PRIVACY CHARACTERIZATION AND QUANTIFICATION IN DATA PUBLISHING

B S Murthy [1], G.Mani Kanta Swamy,

[1]**Assistant professor , PG DEPT,** Dantuluri Narayana Raju College**, Bhimavaram, Andharapradesh**
Email:-suryanarayanamurthy.b@gmail.com
[2]**PG Student of MCA, Dantuluri** Narayana Raju College**, Bhimavaram, Andharapradesh**
Email:-gmanikanta2458@gmail.com

ABSTRACT

The increasing interest in collecting and publishing large amounts of individuals' data to public for purposes such as medical research, market analysis and economical measures has created major privacy concerns about individual' s sensitive information. To deal with these concerns, many Privacy-Preserving Data Publishing (PPDP) techniques have been proposed in literature. However, they lack a proper privacy characterization and measurement. In this paper, we first present a novel multi-variable privacy characterization and quantification model. Based on this model, we are able to analyze the prior and posterior adversarial belief about attribute values of individuals. We can also analyze the sensitivity of any identifier in privacy characterization. Then we show that privacy should not be measured based on one metric. We demonstrate how this could result in privacy misjudgment. We propose two different metrics for quantification of privacy leakage, distribution leakage and entropy leakage. Using these metrics, we analyzed some of the most well-known PPDP techniques such as k-anonymity, l-diversity and t-closeness. Based on our framework and the proposed metrics, we can determine that all the existing PPDP schemes have limitations in privacy characterization. Our proposed privacy characterization and measurement framework contributes to better understanding and evaluation of these techniques. Thus, this paper provides a foundation for design and analysis of PPDP schemes.

1 INTRODUCTION

Nowadays, datasets are considered a valuable source of information for the medical research, market analysis and economical measures. These datasets can include information about individuals that contain social, medical, statistical, and customer data. Many organizations, companies and institutions publish privacy related datasets. While the shared dataset gives useful societal information to researchers, it also creates security risks and privacy concerns to the individuals whose data are in the table. To avoid possible identification of individuals from records in published data, uniquely identifying information such as names and socialsecurity numbers are generally removed from the table.

Literature Survey

K-anonymity: a model for protecting privacy

The solution provided in this paper includes a formal protection model named k-anonymity and a set of accompanying policies for deployment. A release provides k-anonymity protection if the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appears in the release. This paper also examines re-identification attacks that can be realized on releases that adhere to k-anonymity unless accompanying policies are respected. The k-anonymity protection model is important because it forms the basis on which the real-world systems known as Datafly, μ-Argus and k-Similar provide guarantees of privacy protection.

3 IMPLEMENTATION STUDY

EXISTING SYSTEM:

The spate of privacy related incidents has spurred a long line of research in privacy notions for data publishing and analysis, such as k-anonymity, l-diversity and t-closeness, to name a few . A table satisfies k-anonymity if each quasi-identifier attribute in the table is indistinguishable from at least $k - 1$ other quasi-identifier attributes; such a table is called a k-anonymous table.

Disadvantages:

While k-anonymity protects identity disclosure of individuals by linking attacks, it is insufficient to prevent attribute disclosure with side information. By combining the released data with side information, it makes it possible to infer the possible sensitive attributes corresponding to an individual. Once the correspondence between the identifier and the sensitive attributes is revealed for an individual, it may harm the individual and the distribution of the entire table.

Proposed System & alogirtham

All previous approaches to characterize and quantify privacy have only investigated the privacy risk of publishing a sensitive attribute by focusing only on the change of belief of an adversary about the probability distribution of this attribute. However, we believe that any attribute by itself is not sensitive. The sensitivity of an attribute comes from combining it with other attributes.

4.1 Advantages:

1. Privacy-preserving, the publishing technique strictly prohibits any privacy leakage in the published data
2. We focus on instances where different PPDP techniques assume to achieve an intended privacy level.

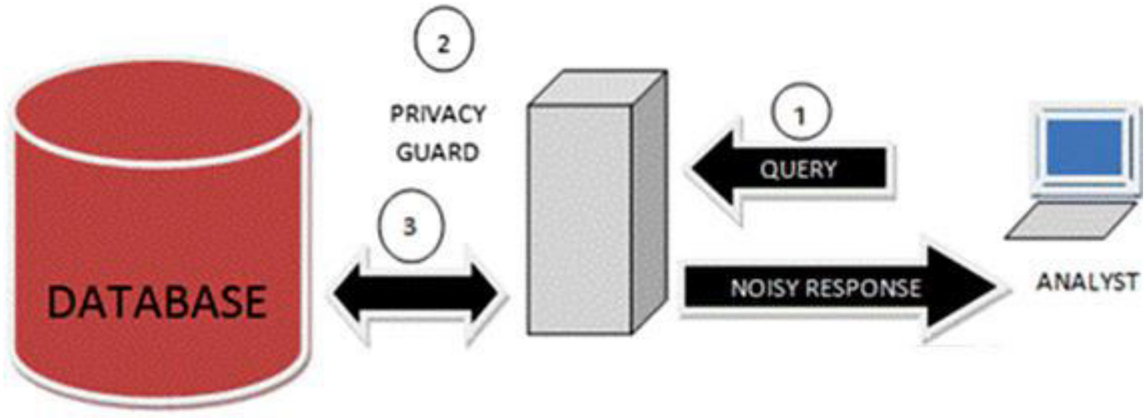


Fig:3.1 System Architecture

IMPLEMENTATION

MODULES:

Admin:

Admin can view users who are registered and admin can authorize users. Admin can see all friend requests information. Along with these details admin can view information of different communities available on network and users who are part of that community and check which community.

