ISSN: 2278 – 7798
International Journal of Science, Engineering and Technology Research (IJSETR)
Volume 2, Issue 2, February 2013

# An Effective Repeated Rule Acquisition Using Rule Ontology from Similar Web Sites based on Genetic Algorithm

**G.Senthil Kumar, K.K.Kanaga Mathan Mohan**

***Abstract*— Knowledge is an important component of mainly Semantic Web applications and ontology. Ontology's are widely used as a model for knowledge depiction and formalization, to characterize user profiles in personalized web information gathering. The ontology can reduce the amount of information and reduce the exertion of utilizing the information in rule acquisition, because it is generalized and specifically reorganized for rule acquisition. Furthermore, the ontology can be accumulated and reused throughout repeated rule acquisition. The main contribution of existing research is that the inclusive and detailed rule composition procedure with examples and its estimation. In our research work we are implementing the optimized rule of acquisition result with our method through the concept of selecting exact parts that contain rules from Web pages to increase the accuracy of result. With that the optimization method is implemented which is referred as the screen method from WebPages with use of genetic Techniques to extract the rule optimally.**

***Index Terms*—Ontology, RuleToOnto, Genetic algorithm, Rule acquisition, Association rule, Best First Search.**

## I. INTRODUCTION

Rule acquisition is as essential as ontology acquisition; still however rule acquisition is at a standstill a bottleneck in the operation of rule-based systems. This is time consuming and difficult, because it wants knowledge proficient as well as domain specialists, and in attendance are statement problems among them. Nevertheless, little bits rules have previously been implied in Web pages, and it is potential to acquire them from Web pages in the same manner as ontology learning. The upper part is the text in a Web page explaining the return policies of Barnes & Noble.com, which is one of the representative online bookstores. The lower part shows acquired rules from the text. We can see that the words with bold font in the upper part are used as the variables and values of the acquired rules in the lower part. That is, most of the rule components already exist in Web pages. It means that we can acquire rules more easily by using an automatic rule acquisition method rather than the old method with domain experts and knowledge experts. As we repeat the rule acquisition process across several sites, we can accumulate rules. However, as the size of rule base increases, it becomes hard to reuse rules. Therefore, we used an ontology named RuleToOnto, which symbolizes the information about rules including terms, rule component types, and rule structures.

Previously, in [2] established the OWLPlugin, a SemanticWeb extension of the Prot´eg´e ontology development platform. The OWL Plugin can be utilized to edit ontologies in the Web Ontology Language (OWL), to access description logic reasoner, and to acquire instances for semantic markup. In [15] have intended a formal ontology for a digital library (DLs) defines the basic concepts, relationships, and axiomatic rules govern the domain of DL. OCL [20] presented the ability to clearly and automatically deal with business rules when building Object-oriented applications.

The ontology can diminish the quantity of information and decrease the effort of utilizing the information in rule acquisition, since it is simplified and purposely reorganized for rule acquisition. In addition, the ontology can be accrued and reclaimed throughout replicated rule attainment. The ontology demonstration is simply unstated and maintainable; combining these attributes with the achievement domain-specialists have had in building libraries of ontologies across different regulations and the reality of a lot of tools which use ontology as the base demonstration. Even though several of the prior representations are proficient of assuring the subsequent requirement of accomplishing difficulty solving during the use of procedural programming we coveted something that reminds you of the method specialists solve problems lacking commanding an additional algorithmic complexity. Professionals are inclined to solve problems through exploring their knowledge base for an appropriate solution. The ease of rule-based demonstration has completed it the majority normally used form for knowledge base systems. The other benefit this structure of representation has is its modularity as rules can be included or eliminated independently of other rules.

The main contribution of the work is:
1. Collecting the general rules from the various websites and given as input to the RuleToOnto algorithm
2. In RuleToOnto the Rule drafts will be created based on the identification of the variable and values and then compose rules criteria
3. In RuleToOnto scheme the "if then else" condition will be checked which will be a main part

***Manuscript received Jan , 2013.***
***G. Senthil Kumar**, Department of Computer Science and Engineering, Anna University Chennai / A.S.L.pauls College of Engineering and Technology / Coimbatore, India, 9940704243*
***K.K. Kanaga Mathan Mohan,** Department of Computer Science and Engineering, Assistant Professory, Anna University Chennai / A.S.L.pauls College of Engineering and Technology, Coimbatore, India/ 8344363775*

4. After this BFS (Best First Search) algorithm is implemented to generate the best matched rules as forming the rule draft's variable instances as a tree structure. If a variable instance is closer to the already chosen instances of a rule than the other instances of the same variable, we assign the instance to the rule. This geographic assumption plays a very important role in our approach.
5. The best solution will be collected by using the optimization algorithm called Genetic [28] approach.

