Online ISSN : 0975-4172
Print ISSN : 0975-4350
DOI : 10.17406/GJCST

Global Journal

of Computer Science and Technology: C

Software & Data Engineering

Guiding Software Developers

Research on Big Data Analytics

Highlights

Automated Adaptation of Object

A Systematic Review of Learning

Discovering Thoughts, Inventing Future

Volume 16 Issue 2 Version 0.1

Global Journal of Computer Science and Technology: C
Software & Data Engineering

Global Journal of Computer Science and Technology: C
Software & Data Engineering

Volume 16 Issue 2 (Ver. 1.0)

Open Association of Research Society

Incorporation No.: 0423089
License No.: 42125/022010/1186
Registration No.: 430374
Import-Export Code: 1109007027
Employer Identification Number (EIN):
USA Tax ID: 98-0673427

Global Journals Inc.

*(A Delaware USA Incorporation with "Good Standing"; **Reg. Number: 0423089**)*

Sponsors: *Open Association of Research Society*
Open Scientific Standards

Publisher's Headquarters office

Global Journals® Headquarters
945th Concord Streets,
Framingham Massachusetts Pin: 01701,
United States of America
USA Toll Free: +001-888-839-7392
USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

Global Journals Incorporated
2nd, Lansdowne, Lansdowne Rd., Croydon-Surrey,
Pin: CR9 2ER, United Kingdom

Packaging & Continental Dispatching

Global Journals
E-3130 Sudama Nagar, Near Gopur Square,
Indore, M.P., Pin: 452009, India

Find a correspondence nodal officer near you

To find nodal officer of your country, please email us at *local@globaljournals.org*

eContacts

Press Inquiries: *press@globaljournals.org*
Investor Inquiries: *investors@globaljournals.org*
Technical Support: *technology@globaljournals.org*
Media & Releases: *media@globaljournals.org*

Pricing (Including by Air Parcel Charges):

For Authors:
22 USD (B/W) & 50 USD (Color)
Yearly Subscription (Personal & Institutional):
200 USD (B/W) & 250 USD (Color)

Integrated Editorial Board
(Computer Science, Engineering, Medical, Management, Natural Science, Social Science)

Dr. Bart Lambrecht
Director of Research in Accounting and FinanceProfessor of Finance
Lancaster University Management School
BA (Antwerp); MPhil, MA, PhD (Cambridge)

Dr. Carlos García Pont
Associate Professor of Marketing
IESE Business School, University of Navarra
Doctor of Philosophy (Management), Massachusetts Institute of Technology (MIT)
Master in Business Administration, IESE, University of Navarra
Degree in Industrial Engineering, Universitat Politècnica de Catalunya

Dr. Fotini Labropulu
Mathematics - Luther College
University of ReginaPh.D., M.Sc. in Mathematics
B.A. (Honors) in Mathematics
University of Windso

Dr. Lynn Lim
Reader in Business and Marketing
Roehampton University, London
BCom, PGDip, MBA (Distinction), PhD, FHEA

Dr. Mihaly Mezei
ASSOCIATE PROFESSOR
Department of Structural and Chemical Biology, Mount Sinai School of Medical Center
Ph.D., Etvs Lornd University
Postdoctoral Training,
New York University

Dr. Söhnke M. Bartram
Department of Accounting and FinanceLancaster University Management SchoolPh.D. (WHU Koblenz)
MBA/BBA (University of Saarbrücken)

Dr. Miguel Angel Ariño
Professor of Decision Sciences
IESE Business School
Barcelona, Spain (Universidad de Navarra)
CEIBS (China Europe International Business School).
Beijing, Shanghai and Shenzhen
Ph.D. in Mathematics
University of Barcelona
BA in Mathematics (Licenciatura)
University of Barcelona

Philip G. Moscoso
Technology and Operations Management
IESE Business School, University of Navarra
Ph.D in Industrial Engineering and Management, ETH Zurich
M.Sc. in Chemical Engineering, ETH Zurich

Dr. Sanjay Dixit, M.D.
Director, EP Laboratories, Philadelphia VA Medical Center
Cardiovascular Medicine - Cardiac Arrhythmia
Univ of Penn School of Medicine

Dr. Han-Xiang Deng
MD., Ph.D
Associate Professor and Research
Department Division of Neuromuscular Medicine
Davee Department of Neurology and Clinical NeuroscienceNorthwestern University
Feinberg School of Medicine