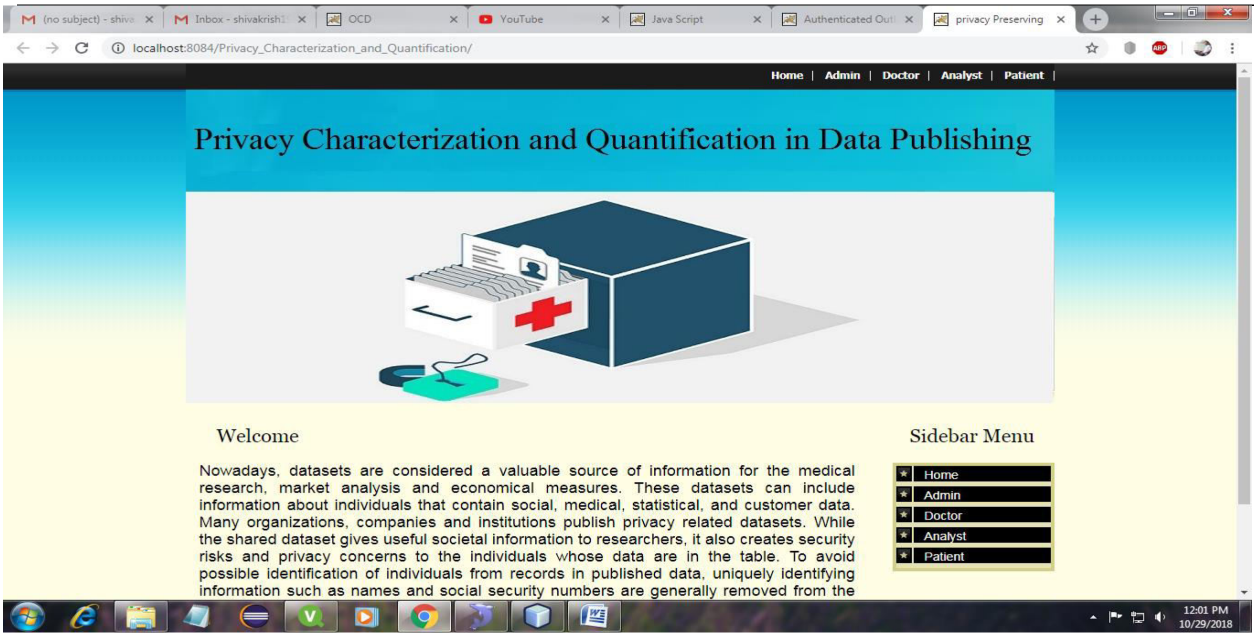
Mountain model:

The *Mountain* model is integral in this research, and is based on modularity, approximate optimization, and graph theory. It sorts the of edges. Owing to the feature of community structures, some chain groups in a community may fall down while surrounding community may rise like mountains.

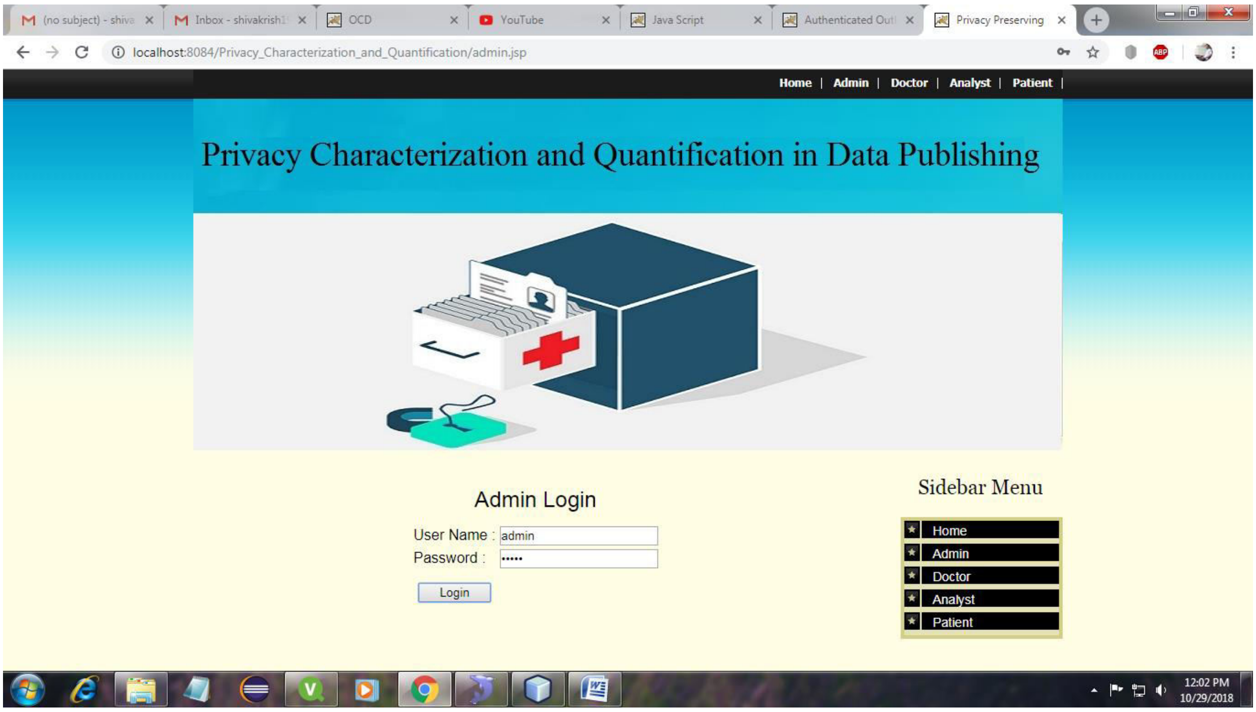
5 RESULTS AND DISCUSSION

Screens Shorts

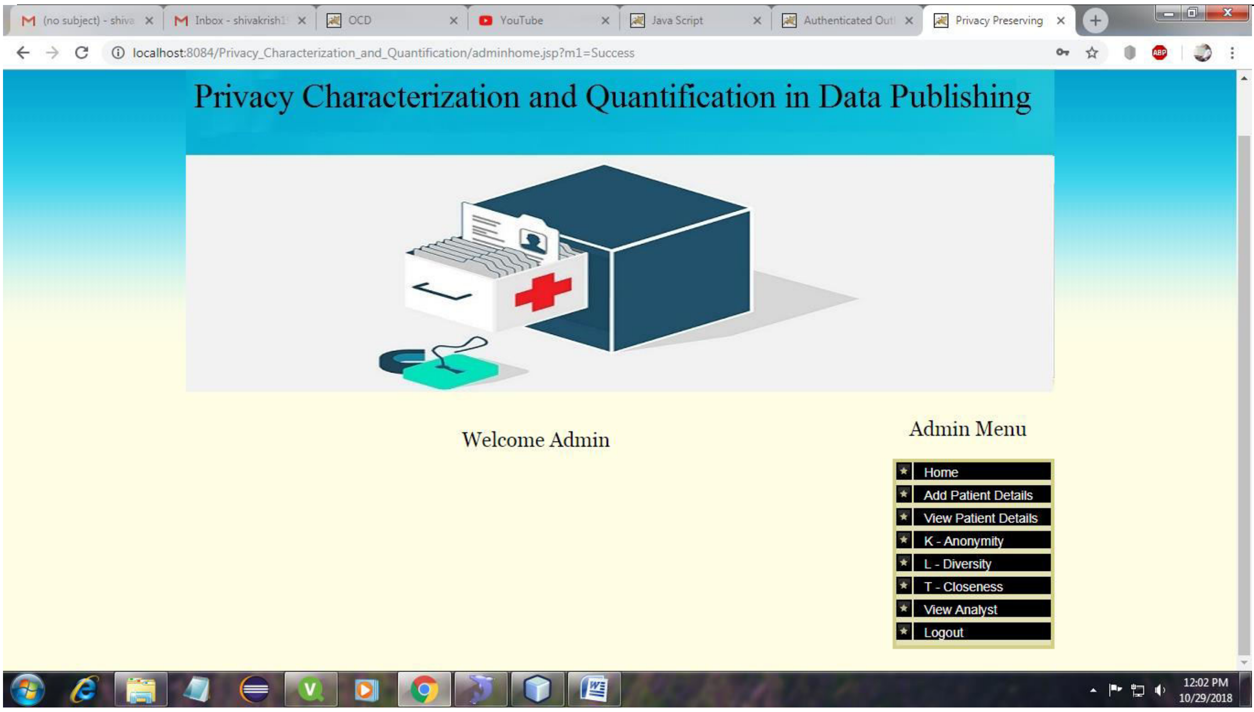
Home:



