AUTHENTICATED OUTLIER MINING FOR OUTSOURCED DATABASES

Sarala [1], D.Dharmendra,

[1]**Assistant professor(HOD) , MCA DEPT,** Dantuluri Narayana Raju College**, Bhimavaram, Andharapradesh**
Email:-vedalasarala21@gmail.com
[2]**PG Student of MSc(cs), Dantuluri** Narayana Raju College**, Bhimavaram, Andharapradesh**
Email:-dulamdharmendra11@gmail.com

ABSTRACT

The Data-Mining-as-a-Service (DMaS) paradigm is becoming the focus of research, as it allows the data owner (client) who lacks expertise and/or computational resources to outsource their data and mining needs to a third-party service provider (server). Outsourcing, however, raises some issues about result integrity: how could the client verify the mining results returned by the server are both sound and complete? In this paper, we focus on outlier mining, an important mining task. Previous verification techniques use an authenticated data structure (ADS) for correctness authentication, which may incur much space and communication cost. In this paper, we propose a novel solution that returns a probabilistic result integrity guarantee with much cheaper verification cost. The key idea is to insert a set of artificial records (ARs) into the dataset, from which it constructs a set of artificial outliers (AOs) and artificial non-outliers (ANOs). The AOs and ANOs are used by the client to detect

1 INTRODUCTION

This dissertation addresses the problem of authentication data that is retrieved through an outsourced storage service: when the repository of the data is not managed directly by the end-user, and the manager of data is not completely trusted, how can data received be proven authentic? This question is the core of several security-related problems underlying any real-life computing application that involves storage of data over a communication or computing structure that can act unreliably. Clearly, data authentication ensuring that received data can be accurately verified to be in its original form is a fundamental problem in the area of information security, for information is valuable only when it is trustworthy. Data authentication captures some primary security needs of today’s computing reality.

Literature Survey

Recently, the relevance of privacy-preserving data mining techniques is thoroughly analyzed and discussed by Matwin (2013). Utilization of specific methods revealed their ability to preventing the discriminatory use of data mining. Some methods suggested that any stigmatized group must not be targeted more on generalization of data than the general population. Vatsalan et al. (2013) reviewed the technique called ‘Privacy-Preserving Record Linkage’ (PPRL), which allowed the linkage of databases to organizations by protecting the privacy.

3 IMPLEMENTATION STUDY

EXISTING SYSTEM:

Researchers have proposed several optimization approaches to improve the efficiency of outlier detection; it is difficult for the data owner who lacks the expertise to exploit these techniques.

Disadvantages:

One major concern is the integrity of the mining results returned by the service provider (server). Given the fact that the server is potentially untrusted, it is necessary for the client to authenticate if the outliers returned by the server are correct. There are many reasons for the server to return wrong results. For example, the server may return wrong mining results accidentally due to software bugs; it may keep part of the mining results to itself intentionally so that it can sell the retained results to the competitors of the client for profit. There also exists a strong financial incentive for the server to reduce the computational cost. For example, the server may execute the outlier mining on a portion of the outsourced dataset, and charge the client for mining of the whole dataset. The primary challenge in authenticating the outsourced outlier mining arises from the weak computational resources at the client side.

Proposed System & alogirtham

We focus on outlier mining, which is to find data objects that do not comply with the general patterns of the majority.

4.1 Advantages:

We proposed a lightweight authentication framework by constructing a set of artificial records (ARs) to catch any unsound or incomplete result with high probability guarantee.

