|  |  |  |
| --- | --- | --- |
|  | MANIPAL INSTITUTE OF TECHNOLOGY  (A constituent Institute of MANIPAL UNIVERSITY)  MANIPAL - 576 104, KARNATAKA, INDIA |  |

Industrial Training

on

Web Application Development using MongoDB

SUBMITTED

BY

Srivatsava Gummalla

Registration no.: 140905292

srivatsavagummalla@gmail.com

Phone: +91-9591311992

Under the Guidance of:

Pandarinath Siddineni

TABLE OF CONTENTS

[ABSTRACT 4](#_Toc488261496)

[1. PROBLEM DEFINITION 5](#_Toc488261497)

[1.1 Introduction 5](#_Toc488261498)

[1.2 Requirements 5](#_Toc488261499)

[1.3 FIND A DOCTOR WEB PORTAL 5](#_Toc488261500)

[1.3.1 NPPES NPI Database 5](#_Toc488261501)

[1.4 Disease code mapping (ICD9 to ICD10) 5](#_Toc488261502)

[1.4.1 ICD Codes 5](#_Toc488261503)

[1.4.2 ICD9 Vs ICD10 5](#_Toc488261504)

[2. STUDY — FULL STACK WEB DEVELOPMENT PLATFORMS 6](#_Toc488261505)

[1.5 Stages of full stack: 6](#_Toc488261506)

[1.5.1 Database: 6](#_Toc488261507)

[1.5.2 server-Side code: 6](#_Toc488261508)

[1.5.3 Client-Side code: 7](#_Toc488261509)

[1.5.4 Middleware: 7](#_Toc488261510)

[1.6 MEAN Stack: 7](#_Toc488261511)

[3. ARCHITECTURE AND DESIGN 8](#_Toc488261512)

[1.7 Development Platform: 8](#_Toc488261513)

[1.8 Bottle MVC framework: (Model-View-Controller): 9](#_Toc488261514)

[1.9 NoSQL Data Bases: 10](#_Toc488261515)

[1.10 Mongo Data Base: 10](#_Toc488261516)

[4. SOFTWARE DEVELOPMENT LIFE CYCLE 13](#_Toc488261517)

[5. DEVELOPMENT METHODOLOGY AND TOOLS 15](#_Toc488261518)

[6. RESULTS AND TESTING 17](#_Toc488261519)

[7. CONCLUSION 21](#_Toc488261520)

[8. REFERENCES 22](#_Toc488261521)

[APPENDIX-A 23](#_Toc488261522)

[APPENDIX-B 24](#_Toc488261523)

# ABSTRACT

*This report describes the work conducted during a three month internship at the Centre for Design at the RMIT University. This internship is part of the master’s curriculum of mechanical engineering.*

*I am Srivatsava Gummalla currently studying in the 7th semester of Computer Science and Engineering. This project is the part of my Internship program which was done in the company Informatics In Context Medical Pvt. Ltd under the guidance of Pandarinath Siddineni. The company develops healthcare software. The current project is a “Find A Doctor” web application. This software has been developed using the full stack programming approach developed using the agile model of software development. This application was designed to locate physicians at a given location with some required specialty.*

*An Excel tool has been developed to modify the SimaPro database using the COM interface. Using this tool, it takes less than half an hour to substitute all electricity processes and associated uncertainty parameters in an ecoinvent 2.2 database containing about 4000 products.*

*A MATLAB LCA tool has been developed which can perform LCA analysis and Monte Carlo simulations. The tool is compatible with the ecoinvent 2.2 database and can import several impact assessment methods. Performance of the tool is more than 100 times faster than SimaPro. Because of its efficiency, the tool can quickly calculate the coefficient of variation per impact category for all the products in an ecoinvent 2.2 database. It would take about 2-3 days calculation time, whereas it would take about one year to calculate the same results in SimaPro. The coefficient of variation could be useful to incorporate uncertainty in streamlined life cycle assessment tools. How this could be useful presented to the user needs to be further investigated.*

*Using generic geographical processes in a LCI database affects the uncertainty and reliability of a life cycle assessment. The coefficient of variation for carbon dioxide emissions for a product can almost double when switching between geographical electricity grids.*

*Some minor projects conducted are also described in this report. This includes development work on Greenfly. This is an online streamlined life cycle assessment tool to design environmentally improved products. It needs to be further developed before it can be released as a final version Land use change values for carbon stocks are calculated for the countries China, Europe, Japan, Mexico and New Zealand. The procedure for calculating these numbers has to be validated.*

*Another minor project was about exploring the capabilities and limitations of openLCA software. openLCA is compatible with the ecoinvent database and all features work well. The calculations are however quite time consuming compared to SimaPro and the software is not very responsive. openLCA is not compatible with the Australasian LCI database.*

# PROBLEM DEFINITION

## Introduction

In the healthcare department, the ICD codes are provided for the diseases and NPI codes are given for each physician. Based on the disease code and the location of the patient the nearby physician with a relevant specialization needs to be suggested. The ICD codes have two versions which are the ICD-9 and the ICD-10. The presently used codes are the ICD-10 but the older ICD-9 codes are still used at some places. A lookup system for the relevant ICD10 code for an ICD9 code is needed.

