positive cases.

8) Counts number of Organisms isolated from OPD

9)Reading Hospital Infection Data From ICUS,PICU,NICU.

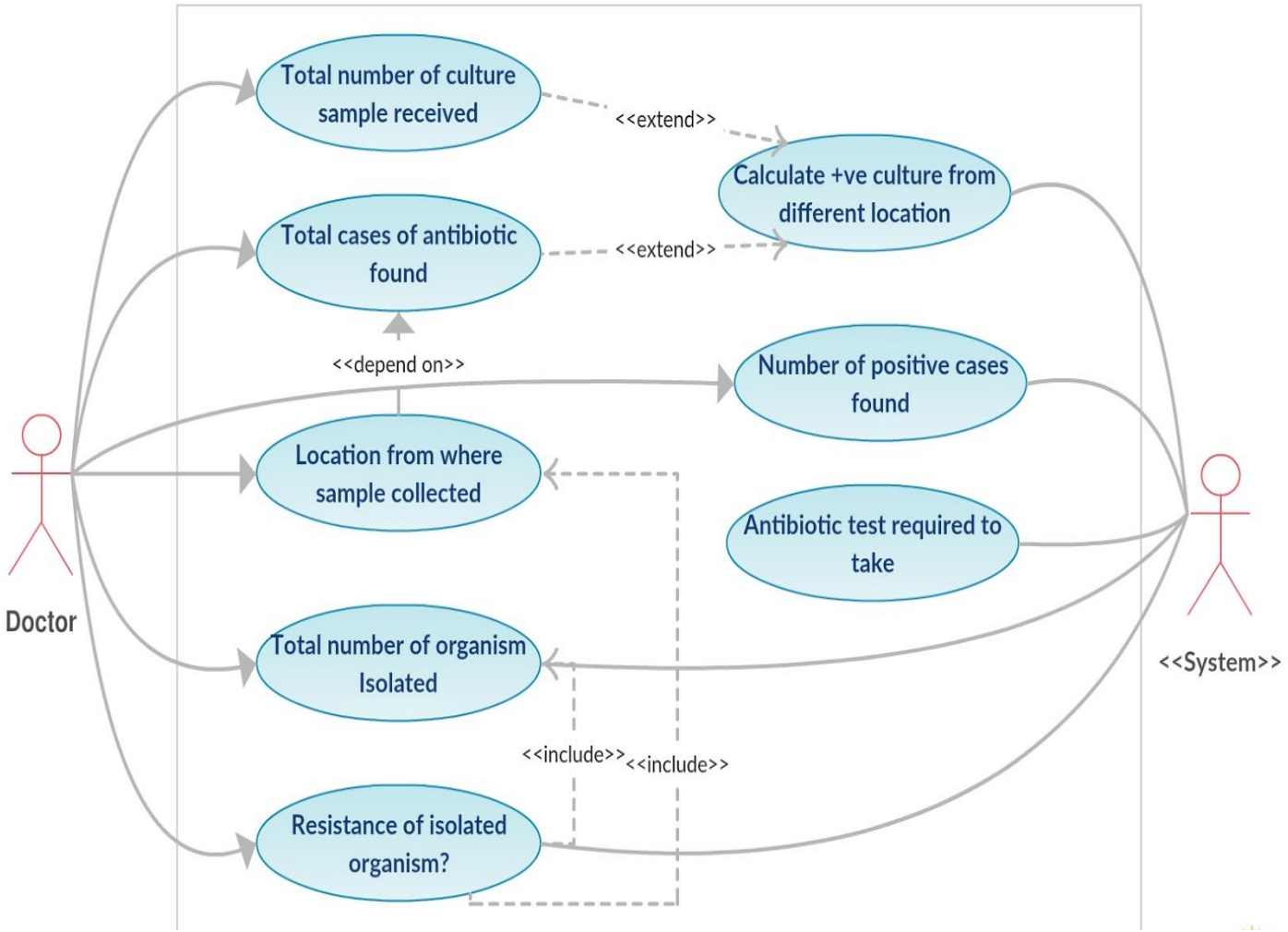
10)Counts Organisms isolated from ICU,cabins and NICU.

3.3 NON-FUNCTIONAL REQUIREMENTS:

Functional requirements define the needs in terms of performance, logical database requirements, design constraints, reliability, availability, security, maintainability and portability.

