ISSN 2250-3153

Semantic Measure Model Based Educational Development System for Using Similarity Attribute in Data Warehouse System

R. Sundara Bharathi, Dr. R. Mala

[*] Research Scholar, Bharathidasan University, Marudupandiyar College, Thanjavur, India
[**] Assistant Professor,Department of Computer Science,Alagappa University College of Arts and Science,Paramakudi

Abstract- Data warehouse (DW) is a group of combined databases planned to maintenance executive decision-making besides problem-solving purposes. The semantic multi attribute model focoused on similarity contained in the data warehouse, in DW iscritical for the certain group of user. The semantic access right in viewing the dimensional schema should be properly specified according to the level of privilege. A set of security controls has been established to protect this crucial asset. It holds both highly comprehensive thenreducedmodern data connecting to numerousgroups, topics, or areas. Semantic Similarity Actions is one of the latestknow-hows that giveentree to various and isolatedmaterialfoundationsattractive the decision-makingcompetences of the group. The Semantic Similarity Measures provides an efficient storing mechanism which can identify the information of interest to the organization. The data repository contains information from multiple sources and presents it in an integrated manner to the end user. In this Semantic Similarity used context the filtration of data is realized previously the data warehouse.

Index Terms- Semantic, Similarity model, educational development, warehouse, metadata

I. Introduction

Data warehousing structure, a huge amount of unemployment was required to provision various choice support surroundings. In superiorworries it was characteristic for shared decision provisionsenvironments to operate separately. Though each disorder served dissimilar users, they frequentlycompulsory abundant of the undistinguishable deposited data.

Exactlycalculating the semantic similarity among words is a significantproblematic in thewarehouse, information retrieval, and normal language processing. warehouse requests such as, public extraction, relation detection, and entity disambiguation, necessitate the ability to preciselyamount the semantic similarity amongst concepts or entities. In material retrieval, one of the main problems is to retrieve a set of forms that is semantically connected to a given manipulator query.

Confidentiality is to confirm that user'scontainer only entree the evidence which they have rights for. In the case of warehouse Meta data replicas, privacy is crucial, since business material is very delicate and can be exposed by performing a simple query. Occasionally, warehouse Meta data models also store evidence regarding private or individualfeatures of those, like documentation data or even spiritual beliefs.

In this case, confidentiality is redefined as privacy. Comparisondemonstratingmethod which we take as our base permits us to perfect complex data warehouse situations and at the same time reflects no severe and comprehensive hierarchies, corrupt dimensions, degenerate facts or many-to-many relationsamid facts and a specificmeasurement with the consistentcharacteristics. Henceforth, we also try to deliver a wider method for modeling a commercial data warehouse, somewhat than amethod aimed at demonstratingtrivial data marts. Lastly, we cover some applicableconditionssuch as long as role orders and section groups for operators, and not objective the typical safety levels of multilevel refuge.

II. RELATED WORK

Performing such complex analysis requires that the meaning of the warehousing repository data be understood entirely, along with its provenance, heritage and lineage, which is achieved by means of metadata [1]. Metadata is not only “data about data”; it has rather a broader sense and role, as it concentrates the totality of information and knowledge of the actual data existent within an enterprise [2]. Metadata captures general and specific characteristics, offers a context and meaning to the raw data and creates the semantic layer of the organization’s informational system [3].

This semantic layer ensures a proper interpretation and understanding of the organization’s data by all the actors involved in its exploitation and usage. The creativity’s informational schemes contain numerouskinds of metadata [4], from commercial and technical metadata, stationary and dynamic metadata, to expressive, structural and directorial. In order to accomplish its functions, metadata has to be properly managed [5].

It also helps to reduce the determinations of data warehouse management and to recover the withdrawal of evidence from the investigativesituation [6]. Inside the data warehousing situation, the metadata composed from numerous sources is deposited in the metadata fountain, which aims to offer reliable and dependableadmission to the data and enablesenquiring and glancingprocesses from the end-user [7].