## Requirements

The project contains two main requirements, first being “developing find a doctor portal for patients and hospitals” and the second is to develop a web portal for ICD9 to ICD10 mapping.

## FIND A DOCTOR WEB PORTAL

Fill the details here:

Why end user need such portal?

Geography: US This application is intended for US customers

### NPPES NPI Database

## Disease code mapping (ICD9 to ICD10)

Fill the details here:

What is ICD codes?

Why ICD-9 to ICD-10 mapping is required?

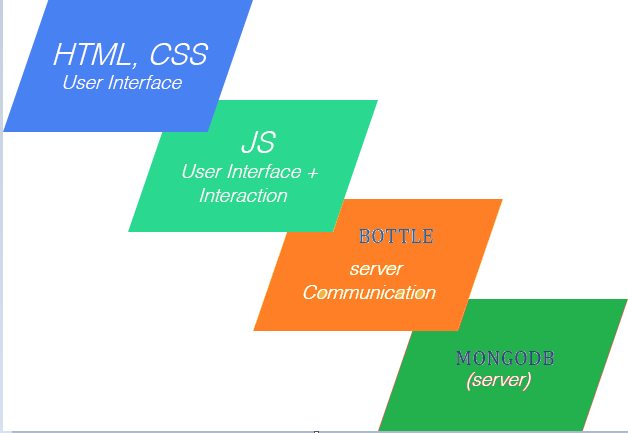
Geography: US This application is intended for US customers

### ICD Codes

### ICD9 Vs ICD10

# STUDY — FULL STACK WEB DEVELOPMENT PLATFORMS

Full stack means a collection of software used in developing a complete web application. It can be broadly classified into front-end and back-end development. The back-end development consists of a server, a database and an application to communicate with the database. The front-end development consists of web pages and client-side code which the user can directly interact with to use the application. There are various kinds of stacks available based on the programming languages that are chosen at the different stages in development. Some of the most common examples of full stack are LAMP, MEAN etc.



## Stages of full stack:

### Database:

Database is used to store all the data related to the service being provided by the application. The data stored in a database needs to be available at all times and it should be able to maintain the consistency of files. In the current project, a NoSQL database was chosen as it emphasizes on horizontal scalability. The database used is MongoDB. This database is a part of MEAN stack. It is easy to use as less restrictions are provided on data storage. The mongod server is used to provide access to the mongo database.

### server-Side code:

The server-side code is the code used to build the backend software of the website. These codes, which are also called scripts, are designed to run on a server and interact with permanent storages like databases. It facilitates the transfer of data and also powers functions in dynamic web applications. It is also used to build the application programming interfaces(APIs). Some of the server-side scripts are PHP, Ruby, Python etc. In this project, the Python’s Bottle MVC framework has been used to develop the server-side scripts.

### Client-Side code:

This code is the code which is used to develop the webpages and is part of the front-end development. It is used to generate the web pages and provide user interaction and interface. The HTML, CSS languages are used to provide the interface and develop static webpages. The JavaScript is used to provide interaction for the user and make the web page dynamic.

### Middleware:

Middleware is computer software that connects software components or applications. It is present along with the server-side scripts in a web application and provides enhanced performance. The software consists of a set of services that allows multiple processes running on one or more machines to interact. The common middleware services are CORBA, DCE etc. In the current project, no middleware has been used.

## MEAN Stack:

One of the full stack technologies which is currently gaining popularity is MEAN stack. This stack uses MongoDB, Express.js, Angular.js and Node.js.



In MEAN stack the MongoDB is a NoSQL database and uses JSON style documents for data representation. The Express.js provides the server framework for web applications. Angular.js is a front-end java script framework to develop the client-side architecture. Node.js is a concurrent JavaScript environment for building scalable and fast web applications. The benefits of using MEAN stack is that the entire code can be written using Java Script. It supports the Model View Controller(MVC) framework. The other advantage is the use of JSON to transfer the data and the huge module library of Node.js.

# ARCHITECTURE AND DESIGN

## Development Platform:

1. Python

Python is an easy to learn and easy to use language for the development of software. It offers high readability and efficiency for web development. The python version 3.6.0 has been used in this project. Python also has a lot of support from the community which was helpful for quickly solving the issues encountered when programming.

1. Bottle

Bottle is a fast, simple and lightweight WSGI micro web-framework for Python. Bottle version 0.12.13 has been used for the server-side scripting in the project. It is used to host the server. It is used in routing and in creating templates. It has a built in HTTP development server and has convenient access to form data, file uploads, cookies, headers and other HTTP-related metadata.

1. MongoDB

MongoDB is a NoSQL and document oriented database. Data is stored in the form of JSON documents in the database. The MongoDB version 3.4.0 has been used in this project. It provides the database for storing documents and a server for accessing the database. This database is used in storing the NPI physician data and the ICD codes. These data are first imported to the database and then is made available through the server.

1. MongoEngine and PyMongo

PyMongo is a Python distribution containing tools for working with MongoDB, and is the recommended way to work with MongoDB from Python. MongoEngine is a document object mapper for working with MongoDB. These software are used in establishing connection to the database and to write queries to the database. The model of the collection that is used to store the records in the database is also designed using MongoEngine.

