



Certain Investigations on Methods Developed for Efficient Discovery of Matching Dependencies

R.Santhya¹, S.Latha², Prof.S.Balamurugan³, S.Charanyaa⁴

Department of IT, Kalaignar Karunanidhi Institute of Technology, Coimbatore, TamilNadu, India^{1,2,3}

Senior Software Engineer Mainframe Technologies Former, Larsen & Tubro (L&T) Infotech, Chennai, TamilNadu,
India⁴

ABSTRACT: This paper details about various methods prevailing in literature for efficient discovery of matching dependencies. The concept of matching dependencies (MDs) has recently been proposed for specifying matching rules for object identification. Similar to the functional dependencies with conditions, MDs can also be applied to various data quality applications such as detecting the violations of integrity constraints. The problem of discovering similarity constraints for matching dependencies from a given database instance is taken into consideration. This survey would promote a lot of research in the area of information mining.

KEYWORDS: Data Anonymization, Matching Dependencies(MDs), Object, Similarity Constraints, Information Mining.

I. INTRODUCTION

Need for publishing sensitive data to public has grown extravagantly during recent years. Recent days have seen a steep rise in preserving data quality in the database community due to the huge amount of “dirty” data originated from different. These data often contain duplicates, inconsistencies and conflicts, due to various mistakes of men and machines. In addition to the cost of dealing with the huge volume of data, manually detecting and removing “dirty” data is definitely out of practice because human proposed cleaning methods may introduce inconsistencies again. Therefore, data dependencies, which have been widely used in the relational database design to set up the integrity constraints. Hence protecting privacy of individuals and ensuring utility of social network data as well becomes a challenging and interesting research topic. In this paper we have made an investigation on the attacks by matching dependencies and possible solutions proposed in literature and efficiency of the same.

II. MINING ASSOCIATION RULES BETWEEN SETS OF ITEMS IN LARGE DATABASES

In this paper the author presented an efficient algorithm that generates all denoting association rules among the item in the database. The author also commented that the past transaction analysis of data is one of the approach to increase or increment the data quality. The basket data is the data which does not contains the details of each product that is been purchased by the customer, instead it contains the information of buying over a period of time particularly on the quality of the data. basket data stores the information about items purchased on the basis of per transaction. Based on the quality of the data increment some decision can be made on it. To improve the quality of the data, the item purchased should be noticed. The reason for the downflow of this database system is that it will not give the required functionality for a user who is willing to take an advantage on the information. In this paper, the author introduce the problem of mining which is a huge collection of basket data type transaction for association rules among the set of items with a little minimal confidence. The author decomposes or split the rules mining problem into two problem



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

1) Find minsupport which is a combination of the items and it have the fractional transaction support above the threshold. This can also called as large itemsets and reset of all the combination that does not meet the threshold value is said be the small itemsets.

2) Assume that Y is the large itemset, if itemset Y is large, then every subset of Y is also large and so thus we use the result of first problem. Let C is a confidence factor, if the ratio is greater than C then the rule is satisfied with the confidence factor C . All the rules derived from Y must satisfy the support constraint because Y satisfies the support constraints. The template algorithm is used for finding large itemsets.

III. :NEIGHBOURHOOD DEPENDENCIES FOR PREDICTION

In this paper, the author introduced the concept of neighbourhood dependency (ND) to explicit regularities like “families with similar size and income, tend to own cars of similar and income”. Rationally, the identification of this thing will be utilized for prediction purposes. The author have also implemented and tested an algorithm for mining NDs. In P-Neighbourhood method, in order to identify the unknown values, the discovered NDs are useful.

IV. DATA QUALITY CONCEPTS, METHODOLOGIES AND TECHNIQUES

In this paper the author provides a systematic introduction to the array of issues related to data quality. It describes the details about the parameters of the data quality like accuracy, completeness and consistency. The information about their importance in the types of data like federated data, web data or time-dependent data. The different data categories classified based on the frequency techniques, methodologies from core data quality research, related fields such as data mining, statistical data analysis and machine learning. The author concludes the book with a critical comparison of tools and practical methodologies to resolve the quality problem.

