

# AUTOMATIC GENERATION AND OPTIMIZATION OF TEST DATA USING HARMONY SEARCH ALGORITHM

Rajesh Kumar Sahoo<sup>1</sup>, Deeptimanta Ojha<sup>1</sup>, Durga Prasad Mohapatra<sup>2</sup> and Manas Ranjan Patra<sup>3</sup>

<sup>1</sup>Department of Computer Engineering, A.B.I.T, Cuttack  
rajesh\_sahoo@rediffmail.com  
deeptimantaojha@gmail.com

<sup>2</sup>Department of Computer Engineering, NIT, Rourkela  
durga@nitrkl.ac.in

<sup>3</sup>Department of Computer Engineering, Berhampur University, Berhampur  
mrpatra12@gmail.com

## **ABSTRACT**

*Software testing is the primary phase, which is performed during software development and it is carried by a sequence of instructions of test inputs followed by expected output. The Harmony Search (HS) algorithm is based on the improvisation process of music. In comparison to other algorithms, the HSA has gain popularity and superiority in the field of evolutionary computation. When musicians compose the harmony through different possible combinations of the music, at that time the pitches are stored in the harmony memory and the optimization can be done by adjusting the input pitches and generate the perfect harmony. The test case generation process is used to identify test cases with resources and also identifies critical domain requirements. In this paper, the role of Harmony search meta-heuristic search technique is analyzed in generating random test data and optimized those test data. Test data are generated and optimized by applying in a case study i.e. a withdrawal task in Bank ATM through Harmony search. It is observed that this algorithm generates suitable test cases as well as test data and gives brief details about the Harmony search method. It is used for test data generation and optimization.*

## **KEYWORDS**

*Harmony search algorithm, meta-heuristics, test case generation, test case optimization, test data.*

## **1. INTRODUCTION**

The test case generation is based on the requirements. It completely ignores the aspect of system execution. Apart from this, the test case design from program code may cause difficult to imbrute. Test cases may not expose the missing functionalities. The proposed approach focuses the redundancy, test cases, and test case optimization challenges. It uses HS optimization

algorithm to optimize the random test cases. Moreover, this proposed method inspires the developers to generate random test cases to improve the design quality. This paper is intended to present the result of the outcome of HS to find the optimum solution in the software construct. Optimization can be represented through the process of finding the best result under the given circumstances which might be used for maximizing or minimizing the local or overall optimum value of a function. The evolutionary algorithms like Harmony search is discussed in this paper. Z. W. Geem [6] introduced Harmony search method in 2009. By this technique, a flawless state of harmony is observed through the musical process. It is correspondent to generate the optimum value through the musical optimization process. The improvisation of music is a technique where the musician plays various musical notes with divergent types of musical instruments and generates the best amalgamation of tunes with frequency.

The rest of the paper is organized as follows. Section 2 discusses basics of test data automation, Section 3 is for literature survey, Section 4 represents the fundamentals of Harmony Search (HS) algorithm, proposed systems, and methodology, Section 5 focuses the simulation results, Section 6 represents discussion and future scope and Section 7 concludes the paper.

## **2. BASIC CONCEPTS**

### **2.1 Automated Test data generation**

Testing is the phenomenon of finding errors after executing the programs. Software testing can be defined by many processes designed sequentially and does not do anything unintended [19]. The objective of software testing is to finalize the application software against the user requirements. It must have good test coverage to test the application software and perform as per the specifications. For generating list of coverage's, the test cases should be designed with maximum possibilities of finding various errors or bugs [20]. The test cases should be very effective and is measured through the number of defects or errors reported. Generation of test data is a method for identifying the data set which satisfies the criteria. Automated generation of test data plays a key factor in software testing. Most of the researchers are used the heuristics approaches for automated generation of test data. Automatic test data generation data helps to minimizing the time and cost in developing test cases.

### **2.2 Overview of Harmony Search**

The harmony search metaheuristic optimization algorithm is based on the music. The harmony search technique is inspired by the musician when he composes the music. Usually harmony consists of different possible amalgamations of music pitches saved in the memory. In this technique first, random solutions are directly stored in the harmony memory based on memory considering rate and pitch adjustment rate, and then pitch adjustment distance will be calculated between different selected random solutions. The best solution is stored in the memory of harmony by discarding the worst solution.