The overall research formed as follows: In section 2 the main part of RuleToOnto is discussed. In section 3 the best first search is explained. In section 4 the genetic approach is discussed briefly. At last the experimental results are explained with that the conclusion and future work.

## II. ONTOLOGY-BASED RULE ACQUISITION PROCEDURE

In this section, we recommend a process which routinely acquires rules through RuleToOnto. In step 1, RuleToOnto is produced from rules which are acquired in another site. In step 2, variables and values are routinely recognized as of the Web page by means of RuleToOnto and the first rule draft is caused. In step 3, rules are automatically invented by combining the identified variables and values. We developed A* algorithm for this purpose. Still, the generated rules may be unfinished. So, the knowledge engineer requires to filter the second rule draft to make it complete in step 4. Rule Ontology Generation RuleToOnto is domain specific knowledge that provides information about rule components and organizations. It is feasible to directly use the rules of the previous system as an alternative of the proposed ontology. Conversely, it requires a large space and extra processes to utilize information on rules, while RuleToOnto is a generalized compact set of information for rule acquisition. Thus, we use RuleToOnto instead of the rules themselves.

While the rule component identification step needs variables, values, and the relationship between them, the rule composition step needs simplified rule structures. Therefore, RuleToOnto represents the IF and THEN parts of each rule by attaching rules with variables with the IF and THEN relations, in addition to essential information about variables, values, and connections between variables and values. The RuleToOnto representation has three object properties has Value, IF and THEN, and three classes, Variable, Value, and Rule, which is an RDF graph produced from the OWL ontology by the RDF validator. We added some axioms in order to utilize ontology suggestion. First, something is a Variable precisely if all the values of the *HasValue* property are instances of the Value class, as shown in the following axiom characterized in the Functional-Style syntax:

```
EquivalentClasses(
:Variable
ObjectAllValuesFrom(:hasValue :Value)
)
```

*Moreover, Rule can only have instances of Variable for its values of if and then properties as follows:*

```
EquivalentClasses(
:Rule
ObjectIntersectionOf(
ObjectAllValuesFrom(:if :Variable)
ObjectAllValuesFrom(:then :Variable)
)
)
```

We also added property cardinality restrictions. An instance of Variable should have at least one Value instance for its *has Value* property, and Rule should have at least on *eVariable* for each of the if and then properties as follows:

```
SubClasses(
:Variable
ObjectMinCardinality(1 :hasValue :Value)
)
SubClasses(
:Rule
ObjectIntersectionOf(
ObjectMinCardinality(1 :if :Variable)
ObjectMinCardinality(1 :then :Variable)
)
)
```

We excluded connectives such as AND and OR from RuleToOnto, because it is hard to represent the complex nested structure of connectives in a simple frame representation, and generalization has no effect if we represent all connectives in the ontology.

Rule component identification has the goal which is to obtain variables and values by comparing parsed words of the given text with the variables and values of RuleToOnto. The concerns and logical processes involved have already been discussed in the previous work. We want to focus on the practical implementation and the new issue, semantic similarity. In the first step, we expanded RuleToOnto by adding synonyms of each term using tool of WordNet. Then, we applied the stemming algorithm to the expanded RuleToOnto in order to normalize the terms in step 2. Also, we parsed and stemmed the Web page for rule acquisition in step 3. In the contrast among the terms of RuleToOnto and the terms of the Web page, we used semantic matching instead of simple string comparison. That find the semantic similarity between two terms, we used the hyponym structure of WordNet. The similarity measure is a reciprocal number of the distance between two terms in the hyponym hierarchy in WordNet, as given by the following equation:

$$semanticsimilarity = \frac{1}{path\_length}$$

## III. RULE COMPONENT IDENTIFICATION

The contribution of this section is the variable instances of rule draft 1 that is an output of the rule component identification step and RuleToOnto. The next step is rule ordering, which produces RuleOrder from the variable instances and the candidate rules. The third step is variable ordering, which generates TotalOrder with the RuleOrder and VariableOrder that is calculated in this step. The last step is best-first search that makes rule draft 2. The input of the Best-First Search (BFS) algorithm is a set of identified variable instances;$VI = \{VI_1, VI_2, \ldots, VI_i, \ldots, VI_n\}$ which is certain from the rule component recognition stage. The output is a set of rule instances, $RI = \{RI_1, RI_2, \ldots, RI_p, \ldots, RI_q\}$ where $RI_p$ is a set of variable instances assigned to the rule. The first job of preparation is extracting rule candidates from RuleToOnto. Each variable of each rule applicant should be matched to the variable instances of $VI$. At this time, the rule candidates are