V. DATA CLEANING AND QUERIES ANSWERING WITH MATCHING DEPENDENCIES AND MATCHING FUNCTIONS

In this paper, the author said that the MDs are proposed as declarative rules for data cleaning and entity resolution. The quality constraints is also concerned, which is declarative in nature and are based on the precise model theoretic semantics. It plays a master role in database with classical integrity constraints. In this paper the author investigated their interaction with the similarity relations in same domains which in turn introduce a partial order of domination among instances. It allows to compare them on the basis of the content in the information and applied to a set of query answers. While considering all the notions, the author defined the class of clean instances for the provided dirty instances. The dirty instances are the intended and admissible instances which can be obtained after enforcing the matching dependencies. The chase like is used to define the clean instances which by turn enforces the MDs. This chase procedure will improve the information content belongs to the domination order. The dirty database was posted with the notion of clean answer to the query which is determined as a pair formed by a lower and an upper bound in the form of information content for the query answers. The author also have studied the intension of monotone query with respect to the domination order and how to relax a query into a monotone one which provided additional informative answer when comparing to the original one. The introduction of domination monotone relational query language utilize the lattice theoretic structure of the domain. The author put an open question to explore the connections with querying databases over partially ordered domains with incomplete or partial information with query relaxation in common and similarity relations based on relational languages.

VI. ADAPTIVE NAME MATCHING IN INFORMATION INTEGRATION

In this paper the author stated that identifying approximately duplicate database records will refer to the same entity is necessary for information integration. The author then compare and combine the methods described and also to learn textual similarity measures for name matching.

VII. A FEASIBILITY AND PERFORMANCE STUDY OF DEPENDENCY INFERENCE

In this paper the author described about the utility of inferring functional dependencies. While a tool is required to aid the database designed in the progress of specifying logical dependencies then the problem occur in the condition



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

of automatic database design. The author concluded in this paper that for practical example relation an frequent implementation of a dependence inference function will inturn lead to acceptable interative response times.

VIII. CONDITIONAL FUNCTIONAL DEPENDENCIES FOR DATA CLEANING

In this paper the author mentioned that the recent satitcs said that dirty data costs US business billions of dollars annually and it is also calculated that for the cleaning of data, a labor intensive and complex process accounts for 30%-80% of the development time in a datawarehouse dw project. These highlight is the cause for the necessarily for the tools which is to clean the data for automatically detecting and removing inconsistencies and bugs in the data. In this paper the author introduces CFDs and shows that CFDs can direct semantics data fundamental to data cleaning. In order to identify the inconsistencies as violation of CFDs, the author developed SQL based techniques. Repair and consistent query answer are the two topics that has been mainly focussed here. Repair is used to find other database which is consistent and quietly vary from the exact or original database.

IX. INCREASING THE EXPRESSIVITY OF CFDs WITHOUT EXTRA COMPLEXITY

In this paper, the author proposed an extended CFDs which is an extension of CFDs. In CFDs, we cannot able to detect the inconsistencies that rise in price but in the case of eCFDs, it is capable of catching inconsistencies that arise in practice. The eCFDs specify patterns of semantically releted values in terms of disjunction and inequity. The eCFD canbe defined by consider a relation schema which is defined over a finite set of attributes and for each attributes, the domain is denoted. The batch detection algorithm is used to find the violation set in eCFD. In addition to that the incremental detection algorithm such as BATCH DETECT and INC DETECT are also developed. These two algorithms will produce SQL queries to find the violations. The BATCHDETECT generates SQL queries and update statements for detectin pattern-constraint violations. As similar to BATCH DETECT algorithm, INC DETECT also generates SQL queries to find the changes to the violation of pattern-constraint and also maintains auxiliary relation in order to reuse previous computations. This INC DETECT algorithm aims to reduce the unwanted recomputation conducted for finding violation. The author also suggested to put our future work to develop algorithms for eliminating eCFD violations and repairing data and to find efficient method for discovering eCFDs automatically from data samples.