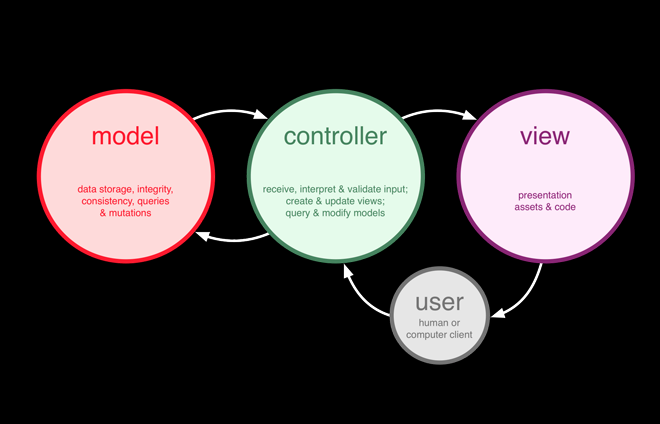
1. HTML, CSS

The HTML is the standard markup language for creating web pages. It is used to design static web pages. CSS is a language that describes the style of an HTML document. In the front-end development, the client-side code has been written using HTML and CSS as the base.

1. JavaScript

JavaScript is a high-level, dynamic, weakly typed, object-based, multi-paradigm, and interpreted client-side programming language. This is used alongside HTML and CSS in creating dynamic and interactive webpages. The jQuery using Ajax has also been implemented in this project. As the Mongo database stores data in JSON format dealing with these objects is easier when using JavaScript.

## Bottle MVC framework: (Model-View-Controller):



The Model-View-Controller (MVC) architectural pattern separates an application into three main components: the model, the view, and the controller. This pattern has been used for the development of the product. The Model component corresponds to all the data-related logic that the user works with. In this project, all the collections present in the database have been created as models using MongoEngine and have been titled starting with the word model. These models are used when the data is being imported to the database. The View component is used for all the user interface logic of the application. It is these views that are visible to the user. The template files of bottle have been created as views and have been titled starting with the word view. These views are written in HTML, CSS and JS. Controllers act as an interface between Model and View components to process all the business logic and incoming requests, manipulate data using the Model component and interact with the Views to render the final output. The controller files in the project have been written using python’s bottle framework. These files have been named starting with the word cnt.

Advantages of using an MVC framework:

1. Faster development process
2. Ability to provide multiple views
3. Modification does not affect the entire model
4. Support for asynchronous technique

## NoSQL Data Bases:

NoSQL database provides a mechanism for the storage and retrieval of data that is modeled in means other than the tabular relations.

NoSQL Database Types:

* Document databases pair each key with a complex data structure known as a document. Documents can contain many different key-value pairs, or key-array pairs, or even nested documents.
* Graph stores are used to store information about networks of data, such as social connections. Graph stores include Neo4J and Giraph.
* Key-value stores are the simplest NoSQL databases. Every single item in the database is stored as an attribute name (or 'key'), together with its value. Examples of key-value stores are Riak and Berkeley DB. Some key-value stores, such as Redis, allow each value to have a type, such as 'integer', which adds functionality.
* Wide-column stores such as Cassandra and HBase are optimized for queries over large datasets, and store columns of data together, instead of rows.

The Benefits of NoSQL:

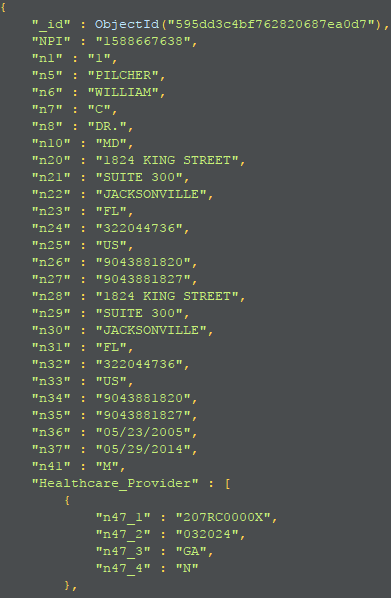
When compared to relational databases, NoSQL databases are more scalable and provide superior performance, and their data model addresses several issues that the relational model is not designed to address:

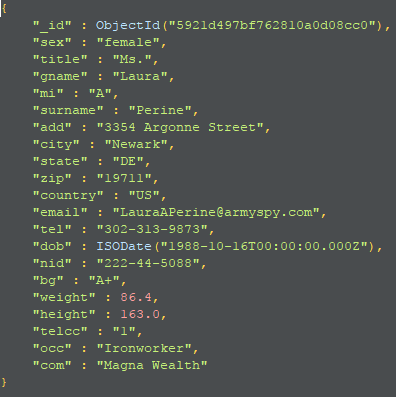
* Large volumes of rapidly changing structured, semi-structured, and unstructured data
* Agile sprints, quick schema iteration, and frequent code pushes
* Object-oriented programming that is easy to use and flexible
* Geographically distributed scale-out architecture instead of expensive, monolithic architecture

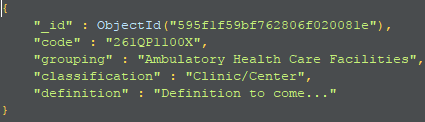
## Mongo Data Base:

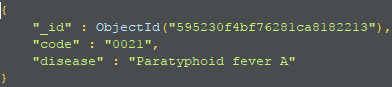
MongoDB is a document database with the scalability and flexibility that is required for the querying and indexing. MongoDB stores data in flexible, JSON like documents. JSON means Java Script Object Notation. It is a lightweight data-interchange format that is easy for humans to read and write. JSON has two structures which are a collection of name/value pairs and an ordered list of values. As MongoDB uses JSON objects, the data stored is easily readable and is represented in key/value pairs. MongoDB is a distributed database at its core, so high availability, horizontal scaling, and geographic distribution are built in and easy to use. In MongoDB, index can be made on any attribute and it increases the speed of queries using that attribute.