The initiative data warehouse buildingsignifies the most comprehensive and multifaceted architectural type in the data warehousing atmosphere and is the consequence of a complete enterprise-wide analysis of data necessities [8]. Its key goal is to provide an integrated data foundation, defined by atomic level data maintained in normalized process in the information warehouse, then to allow the structure of several multidimensional opinions of combined data complete from the fundamental data warehouse [9]. Warehouse team needs to comprehend the metadata about the foundationschemes to build the data warehouse expansion. The model suggests that a sheet of metadata has to be about the foundationschemes [10].

III. SEMANTIC SIMILARITY MEASURE FOR EDUCATIONAL GROWTH SYSTEM

A novel Semantic Similarity Measures algorithm (SSMA)described for multi attribute based ware housing for optimal combination page counts and its co-occurrence measures and lexical pattern based warehousing is learned using Educational Growth system.Thencharacterize the quality set over a universe of properties that a decrypt needs to have with a specific end goal to decode the figure message, and uphold it on the substance. In this way, every client with an alternate arrangement of credits is permitted to decode distinctive bits of information per the security strategy.

We formally define a SM as a function:

$\sigma k: Ek \times Ek \rightarrow \mathcal{R}$

with Ek the set of elements of type $k \in K$ and K, the various types of elements which can be compared regarding their semantics, e.g., K = {words, concepts, sentences, texts, web pages, instances annotated by concepts, …}, and

$\mathcal{R} = \{[0,1], \mathbb{R}+, \{a, b, c \dots\}\}$.

This expression can be generalized to take into account the comparison of elements of different types. This could be interesting to evaluate entailment of texts or to compare words and concepts to mention a few. However, in this restrict our study to the comparison of pairs of elements of the same nature. We stress that similarity measure must implicitly or explicitly take advantage of semantic evidences. As an example, as we said in the introduction, measures comparing words through their syntactical similarity cannot be considered to be similarity measure; recall that semantics refers to evidences regarding the meaning or the nature of compared elements.