X. EXTENDING DEPENDENCIES WITH CONDITIONS

In this paper, the author introduced a class of conditional inclusion dependencies (CINDs) which is an extension of traditional inclusion dependencies (INDs) and is done by enforcing combining of related data values. In this paper, the author showed that the proposed CINDs is not only utilized for cleaning the data but also used in contextual schema matching. The first contribution of this paper is towards the notion of CINDs. The CIND is defined as a pair that consists of IND and a pattern tablean where the tablean enforces combining of semantically related data values.

The second contribution is towards the techniques for reasoning about the CINDs. In this, first has to check that the CINDs are consistent without having any conflicts. Other decision problem associated with CINDs is the implication problem that is used to decide whether a set of CINDs entails another CINDs. The PSPACE-complete is the implication for traditional INDs. The third contribution is about the survey of the interaction among CINDs and CFDs. If the CINDs and CFDs are together then the consistency problem is also shown in this paper. The fourth contribution contains a set of algorithm which is used for checking the consistency of CFDs and CINDs.

The author concluded that the CINDs together with CFDs may lead to promising tools for cleaning data and for finding quality schema matches.

The author put forth his future work for checking if better complexity results can be obtained by considering adding assumptions like acyclicity of CINDs.

XI. SEARCHING FOR DEPENDENCIES AT MULTIPLE ABSTRACTION LEVELS

In this paper the author introduces the concept of roll up dependency (RUD) Which is an extension of functional dependencies with generalization hierarchies. The problem addressed in this paper is that for a given relational table with multiple attributes, every attributes in the table will take the value from a specified domain. The domain values are listed into generalization. The problem is to identify the roll up dependencies that are undergone with high confidence. A roll-up dependency (RUD) is an significance among two generalized schema of the same



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

underlying relation schemas. In this paper the author addressed the RUDMINE problem which is to discover RUPs whose support and confidence exceed certain specified threshold value. It is also shown that the problem is NP-hard in the schema size, but polynomial in the number of tuples. The author put forth his future work to use of RUDs in multidimensional database design.

XII. DISCOVERING DATA QUALITY RULES

In this paper the author maintained that the dirty data is a serious problem for leading a business to wrong decision making, inadequate daily operations and also waste of time and money. The dirty data will arise when domain constraints and business rules are in turn in the need to protect data consistency and accuracy.

In this paper the author propose a new data driven tool which can be utilized within a particular organization data quality management process to suggest various possible rules and to find applicable and non applicable records. The author focuses on the discovery of context dependent rules because the data quality rules are contextual. The author presented a data driven tool which identifies CFDs threshold over a provided data instance which is useful in data cleaning and towards enforcing semantic data inconsistency. An algorithm is presented which search the approximate conditional rules and find exceptions to these rules that are dirty.

XIII. INTEGRATION OF HETEROGENEOUS DATABASES WITHOUT COMMON DOMAINS USING QUERIES BASED ON TEXTUAL SIMILARITY

In this paper the author mentioned that in many databases, it contains "name constraints" such as personal names, place names etc. in the previous work, it is assumed that the local contents can be aligned to an proper global domain with the help of normalization. but in certain circumstances this assumption does not hold. While identifying, if two name constants are same then it need detailed information of the world. In this paper the author rejected the assumption where the global domains can be built easily and replace the assumption by assigning the names are provided in natural language text instead of others. The author then proposes the logic called WHIRL that reasons explicitly about the similarity of local names which is measured using the vector-space model. Recently data integration system use domain-specific rules to normalize entity names and used the normalized versions of these names as keys.

In this paper the author shown that the accuracy of WHIRL's similarity joins are better while comparing to hand-coded integration schemes which is based on normalization. The author put forth the future work to handle the WHIRL in a distributed fashion.

XIV. IMPROVING DATA QUALITY

In this paper the author proposed a framework for improving the quality of the data based on CFDs. He shown that the problem for identifying optimal repairs and also for incrementally finding that optimal repairs are both NP-Complete, and heuristic algorithm is developed for both the problems. In order to increase the accuracy of the data, a statistical method is proposed which guarantees to find a repair that is above a predefined accuracy rate with high confidence.

The author put forth his further works to be carried on as to efficiently clean real-life data by consistency of both CFDs and inclusion dependencies, to study about the effective methods to automatically identifying useful CFDS from the data and to explore conditional constraints beyond CFDs.