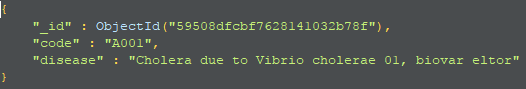
In a single MongoDB server, there can be multiple databases. A database is a physical container for collections, with its own set of files in the file system. A collection is a group of MongoDB documents. Documents within a collection can have different fields as it does not enforce any schema. All documents of a related purpose are generally stored in a single collection. In this project, the database was named as ClinicalData. In it collections for the physician data (NPI data), ICD-9 codes, ICD-10 codes and patient data (randomly generated names and information) were made. Data import files were written with command line parameters to import these collections to the database. The data models for the collections were written in MongoEngine and have been used to form the key/value pairs and insert each document to database. MongoDB is horizontally scalable and the capacity can be increased by connecting multiple hardware or software entities so that they work as a single logical unit. Because of this MongoDB is suitable for handling large amounts of data spread across multiple servers. The physician collection in the ClinicalData database consists of more than 4.8 million records. This collection can be handled with ease when using the MongoDB.

Sample records from the various collections in the database:



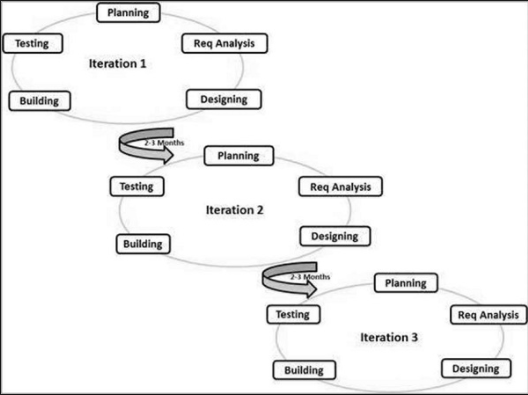




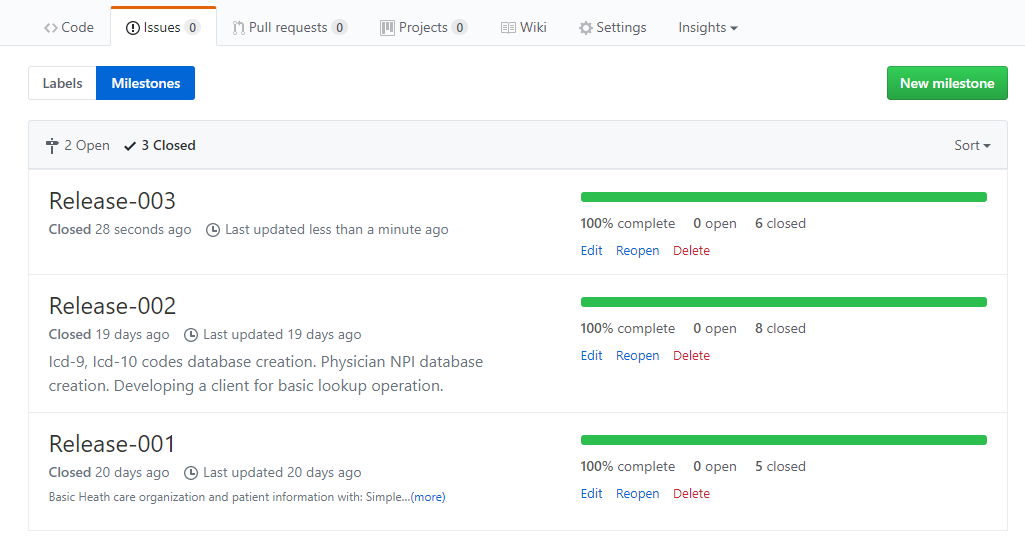


# SOFTWARE DEVELOPMENT LIFE CYCLE

Software Development Life Cycle (SDLC) is a process used by the software industry to design, develop and test high quality software. It is composed of a number of clearly defined and distinct work phases which are used by systems engineers and systems developers to plan for, design, build, test, and deliver information systems. There are various SDLC models that are used in the development of software. The Agile software development model was used to develop this software product.