just rule templates with variables to which the variable instances are not yet assigned. Therefore, we use variables to denote the rule candidates. Rule candidates are indicated as a set $RC = \{R1; R2; \ldots; Rj; \ldots; Rl\}$, where a rule candidate $Rj$ is $\{Vj_1, Vj_2; \ldots; Vj_k; \ldots; Vj_m\}$ and $Vj_k$ every of this is contested to one or more variable instances of $VI$. In a while, rule composition generates rule instances by conveying variable instances to rule candidates.

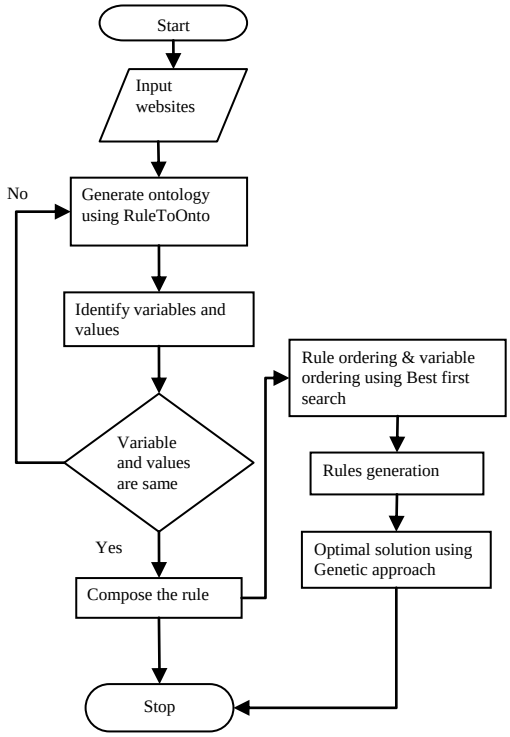


Fig1: Flowchart for whole process

## IV. Genetic Algorithm For Optimal Rule Acquisition

Genetic Algorithm is found on Charles Darwin's theory of 'The endurance of the fittest'. Algorithm is started with a set of results (represented by chromosomes) called population. Solutions from one population are taken and used to form a new population. This is aggravated by a hope, that the new population will be better than the old one. Solutions which are chosen to form new solutions (offspring) are selected according to their fitness [27] the more appropriate they are the more chances they have to reproduce. If the fitness of the new individuals is better than the fitness of the individuals in the previous generation, the individuals are replaced. This is approved out till the extinction condition is achieved. The chromosome should in several ways enclose information about solution which it represents. The most used way of encoding is a binary string. Each chromosome has one binary string. Each bit in this string can characterize some attribute of the solution. Or the whole series can represent a number. But there are many other ways of encoding.

### A. Methodology

A new population is first initialized. For every individual in the population, a fitness [27] function is applied and the fitness [27] is intended. Then based on the crossover and mutation charge, the crossover and mutation functions are performed. The new individual's obtained are again subjected to the fitness function. If the fitness of the new those are better than the fitness of the individuals in the preceding generation, the individuals are replaced. This is carried out till the termination stipulation is reached.

### B. Selection of Individuals

When you submit your final version, after your paper has been accepted, prepare it in two-column format, including figures and tables. Crossover decides on genes from parent chromosomes and makes a new offspring. The simplest method how to do this is to select randomly some crossover point and everything before this point copy from a first parent and then everything after a crossover point copy from the second parent. Ordinary form of crossover is single point crossover where arbitrarily one position in the chromosomes is select and child 1 is head of chromosome of close relative 1 with tail of chromosome of parent 2 and child 2 is head of 2 with tail of 1. There are other ways to create crossover, for example we can choose more crossover points. Crossover can be rather difficult and depends on encoding of the encoding of chromosome.