XV. CONCLUSION AND FUTURE WORK

This paper detailed about various methods prevailing in literature for efficient discovery of matching dependencies. The concept of matching dependencies (MDs) has recently been proposed for specifying matching rules for object identification. Similar to the functional dependencies (with conditions), MDs can also be applied to various data quality applications such as detecting the violations of integrity constraints. The problem of discovering similarity constraints for matching dependencies from a given database instance is taken into consideration. This survey would promote a lot of research in the area of information mining.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

REFERENCES

1. Shaouxu Song, Lei Chen, "Efficient discovery of similarity constraints for matching dependencies", Data & Knowledge Engineering, Elsevier, 2013.
2. S. Abiteboul, R. Hull, V. Vianu, Foundations of Databases, Addison-Wesley, 1995.
3. R. Agrawal, T. Imielinski, A.N. Swami, Mining association rules between sets of items in large databases, SIGMOD Conference, 1993, pp. 207–216.
4. R. Bassée, J. Wijsen, Neighborhood dependencies for prediction, PAKDD, 2001, pp. 562–567.
5. C. Batini, M. Scannapieco, Data quality: concepts, methodologies and techniques, Data-Centric Systems and Applications, Springer, 2006.
6. L.E. Bertossi, S. Kolahi, L.V.S. Lakshmanan, Data cleaning and query answering with matching dependencies and matching functions, ICDT, 2011, pp. 268–279.
7. M. Bilenko, R.J. Mooney, W.W. Cohen, P. Ravikumar, S.E. Fienberg, Adaptive name matching in information integration, IEEE Intelligent Systems 18 (5) (2003) 16–23.
8. D. Bitton, J. Millman, S. Torgersen, A feasibility and performance study of dependency inference, ICDE, 1989, pp. 635–641.
9. P. Bohannon, W. Fan, F. Geerts, X. Jia, A. Kementsietsidis, Conditional functional dependencies for data cleaning, ICDE, 2007, pp. 746–755.
10. L. Bravo, W. Fan, F. Geerts, S. Ma, Increasing the expressivity of conditional functional dependencies without extra complexity, ICDE, 2008, pp. 516–525.
11. L. Bravo, W. Fan, S. Ma, Extending dependencies with conditions, VLDB, 2007, pp. 243–254.
12. T. Calders, R.T. Ng, J. Wijsen, Searching for dependencies at multiple abstraction levels, ACM Transactions on Database Systems 27 (3) (2002) 229–260.
13. F. Chiang, R.J. Miller, Discovering data quality rules, PVLDB 1 (1) (2008) 1166–1177.
14. W.W. Cohen, Integration of heterogeneous databases without common domains using queries based on textual similarity, SIGMOD Conference, 1998, pp. 201–212.
15. G. Cong, W. Fan, F. Geerts, X. Jia, S. Ma, Improving data quality: consistency and accuracy, VLDB, 2007, pp. 315–326.
16. A.K. Elmagarmid, P.G. Ipeirotis, V.S. Verykios, Duplicate record detection: a survey, IEEE Transactions on Knowledge and Data Engineering 19 (1) (2007) 1–16.
17. W. Fan, Dependencies revisited for improving data quality, PODS, 2008, pp. 159–170.
18. W. Fan, H. Gao, X. Jia, J. Li, S. Ma, Dynamic constraints for record matching, The VLDB Journal (2010) 1–26.
19. W. Fan, F. Geerts, L.V.S. Lakshmanan, M. Xiong, Discovering conditional functional dependencies, ICDE, 2009, pp. 1231–1234.
20. W. Fan, J. Li, X. Jia, S. Ma, Reasoning about record matching rules, PVLDB, 2009.
21. W. Fan, J. Li, S. Ma, N. Tang, W. Yu, Interaction between record matching and data repairing, SIGMOD Conference, 2011, pp. 469–480.