Dr. Pina C. Sanelli
Associate Professor of Public Health
Weill Cornell Medical College
Associate Attending Radiologist
NewYork-Presbyterian Hospital
MRI, MRA, CT, and CTA
Neuroradiology and Diagnostic Radiology
M.D., State University of New York at Buffalo,School of Medicine and Biomedical Sciences

Dr. Roberto Sanchez
Associate Professor
Department of Structural and Chemical Biology
Mount Sinai School of Medicine
Ph.D., The Rockefeller University

Dr. Wen-Yih Sun
Professor of Earth and Atmospheric SciencesPurdue University Director
National Center for Typhoon and Flooding Research, Taiwan
University Chair Professor
Department of Atmospheric Sciences, National Central University, Chung-Li, TaiwanUniversity Chair Professor
Institute of Environmental Engineering, National Chiao Tung University, Hsin-chu, Taiwan.Ph.D., MS The University of Chicago, Geophysical Sciences
BS National Taiwan University, Atmospheric Sciences
Associate Professor of Radiology

Dr. Michael R. Rudnick
M.D., FACP
Associate Professor of Medicine
Chief, Renal Electrolyte and Hypertension Division (PMC)
Penn Medicine, University of Pennsylvania
Presbyterian Medical Center, Philadelphia
Nephrology and Internal Medicine
Certified by the American Board of Internal Medicine

Dr. Bassey Benjamin Esu
B.Sc. Marketing; MBA Marketing; Ph.D Marketing
Lecturer, Department of Marketing, University of Calabar
Tourism Consultant, Cross River State Tourism Development Department
Co-ordinator , Sustainable Tourism Initiative, Calabar, Nigeria

Dr. Aziz M. Barbar, Ph.D.
IEEE Senior Member
Chairperson, Department of Computer Science
AUST - American University of Science & Technology
Alfred Naccash Avenue – Ashrafieh

Contents of the Issue

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 16 Issue 2 Version 1.0 Year 2016
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Research on Big Data Analytics

By Saloni Jain

Lecturer in RG PG College, India

Abstract - This paper gives an insight in the scope of Big Data in the field of Geoscience and which scripting language is acceptable for Big Data.Big Data transforms traditional information to customary conviction of how data should be aggregated, processed, analysed and stored. Big data is rudimentary transfiguring world of science. The large volume of data is posing a great menace on scientists.As data volumes is increasing with time there is complication in transferring Big Data. Thus, it is vital to reinforce sustainable infrastructure, correct analysis of dataand reduction in data around Geoscience Big Data. The prominence of the growth on sharing of data has led to new inventions and it's variations.

Keywords: *big data, geoscience, python.*

GJCST-C Classification : *H.2.8*

Strictly as per the compliance and regulations of:

Research on Big Data Analytics

Saloni Jain

Abstract- This paper gives an insight in the scope of Big Data in the field of Geoscience and which scripting language is acceptable for Big Data.Big Data transforms traditional information to customary conviction of how data should be aggregated, processed, analysed and stored. Big data is rudimentary transfiguring world of science. The large volume of data is posing a great menace on scientists.As data volumes is increasing with time there is complication in transferring Big Data. Thus, it is vital to reinforce sustainable infrastructure, correct analysis of dataand reduction in data around Geoscience Big Data. The prominence of the growth on sharing of data has led to new inventions and it's variations.

Keywords: big data, geoscience, python.

I. Introduction

Big Data or Data Integration is basically related with interoperability of data. Big Data deals with divergent fields such as:

1. Substantial data movement
2. Replication of data
3. Synchrony of data
4. Transmutation of data

Geoscience is the application and exploration of Earth's minerals, soil, water and energy resources. The variability in Earth sciences in any area can be shown in both spatial and temporal variations.

II. Analysis of Big Data

Prior to 2012 U.S was the largest single contributor to global data.

The emerging markets are showing the largest increases in data growth. In 2012, the amount of information stored worldwide exceeded 2.8 zetabytes. By 2020, the total amount of data stored is expected to be 50 times greater than today.

What good is all of this data? Data is raw, unrecognized facts that is in and of itself worthless. Information is potentially valuable concepts based on data. Knowledge is what we understand based on information. Wisdom is effective use of knowledge in decision making.