In the agile model, the product is developed in small incremental builds. These builds are called sprints and at the end of each sprint a working product is developed. A major task is broken down into many minor tasks and these tasks are completed iteratively. This model focuses on adaptability and customer satisfaction by rapid delivery of working software product. Scrum framework has been used to manage the agile model of software development. Task for each sprint is assigned by the scrum master and the product has been developed in four sprints. At the end of each sprint functional code has been released. Each sprint in the development consists of planning, requirement specification/analysis, design, coding and testing. The planning stage is used for cost/effort estimation. A project plan is developed which also assigns the time required for other stages in development. The requirement analysis stage is used to gather all the requirements that are expected from the software product. 30% of the time in the sprint is assigned to this phase. System requirement specification document is formed in this stage. In the design phase the low level design, high level design, the database design are decided. 20% of the time is allocated for this phase. The source files for the product based on the design given in the design phase. The code written needs to follow the coding guidelines which have been agreed upon prior to the development of the software. 30% of the time is allocated for this phase. In the testing phase, the test cases are written and these test cases are automated using a script. The unit testing and acceptance testing is carried out on each module and integration testing is carried out when two modules are integrated. These tests are carried out based on the system requirement specification document developed in the requirement phase. At the end of each sprint the product developed is analyzed by the scrum master and the next sprint is decided. At the end of development of the software product system testing is done and the final source code and documents are released. This method of development has been followed in building the “Find A Doctor” web application. The GitHub issue tracking and management tool was used in assigning the tasks and the deadline for each sprint.



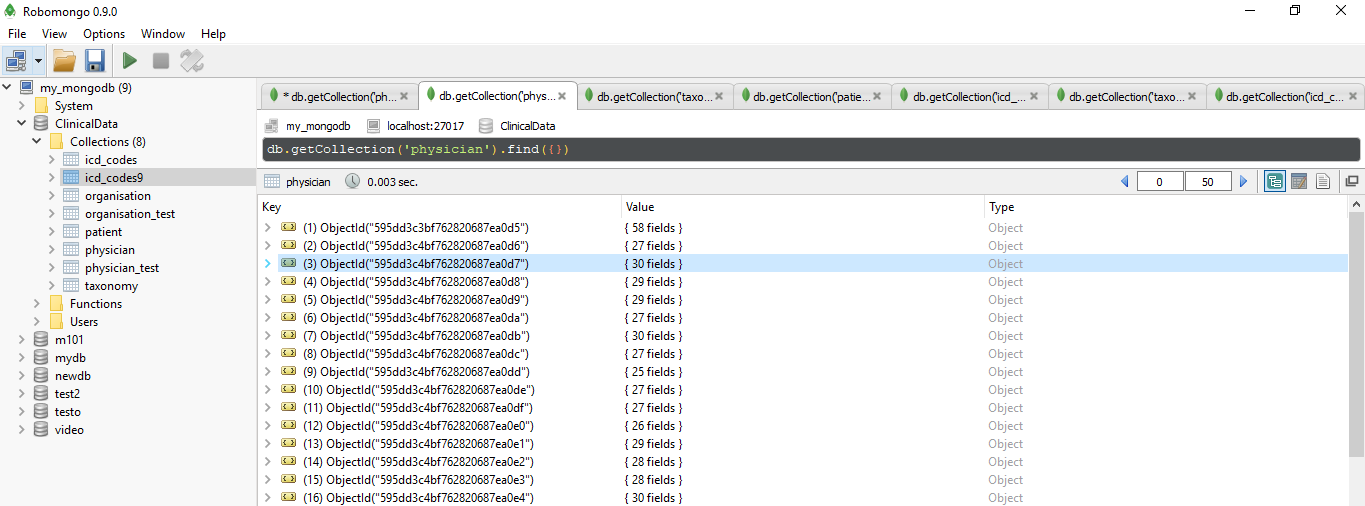
# DEVELOPMENT METHODOLOGY AND TOOLS

1. GitHub: Version Control and Configuration Management

GitHub is a web based version control repository and Internet hosting service. It provides access control, source code management and is used to host open source software products. It offers public and private repositories. All the versions of the code at each commit are preserved from the time the code was added to the repository. In case of errors or bug tracking, the previous versions of the code can be accessed and retrieved. The URL for the current GIT project is <https://github.com/Srivatsava96/Internship>.