22. W. Fan, S. Ma, Y. Hu, J. Liu, Y. Wu, Propagating functional dependencies with conditions, PVLDB 1 (1) (2008) 391–407.
23. P.A. Flach, I. Savnik, Database dependency discovery: a machine learning approach, AI Communications 12 (3) (1999) 139–160.
24. J. Gardezi, L.E. Bertossi, I. Kiringa, Matching dependencies with arbitrary attribute values: semantics, query answering and integrity constraints, LID, 2011, pp. 23–30.
25. C. Giannella, E.L. Robertson, On approximation measures for functional dependencies, Information Systems 29 (6) (2004) 483–507.
26. L. Golab, H.J. Karloff, F. Korn, A. Saha, D. Srivastava, Sequential dependencies, PVLDB 2 (1) (2009) 574–585.
27. L. Golab, H.J. Karloff, F. Korn, D. Srivastava, B. Yu, On generating near-optimal tableaux for conditional functional dependencies, PVLDB 1 (1) (2008) 376–390.
28. L. Gravano, P.G. Ipeirotis, N. Koudas, D. Srivastava, Text joins in an rdbms for web data integration, WWW, 2003, pp. 90–101.
29. Y. Huhtala, J. Kärkkäinen, P. Porkka, H. Toivonen, Efficient discovery of functional and approximate dependencies using partitions, ICDE, 1998, pp. 392–401.
30. Y. Huhtala, J. Kärkkäinen, P. Porkka, H. Toivonen, Tane: an efficient algorithm for discovering functional and approximate dependencies, The Computer
31. I.F. Ilyas, V. Markl, P.J. Haas, P. Brown, A. Aboulnaga, Cords: automatic discovery of correlations and soft functional dependencies, SIGMOD Conference, 2004, pp. 647–658.
32. R.S. King, J.J. Legendre, Discovery of functional and approximate functional dependencies in relational databases, JAMDS 7 (1) (2003) 49–59.
33. J. Kivinen, H. Mannila, Approximate inference of functional dependencies from relations, Theoretical Computer Science 149 (1) (1995) 129–149.
34. N. Koudas, A. Saha, D. Srivastava, S. Venkatasubramanian, Metric functional dependencies, ICDE, 2009, pp. 1275–1278.
35. S. Kramer, B. Pfahringer, Efficient search for strong partial determinations, KDD, 1996, pp. 371–374.
36. S.E. Madnick, H. Zhu, Improving data quality through effective use of data semantics, Data & Knowledge Engineering 59 (2) (2006) 460–475.
37. H. Mannila, K.-J. Räihä, Design of Relational Databases, Addison-Wesley, 1992.
38. H. Mannila, K.-J. Räihä, Algorithms for inferring functional dependencies from relations, Data & Knowledge Engineering 12 (1) (1994) 83–99.
39. A. McCallum, K. Nigam, L.H. Ungar, Efficient clustering of high-dimensional data sets with application to reference matching, KDD, 2000, pp. 169–178.
40. G. Navarro, A guided tour to approximate string matching, ACM Computing Surveys 33 (1) (2001) 31–88.
41. B. Pfahringer, S. Kramer, Compression-based evaluation of partial determinations, KDD, 1995, pp. 234–239.
42. T. Scheffer, Finding association rules that trade support optimally against confidence, Intelligent Data Analysis 9 (4) (2005) 381–395.
43. J.C. Schlimmer, Efficiently inducing determinations: a complete and systematic search algorithm that uses optimal pruning, ICML, 1993, pp. 284–290.
44. S. Song, L. Chen, Discovering matching dependencies, CIKM, 2009, pp. 1421–1424.
45. S. Song, L. Chen, Differential dependencies: reasoning and discovery, ACM Transactions on Database Systems 36 (4) (2011).
46. S. Song, L. Chen, P.S. Yu, On data dependencies in dataspaces, ICDE, 2011, pp. 470–481.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

