Multi-Source Medical Data Integration and Mining for Healthcare Services

Shaik Rasheed[1], B. Murali[2]
[1]*PG Student, CSE, Quba College of Engineering & Technology*
[2]*Associate professor, CSE, Quba College of Engineering & Technology*

***Abstract*— In this project the advent of Internet of Health (IoH) age, traditional medical or healthy services are gradually migrating to the Web or Internet and have been producing a considerable amount of medical data associated with patients, doctors, medicine, medical infrastructure and so on. Effective fusion and analyses of these IoH data are of positive significances for the scientific disaster diagnosis and medical care services. However, IoH data are often distributed across different departments and contain partial user privacy. Therefore, it is often a challenging task to effectively integrate or mine the sensitive IoH data, during which user privacy is not disclosed. So, to overcome this difficulty, we introduce IOH, where the IOH can upload medicines to the particular patient without leaking the patients data and the admin can view the medicines after he got an acceptance from IOH. Patients also can view doctors and they can consult the doctors only after the acceptance from admin and doctor.**

***Index Terms*— Data Fusion, Distributed Data Management, Privacy-Preserving Data Mining, Access Control Mechanisms, Secure Data Upload/Download**

I. INTRODUCTION

Generally, most of historical IoH data records contain valuable information especially for the medical or healthy agencies, such as the past disease of a patient at a time point. Mining and analyzing such historical IoH data records can contribute much to doctors' scientific and reasonable diagnosis and treatment decision-makings, as well as disaster trend prediction and precaution. Therefore, it is of emergent necessity to collect, integrate, fuse and analyze these multi-source IoH data records for high-quality healthcare services suitable for patients.

However, historical IoH data records from patients often contain sensitive patient privacy (e.g., blood pressure, temperature) as a patient is often not willing to let others know his or her historical disasters. Therefore, the patients or the stakeholders of historical IoH data records dare not disclose their IoH data records to the public. In addition, they lack sufficient incentive for IoH data records sharing with others The above two concerns significantly block the utilization of historical IoH data records. As a consequence, although many hospitals or other medical & healthy agencies have accumulated a considerable amount of historical IoH data records, they seldom release the data to the outside due to privacy concerns. Furthermore, the historical IoH data records are often distributed across different platforms or agencies, the integration and fusion of which further increases the privacy disclosure concerns. Considering all the issues, we have proposed a new method to solve the issues. Here IOH will IOH can upload medicines and they will be stored in the encryption form using AES algorithm and to the particular patient without leaking the patients data and the admin can view the medicines after he got an acceptance from IOH. Patients also can view doctors and they can consult the doctors only after the acceptance from admin and doctor.

II. LITERATURE SURVEY

2.1 INTRODUCTION:

[1] F. Casino, J. Domingo-Ferrer, C. Patsakis, D. Puig, and A. Solanas, ''A k-anonymous approach to privacy preserving collaborative filtering,'' J. Comput. Syst. Sci., vol. 81, no. 6, pp. 1000–1011, Sep. 2015.

Collaborative Filtering (CF) is a recommender system which is becoming increasingly relevant for the industry. Current research focuses on Privacy Preserving Collaborative Filtering (PPCF), whose aim is to solve the privacy issues raised by the systematic collection of private information. In this, we propose a new micro aggregation-based PPCF method that

distorts data to provide k-anonymity, whilst simultaneously making accurate recommendations. Experimental results demonstrate that the proposed method perturbs data more efficiently than the well-known and widely used distortion method based on Gaussian noise addition.

Summary: From this, I have learned that, PPCF provides accurate recommendations estimated from perturbed data whilst guaranteeing user *k*-anonymity.

[2] Z. Xia, X. Wang, X. Sun, and Q. Wang, ‘‘A secure and dynamic multikeyword ranked search scheme over encrypted cloud data,’’ IEEE Trans. Parallel Distrib. Syst., vol. 27, no. 2, pp. 340–352, Feb. 2016.

Due to the increasing popularity of cloud computing, more and more data owners are motivated to outsource their data to cloud servers for great convenience and reduced cost in data management. However, sensitive data should be encrypted before outsourcing for privacy requirements, which obsoletes data utilization like keyword-based document retrieval. In this, we present a secure multi-keyword ranked search scheme over encrypted cloud data, which simultaneously supports dynamic update operations like deletion and insertion of documents. Specifically, the vector space model and the widely-used TF x IDF model are combined in the index construction and query generation. We construct a special tree-based index structure and propose a “Greedy Depth-first Search” algorithm to provide efficient multi-keyword ranked search. The secure kNN algorithm is utilized to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vectors. In order to resist statistical attacks, phantom terms are added to the index vector for blinding search results.