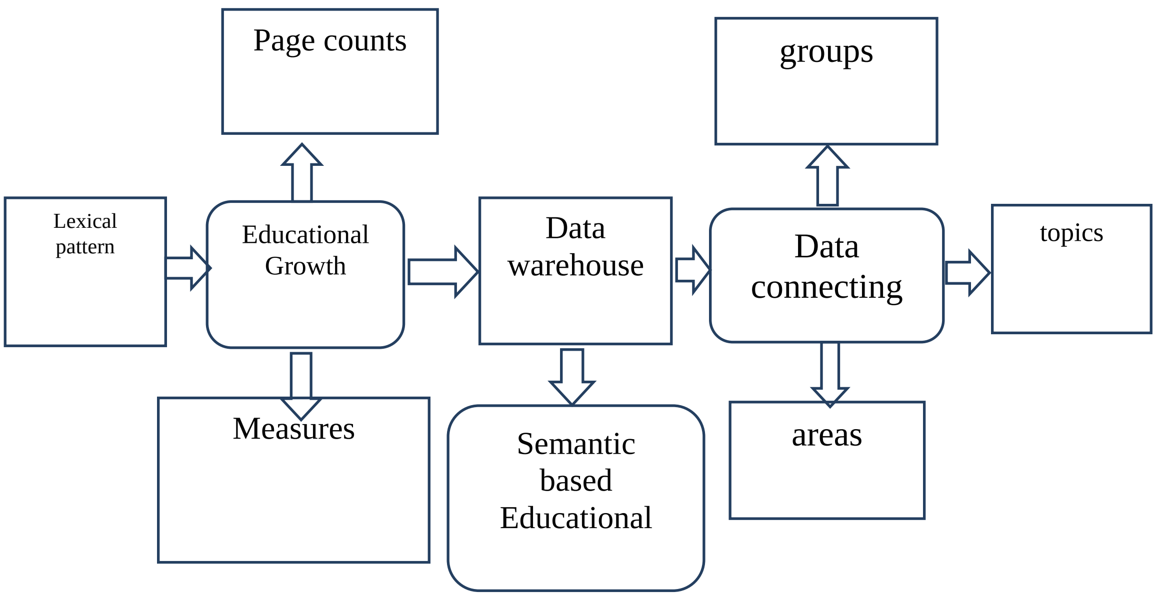


Figure 3.1 over all architecture diagram

A data warehouse is secondhand for retrieving brainy information from huge files. Presently these databases are dispersed across the world. Dispersed data necessity be recovered from multiple locations in to the data warehouse, so here is an obligation for a secure broadcast and upholding confidentiality.We develop our proposal according to a bottom-up algorithm for searching association rules. Our algorithm consists of an adaptation of the traditional sematic algorithm to multidimensional data. In addition, in order to focus on the discovered associations and validate them, we provide a visual representation based on the graphic semiology principles. Such a representation consists of a graphic encoding of frequent patterns and association rules in the same multidimensional space as the one associated with the mined data cube.

3.1 Semantic based Educational Development system:

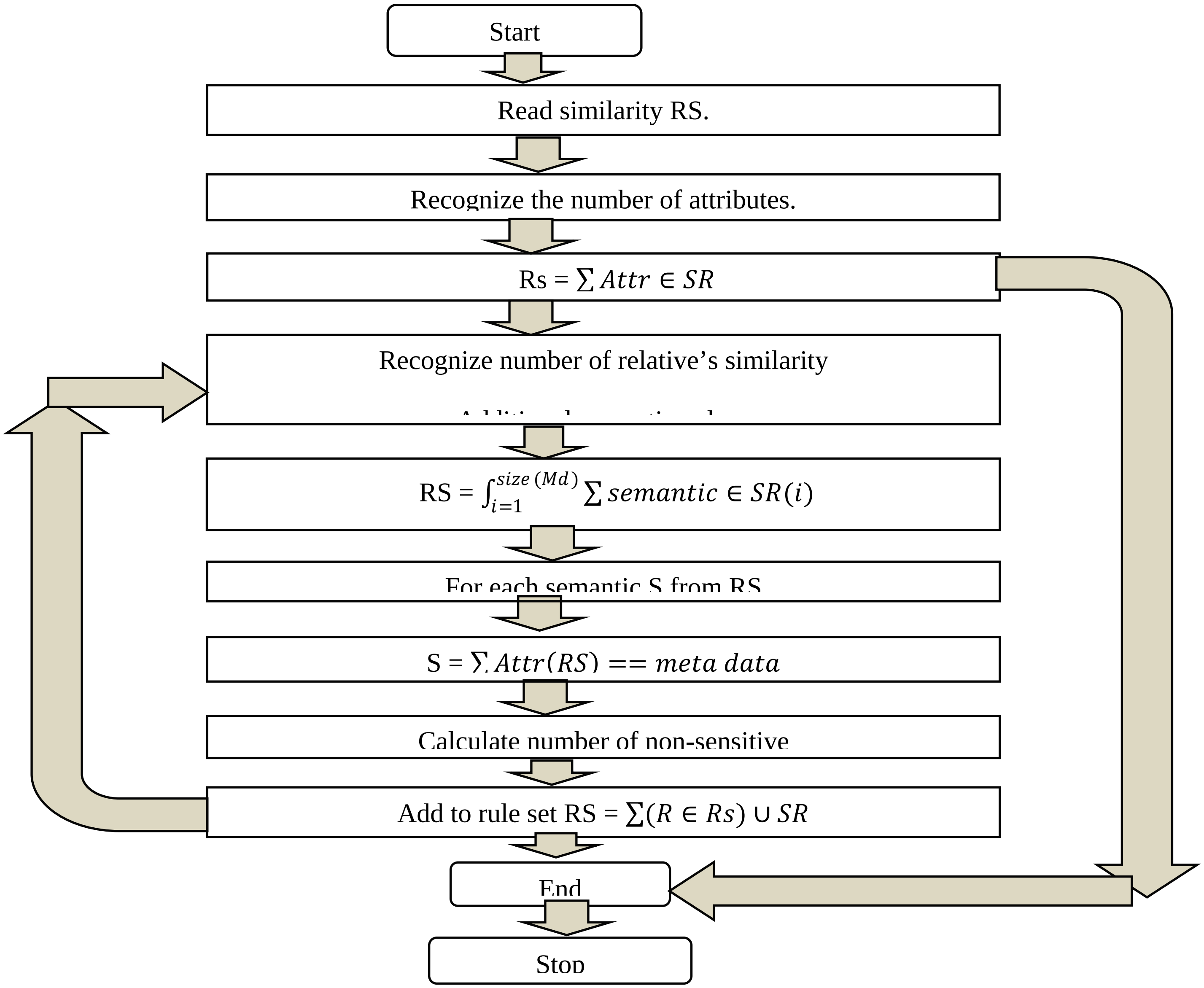
Data warehouse technology contains tools for repeatedly and logicallychanginga large quantity of data in to informationapplicable to the obligation of an end user. Semantic Informationexposed by data warehouse can also recovercomplexevidence about entitiescooperating the individual's right to privacy.Semantic based educational development system are perhaps even more semantic in a data warehouse then by clarification a data warehouse contains data united from manifold sources, and therefore from the viewpoint of a hateful separate trying to reduced info a data warehouse container be one of the utmost rewarding boards in an initiative.

