Abstract

The word BIG DATA is nothing but huge amount of data generated through all the sources including social networking sites like facebook, twitter, Instagram etc. this data sometimes may be repetitive that is same person can have record in more than one databases, whereas it is belonging to a single person so those records should be merged. Also sometimes a situation may occur where you need entire history of a person in this case record linkage will make it possible. Many Researches has been done for efficiently linking the records as record linking is becoming important day-by-day since it increases the quality of the data. In this research we are going to focus on algorithm for efficiently linking the records and keeping records secure. The software called FEBRL is used for comparing our algorithm efficiency with previously defined algorithms.

References

1. peters christen,"A Survey of Indexing Techniques for Scalable Record Linkage and
Deduplication”IEEE Transaction on Knowledge and Data Engineering, vol 24, No. 9, September 2012

2. Gayan Prasad Hettiarachchi, Dhammika Suresh Hettiarachchi, Nadeeka Nilmini Hettiarachchi, Azusa Ebisuya, ”Next Generation Data Classification and Linkage”, Osaka University of Tokyo, Japan

3. Timothy C Heavens[1], James C. Bezdek[2], Marimuthu Palsiwami[2], ”Scalable Single Linkage Hierarchical Clustering for BigData”, University of Melbourne, USA, Australia

4. Liang Jin, Chen Li, Sharad Mehrotra, ”Efficient Record Linkage in Large Datasets”, University of California, Irvine

5. Rainer Schnell, ”An Efficient Privacy Preserving Record Linkage Technique for Administrative data and census”, Statistical Journal of (IAOS), 2014

6. Peter Christen, ”Overview and taxonomy of Techniques for Privacy-preserving Record Linkage ” (JSM) Joint Statistical Meeting, August 2013

7. Rohan Baxtor, Peter Christen, Tim Churches, ”A Comparison for Fast Blocking methods for Record Linkage”, Australian National University, Australia

8. Dimitrios Karapiperis, Aris Gkoulalas-Divanis, Vassilios S. Verykios, ”Summarization Algorithms for Record Linkage” Hellenis Open University Patras Greece 2005

**Index Terms**

Computer Science Distributed Systems

**Keywords**

Efficient, Big data, Records, Clustering