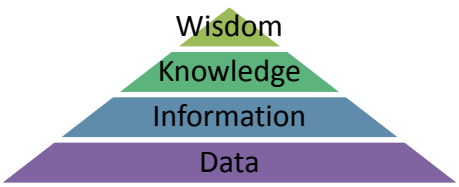

Figure 1 : Analysis of Big Data

III. Literature Review

There are many studies wherein many scientists have studied Big Data by inventing customized tools have been developed using various scripting languages. An overview of such studies is discussed in this section.

Azza Abouzeid et al devised a paper entitled "Ha doop DB: An Architectural Hybrid of Map Reduce and DBMS Technologies for Analytical Workloads "This paper elaborates on how Hadoop DB is able to approach the performance of parallel data systems and how Hadoop works in het erogenous environments.

Jerome Boulon et al have discussed about "Chukwa: A large-scale monitoring system" used for monitoring and analysing large distributed systems.

Jeffrey Dean et al have elaborated on "MapReduce: Simplified Data Processing on Large Clusters" which is a programming model and is used processing and generating large data sets. Two functions are used: map function and reduce function.

Tom Narock and Pascal Hitzler discussed about "Crowd sourcing Semantics for Big Data in Geosciences Applications" i.e. how semantic algorithms have been used for achieving accurate data .

Sanjay Ghemawat et al discussed on "The Google File System" which is a scalable distributed file system for large distributed data-intensive applications. It enhances the performance while analysing large clusters of data and provides great performance when dealing with large number of clients.

a) Technique used

Big Data can be coded in many different languages such as C, C++, Python. However, most suitable language considered for coding is Python. Python is said to be multi-model programming language. It authorize programmers to acquire various methodology of programming: object-oriented and structured programming which is fully sustained by Python. Python offers diverse language characteristics which stimulates functional programming and aspect-oriented programming.

There are many factors that favours Python as a language to code for Big Data. In modern times plenty of API's and libraries have been advanced for Python. In research also Python has a lot to implement ranging from networking to GUI development. Thus the interaction among systems has been highly enriched even though it remains a formidable task in many programming languages.

Author: R.G.P.G College Meerut Dept of Information Technology India.
e-mail: salonij91@gmail.com

Libraries in Python which are used for Big Data coding are PyDoop and SciPy. PyDoop offers an API for writing Hadoop programs in Python and is used for Map Reduce also.

Pydoop recommends diverse features which are usually not found in other Python libraries for Hadoop like MapReduce library which enables users to combine and partition data sets, easily installed library and can be used freely.

SciPy is an open source library that is offered by Python for all the users aiming to do scientific computations. This library furnish various modules such as ODE(Ordinary Differential Equations), FFT(Fast Fourier Transformation),optimization which finds application in the field of science and engineering.