47. U. ul Hassan, S. O'Riain, E. Curry, Leveraging matching dependencies for guided user feedback in linked data applications, Proceedings of the Ninth International Workshop on Information Integration on the Web, IIWeb '12, ACM, New York, NY, USA, 2012, pp. 5:1–5:6.
48. H. Wang, R. Liu, Privacy-preserving publishing microdata with full functional dependencies, Data & Knowledge Engineering 70 (3) (2011) 249–268.
49. C.M. Wyss, C. Giannella, E.L. Robertson, Fastfids: a heuristic-driven, depth-first algorithm for mining functional dependencies from relation instances —extended abstract, DaWaK, 2001, pp. 101–110.
50. B.Powmeya , Nikita Mary Ablett ,V.Mohanapriya,S.Balamurugan,"An Object Oriented approach to Model the secure Health care Database systems,"In proceedings of International conference on computer , communication & signal processing(IC³SP)in association with IETE students forum and the society of digital information and wireless communication,SDIWC,2011,pp.2-3
51. Balamurugan Shanmugam, Visalakshi Palaniswami, "Modified Partitioning Algorithm for Privacy Preservation in Microdata Publishing with Full Functional Dependencies", Australian Journal of Basic and Applied Sciences, 7(8): pp.316-323, July 2013
52. Balamurugan Shanmugam, Visalakshi Palaniswami, R.Santhya, R.S.Venkatesh "Strategies for Privacy Preserving Publishing of Functionally Dependent Sensitive Data: A State-of-the-Art-Survey", Australian Journal of Basic and Applied Sciences, 8(15) September 2014.
53. S.Balamurugan, P.Visalakshi, V.M.Prabhakaran, S.Charyana, S.Sankaranarayanan, "Strategies for Solving the NP-Hard Workflow Scheduling Problems in Cloud Computing Environments", Australian Journal of Basic and Applied Sciences, 8(15) October 2014.
54. Charanyaa, S., et. al., , A Survey on Attack Prevention and Handling Strategies in Graph Based Data Anonymization. International Journal of Advanced Research in Computer and Communication Engineering, 2(10): 5722-5728, 2013.
55. Charanyaa, S., et. al., Certain Investigations on Approaches for Protecting Graph Privacy in Data Anonymization. International Journal of Advanced Research in Computer and Communication Engineering, 1(8): 5722-5728, 2013.
56. Charanyaa, S., et. al., Proposing a Novel Synergized K-Degree L-Diversity T-Closeness Model for Graph Based Data Anonymization. International Journal of Innovative Research in Computer and Communication Engineering, 2(3): 3554-3561, 2014.
57. Charanyaa, S., et. al., , Strategies for Knowledge Based Attack Detection in Graphical Data Anonymization. International Journal of Advanced Research in Computer and Communication Engineering, 3(2): 5722-5728, 2014.
58. Charanyaa, S., et. al., Term Frequency Based Sequence Generation Algorithm for Graph Based Data Anonymization International Journal of Innovative Research in Computer and Communication Engineering, 2(2): 3033-3040, 2014.
59. V.M.Prabhakaran, Prof.S.Balamurugan, S.Charanyaa," Certain Investigations on Strategies for Protecting Medical Data in Cloud", International Journal of Innovative Research in Computer and Communication Engineering Vol 2, Issue 10, October 2014
60. V.M.Prabhakaran, Prof.S.Balamurugan, S.Charanyaa," Investigations on Remote Virtual Machine to Secure Lifetime PHR in Cloud ", International Journal of Innovative Research in Computer and Communication Engineering Vol 2, Issue 10, October 2014
61. V.M.Prabhakaran, Prof.S.Balamurugan, S.Charanyaa," Privacy Preserving Personal Health Care Data in Cloud" , International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 2, October 2014
62. P.Andrew, J.Anish Kumar, R.Santhya, Prof.S.Balamurugan, S.Charanyaa, "Investigations on Evolution of Strategies to Preserve Privacy of Moving Data Objects" International Journal of Innovative Research in Computer and Communication Engineering, 2(2): 3033-3040, 2014.
63. P.Andrew, J.Anish Kumar, R.Santhya, Prof.S.Balamurugan, S.Charanyaa, " Certain Investigations on Securing Moving Data Objects" International Journal of Innovative Research in Computer and Communication Engineering, 2(2): 3033-3040, 2014.
64. P.Andrew, J.Anish Kumar, R.Santhya, Prof.S.Balamurugan, S.Charanyaa, " Survey on Approaches Developed for Preserving Privacy of Data Objects" International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 2, October 2014
65. S.Jeevitha, R.Santhya, Prof.S.Balamurugan, S.Charanyaa, " Privacy Preserving Personal Health Care Data in Cloud" International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 2, October 2014.
66. K.Deepika, P.Andrew, R.Santhya, S.Balamurugan, S.Charanyaa, "Investigations on Methods Evolved for Protecting Sensitive Data", International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 4, December 2014.
67. K.Deepika, P.Andrew, R.Santhya, S.Balamurugan, S.Charanyaa, "A Survey on Approaches Developed for Data Anonymization", International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 4, December 2014.
68. S.Balamurugan, S.Charanyaa, "Principles of Social Network Data Security" LAP Verlag, Germany, ISBN: 978-3-659-61207-7, 2014
69. S.Balamurugan, S.Charanyaa, "Principles of Scheduling in Cloud Computing" Scholars' Press, Germany,, ISBN: 978-3-639-66950-3, 2014
70. S.Balamurugan, S.Charanyaa, "Principles of Database Security" Scholars' Press, Germany, ISBN: 978-3-639-76030-9, 2014