Algorithm:
Input: Meta data Md.
Output: semantic Rule set SR.
Start
Read similarityRS.
For each similarity database RS
Recognize the number of attributes.
Rs = $\sum Attr \in SR$
Recognize number of relative'ssimilarity
Additionalsemantic values
RS = $\int_{i=1}^{size\,(Md)} \sum semantic \in SR(i)$
For each semanticS from RS
Divide number of semantic
S = $\sum Attr(RS) == meta\ data$
Calculate number of non-sensitive
Add to rule set RS = $\sum(R \in Rs) \cup SR$
End
End
Stop.

Above algorithm is show similarity based Meta data analysis for all educational development process in ware house. Such measures take into account the frequency distribution of different attribute values in a given data set to define similarity between two categorical attribute values. In this paper, we study a variety of similarity measures proposed in diverse research fields ranging from statistics to ecology as well as many of their variations. Each measure uses the information present in the data uniquely to define similarity.

The distinction between approaches that can and cannot be assimilated to similarity is sometime narrow; there is no clear border distinguishing non-semantics to semantic-augmented approaches, but rather a range of approaches. Some explanations can be found in the difficulty to clearly characterize the notion of Meaning. For instance, someone can say that measures used to evaluate lexical distance between words, such as edit distances, capture semantic evidences regarding the meaning of words. Indeed, the sequence of characters associated to a word derives from its etymology which is sometimes related to its meaning,

Flow char for Semantic based Educational Development system



The personalization process is based on the user knowledge about data aggregation, which is the basis of data organization in multidimensional models. Then, our key idea consists of generating new analysis axes by dynamically creating new dimension hierarchies or extending existing old ones. More precisely, we define new granularity levels inside the data warehouse schema. This provides a real time evolution of dimension hierarchies to cope with personalized analysis needs.