## **3. LITERATURE SURVEY**

Biswal et al.[1] described how the activity diagram is derived from the test scenarios by converting into a control flow graph (CFG) where every node explains an activity and the edges of the node with the activities of control flow. According to Geem et al. [6, 7] harmony search a

metaheuristic population-based algorithm where multiple harmonies are used in parallel which gives better performance with high efficiency. Geem [7] focused the Pitch Adjustment Rate (PAR) function which is used in simulated annealing and it also increases the robustness of the algorithm. It is highly reliable. Geem et al. [8] described how the parameter like Pitch adjustment rate (PAR) increases in a linear way with the number of generations while the bandwidth (BW) decreases exponentially in Harmony Search algorithm. Manjarres et al. [5] focused described various applications of HS like industry, power systems, construction design, and information technology. Das et al. [3] described the background of the power of Harmony Search (HS) which gives the better solutions. The major drawback is user must specify the minimum and maximum bandwidth values. It is difficult to deduce the program which is dependent. Chen et al. [2] proposed a concept which is having the relation between nodes of control flow graph (CFG). It also defined the fitness function for evaluating the generation of test data by using Genetic Algorithm (GA) approach which performed the random technique. Ghiduk et al. [10] described how the test data are generated by using a search based technique but in case of random search techniques the optimal test cases are not generated effectively and efficiently. Kirkpatrick et al. [11] focused on a key factor like the Pitch adjustment which is diversified randomly in Harmony Search method. Mahdavi et al. [12] explained how HS algorithm is used in clustering problems of web pages through continuous and discrete representation of data. Mira et al. [13] focused on the parameters which required for improving the efficiency of an HS algorithm in a particular problem where diversification and intensification are available. Omran et al. [14] focused on the how the Global Harmony Search (GHS) is simulated through Particle Swarm Optimization (PSO). This paper also explained how the pitch adjustment of Harmony Search technique can imitate by harmony memory through the harmony vector. It also replaces bandwidth (BW) and adds to the dimension of HS. Peng et al. [15] proposed a novel algorithm which upgrades the search based test case generation by using an adaptive genetic algorithm (AGA) and it gives better results in the software testing tool. Suresh et al. [16] described a genetic algorithm used for test data generation and it also generates the basic path. In this case the genetic algorithm combines the features of local and global test data optimization. According to this paper sequence diagram is converted into control flow graph and generates the optimal test cases using genetic algorithm. Yang et al. [17] focused on HS algorithm which is diversified with the randomization operation of pitch adjustment rate through the probability. Existing pitch (or solution) is saved in the Harmony Memory. It also gives the global search solutions. Yang et al. [18] proposed an approach to generate the test data in a single specific path based genetic algorithm (GA). In this case the effectiveness of the fitness function and performance are evaluated.

## **4. PROPOSED SYSTEM**

We have proposed a methodology for generating test cases for withdrawal system of an ATM machine and test cases are optimized by Harmony search algorithm (HSA). This method is used for evaluating its efficiency and effectiveness for generating the test cases and to maximize to achieve the goal.

### **4.1. Necessity of Proposed System:**

The proposed system is intended to generate an automatic and optimized test case with existing approaches of Harmony Search. Optimized test cases may not be helpful in the testing process. It may be required to differentiate between the various test cases. First of all the system may be initialized with harmony memories. Each harmony memory maintains its own prevailing location

on the basis of which the test cases may be generated. Harmony Search has a memory that helps to maintain the solutions by harmonies. This paper also finds out the effectiveness of the proposed approach through the number of test cases or test data.

#### 4.1.1. Harmony Search Algorithm (HSA)

The main principle and working of proposed approach:

Harmony search is a meta heuristic population-based optimization technique. Through this technique, the problem is represented through different test cases. The quality of each test case is calculated through the fitness value of the problem. The working principle of harmony search technique is inspired by the musician when he composes the music; a musician usually tries different combinations of music pitches which are stored in the harmony memory. Perfect harmony search needs a correspondent for generating the process to find the solution which is optimal. The main steps of a harmony search method are:

- Generate random solutions which will be stored in the Harmony Memory (HM).
- Select a random solution from Harmony Memory based factors that are Harmony Memory Considering Rate (HMCR) and Pitching Adjustment Rate (PAR).
- Apply adjustments pitch distance to the selected random solution.
- Compare the fitness function values of the mutated solution or newly formed solution with the worst solution.
- If better solution is found then the worst solution is substituted with the mutated solution available in the harmony memory otherwise the solution is discarded.
- Remember the solution which is best so far.