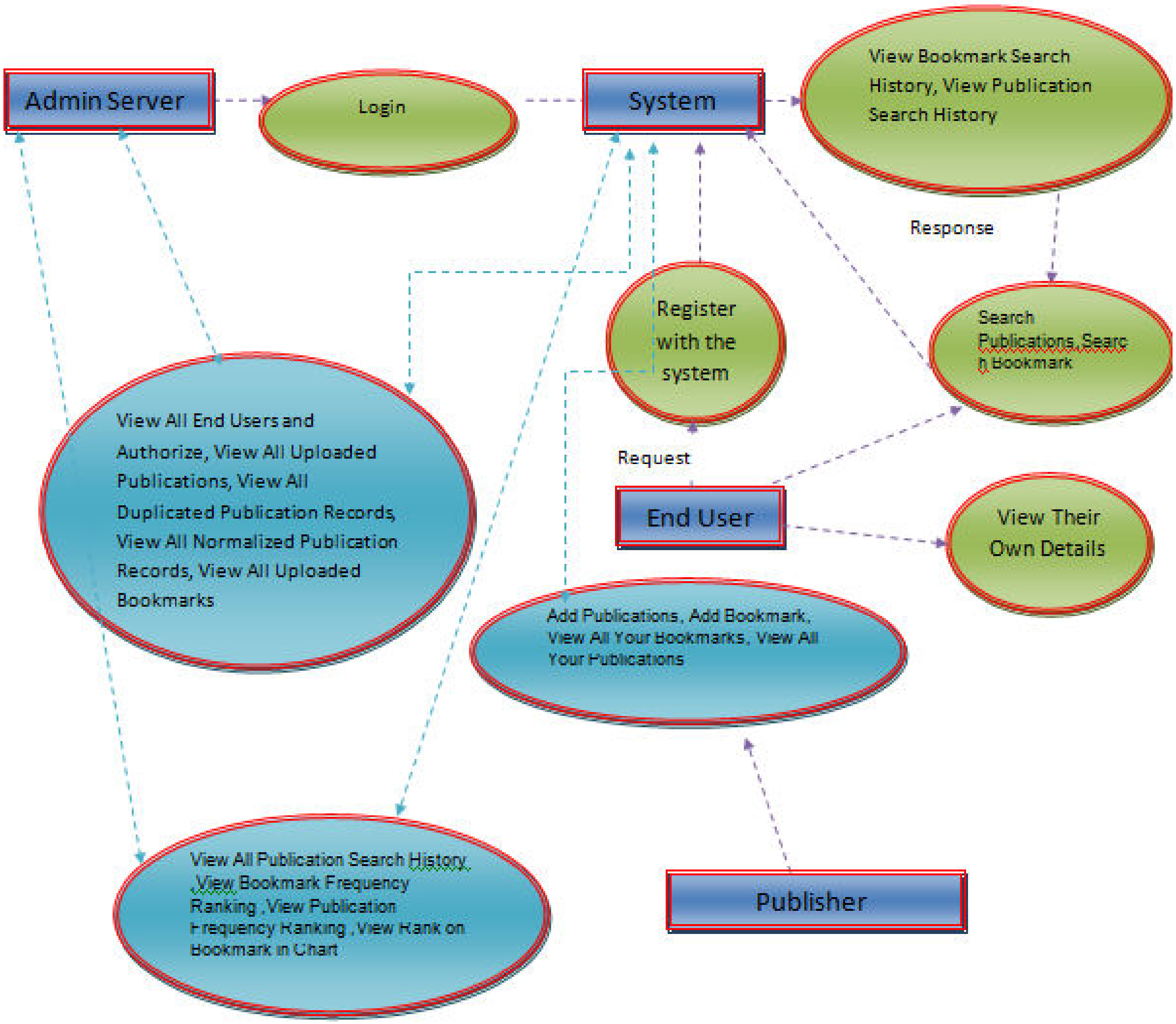

Admin Server
Login
System
View Bookmark Search History, View Publication Search History
Response
Search Publications,Searc h Bookmark
Register with the system
Request
End User
View Their Own Details
View All End Users and Authorize, View All Uploaded Publications, View All Duplicated Publication Records, View All Normalized Publication Records, View All Uploaded Bookmarks
Add Publications, Add Bookmark, View All Your Bookmarks, View All Your Publications
View All Publication Search History ,View Bookmark Frequency Ranking ,View Publication Frequency Ranking ,View Rank on Bookmark in Chart
Publisher

Fig:3.1 System Architecture

IMPLEMENTATION

MODUELS:

Admin

In this module, the Admin has to login by using valid user name and password. After login successful he can do some operations such as View All End Users and Authorize, View All Uploaded Publications, View All Duplicated Publication Records, View All Normalized Publication Records