### C. Crossover Operator

Crossover decides on genes from parent chromosomes and makes a new offspring. The simplest method how to do this is to select randomly some crossover point and everything before this point copy from a first parent and then everything after a crossover point copy from the second parent. Ordinary form of crossover is single point crossover where arbitrarily one position in the chromosomes is select and child 1 is head of chromosome of close relative 1 with tail of chromosome of parent 2 and child 2 is head of 2 with tail of 1. There are other ways to create crossover, for example we can choose more crossover points. Crossover can be rather difficult and depends on encoding of the encoding of chromosome.

### D. Mutation Operator

Mutation changes randomly the new offspring. For binary encoding we can switch a few randomly selected bits from 1 to 0 or from 0 to 1. Mutation provides a small amount of random search, and helps ensure that no point in the search has a zero probability of being checked. A fitness [27] function is an exacting type of objective function that advises the optimality of a chromosome in a genetic algorithm [28], so that the exacting chromosome may be ranked beside all the other chromosomes.

### E. Mutation Operator

An ideal fitness purpose associates closely with the algorithm's goal, and up till now might be computed quickly. Speed of execution is extremely significant, as a typical genetic algorithm must be iterated many times in order to turn out a practical result for a non-trivial problem. This paper adopts minimum support and minimum confidence for filtering rules. Then correlative degree is established in rules which gratify minimum support degree and minimum confidence-degree. Following support degree and confidence-degree are synthetically in use addicted to account, fit degree purpose is defined as pursues.

$$Fitness(X) = R_S \cdot \frac{Supp(X)}{Supp_{min}} + R_C \cdot \frac{Conf(X)}{Conf_{min}}$$

In the above formula, $R_S + R_C = 1\ (R_S \geq 0,\ R_C \geq 0)$ and $Supp_{min}$, $Conf_{min}$ are individual values of minimum support and minimum confidence. By all emergences if the $Supp_{min}$ and $Conf_{min}$ are put to higher values, then the value of fitness [27] function is also establish to be high.

## V. EXPERIMENTAL RESULTS

### A. Recall Rate

This graph shows the recall rate of existing and proposed system based on two parameters of recall and dataset. From the graph we can see that, when the number of dataset is improved the recall rate also improved in proposed system but when the number of dataset is improved the recall rate is reduced in existing system than the proposed system. From this graph we can say that the recall rate of proposed system is increased which will be the best one. In this graph we have chosen two parameters called dataset and recall which is help to analyze the existing system and proposed systems on the basis of recall. In X axis the iteration parameter has been taken and in Y axis recall parameter has been taken. . From this graph we see the recall rate of the proposed system is in peak than the existing system. Through this we can conclude that the proposed system has the effective recall. The values are shown in Table 2:

Table 1: Recall Rate

| S NO | Number of dataset | Proposed system | Existing system |
|---|---|---|---|
| 1 | 10 | 256 | 238 |
| 2 | 20 | 219 | 192 |
| 3 | 30 | 177 | 148 |
| 4 | 40 | 135 | 113 |
| 5 | 50 | 96 | 72 |
| 6 | 60 | 58 | 41 |

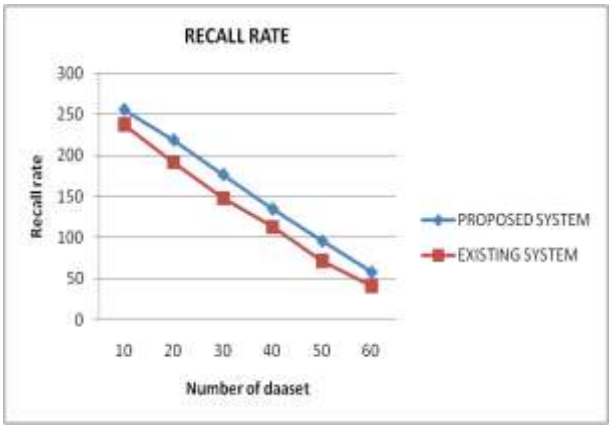


Fig 2: Recall Rate

### B. Precision Rate

This graph shows the precision rate of existing and proposed system based on two parameters of precision and the dataset. From the graph we can see that, when the number of dataset is improved the precision also improved in proposed system but when the number of dataset is improved the precision is reduced somewhat in existing system than the proposed system. From this graph we can say that the precision of proposed system is increased which will be the best one. In this graph we have chosen two parameters called dataset and precision which is help to analyze the existing system and proposed systems. The precision parameter will be the Y axis and the dataset parameter will be the X axis. The blue line represents the existing system and the red line represents the proposed system. From this graph we see the precision of the proposed system is higher than the existing system. Through this we can conclude that the proposed system has the effective precision. The values are shown in Table 2:

Table 2: Precision Rate

| SNO | Number of dataset | Proposed system | Existing system |
|---|---|---|---|
| 1 | 10 | 9 | 4 |
| 2 | 20 | 14 | 10 |
| 3 | 30 | 19 | 16 |
| 4 | 40 | 25 | 21 |
| 5 | 50 | 31 | 27 |
| 6 | 60 | 36 | 33 |
| 7 | 70 | 41 | 37 |

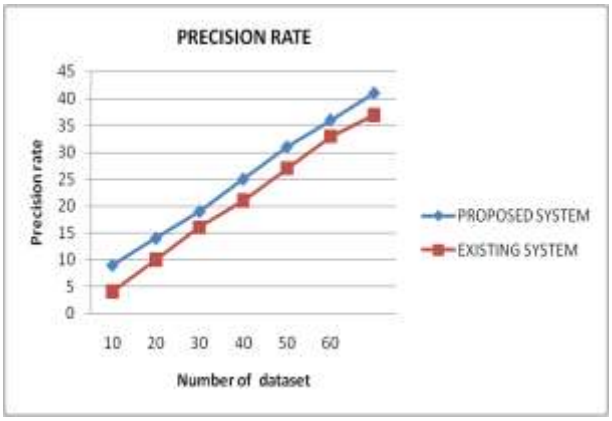


Fig 3: Precision Rate

## VI. CONCLUSION

In this work, a Rule ontology representation is proposed for knowledge illustration and reasoning over user profiles. In this research the effective rule acquisition goal is achieved through the RuleToOnto methodology. The effective BFS algorithm is implemented for the semantic rule generation and composition of rules. The results show that this ontology model is successful using the method called genetic algorithm. Genetic Algorithms have been used to work out complicated optimization troubles in a number of fields and have proved to produce optimum results in mining rules. Values of minimum support, minimum confidence and population size makes a decision ahead the accuracy of the system than other GA parameters. The optimum importance of crossover rate directs to earlier convergence while playing minimum role in achieving better accuracy.

## REFERENCES

[1] Knublauch, Fergerson H., Noy N., and Musen M., 2004. “The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications”.Proc Third ISWC (ISWC 2004), Hiroshima, Japan, pp.229-243.

[2] Mei J.”Semantic Web Ontology and rules”. Beijing: Beijing University. Peachavanish R., Karimi H., 2007. Ontological

Engineering for Interpreting Geospatial Queries. Transactions in GIS , 11(1):pp.115 – 130., 2007

[3] Smith M. Welty C., McGuinness D., 2004."OWL Web Ontology Language" Guide.http://www.w3.org/TR/owl-guide/.

[4] Haarslev V. and M¨oller R., 2001."RACER System Description".In Proceedings of the 1st International Joint Conference on Automated Reasoning (IJCAR), LNCS 2083, 701–706.

[5] Horrocks I., Harold Boley P.,Tabet S., Grosof B., Dean K., 2004.SWRL: "A Semantic Web Rule Language Combining OWL and RuleML".http://www.w3.org/Submission/SWRL/.

[6] Horrocks I., Sattler U., and Tobies S., 1999. "Practical reasoning for expressive description logics". Proceedings of the 6th International Conference on Logic for Programming and Automated Reasoning (LPAR'99), number 1705 in Lecture Notes in Artificial Intelligence, pages 161–180.

[7] Wang H., Cao H., 2007. "SWRL Ontology-based and space relations with the reasoning methods" that [J]. Micro-electronics and computers, 24(7): pp.166-169.

[8] Xu B., Ye P., 2007. "Knowledge method Research". Intelligence science, 25(5): pp.690-694.

[9] Beijing," The International Archives of the Photogrammetry", Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B2. 2008

[10] 2. M. Jarrar, J. Demey, R. Meersman. "On Using Conceptual Data Modeling for Ontology Engineering". In:S. Spaccapietra et al. (eds.), Journal on Data Semantics. LNCS, Vol. 2800. Berlin/Heidelberg: Springer, 2003, pp. 185–207.

[11] Y . Wand, V. C. Storey, R. Weber. "An ontological analysis of the relationship construct in conceptual modeling". ACM Transactions on Database Systems (TODS), Vol. 24(4), 1999, pp. 494–528.

[12] T. R. Gruber. "Toward Principles for the Design of Ontologies for Knowledge Sharing". International Journal of Human and Computer Studies, Vol. 43(4–5), 1995, pp. 907–928.