In this case, mainly two parameters are used. They are HMCR and PAR. Harmony Memory Considering Rate (HMCR) is described as the probability of selecting a particular component or solution available in HM. Pitching Adjust Rate (PAR) describes the probability distribution of the candidate solution for mutation purpose through HM.

The value of Pitching Adjust Rate (PAR) can be calculated as follows.

$$PAR = (PAR_{max} - PAR_{min}) / (it_{max}) * it + PAR_{min} \quad (1)$$

Where  $PAR_{max}$  → maximum pitch adjusting rate

$PAR_{min}$  → minimum pitch adjusting rate

$it_{max}$  → given maximum iterations

$it$  → current iteration number

The bandwidth factor 'bw' can be calculated as.

$$bw = bw_{max} * \exp(\text{coef} * it) \quad (2)$$

Here the value of 'coef' is evaluated as follows:

$$\text{coef} = \log (b w_{\min} / b w_{\max} / i t_{\max}) \quad (3)$$

Where  $b w$  = bandwidth  
 $b w_{\min}$  = minimum bandwidth  
 $b w_{\max}$  = maximum bandwidth

The bandwidth factor  $b w$  is used to alter the value of existing solution available in Harmony Memory. This can be done as follows:

$$x_{\text{new}} = x_{\text{old}} + \text{rand} (0,1) * b w \quad (4)$$

Where  $x_{\text{new}}$  = new solution  
 $x_{\text{old}}$  = old existing solution  
 $\text{rand} (0,1)$  = a random value ranging in between 0 and 1.

Harmony Search is a meta heuristic search method which is generally used to optimize problems whose solution may be analyzed in an n-dimensional space. According to Harmony Search Algorithm, each musician plays a musical note to generate the best possible harmony altogether. A new solution is generated for a musical note with the help of harmony factors like HMCR (Harmony Memory Considering Rate), PAR (Pitch Adjustment Rate) and  $b w$  (Bandwidth). Each musical note usually possesses its current position. After iteration, the worst harmony is substituted by a new better solution. The flow chart of Harmony Search is depicted in Figure-1.

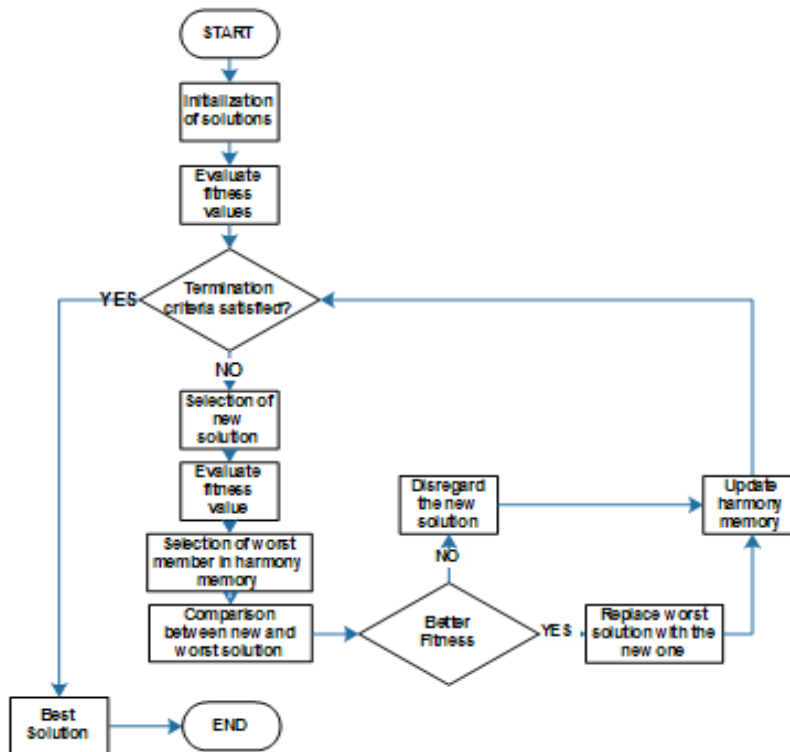


Figure 1: Flow chart of Harmony Search

#### 4.2. Pseudo code for Harmony Searches for generating test data

The number of generation is initialized.