BIOGRAPHY

R.Santhya and S.Latha are currently pursuing their B.Tech. degree in Information Technology at KalaingarKarunanidhi Institute of Technology, Coimbatore, Tamil Nadu, India. Their areas of research interests include Network Security, Cloud Computing and Database Security.



Prof.S.Balamurugan obtained his B.Tech degree in Information Technology from P.S.G. College of Technology, Coimbatore, Tamil Nadu, India and M.Tech degree in Information Technology from Anna University, Tamil Nadu, India respectively. He is currently working towards his PhD degree in Information Technology at P.S.G. College of Technology, Tamil Nadu, India. At present he holds his credit **65 papers International Journals and IEEE/ Elsevier International Conferences**. He is currently working as Assistant Professor in the Department of Information Technology, Kalaingar Karunanidhi Institute of Technology, Coimbatore, Tamil Nadu, India affiliated to Anna University TamilNadu, India. He is **State Rank holder** in schooling. He was **University First Rank holder** M.Tech. Semester Examinations at Anna University, Tamilnadu, India. He served as a Joint Secretary of IT Association, Department of Information Technology,

PSG College of Technology, Coimbatore, Tamilnadu, India. He is the **recipient of gold medal and certificate of merit** for best journal publication by his host institution **consecutively for 3 years**. Some of his professional activities include invited Session Chair Person for two Conferences. He has guided 16 B.Tech projects and 2 M.Tech. projects. He has won a best paper award in



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

International Conference. His areas of research interest accumulate in the areas of Data Privacy, Database Security, Object Modeling Techniques, and Cloud Computing. He is a life member of ISTE,CSI. **He has authored a chapter in an International Book "Information Processing" published by I.K. International Publishing House Pvt. Ltd, New Delhi, India, 978-81-906942-4-7. He is the author of 3 books titled "Principles of Social Network Data Security", ISBN: 978-3-659-61207-7, "Principles of Scheduling in Cloud Computing" ISBN: 978-3-639-66950-3, and "Principles of Database Security", ISBN: 978-3-639-76030-9.**



S.Charanyaa obtained her **B.Tech** degree in Information Technology and her **M.Tech** degree in Information Technology from Anna University Chennai, Tamil Nadu, India. She was **gold medalist** in her B.Tech. degree program. She has to her credit **27 publications in various International Journals and Conferences**. Some of her outstanding achievements at school level include **School First Rank holder in 10th and 12th grade**. She was working as Software Engineer at Larsen & Turbo Infotech, Chennai for 3 years where she got promoted as Senior Software Engineer and worked for another 2 years. She worked at different verticals and worked at many places including Denmark, Amsderdam handling versatile clients. She is also the recipient of **best team player award for the year 2012 by L&T**. Her areas of research interest accumulate in the areas of

Database Security, Privacy Preserving Database, Object Modeling Techniques, and Cloud Computing. **She is the author of 3 books titled "Principles of Social Network Data Security", ISBN: 978-3-659-61207-7, "Principles of Scheduling in Cloud Computing" ISBN: 978-3-639-66950-3, and "Principles of Database Security", ISBN: 978-3-639-76030-9.**