1.2 A similarity measure for attribute model:

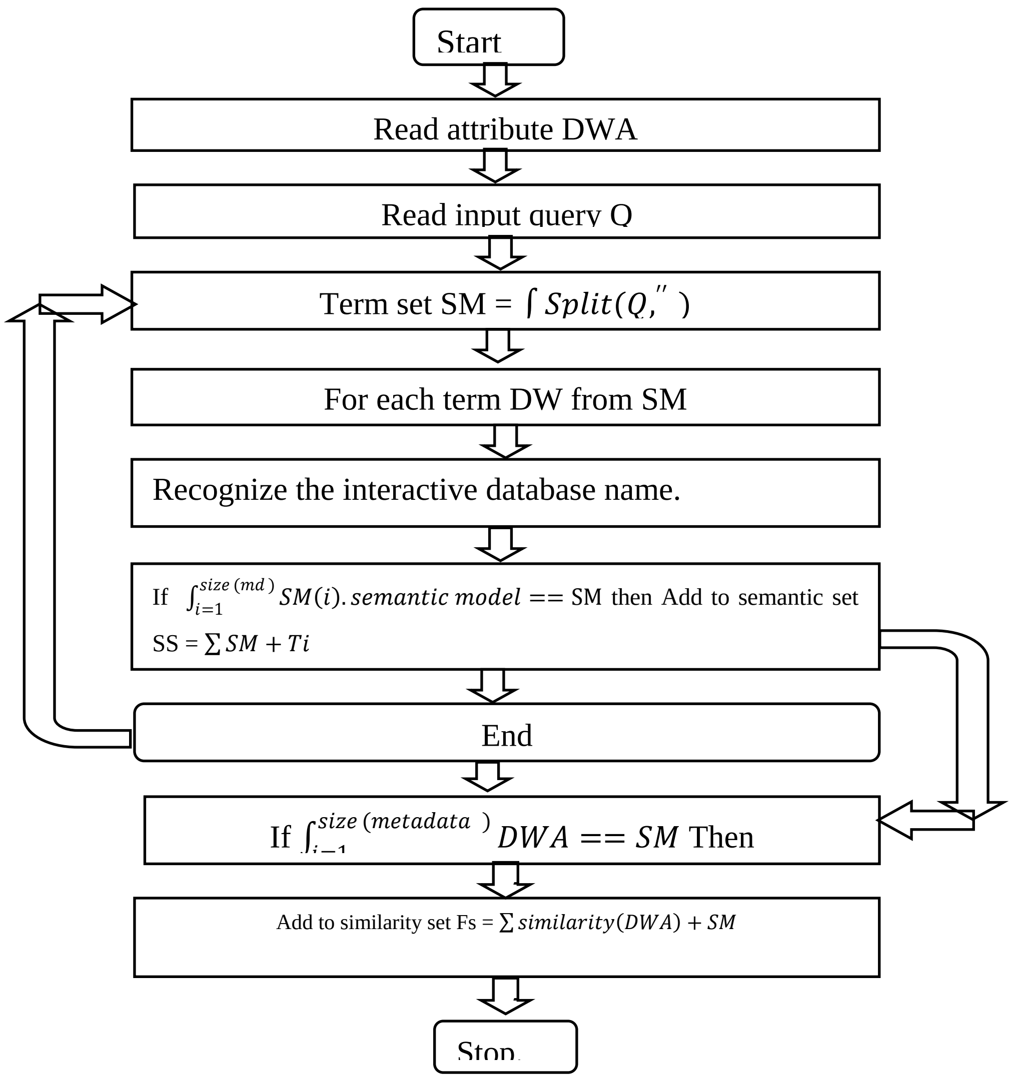
The similarity measure for DWs encompasses several creativities to comprise attribute modelin the DW enterprise. In the define an example model for DW safety based on metadata, its main goal is to decrease user enquiries to only those statistics which remain to be understood by that user. However, this fixes not permit the requirement of complex limits of confidentiality such as deny-allow admittance to a special user uniting groups and security restraints.To achieve the value of a data warehouse, input data must be transformed into an analysis-ready format. In the case of numerical data, data warehousing systems often provide tools to assist in this process. Unfortunately, standard tools are inadequate for producing a relevant analysis axis when data are complex. In such cases, the data warehousing process should be adapted in response to evolving data and information requirements. In a data warehousing process, the data integration phase is crucial. Data integration is a hard task that involves reconciliation at various levels: data models, data schema, data instances, and semantics.

ISSN 2250-3153

Algorithm:

Input: Data warehouse attribute DWA,
Output: semantic model SM
Start
Read attribute DWA.
Read input query Q.
Term set SM = $\int Split(Q,'')$
For each term DW from SM
Recognize the interactive database name.
If $\int_{i=1}^{size(md)} SM(i).semantic\ model == $ SM then
Add to semantic set SS = $\sum SM + Ti$
End
If $\int_{i=1}^{size(metadata)} DWA == SM$ Then
Add to similarity set Fs = $\sum similarity(DWA) + SM$
End.
Stop.