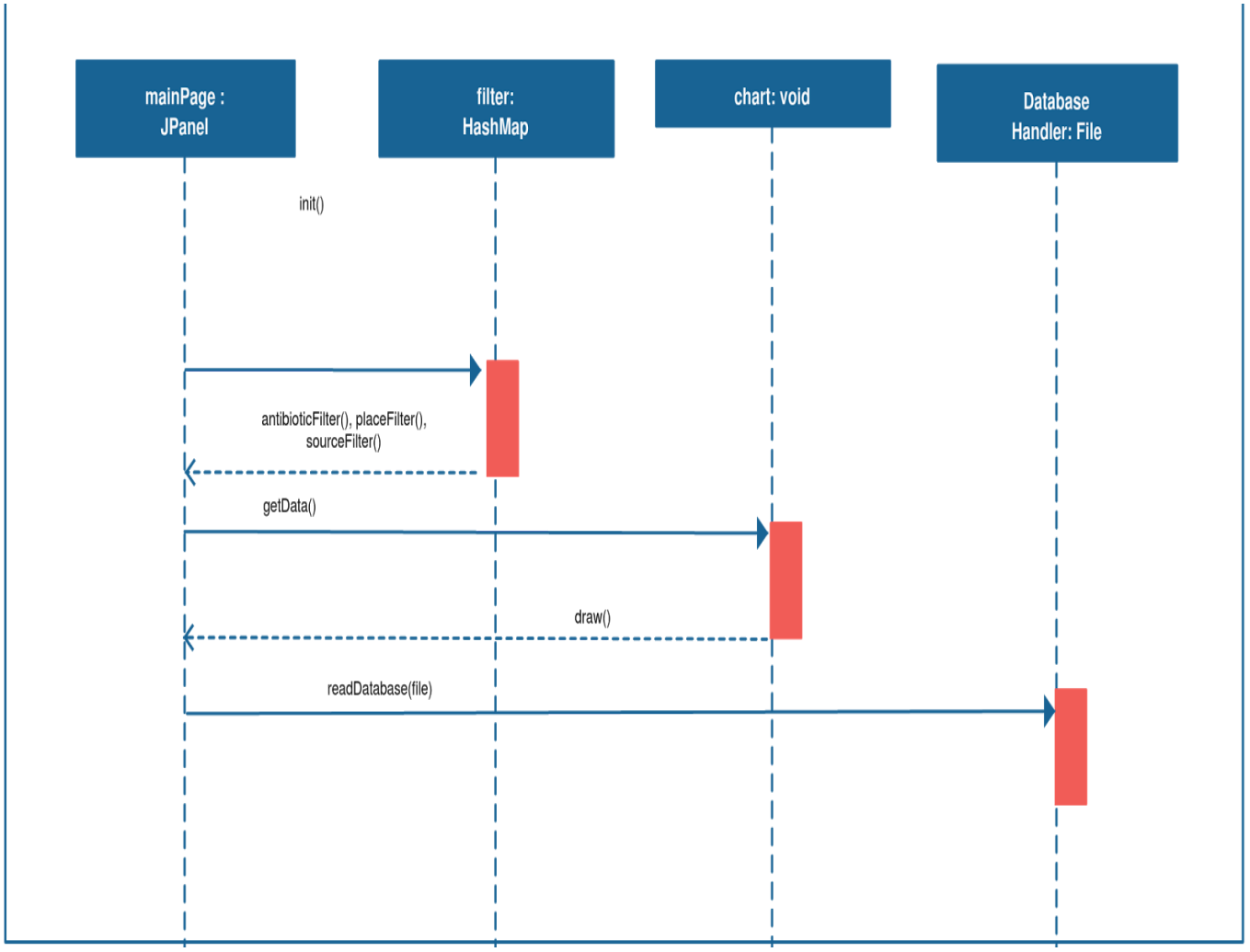
USE CASE DIAGRAM

A use case is a methodology used in system analysis to identify, clarify, and organize system requirements. In this context, the term "system" refers to something being developed or operated, such as a mail-order product sales and service Website. Use case diagrams are employed in UML (Unified Modeling Language), a standard notation for the modeling of real-world objects and systems.



SEQUENCE DIAGRAM

A **sequence diagram** shows object interactions arranged in time **sequence**. It depicts the objects and classes involved in the scenario and the **sequence** of messages exchanged between the objects needed to carry out the functionality of the scenario.



REFERENCES

1.Auffray, C., Balling, R., Barrosso, I., Bencze, L., Benson, M., Bergeron, J., … Bock, C. (2016, June 23). Making sense of big data in health research: Towards an EU action plan. US National Library of Medicine National Institute of Health, 8–71. http://dx.doi.org/10.1186/s13073-016-0323-y

2.Reddy, C. K., & Aggarwal, C. C. (2015). Healthcare Data Analytics (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series). Boca Raton, Florida: Chapman and Hall/CRC.

3.Proprietary software developed by World Health Organization for use by medical organization. http://www.whonet.org/software.html

4.Analysis of Medical Data Research Perspective Nancy B. Clark. M.Ed. Director of Medical Informatics Education FSU College of Medicine Spring 2004 http://www.med.fsu.edu/informatics