[13] J. Trinkunas, O. Vasilecas. "Ontology Transformation: from Requirements to a Conceptual Model". Acta Universitatis Latviensis [Latvijas Universitates Raksti], Computer Science and Information Technologies,Vol. 751. University of Latvia, 2009, pp. 54–68.

[14] E. Bozsak et al. KAON ," Towards a Large Scale Semantic Web". In: K. Bauknecht et al. (eds.), Proc.of the Third International Conference on E-Commerce and Web Technologies (EC-Web 2002). LNCS,Vol. 2455. London: Springer-Verlag, 2002, pp. 304–313.

[15] M. A. Goncalves, L. T. Watson, E. A. Fox. "Towards a Digital Library Theory: A Formal Digital Library Ontology". International Journal on Digital Libraries, Vol. 8(2), 2008, pp. 91–114.

[16] "OMG: Ontology Definition Metamodel", 2005. Available: http://www.omg.org/docs/ad/05-08-01.pdf.Accessed September, 2008.

[17] T. Morgan. "Business Rules and Information Systems: Aligning IT with Business Goals". Boston: Addison Wesley, 2002.

[18] B. von Halle. "Business Rules Applied: Building Better Systems Using the Business Rules Approach". New York: John Wiley & Sons, 2002.

[19] R. G. Ross." Principles of the Business Rule Approach". Addison Wesley, 2003.

[20] B. Demuth, H. Hussmann, S. Loecher. "OCL as a Specification Language for Business Rules in Database Applications". In: M. Gogolla, C. Kobryn (eds.), Proc. of the 4th International Conference on the Unified Modeling Language, Modeling Languages, Concepts, and Tools (UML 2001). LNCS, Vol. 2185. London:Springer-Verlag, 2001, pp. 104–117.

[21] R. G. Ross. The Business Rule Book. Classifying, "Defining and Modeling Rules". Houston: Business Rules Solutions Inc., 1997.

[22] L. Boyd. CDM RuleFrame – the Business Rule Implementation Framework That Saves You Work. In: Proc. of ODTUG 2001. Available: http://www.dulcian.com/odtug_conference.htm. Accessed November, 2006.

[23] I. Horrocks, P. F. Patel-Schneider, H. Boley et al. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. W3C document, 2004. Available: http://www.w3.org/Submission/SWRL/. Accessed September, 2009.

[24] V. Zacharias. Technical Report: "Development and Verification of Rule Based Systems – a Survey of Developers". Technical Report, 2008.Available:http://vzach.de/papers/2008_SurveyTechReport.pdf.Accessed May, 2009.

[25] C. Golbreich, A. Imai. "Combining SWRL rules and OWL ontologies with Protégé OWL plugin", Jess and Racer. In: Proc. of the 7th International Protégé Conference, 2004. Available: http://galimed.med.univrennes1. fr/lim/doc_92.pdf. Accessed May, 2009.

[26] H . Herbst et al. "The Specification of Business Rules: A Comparison of Selected Methodologies". In:A. A. Verrijn-Stuart, T. W. Olle (eds.), Methods and Associated Tools for the Information System Life Cycle. New York: Elsevier, 1994, pp. 29–46.

[27] Gonzales, E., Mabu, S., Taboada, K., Shimada, K., Hirasawa,K., "Mining Multi-class Datasets using Genetic Relation Algorithm for Rule Reduction", IEEE Congress on Evolutionary Computation,CEC'09 , pp. 3249 – 3255, 2009.

[28] Xian-Jun Shi, Hong Lei, "A Genetic Algorithm-Based Approach for Classification Rule Discovery", International Conference on Information Management, Innovation Management and Industrial Engineering, ICIII '08, Volume: 1, pp. 175 – 178, 2008.

.

**G.Senthil Kumar** has done BE (CSE), from Anna University Chennai in 2009. Currently doing his ME(Computer Science Engineering) final Year under Anna University Chennai, A.S.L.Pauls College of Engineering and Technology in Coimbatore. He published this paper in International conference on Innovations in Communication, Information and Computing (ICICIC'13) Sasurie College of Engineering, Tirupur, Tamil Nadu, India. January 9 - 11, 2013. His research area is data mining.

**K.K.Kanaga Mathan Mohan** received BE (Computer Science Engineering) and ME.(CSE), from Anna University Chennai. Currently working as an Asst. Professor in Dept. of Computer Science Engineering, A.S.L.pauls College of Engineering and Technology, Coimbatore. His research area is data mining, image processing.