III. SYSTEM ANALYSIS

3.1 EXISTING METHOD

In existing we will update the suggested PDFM method by considering the possible diversity of data types and data structure. In addition, how to fuse multiple existing privacy solution for better performances is still an open problem that requires intensive and continuous study.

DRAWBACKS:

- Less accuracy.
- Low performance.
- Less flexibility.

PROPOSED METHOD

In our proposed system, we can secure all the outsourcing data that is transferring across all the fields. We introduce IOH, where the IOH can upload medicines to the particular patient without leaking the patients data and the admin can view the medicines after he got an acceptance from IOH. Patients also can view doctors and they can consult the doctors only after the acceptance from admin and doctor.

Advantages:

- High accuracy.
- More performance.
- High flexibility.

3.3.SYSTEM REQUIREMENTS

3.3.1.HARDWARE

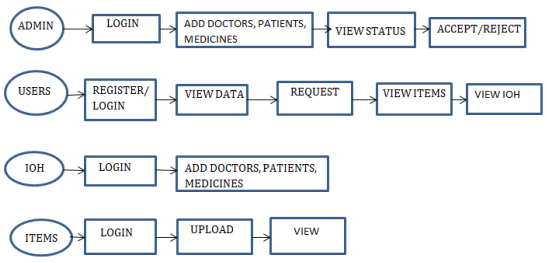
Hardware System Configuration:-

- Processor - I3/Intel Processor
- RAM - 4GB (min)
- Hard Disk - 160GB

Software System Configuration:-

- Operating System : Windows 7/8/10
- Application Server : Tomcat 9.0
- Front End : HTML, JSP
- Scripts : JavaScript.
- Server side Script : Java Server Pages.
- Database : My SQL 6.0
- Database Connectivity : JDBC.

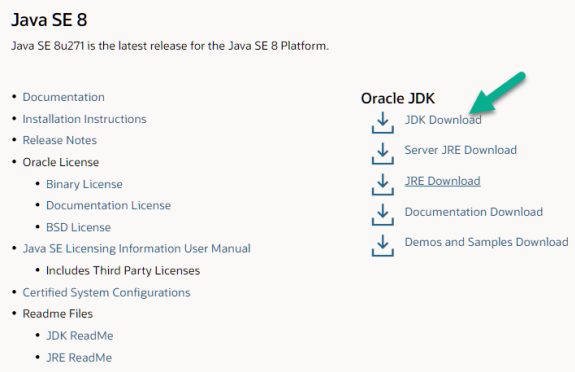
IV. SYSTEM ARCHITECTURE



V. SOFTWARE INSTALLATION FOR JAVA PROJECTS

This Java Development Kit(JDK) allows you to code and run Java programs. It's possible that you install

multiple JDK versions on the same PC. But Its recommended that you install only latest version.
How to install Java for Windows
Following are the steps for JDK 8 free download for 32 bit or JDK 8 download 64 bit and installation
Step 1) Go to link. Click on JDK Download for Java



VI. SYSTEM ANALYSIS

MODULES:
1. Admin
Login:
Admin can login with his credentials.
View Doctors:
Admin can view all the doctors.
Add Doctors:
Admin can add the doctors.
View patients:
Admin can view patients.
Add patients:
Admin can add patients.
View medicines:
Admin can view all the medicines added by IOH and can send a request to IOH.
View status:
Admin can view the status whether IOH accepted/rejected the request. If accepted, admin can view such medicines.
View request:
Admin can view the request sent by the patients and admin can inform to doctor for an appointment.
View status:
Admin can view the status whether the doctor accepted/rejected the request. If accepted, admin can intimate to the patient about the doctor's appointment.
2. Doctor:
Register/login:
Doctor can register by entering his details or if he already a registered candidate he can login with his valid credentials.
View appointments:
Doctors can view all the requests and he can generate a date for appointment by accepting the request or if doctor is unavailable doctor can reject the request.