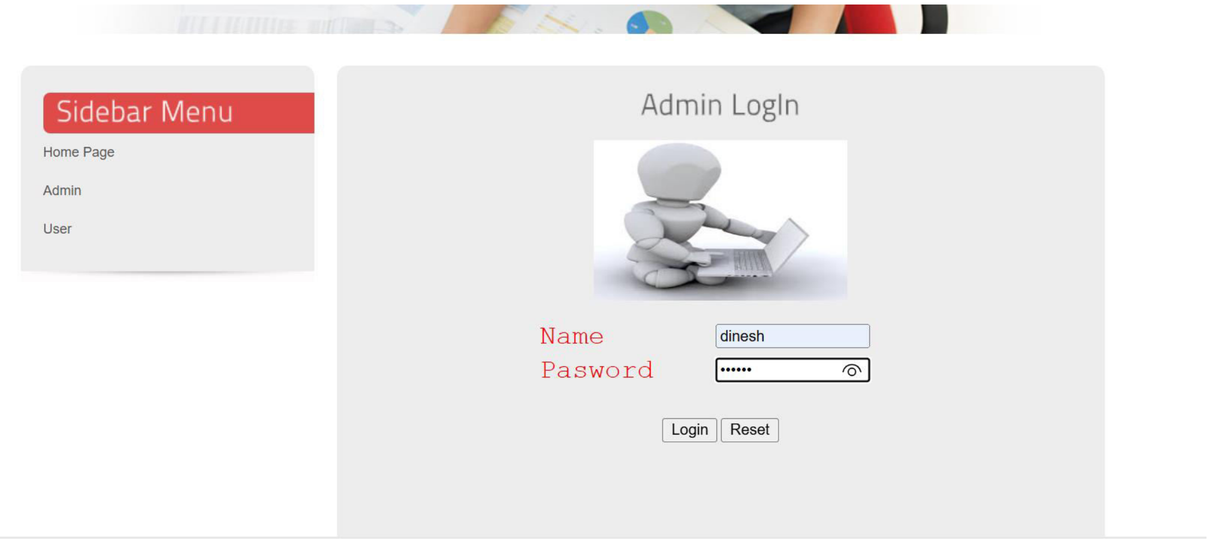
, View All Uploaded Bookmarks, View All Bookmark Search History, View All Publication

Search History, View Bookmark Frequency Ranking, View Publication Frequency Ranking, View Rank on Bookmark in Chart, View Rank on Publication in Chart.

5 RESULTS AND DISCUSSION

SCREENSHOTS:

FIG:5.3.1 ADMIN LOGIN



Name dinesh

Pasword ••••••

FIG:5.1 ADMINLOGIN

FIG:5.3.2 ADMIN MENU

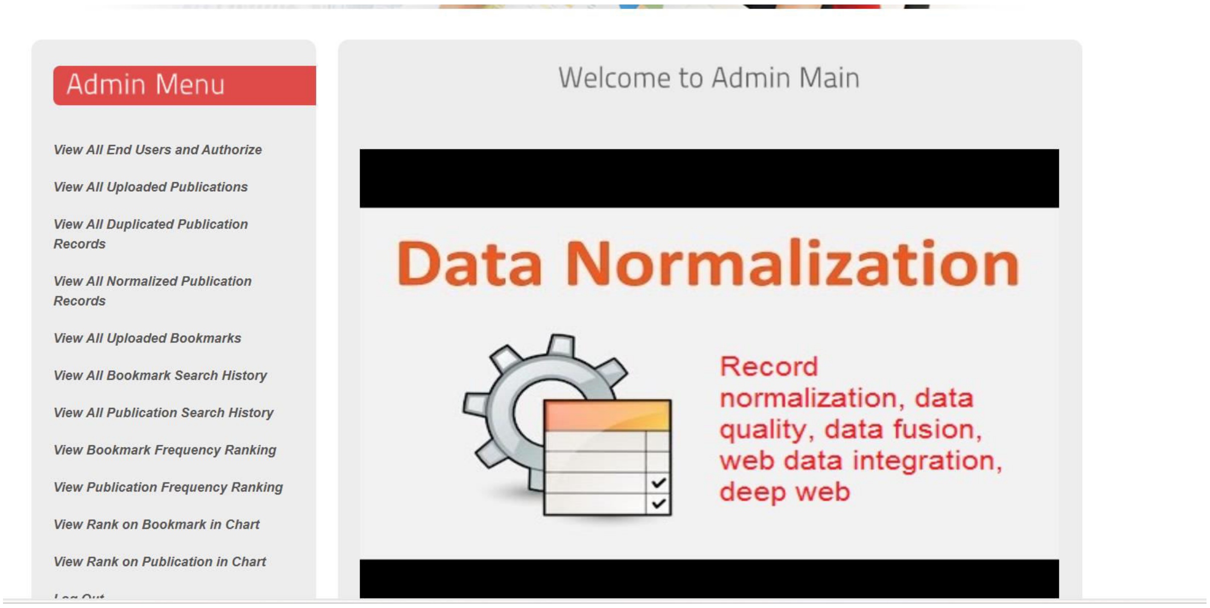
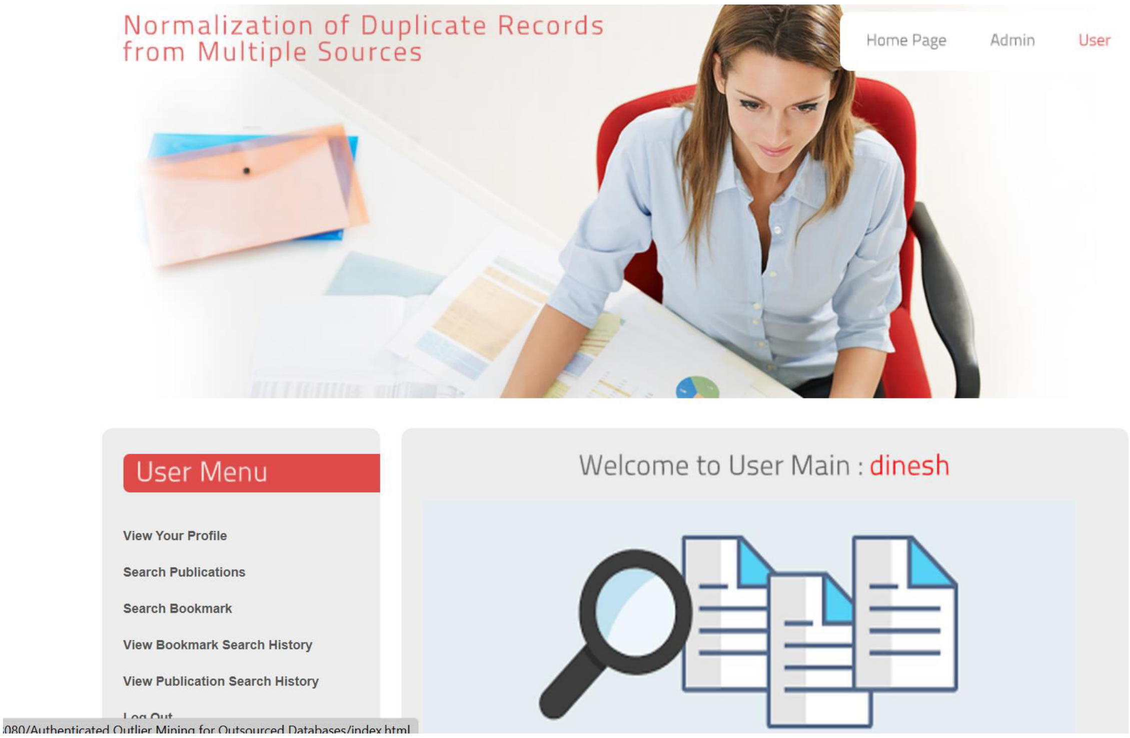

Admin Menu
View All End Users and Authorize
View All Uploaded Publications
View All Duplicated Publication Records
View All Normalized Publication Records
View All Uploaded Bookmarks
View All Bookmark Search History
View All Publication Search History
View Bookmark Frequency Ranking
View Publication Frequency Ranking
View Rank on Bookmark in Chart
View Rank on Publication in Chart
Welcome to Admin Main
Data Normalization
Record normalization, data quality, data fusion, web data integration, deep web