The population size is initialized

Generate Initial Population and store it in the Harmony Memory

Evaluate the initial fitness function value 'fx' of all candidate solutions

$$f_x = 1 / (\text{abs}(\text{net\_bal} - \text{wd\_amt}) + \epsilon)^2$$

where  $\epsilon$  varies from 0.1 to 0.9

Find the initial best and worst solution

Initialize different HARMONY parameters

i.e.,  $\text{PAR}_{\min}$ ,  $\text{PAR}_{\max}$ ,  $\text{bw}_{\min}$ ,  $\text{bw}_{\max}$ , HMCR

While Generation < max\_iteration do

Evaluate the values of PAR, bw

If  $\text{rand}(0, 1) \leq \text{HMCR}$  then

Choose a random value from Harmony Memory i.e., x

If  $\text{rand}(0, 1) \leq \text{PAR}$  then

Adjusting the value of x by using the following equation:

$$x_{\text{new}} = x + \text{rand}(0, 1) * \text{bw}$$

End if

else

Choose a random value with the help of following equation:

$$x_{\text{new}} = \text{min\_value} + \text{rand}(0, 1) * (\text{max\_value} - \text{min\_value})$$

End if

Checking the boundary condition of the new solution

Evaluate the fitness function value of the new solution

If  $\text{fitness}(\text{new solution}) \geq \text{fitness}(\text{worst solution})$

Accept the new solution

Replace the worst solution with the new solution available in the Harmony Memory.

End if

Update current best solution and worst solution

Generation (t) = Generation (t) + 1

End While

Select the best fitness with function value

#### 4.3. Methodology:

For Mathematical function

$$f(x) = 1 / (\text{abs}(\text{suc\_bal}) + \epsilon)^2 \quad (5)$$

Where  $0.1 \leq \epsilon < 0.9$  (taking  $\epsilon$  -value because overflow condition due to infinity).

Here Successive Amount (suc\_amt) is defined as :

$$\text{suc\_bal} = \text{net\_bal} - (\text{wtd\_amt} - \text{min\_bal}) \quad (6)$$

Where net\_bal = current account balance  
min\_bal= minimum bank balance limit

Initially, each solution is initialized with a musical note. The musician will search for optimal solutions by changing the pitch and bandwidth of their musical note. It will keep track of best note in the population and upgrade its solution. The optimal solution is used to maximize the mathematical function  $f(x)$  which may be implemented in Harmony Search using MATLAB-7.0.as shown in Table-1.This table primarily focuses on the musician's attempt to generate the best note in the search space.

Table 1: Fitness Function Value for each sample space or test data

Iteration Number	Test Cases/Test Data	Fitness Function Value
0	4000	5.9488e-010
10	5700	6.4746e-010
20	7300	7.0357e-010
40	10800	8.5496e-010
60	10800	8.5496e-010
80	14300	1.061e-009
100	16400	1.2225e-009
120	18700	1.4457e-009
140	20800	1.7075e-009
160	24300	2.3338e-009
180	26600	2.9537e-009
200	29800	4.3282e-009
220	31000	5.1018e-009
240	33700	7.8313e-009
260	35400	1.085e-008
280	39900	3.8445e-008
300	43500	4.4438e-007
320	43800	6.943e-007
340	44000	9.9976e-007
350	44000	9.9975e-007

In this case, 20 numbers of sample test cases are considered. The function value depends on upon the parametric values of the input variables. It was found that the solution reaches its optimum value after 325 iterations.

## 5. SIMULATION RESULTS

The proposed approach generates the automated test cases through test data for Bank ATM using Harmony Search Algorithm. The figure-2 shows the relation between two variable quantities which are fitness function value range and test data measured along one of a pair of axis represented in table-1.