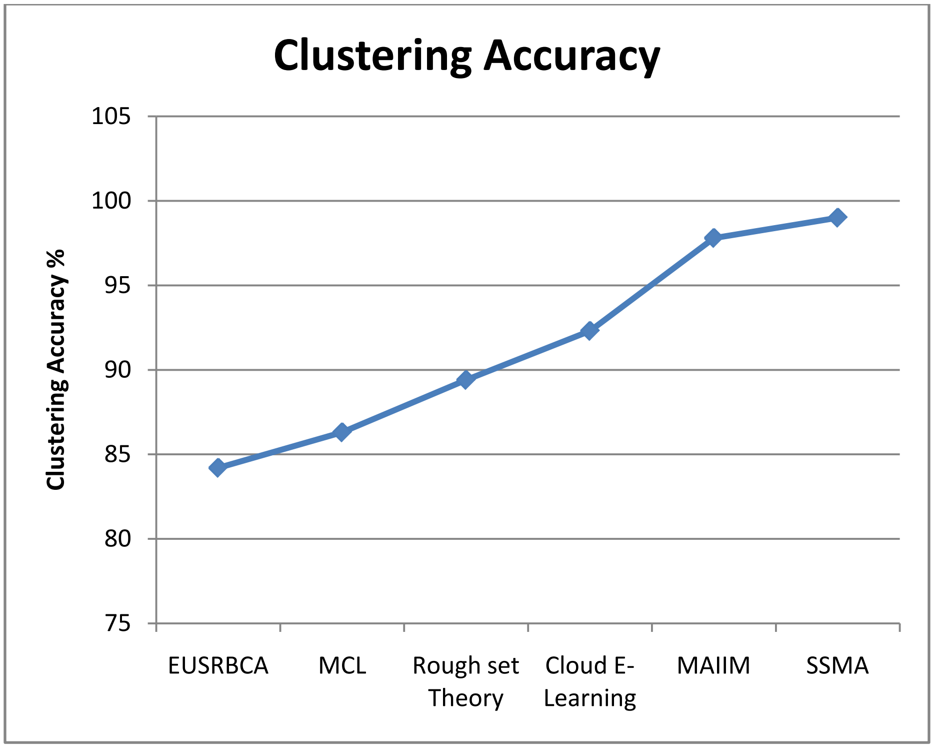
The clean is basically recycled to encrypt the collected data which is awaiting from various active data, dispersed data and external data sources and after encryption the correspondence data permits into the data warehouse and the determined data is deposited in data warehouse.Data rules instantiate structure rules, meaning that they define the aggregation link on data themselves. To build a new level, the user defines one structure rule, and various data rules. Data rules define the various instances to be created in the new level and the condition that defines the link with the instances of the lower level. Each data rule corresponds to one instance of the level to be created.To design and build data warehouses, traditional data and goal-driven approaches bear the latent risk of not meeting user requirements. Therefore, user-driven developing approaches seem promising for successful completion of data warehouse projects.



Active semantic documentationfurther extends the scope of clinical decision support by gettingits semantic features from automatic semantic annotationto automatic decision making on the text of the archive byapplying logically relevant principles to the segments of thearchive.Semantic annotations can likewise be used for semantic reports to give a comprehensibility support to customers.In this paper, the benefits of using ontology and semanticannotations are tested in terms of providing an interpretativelayer between the text generated by clinicians andits display to consumers in the context of similarity.

IV. RESULTS AND DISCUSSION

The prearrangedmethod has been appreciated and projected for its competence. The consequencedemonstrations that the planned method has formedeffectual results in all the effects of data warehousing then the techniquerecovers the routine of educational scheme. The method can be easily adapted to rank and support the online education and E-learning approaches. Also the method has been focused on supporting various other solutions like business intelligence, Administrative support and many more. Accuracy is not really a reliable metric for the real performance of a classifier when the number of samples in different classes vary greatly (unbalanced target) because it will yield misleading results.