FIG:5.2 ADMIN MENU

FIG:5.3.3 USER MENU

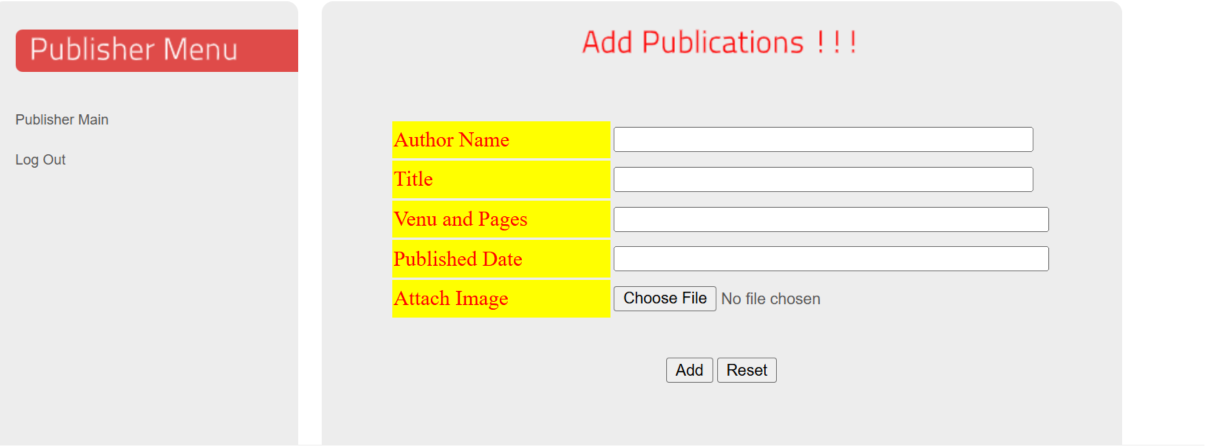

Normalization of Duplicate Records from Multiple Sources

Welcome to User Main : dinesh

Home Page
Admin
User
User Menu
View Your Profile
Search Publications
Search Bookmark
View Bookmark Search History
View Publication Search History
Log Out
080/Authenticated Outlier Mining for Outsourced Databases/index.html

FIG:5.3 USER MENU

FIG:5.3.4 PUBLISHIER MENU


Add Publications ! ! !

Author Name	
Title	
Venu and Pages	
Published Date	
Attach Image	Choose File No file chosen

FIG:5.4 PUBLISHIER MENU

6. CONCLUSION AND FUTURE WORK

CONCLUSION

An inclusive overview on PPDM techniques based on distortion, associative classification, randomization, distribution, and k-anonymization is presented. It is established that PPDM is appeared progressively common due to easy sharing of privacy sensitive data for analysis. The notable advantages and obvious disadvantages of current studies are emphasized. Presently, Big Data are often shared across sectors such as health, military and others, and transverses across Business-to-Businesses, Entities-to-Entities and Government-to-Government. Thus, the preservation of privacy against disclosure and attacks are of critical concern. Several big organizations and governments worldwide being totally dependent on information communications via internet expressed grave concerns over privacy issues. Consequently, the rapid development of IT faced new challenges to PPDM. Data mining possesses being the capability to extract and mine vast sea of interesting patterns or knowledge from a huge amount of data requires absolute security. The main idea of PPDM is to incorporate the traditional data mining techniques in transforming the data to mask sensitive information. The major challenge is to efficiently transform the data and recover its mining outcome from the transformed one.

7. REFRENCES

- [Aggarwal G, Bawa M, Ganesan P (2005) Two can keep a secret: a distributed architecture for secure database services. CIDR
- Agrawal R, Srikant R (2000) Privacy-preserving data mining. In: Proceedings of the 2000 ACM SIGMOD international conference on management of data—SIGMOD ’00, vol 29, no 2. pp 439–450. http://doi.org/10.1145/342009.335438
- Aggarwal CC, Yu PS (2008) A general survey of privacy-preserving data mining models and algorithms. In: Privacy preserving data mining, Chap 2. Springer, New York, pp 11–52. http://doi.org/10.1007/978-0-387-48533
- Arunadevi M, Anuradha R. Privacy preserving outsourcing for frequent itemset mining. *Int J Innov Res Comp Commun Eng*. 2014;2(1):3867–3873. [Google Scholar]
- Baotou T (2010) Research on privacy preserving classification data mining based on random perturbation Xiaolin Zhang Hongjing Bi. 1–6

- Belwal R, Varshney J, Khan S (2013) Hiding sensitive association rules efficiently by introducing new variable hiding counter. In: IEEE international conference on service operations and logistics, and informatics, 2008, IEEE/SOLI 2008, vol 1, pp 130–134. doi:10.1109/SOLI.2008.4686377