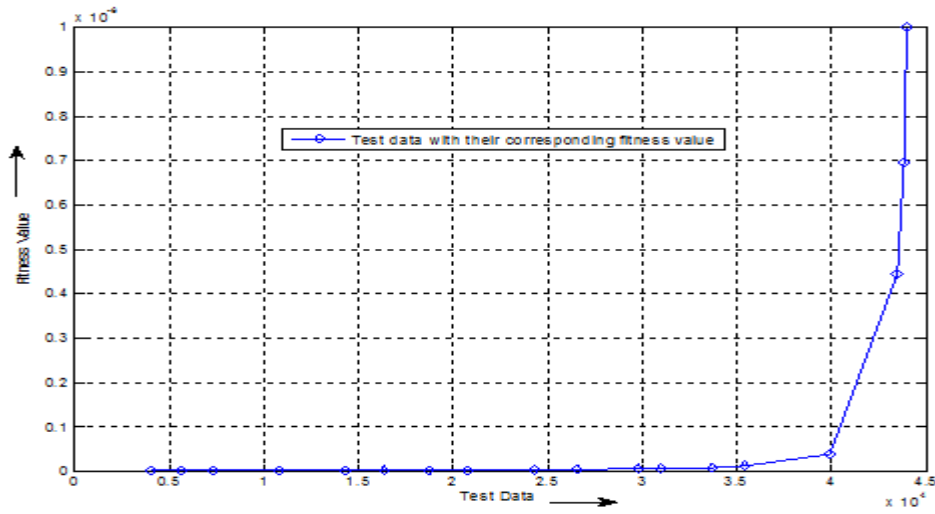


Figure 2: Graphical representation of test data and fitness function value for table-1

The proposed approach generates the test data for Bank ATM’s withdrawal operation using Harmony Search Algorithm. Table-2 represents the range of fitness value with different test data and also it gives the individual candidate solution according to the range of fitness value in terms of percentage.

Table 2: %of test data in terms of maximum fitness value

Fitness Value Range	% of Test data
$0 \leq f(x) < 0.3$	25
$0.3 \leq f(x) < 0.7$	55
$0.7 \leq f(x) < 1.0$	20

The above table shows that around 10% test cases or test data are having the higher fitness function  $f(x)$  value and lies in between 0.7 and 1.0. Figure-2 shows a pictorial representation of the relation of two variable quantities like percentage of test data and fitness value range.

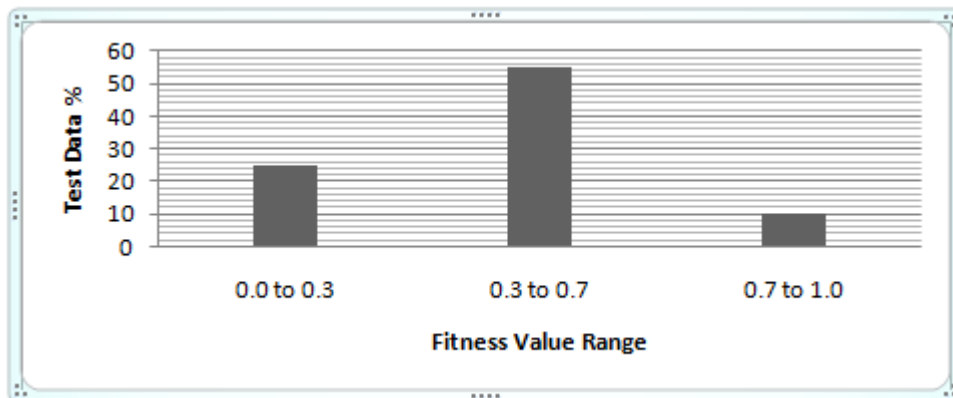


Figure 2: Graphical representation of % test data and fitness value range for table-2



## 6. DISCUSSION AND FUTURE SCOPE

While considering the mathematical function  $f_x=1/(\text{abs}(\text{net\_bal-wd\_amt})+\varepsilon)^2$ , where  $\varepsilon$  varies from 0.1 to 0.9, along with each member is initialized with a harmony. It has been found that optimality of solution keeps track of best and worst member in the harmony memory and updates its solution accordingly. By considering some sample test cases it has been observed that the function value depends upon the parametric values of the input variables like Harmony Memory Considering Rate, Pitching Adjusting Rate and the Bandwidth. The proposed approach generates the test data for small applications. The future approach to this work could enhance the test data generation for large programs automatically. The different parameters could be added to this approach which gives more optimized test cases and also increases the efficiency of Harmony Search (HS) technique. Another perspective area could be the randomly generated test data by using various paths according to the control flow graph (CFG). Test Cases can be generated by using various kinds of meta heuristic algorithms like GA, FA, PSO, BCO etc. The test data generated by using HS algorithm is compared with test data generated by PSO and it was found that HS produces optimal result in very less time and with more accuracy.

## 7. CONCLUSION

Harmony search algorithm (HSA) is a very important tool for optimization of test cases or test data. It has been diversified the problems in a very effective manner for generating the test data automatically. In this paper, HS algorithm has been discussed to generate the test cases which are optimized by taking an example of withdrawal operation of an ATM machine. This paper also describes the fundamental notions of HSA, how the random test cases are generated and finding the optimal solution to maximize the problem. This paper will inspire researchers to work on HSA by applying in computer science engineering area to generate the effective automated test cases.