b) Snapshots of Coding

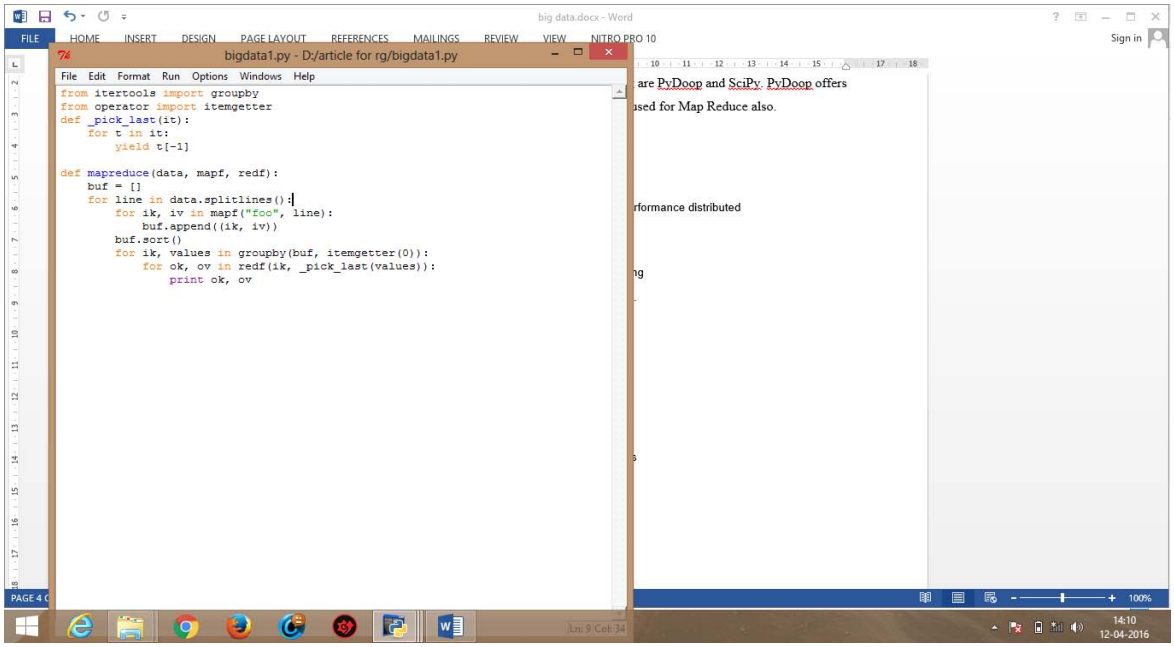

Figure : Snapshot of Program 1

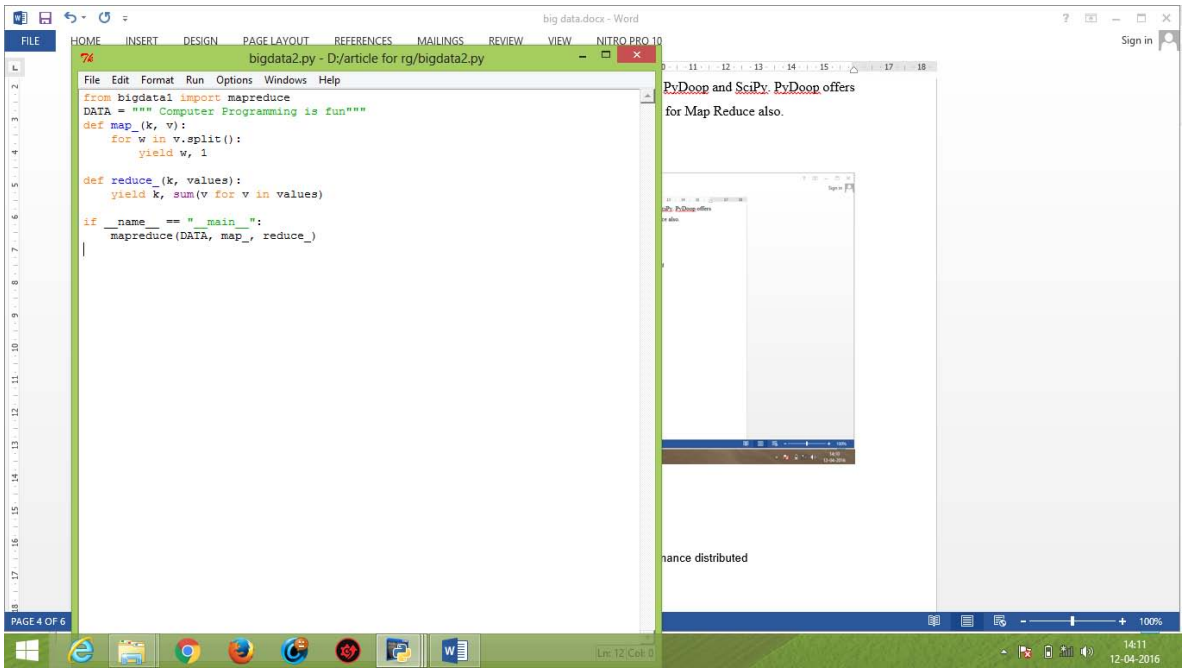

Figure : Snapshot of Program 2

c) Conclusion And Future Scope

Big Data analyses large clusters of data and provides users with accurate and refined search. In this case whatever text the user enters the tool will count the words and displays the output for the user as shown in the snapshot.