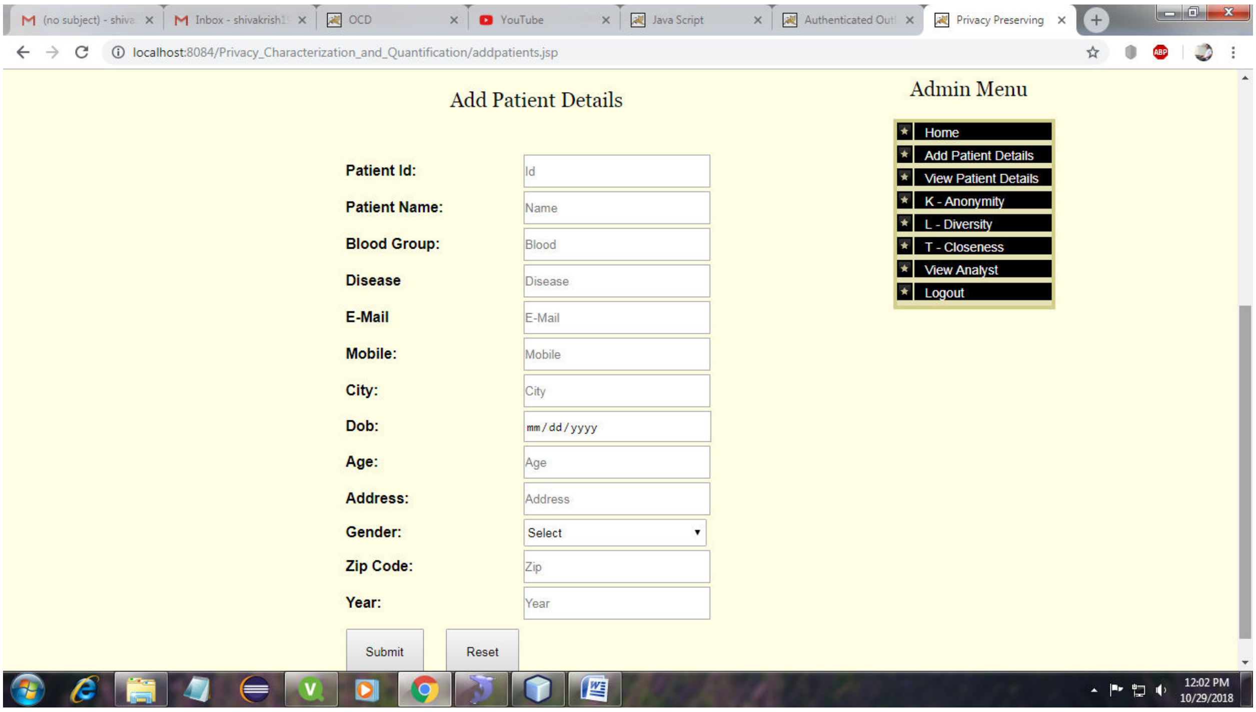
Admin login:



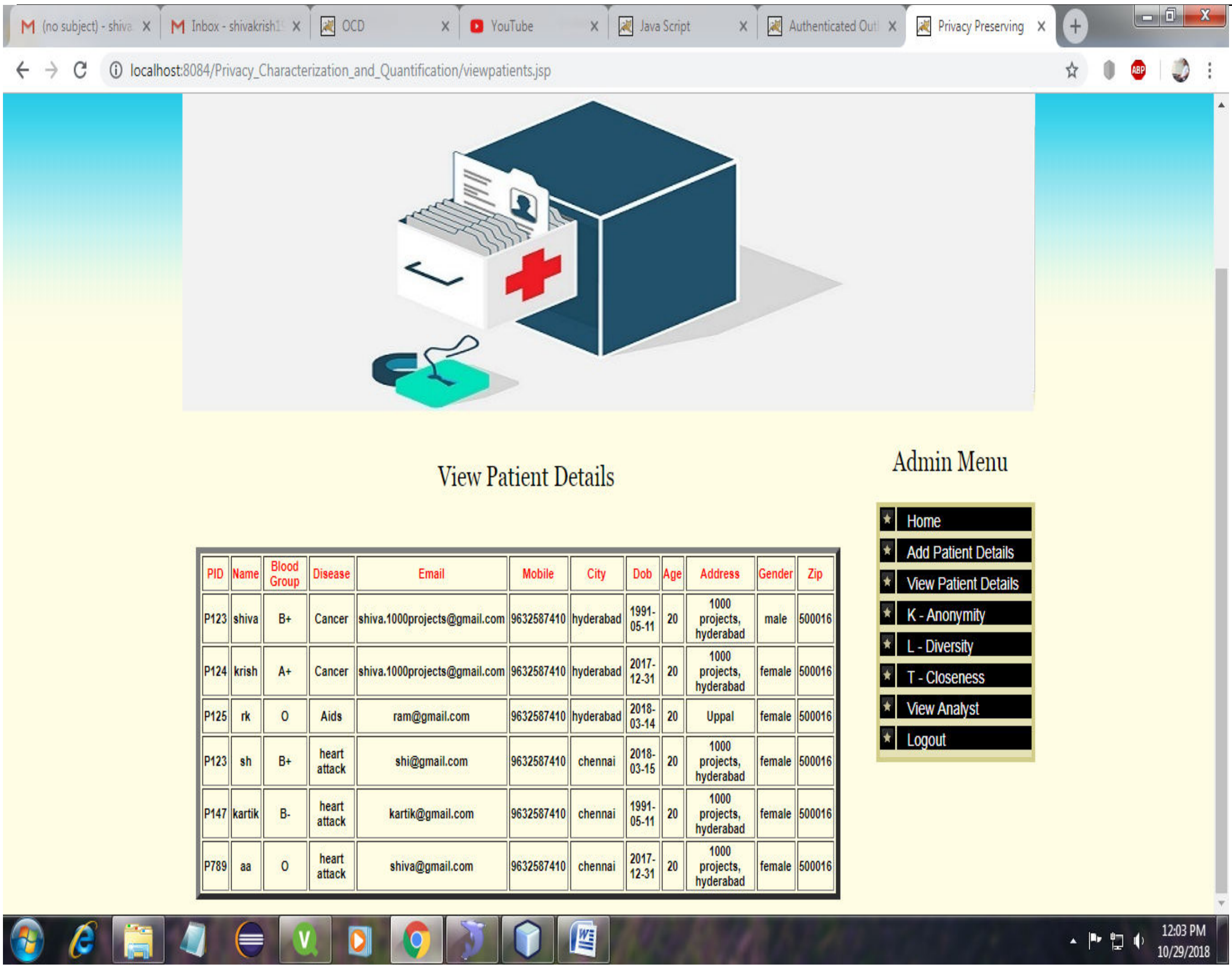
Admin home:



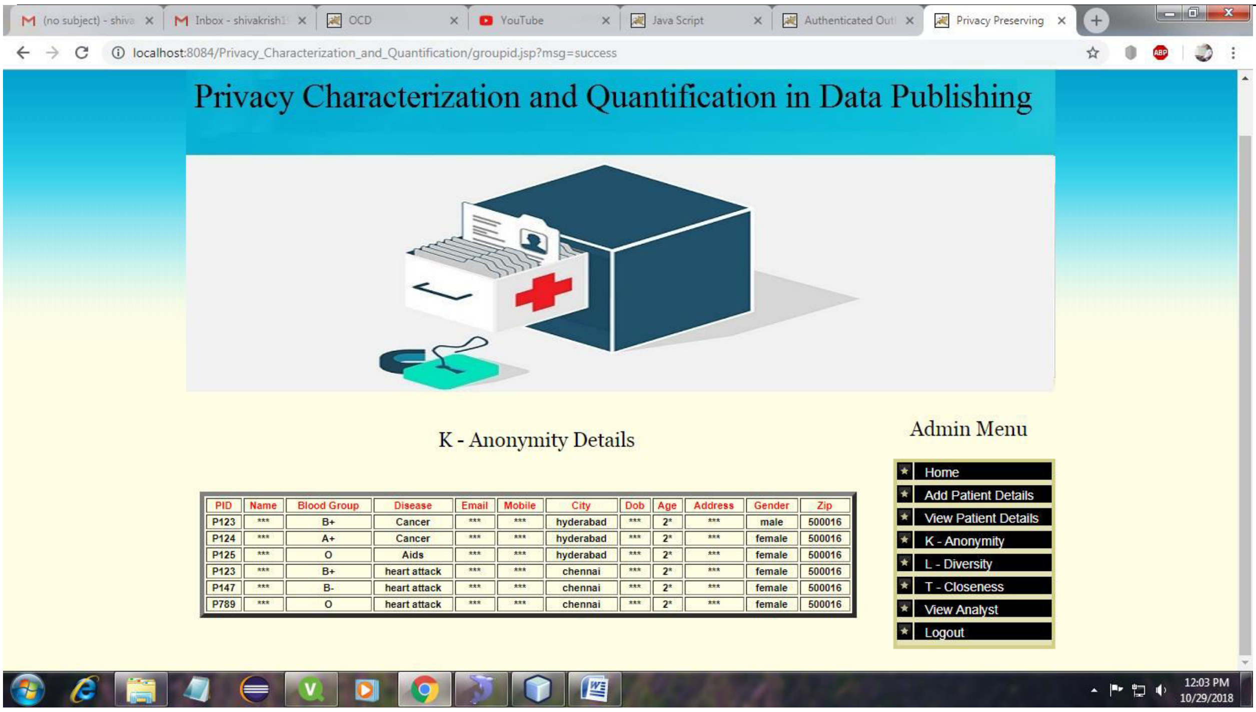
Add patient Details:



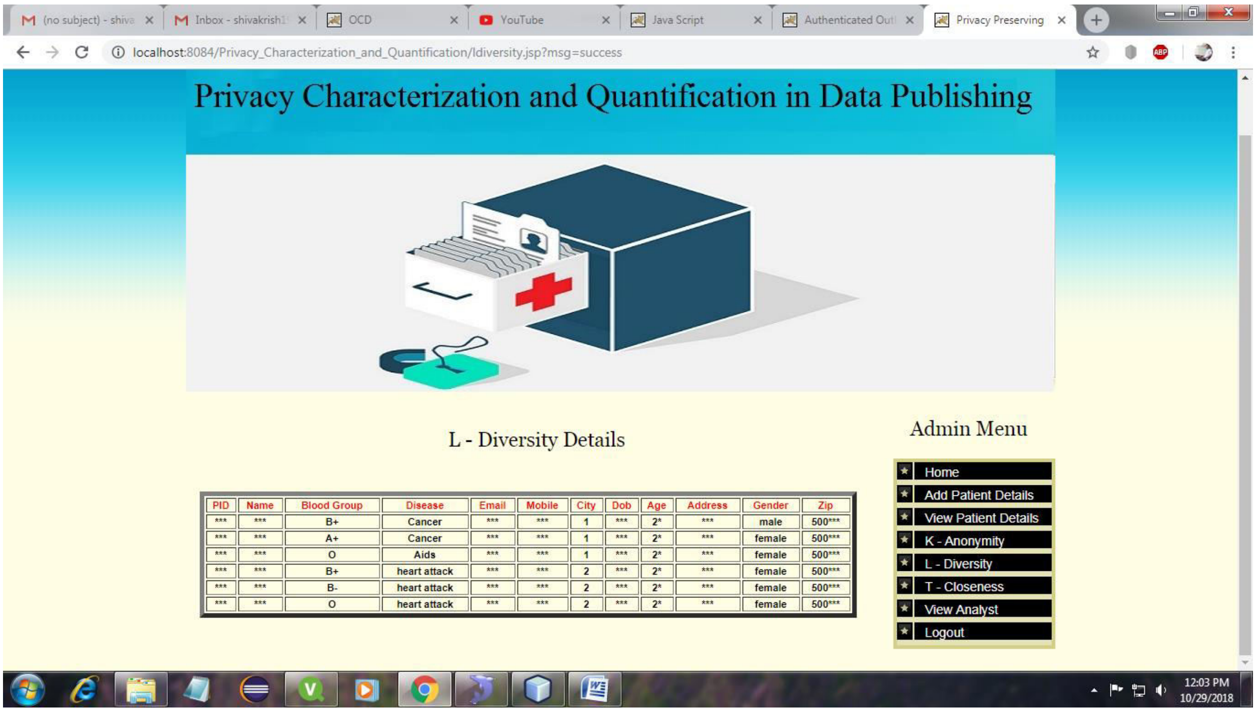
View Patient Details:


localhost:8084/Privacy_Characterization_and_Quantification/viewpatients.jsp
View Patient Details
PID Name Blood Group Disease Email Mobile City Dob Age Address Gender Zip
P123 shiva B+ Cancer shiva.1000projects@gmail.com 9632587410 hyderabad 1991-05-11 20 1000 projects, hyderabad male 500016
P124 krish A+ Cancer shiva.1000projects@gmail.com 9632587410 hyderabad 2017-12-31 20 1000 projects, hyderabad female 500016
P125 rk O Aids ram@gmail.com 9632587410 hyderabad 2018-03-14 20 Uppal female 500016
P123 sh B+ heart attack shi@gmail.com 9632587410 chennai 2018-03-15 20 1000 projects, hyderabad female 500016
P147 kartik B- heart attack kartik@gmail.com 9632587410 chennai 1991-05-11 20 1000 projects, hyderabad female 500016
P789 aa O heart attack shiva@gmail.com 9632587410 chennai 2017-12-31 20 1000 projects, hyderabad female 500016
Admin Menu
Home
Add Patient Details
View Patient Details
K - Anonymity
L - Diversity
T - Closeness
View Analyst
Logout