3. Patients:
Register/login:
Patients can register by entering his details or if he already a registered candidate he can login with his valid credentials.
View doctors:
Patients can view doctors and they can send the request if they want an appointment to admin.
View status:
Patients can view the status of their request and they can check whether their request is accepted or not and the appointment date.
4. IOH:
Login:
IOH can login with their valid credentials.
View:
IOH can view all the patient's diseases.
Upload medicines:
IOH can upload the medicines based on patient's diseases.
View request:
IOH can view the request sent by the admin and the IOH can accept/reject the request.

VII. SYSTEM DESIGN

UML DIAGRAMS
UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.
The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.
The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as

well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

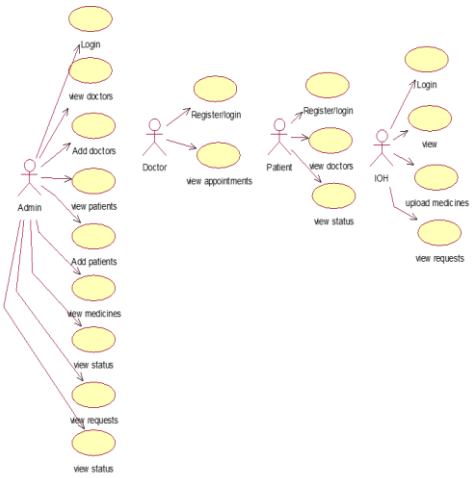
GOALS:

The Primary goals in the design of the UML are as follows:

1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.

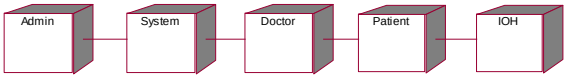
USE CASE DIAGRAM:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted



DEPLOYMENT DIAGRAM:

Deployment diagram represents the deployment view of a system. It is related to the component diagram. Because the components are deployed using the deployment diagrams. A deployment diagram consists of nodes. Nodes are nothing but physical hardware's used to deploy the application.



ACTIVITY DIAGRAM:

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

VIII SYSTEM TESTING

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.
Output : identified classes of application outputs must be exercised.

IX SCREENSHOTS

Home page:

Admin login page:

X.CONCLUSION

Effective fusion and analyses of IoH data are of positive significances for scientific disaster diagnosis and medical care services. However, the IoH data produced by patients are often distributed across different departments and contain partial patient privacy. Therefore, it is often a challenging task to effectively integrate or mine the sensitive IoH data without disclosing patient privacy. To tackle this challenge, we bring forth a novel multi-source medical data integration and mining solution for better healthcare services, named PDFM. Through PDFM, we can search for similar medical records in a time-efficient and privacy-preserving manner, so as to provision patients with better medical and health services. The experiments on a real dataset prove the feasibility of PDFM.

REFERENCES

[1] S. Din and A. Paul, ‘‘Smart health monitoring and management system: Toward autonomous wearable sensing for Internet of Things using big data analytics,’’ Future Gener. Comput. Syst., vol. 111, p. 939, Feb. 2020.

[2] N. C. Benda, T. C. Veinot, C. J. Sieck, and J. S. Ancker, ‘‘Broadband Internet access is a social determinant of health!,’’ Amer. J. Public Health, vol. 110, no. 8, pp. 1123–1125, Aug. 2020.

[3] E. Sillence, J. M. Blythe, P. Briggs, and M. Moss, ‘‘A revised model of trust in Internet-based health information and advice: Cross-sectional questionnaire study,’’ J. Med. Internet Res., vol. 21, no. 11, Nov. 2019, Art. no. e11125.

[4] K. Szulc and M. Duplaga, ‘‘The impact of Internet use on mental wellbeing and health behaviours among persons with disability,’’ Eur. J. Public Health, vol. 29, no. 4, pp. 185–425, Nov. 2019.

[5] T. Peng, Y. Lin, X. Yao, and W. Zhang, ‘‘An efficient ranked multi-keyword search for multiple data owners over encrypted cloud data,’’ IEEE Access, vol. 6, pp. 21924–21933, 2018.

[6] H. Dai, Y. Ji, G. Yang, H. Huang, and X. Yi, ‘‘A privacy-preserving multi-keyword ranked search over encrypted data in hybrid clouds,’’ IEEE Access, vol. 8, pp. 4895–4907, 2020.

[7] T. V. Xuan Phuong, G. Yang, W. Susilo, F. Guo, and Q. Huang, ‘‘Sequence aware functional encryption and its application in searchable encryption,’’ J. Inf. Secur. Appl., vol. 35, pp. 106–118, Aug. 2017.