Graph 1: Comparison of clustering accuracy produced

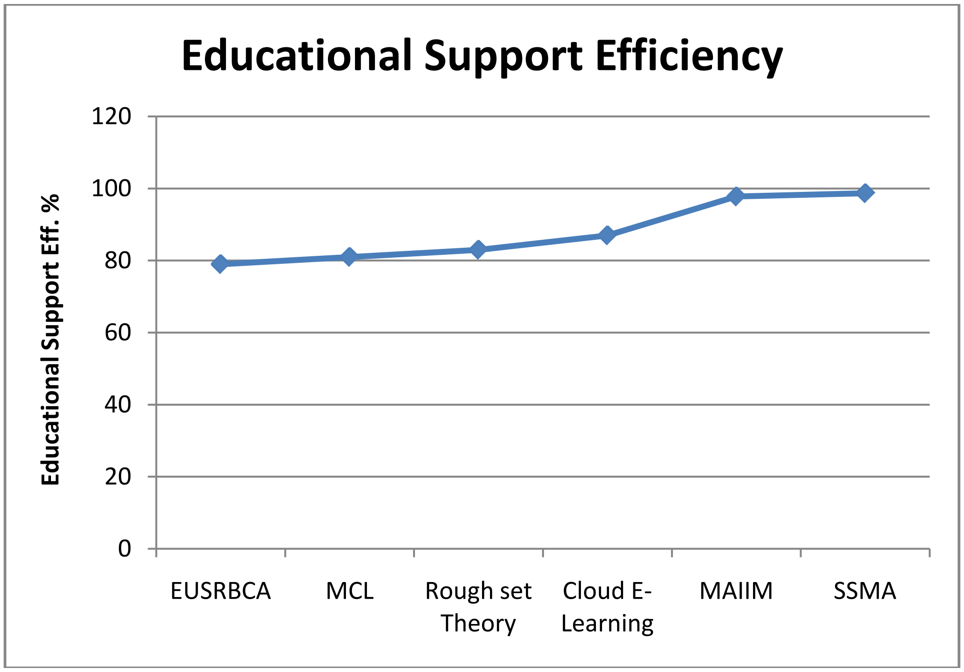
Techniques	Clustering Accuracy%
EUSRBCA	84.5
MCL	86.9
Rough set Theory	89.9
Cloud E-Learning	93.5
MAIIM	96.4
SSMA	98.9

Table 1 clustering accuracy in different algorithm

The Graph 1, demonstrations the evaluation of clustering correctnesstwisted by dissimilarapproaches on warehousing the data. It demonstrations clearly that the planned method has twisted amoreeffective result than other approaches.This recommends denial of any characteristic or any single customer in a quality social event would impact all customers in the get-together. It may achieve clustering of the windows of clustering.That the plan just requires just a single authenticator for each square, it has two serious downsides. To begin with, since the confirmation procedure requires mystery material, there will be security issues while stretching.

Techniques	Educational Support %
EUSRBCA	79.2
MCL	81.6
Rough set Theory	83.2
Cloud E-Learning	87.9
MAIIM	97.8
SSMA	98.8

Table 2 Educational growth support efficiency



\

Graph 2: Evaluation of educational support efficiency

The Graph 2, demonstrations the evaluation of educational support efficiency formed by dissimilar methods and it demonstrationsobviously that the planned method has formed more provision than other methods.At least one of these credits will be utilized to approve the personality of the client and hence the approach resolver comprises of various subcomponents to confirm the demand of the client in view of his traits and entitle the gave character related standards and arrangements

V. CONCLUSION

The method computes the semantic similarity measure based algorithm, and using the measure computed the method validate the user request to maintain integrity. The planned method recovers the routine of the query processing and improves clustering efficiency also. In addition to improving the accuracy, educational support efficiency attains to enhance the future process to produce best results.

REFERENCES

[1] N. Katic G. Quirchmayr2 J. Schiefer M. Stolba AM. Tjoa ."A Prototype Model for Data Warehouse Security Based on Metadata", in proc,IEEE,2000,vol no:5.

[2] Sajjad Solat, "A priority based approach for protection coordination in distribution system",in proc,IEEE,2015,vol no :6.