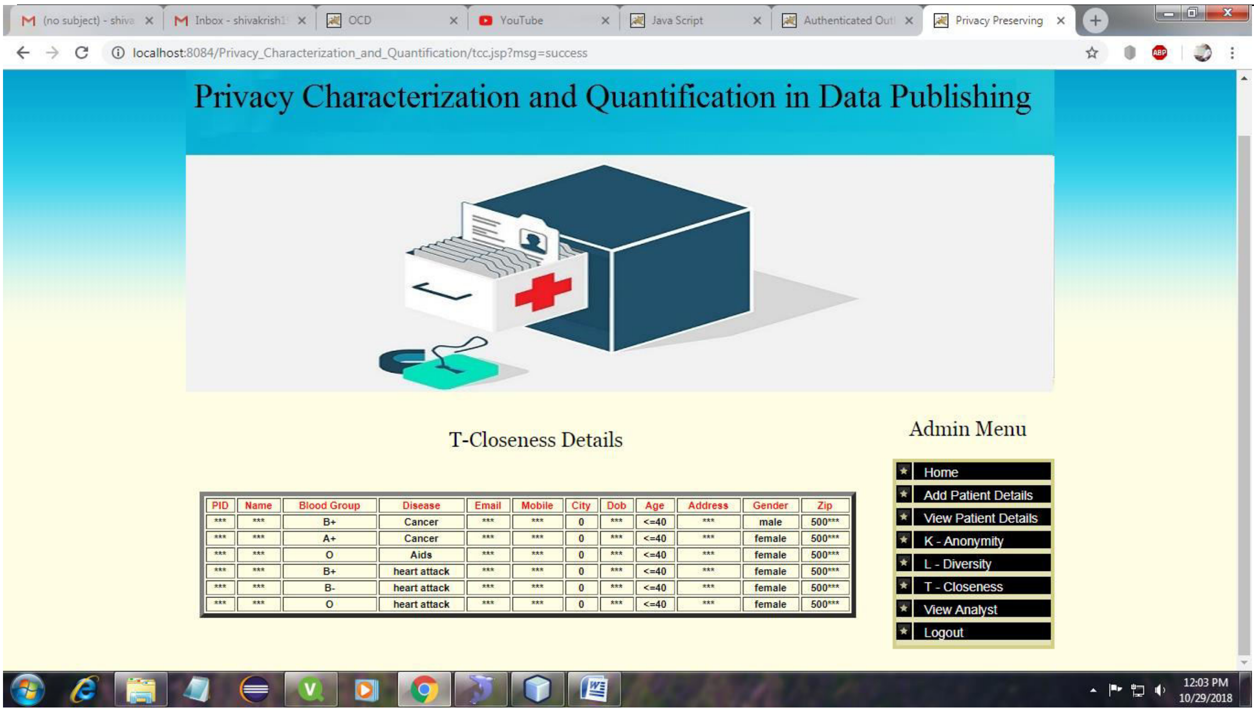
K – Anonymity:



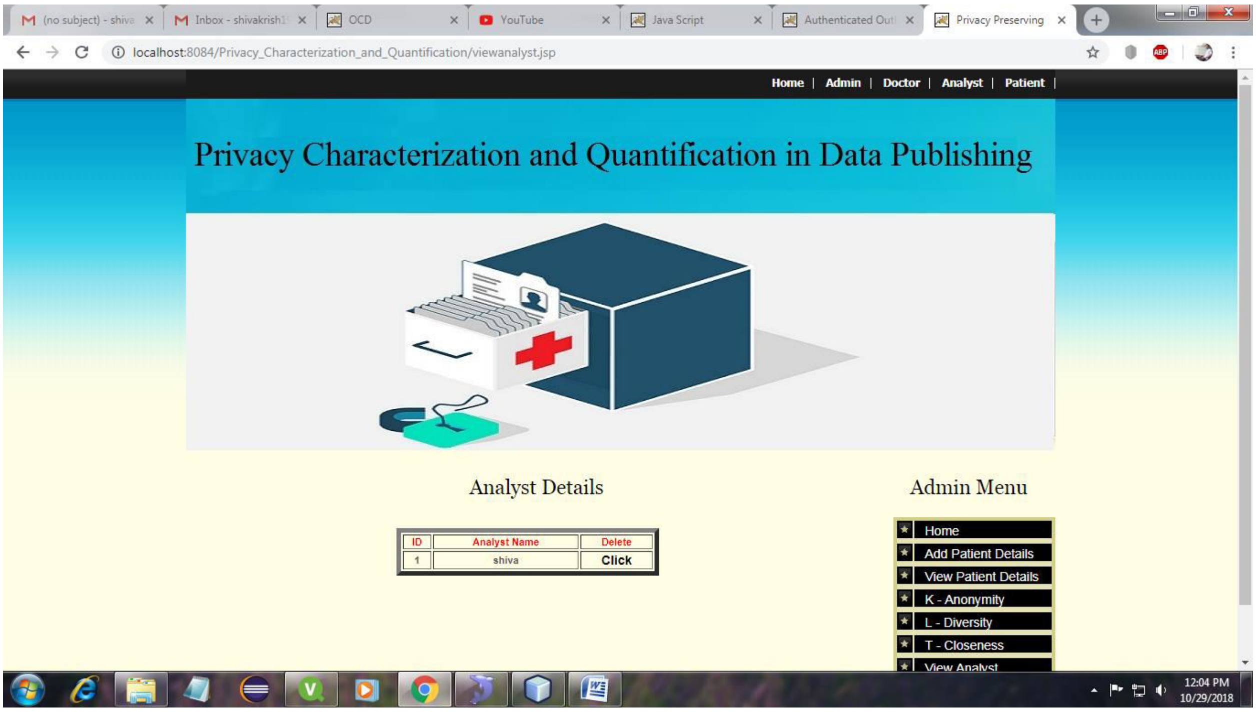
L – Diversity:



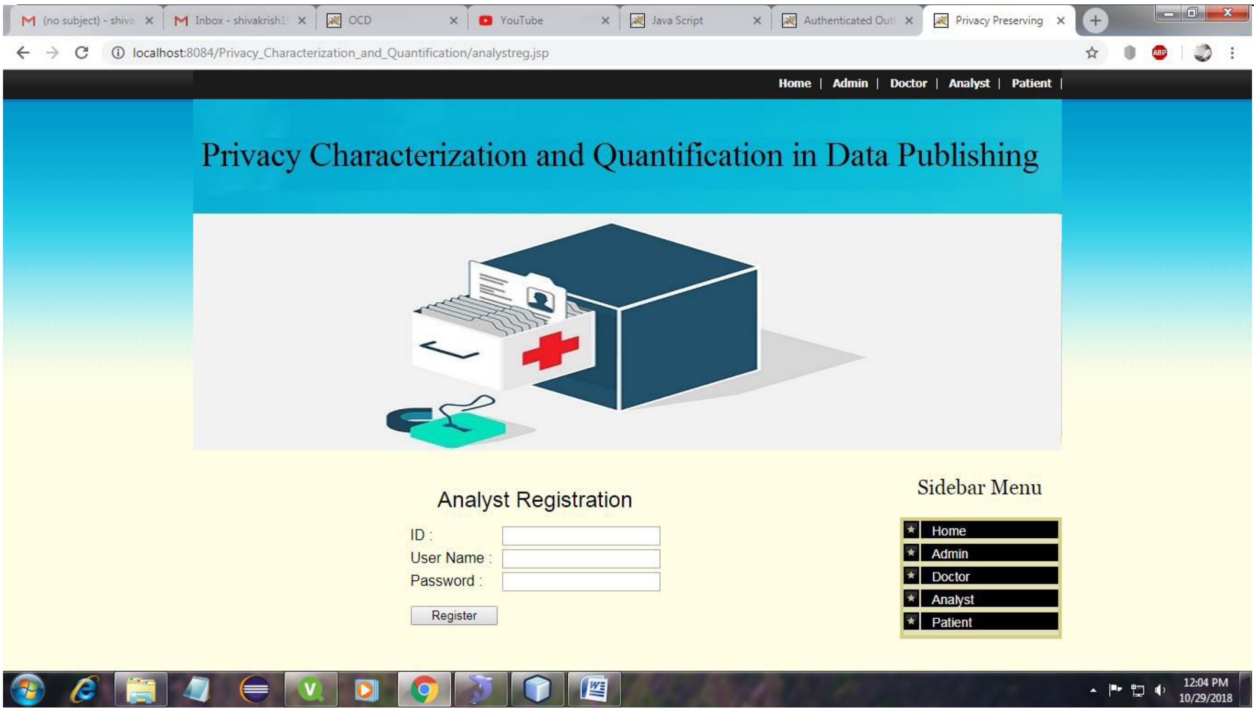
T – Closeness:



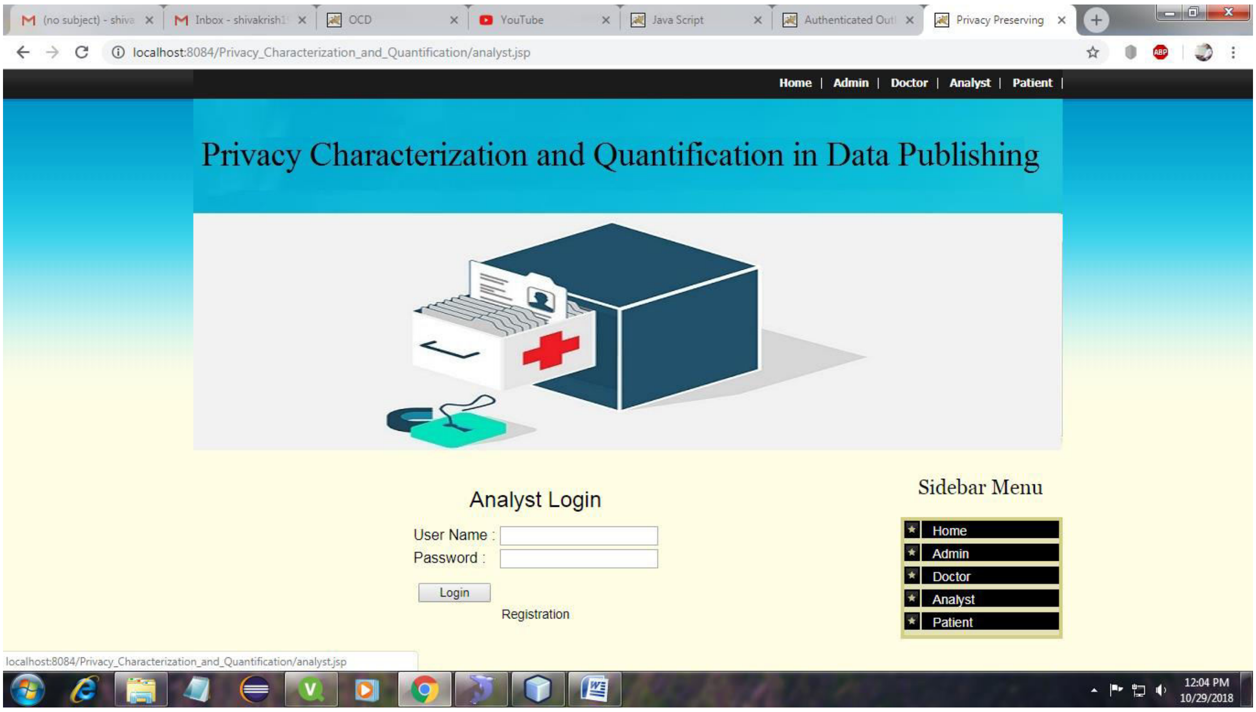
View Analyst:



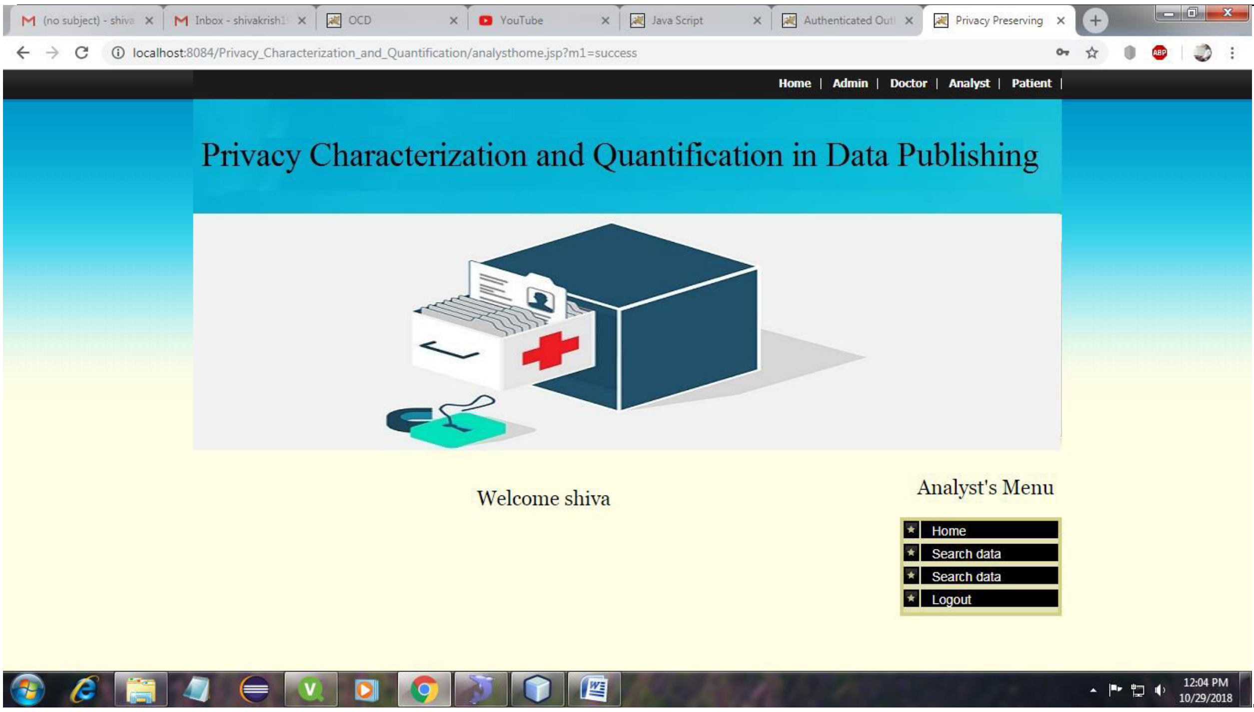
Analyst Registration:



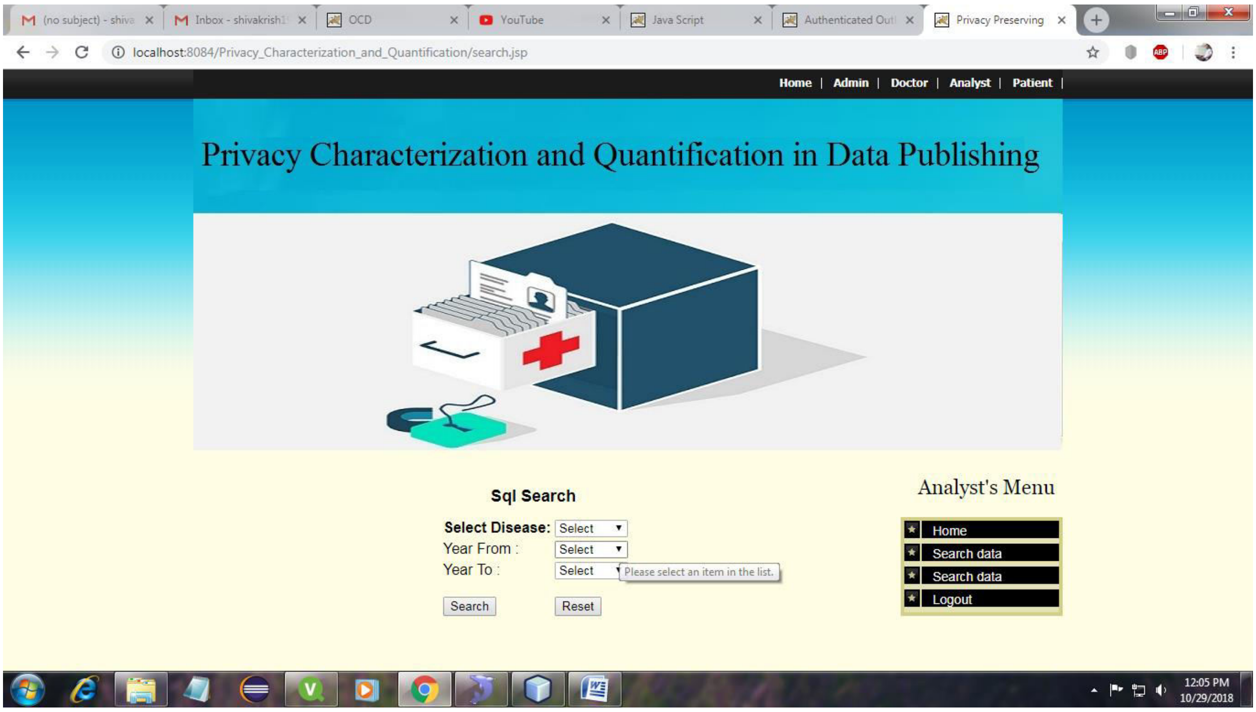
Analyst Login:



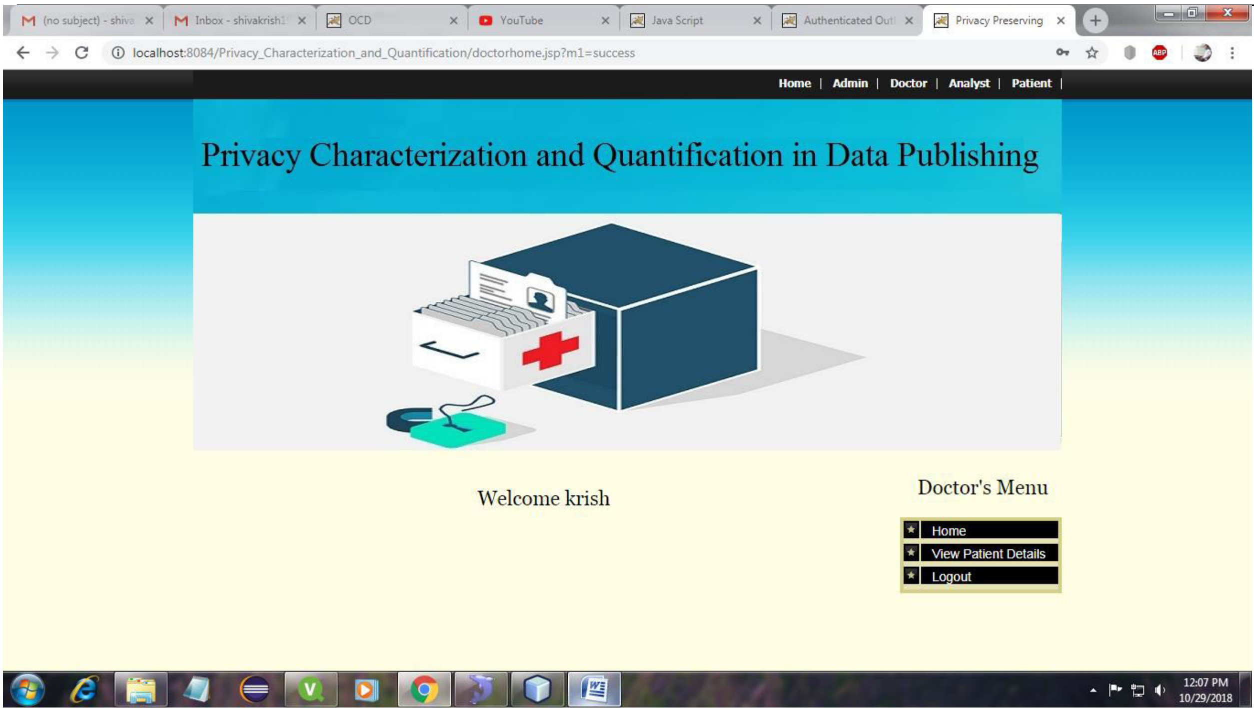
Analyst Home:



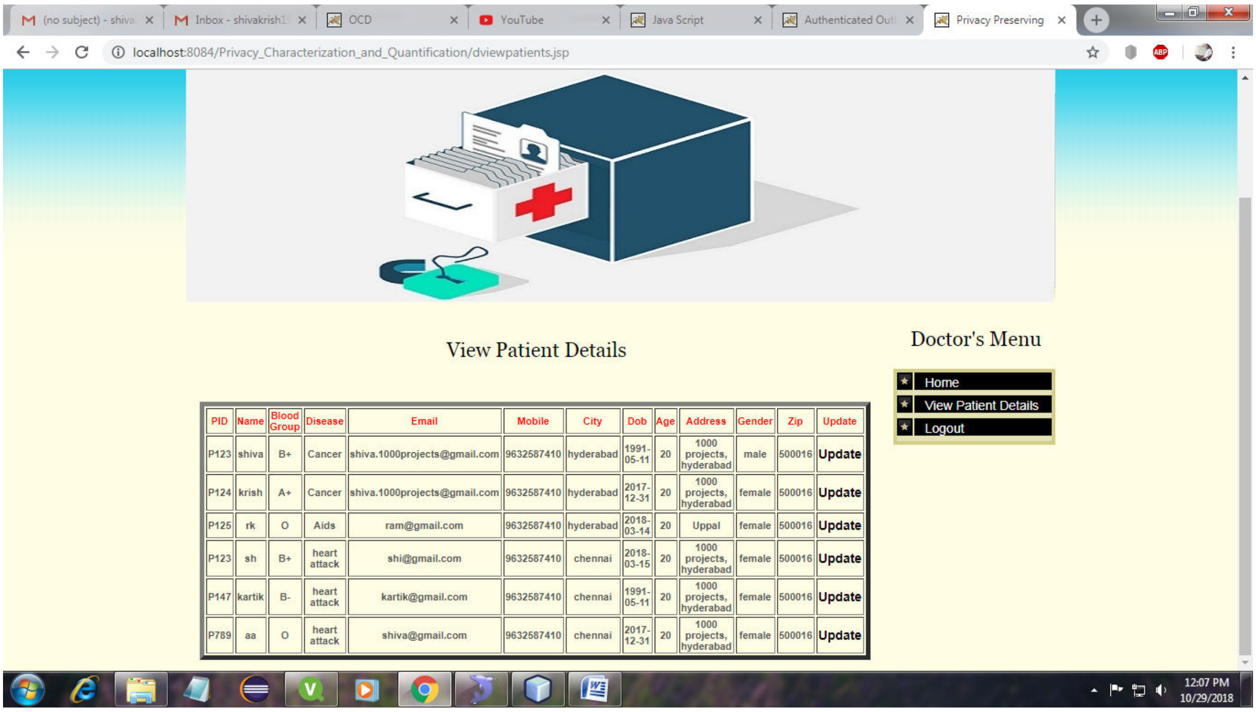
SQL Search:



Doctor home:



Patient Details:



6. CONCLUSION AND FUTURE WORK

CONCLUSION

In this paper, we introduced comprehensive characterizationand novel quantification methods of privacy to dealwith the problem of privacy quantification in privacypreservingdata publishing. In order to consider the privacyloss of combined attributes, we presented data publishingas a multirelational model. We re-defined the prior andposterior beliefs of the adversary. The proposed model andadversarial beliefs contribute to a more precise privacycharacterization and quantification. Supported by insightfulexamples, we then showed that privacy could not be quantifiedbased on a single metric. We proposed two differentprivacy leakage metrics. Based on these metrics, the privacyleakage of any given PPDP technique could be evaluated.Our experiments demonstrate how we could gain a betterjudgment of existing techniques and help analyze theireffectiveness in reaching privacy.

7. REFRENCES

- **1** L. Sweeney, " k-anonymity: A model for protecting privacy," International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, vol. 10, no. 5, pp. 557– 570, 2002.
- L. Sweeney, " Uniqueness of simple demographics in the U.S. population," 2000.
- Narayanan and V. Shmatikov, " Robust de-anonymization of large sparse datasets," in Security & Privacy, pp. 111– 125, 2008.
- Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, " `-diversity: Privacy beyond k-anonymity," ACM Trans. Knowl. Discov. Data, vol. 1, Mar. 2007.
- N. Li, T. Li, and S. Venkatasubramanian, " t-closeness: Privacy beyond k-anonymity and l-diversity," in ICDE, pp. 106– 115, 2007.
- N. Li,W. Qardaji, D. S. Purdue, Y.Wu, andW. Yang, " Membership privacy: A unifying framework for privacy definitions," in CCS,(Berlin, Germany), 2013.
- Wagner and D. Eckhoff, " Technical privacy metrics: a systematic survey," CoRR, vol. abs/1512.00327, 2015.
- J. Zhang, G. Cormode, C. M. Procopiuc, D. Srivastava, and X. Xiao, " Privbayes: Private data release via bayesian networks," in Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, SIGMOD ' 14, (New York, NY, USA), pp. 1423– 1434, ACM, 2014.
- M. G¨ otz, S. Nath, and J. Gehrke, " Maskit: Privately releasing user context streams for personalized mobile applications," inProceedings of the 2012 ACM SIGMOD International Conference onManagement of Data, SIGMOD ' 12, (New York, NY, USA), pp. 289– 300, ACM, 2012.

- Y. Rubner, C. Tomasi, , L. J., and Guibas, " The earth mover' s distance as a metric for image retrieval," International Journal of Computer Vision, vol. 40, no. 2, pp. 99– 121, 2000.