## REFERENCES

- [1] Biswal.B.N., Nanda.P., Mohapatra.D.P.,(2008),"A Novel Approach for Scenario-Based Test Case Generation", International Conference on Information Technology,pp.244-247.
- [2] Chen Yong, Zhong Yong, Tingting Shi, Liu Jingyong,(2009), "Comparison of Two Fitness Functions for GA-based Path-Oriented Test Data Generation", Fifth International Conference on Natural Computation, Vol.4,pp.177-18.
- [3] Swagatam Das, Arpan Mukhopadhyay, Anwit Roy, Ajith Abraham, and Bijaya K. Panigrahi,(2011),"Exploratory Power of the Harmony Search Algorithm: Analysis and Improvements for Global Numerical Optimization", IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics. Vol.41, No.1, pp. 89-106.
- [4] Desikan.S. and Ramesh.G,Software testing principles & practices, Pearson Education,2007.
- [5] D. Manjarres, I. Landa-Torres, S. Gil-Lopez et al., (2013)," A survey on applications of the harmony search algorithm", Engineering Application in Artificial Intelligence. 26(8), pp.1818–1831.
- [6] Geem, Z.W.,(2009),"Harmony search algorithms for structural design optimization, in Studies in computational intelligence", Springer Berlin Heidelberg: Berlin, Heidelberg. 23(9), pp. 228-242.

- [7] Geem, Z.W.,(2007),“Harmony search algorithm for solving Sudoku”, in Proceedings of the 11th international conference, KES 2007 and XVII Italian workshop on neural networks conference on Knowledge-based intelligent information and engineering systems: Part I.,Springer-Verilog: Vietri Sul Mare, Italy.
- [8] Geem, Z,(2010),“State-of-the-Art in the Structure of Harmony Search Algorithm”, in Recent Advances In Harmony Search Algorithm, Springer Berlin / Heidelberg, pp. 1-10.
- [9] Geem, Z.W., and K.-B. Sim,(2010),” Parameter-setting-free harmony search algorithm”, Applied Mathematics and Computation, Vol.217 (8):, pp. 3881-3889.
- [10] Ghiduk, Ahmed S, and Girgis, Moheb R.,(2010), “Using Genetic Algorithms and Dominance Concepts for Generating Reduced Test Data”, Informatica (Slovenia), Volume 34, pp.377-385.
- [11] Kirkpatrick, S., “Optimization by simulated annealing: Quantitative studies”, Journal of Statistical Physics, 34(5-6):1984, pp. 12.
- [12] Mahdavi, M. and H. Abolhassani,(2009), “Harmony K-means algorithm for document clustering”, Data Mining and Knowledge Discovery, 18(3): pp. 370-39.
- [13] Mira, J., J. Álvarez, and X.-S. Yang,(2005),” Engineering Optimizations via Nature-Inspired Virtual Bee Algorithms, in Artificial Intelligence and Knowledge Engineering Applications: A Bioinspired Approach”, SpringerBerlin / Heidelberg, pp. 317-323.
- [14] Omran, M.G.H. and M. Mahdavi,(2008),” Global-best harmony search”,Applied Mathematics and Computation, 198(2): pp. 643-656.
- [15] Peng Lin, Xiaolu Bao, Zhiyong Shu, Xiaojuan Wang, Jingmin Liu,(2012),” Test Case Generation Based on Adaptive Genetic Algorithm”,IEEE,978-1-4673-0788-8/12/\$31.00.
- [16] Y.Suresh and S.Rath,(2013), “A genetic Algorithm based Approach for Test Data Generation in Basis Path Testing”, The International Journal of Soft computing and software Engineering, vol.3, issue.3.
- [17] Yang, X.-S.,(2008),” Nature-Inspired Metaheuristic Algorithms”,Springer-Verilog, pp. 73-80.
- [18] Yang Cao, Chunhua Hu and Luming Li,(2009),” An Approach to Generate Software Test Data for a Specific Path automatically with Genetic Algorithm”, International Conference on Reliability, Maintainability and Safety,pp.888-892.
- [19] Glenford J.Myers,(2004), The art of software testing, 2nd ed.: Wiley.
- [20] S.Kuppuraj and S.Priya,(2012),”Search-Based Optimization for Test Data Generation Using Genetic Algorithms,” in Proc of the 2nd International Conference on Computer Applications, pp.201-205.