1. RoboMongo, Version-0.9.0: DB administration and Visualization

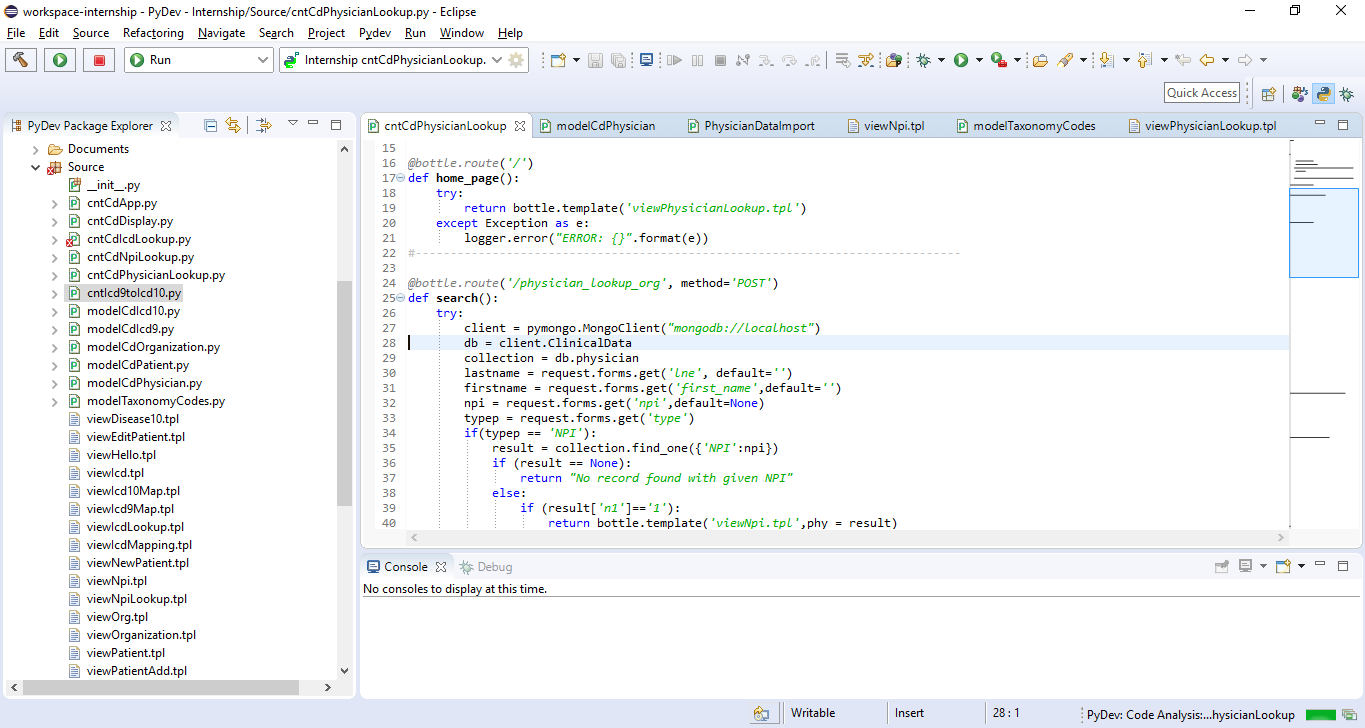
RoboMongo is a MongoDB management tool. We can use it to connect to the mongo server and view the database. We can run queries on the collections present in the database. The documents present in the collection can be visualized in the JSON representation. We can add, modify and drop collections present in the database. This tool is used to check whether the data has been correctly imported to the database.





1. Eclipse IDE Version: Neon.3 Release (4.6.3):

Eclipse is an integrated development environment(IDE) used in programming. An IDE is a software application that provides comprehensive facilities to computer programmers for software development. This project, has used the Python programming language which was written in the Eclipse IDE. Eclipse IDE offers features such as code completion, resource management, debugging tools, compile and build the object files. These features are highly useful in developing the software. Navigation through the file system is made easier. All the files can be viewed and managed on the same screen.