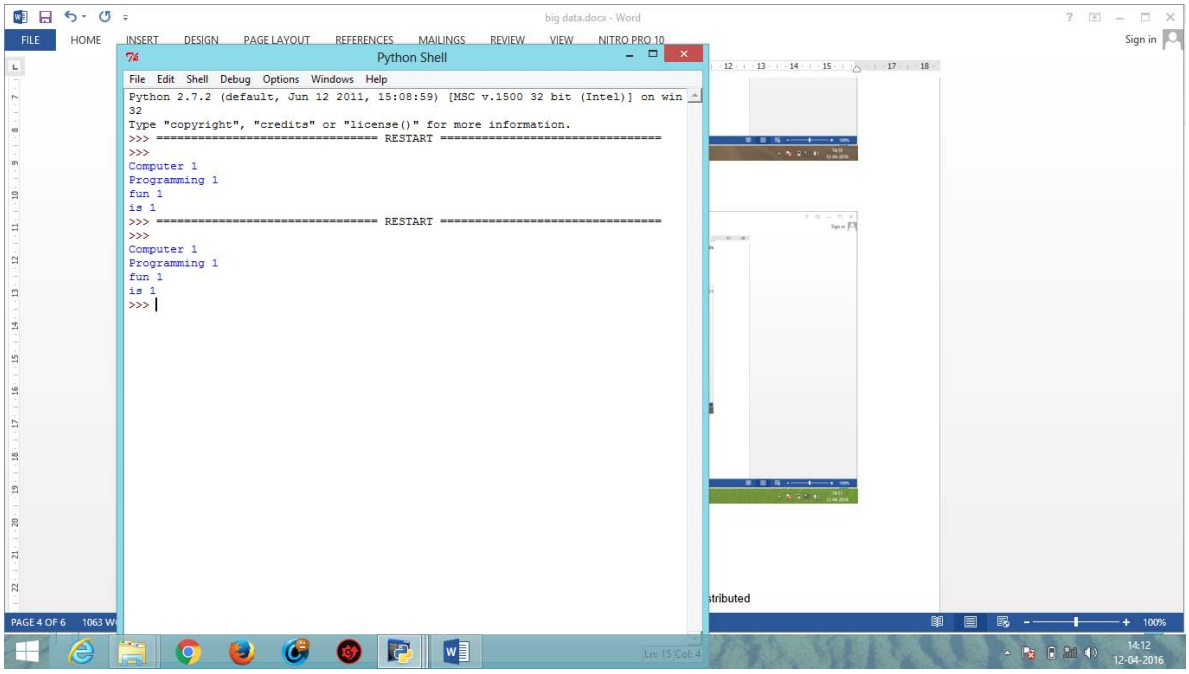

Figure : Snapshot of Execution 1

If the user manipulates the text then the output will be modified accordingly as indicated.