[3] Chung-Hsien Kuo, Han-Pang Huang,"The Development of PC-Based Highly Model-Mixed Engine Dispatching Systems", in proc,IEEE,2007,vol no:8.

[4] Wang Hui1, Wang Jing, Zheng Tao,"Analysis of Decision Tree Classification Algorithm Based on Attribute Reduction and Application in Criminal Behavior", in pro,IEEE,2010.vol no:7.

[5] Jian Guan, " A Multi-Solver Scheme Based on Combined Field Integral Equations for Electromagnetic Modeling of Highly Complex Objects "in proc,IEEE,2016,vol no:9.

[6] H. Baali, M. J. E Salami, R. Akmeliawati and A. M. Aibinu,"Analysis of the ECG Signal using SVD-Based Parametric Modelling Technique",in proc,IEEE,2011,vol no:5.

[7] H. Baali, R. Akmeliawati, M. J. E. Salami, A. Khorshidtalab," ECG Parametric Modeling Based on Signal Dependent Orthogonal Transform",in proc ,IEEE,2014,vol no:6.

[8] Sara Reese Hedberg," AI fools for businessm process modeling",in proc,IEEE,2000,vol no:8.

[9] David van Oheimb Siemens," Formal security analysis in industry, at the example of Electronic Distribution of aircraft Software (EDS)", in proc,IEEE,2007,vol no:7.

[10] Juan Manuel Pe´ rez, Rafael Berlanga, Marı´a Jose´ Aramburu," Integrating Data Warehouses with Web Data: A Survey,"in proc,IEEE,2010,vol no:5.

[11] Rizwan Alam Siddiqui, Mehnaz Adnan," An Ontology Based Framework For Navigation Support in Electronic Discharge Summaries", in proc,IEEE,2015,vol no:5.

[12] Sondre Skatter, Stephan Olesinski," Energy-Resolving Photon-Counting Detectors in CT Based Explosive Detection Systems (EDS)",in proc,IEEE,2014,vol no:9.

[13] YIN Hang, CHANG Guiran,"A Cold-start Recommendation Algorithm Based on New User's Implicit Information and Multi-Attribute Rating Matrix", in proc ,IEEE,2015,vol no:7.

[14] Shi Ziyan," The Research of the Training Assistant Decision-making support System Based on Data Mining Technology",in proc,IEEE, 2016, vol no:4.

[15] P. Stasa, F. Benes, J. Svub, Z. Neustupa, "Monitoring and evidence of cars in private objects". In Proceedings of the 20 14 15tl' International Carpathian Control Conference, ICCC 20 14. Los Alamitos IEEE Computer Society, 20 14, s. 545-550.

[16] Vaculik, 1. Michalek, P. Kolarovszki, "Principles of selection, implementation and utilization of RFID in supply chain management" in Promet - Traffic & Transportation, Volume: 2 1, Issue: 1, Pages: 4 1-48, Published: 2009; ISSN 0353-5320.

[17] S. Amritpal, Nitin Umesh "Implementing Log Based Security in Data Warehouse", International Journal of Advanced Computer, 2013

[18] Oracle Corporation, "Oracle Advanced Security Transparent Data Encryption Best Practices", Oracle White Paper, July 2010.

[19] E. Lo, N. Cheng, and W. Hon, "Generating Databases for Query Workloads", Int. Conf. on Very Large DataBases (VLDB), 2010.

[20] Santos, R. J., Bernardino J., Vieria, "Balancing Security and Performance for Enhancing Data Privacy in Data Warehouses", International Joint Conference Of IEEE TrustCom-11/IEEE ICESS- 11/FCST -11, 2011.

AUTHORS

First Author – R. Sundara Bharathi, M.Sc., M.C.A., M.Phil. Research Scholar, Bharathidasan University, Marudupandiyar College, Thanjavur, India, bharathikartsphd@gmail.com
Second Author – Dr. R. Mala, M.C.A., M.Phil., Ph.D. Assistant Professor,Department of Computer Science,Alagappa University College of Arts and Science,Paramakudi.