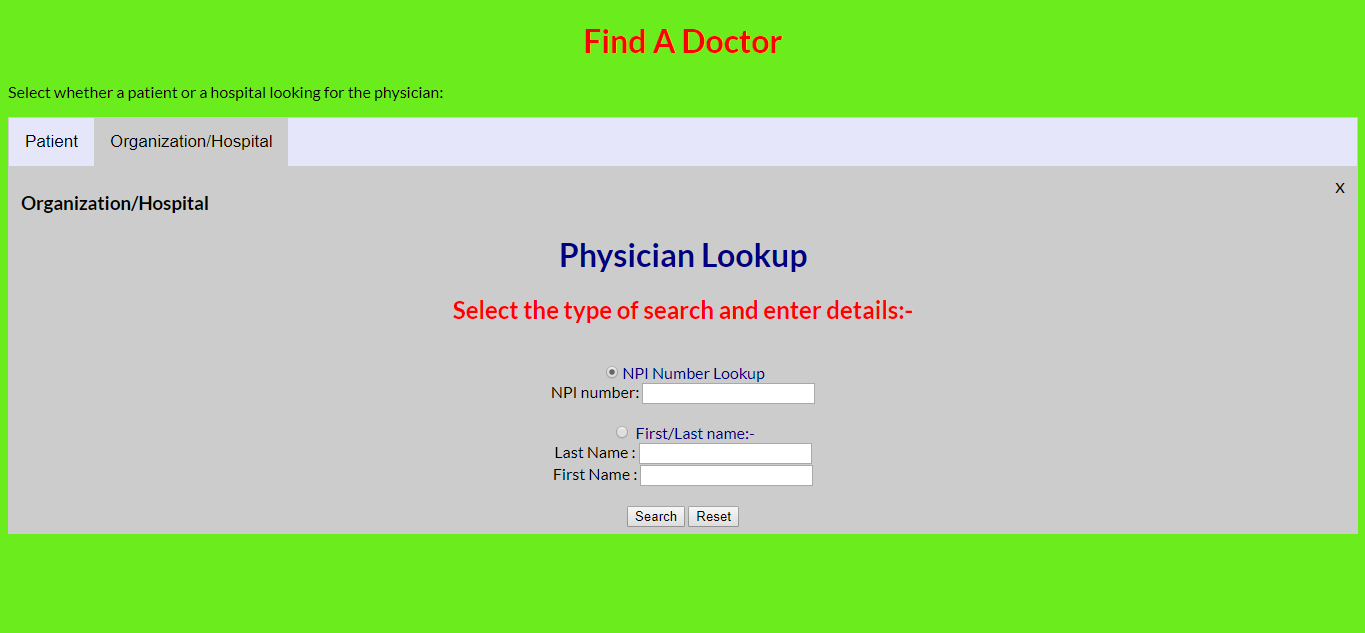


1. Task Assignment and Issue Tracking: Git Issue:

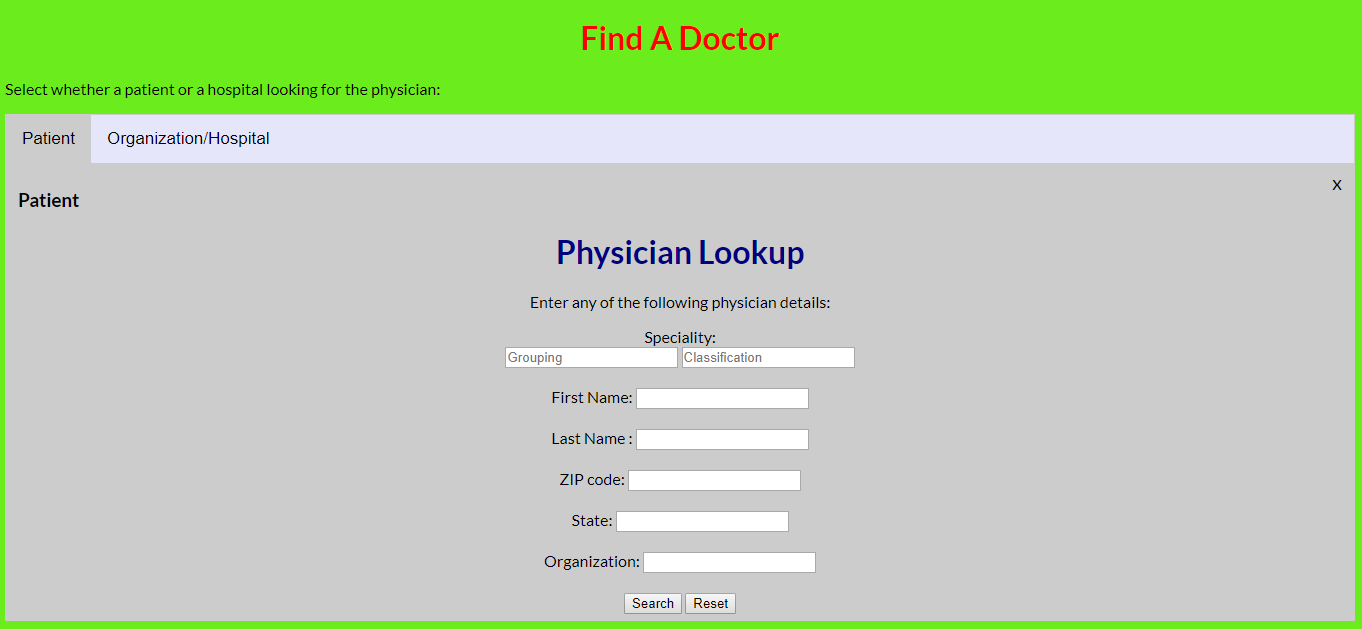
The entire project has been broken down into small tasks and each task has been created as an issue in GitHub issues. These issues have been given the label task and the label verify is added when the task is done. The commits to the source code can be associated with the issues in Git. These commits are then verified and the issue is closed. Comments for each issue can be written by any of the collaborator for the project which can be used to know the status of the issue between the collaborators. Each issue can be associated with a milestone. These milestones can be given the deadlines and they are closed when all the issues pertaining to the milestone have been closed. These milestones can be used to keep track of the sprints in the agile model of development.

# RESULTS AND TESTING

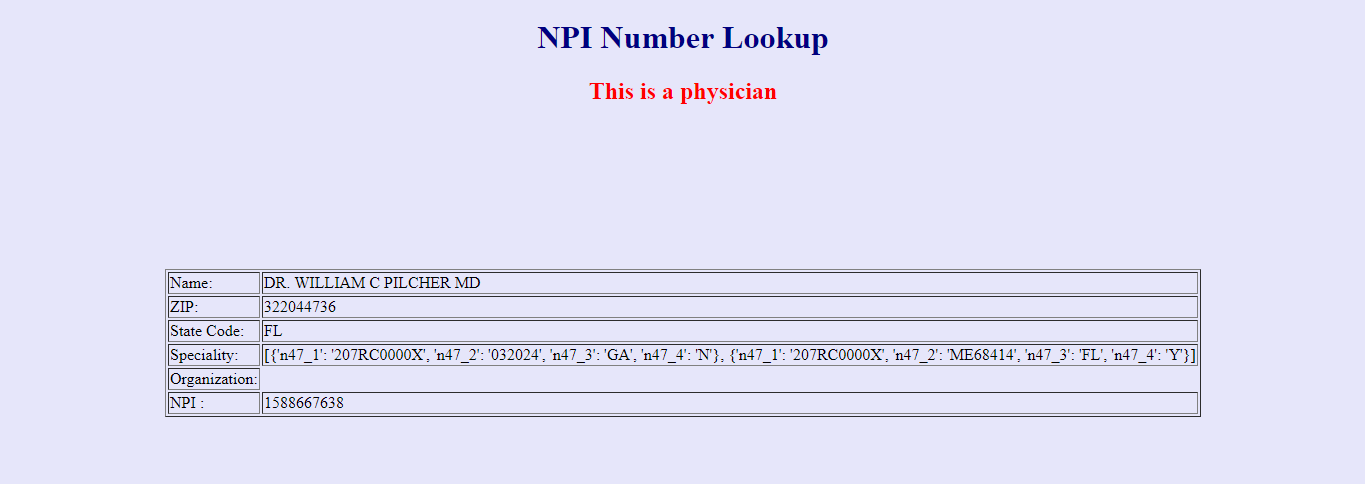
1.Physician Search for organizations based on the NPI number or the first/last name



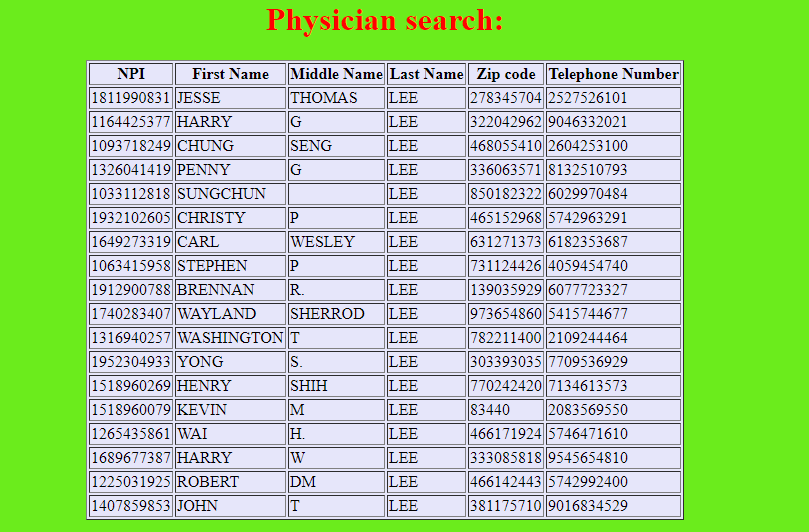
2.Physician Search for individuals based on specialty and location



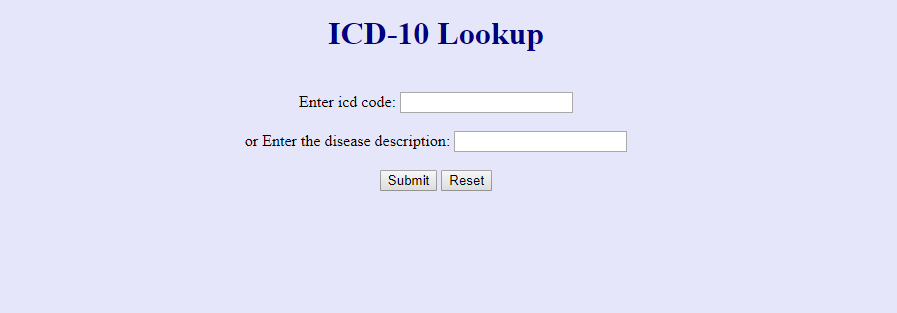
3. Display of physician details based on NPI number

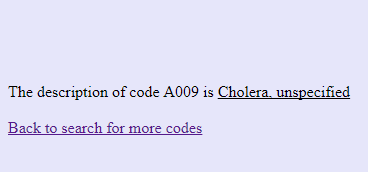


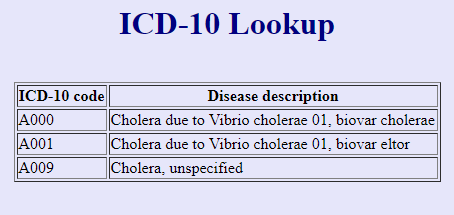
4.Physician search given first name or last name



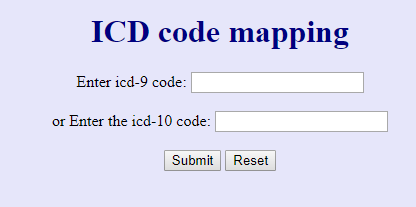
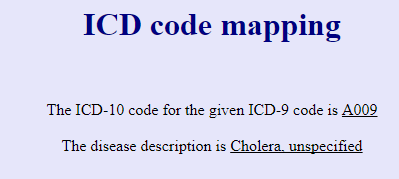
5.ICD code lookup:







6.ICD9 to ICD10 mapping:



# CONCLUSION

Healthcare is an important service in the modern human society. To bring the services of the physicians to the patient is the purpose of this project.

# REFERENCES

1. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
2. I.S. Jacobs and C.P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.
3. K. Elissa, “Title of paper if known,” unpublished.
4. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
5. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
6. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

# APPENDIX-A

# APPENDIX-B

**INFORMATION**

**(Regarding Industrial training)**

**IV year B.Tech. (CSE)**

1. Industrial training could be training, internship, to be completed before 7th semester for a minimum period of 4 weeks
2. Certificate from the organization citing the date of commencement of the training and the end date (minimum of 4 weeks) is essential.
3. The certificate should preferably contain the word **“industrial training”** in it.
4. It should be attested by the organization head or any concerned authority.

The following information should be present in the industrial training report.

1. Front page
2. Copy of the Certificate from the company in the company letterhead
3. Abstract ( 1 page)
4. Details of the organization (1-3 pages)
5. Information acquired during the study period (10-20 pages)
6. Conclusion (1 page)
7. References in IEEE format

Font size 12, spacing 1.5 with proper justification.

Industrial training coordinator H.O. D ( Dept. of CSE)

Shanthi P.B