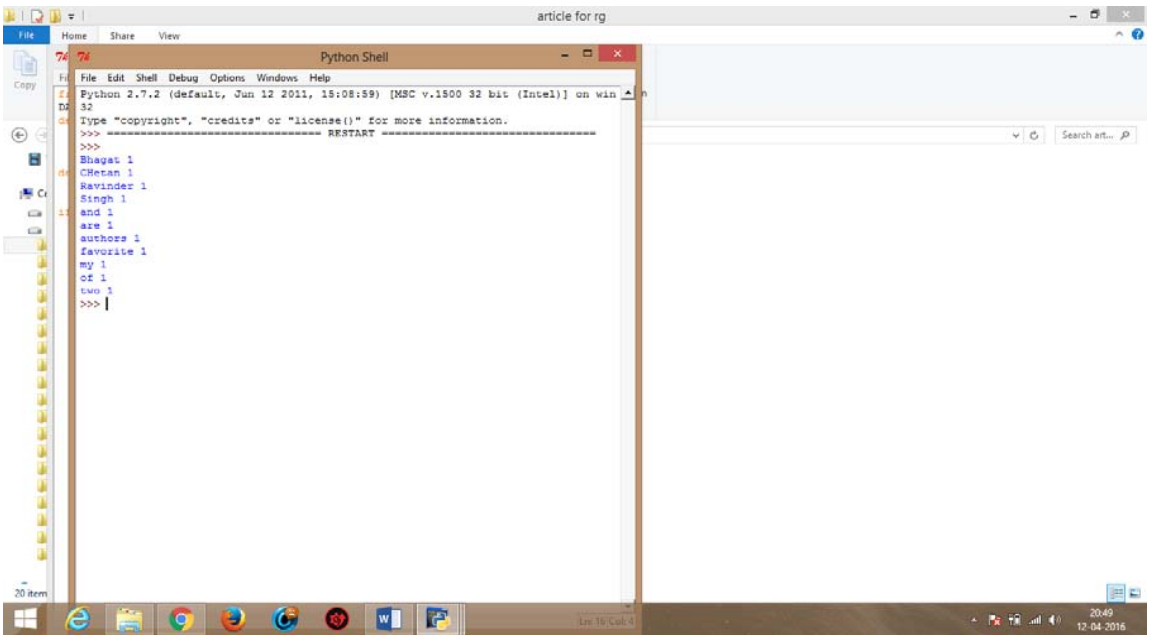

Figure : Snapshot of Execution 2

The MapReduce provides a framework where large volumes of data can be analysed. The tool can be extended futhur by increasing the volume of data supplied as well as some other scripting language can be adopted by the scientists to enhance the power of Big Data and thus make new discoveries in this discipline. Big Data is emerging as a powerful technique in recent years and provides solutions to the challenges of merging data thus making a mark in manifold fields like banking, health care, education which will involve whole world at large.

References Références Referencias

1. Lynch, C., (2008), Big data: How do your data grow?, Nature,455, 28-29, doi:10.1038/455028a.
2. R. Chaiken, B. Jenkins, P.-A. Larson, B. Ramsey, D. Shakib, S. Weaver, and J. Zhou. Scope: Easy and efficient parallel processing of massive data sets. In Proc. of VLDB, 2008.
3. G. Czajkowski. Sorting 1pb with mapreduce. googleblog.blogspot.com/2008/11/ sorting-1pb-with-mapreduce.html .
4. J. Dean and S. Ghemawat. MapReduce: Simplified Data Processing on Large Clusters. In OSDI, 2004.
5. Jeffrey Dean and Sanjay Ghemawat. "MapReduce: Simplified Data Processing on Large Clusters." In Communications of the ACM, Volume 51, Issue 1, pp. 107-113, 2008.
6. Matthew L. Massie, Brent N. Chun, and David E. Culler. "The Ganglia Distributed Monitoring System: Design, Implementation, and Experience". In Parallel Computing Volume 30, Issue 7, pp 817-840, 2004.
7. Luiz A. Barroso, Jeffrey Dean, and UrsH¨olzle. Web search for a planet: The Google cluster architecture. *IEEEMicro*, 23(2):22.28, April 2003.
8. Miller, H. G., and P. Mork, (2013), From Data to Decisions: A Value Chain for Big Data, *IT Professional*, vol. 15, no. 1, pp. 57-59, doi:10.1109/MITP.2013.11

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 16 Issue 2 Version 1.0 Year 2016
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

A Fitness Function for Search-based Testing of Java Classes, which is based on the States Reached by the Object under Test

By Ina Papadhopulli & Elinda Meçe
Polytechnic University of Tirana, Albania

Abstract- Genetic Algorithms are among the most efficient search-based techniques to automatically generate unit test cases today. The search is guided by a fitness function which evaluates how close an individual is to satisfy a given coverage goal. There exists several coverage criteria but the default criterion today is branch coverage. Nevertheless achieving high or full branch coverage does not imply that the generated test suite has good quality. In object oriented programs the state of the object affects its behavior. Thereupon, test cases that put the object under test, in new states are of interest in the testing context. In this article we propose a new fitness function which takes into consideration three factors for evaluation: the approach level, the branch distance and the new states reached by a test case. The coverage targets are still the branches, but during the search, the state of the object under test evolves with the scope to produce individuals that discover interesting features of the class and as a consequence can discover errors. We implemented this fitness function in the eToc tool. In our experiments the usage of the proposed fitness function towards the original fitness function results in a relative increase of 15.6% in the achieved average mutation score with the cost of a relative increase of 12.6% in the average test suite size.

Keywords: structural testing, test case generation, search based software testing, fitness function, object state, coverage criteria, mutation score.

GJCST-C Classification : D.1.1, D.1

Strictly as per the compliance and regulations of:

A Fitness Function for Search-based Testing of Java Classes, which is based on the States Reached by the Object under Test

Ina Papadhopulli [α] & Elinda Meçe [σ]

Abstract- Genetic Algorithms are among the most efficient search-based techniques to automatically generate unit test cases today. The search is guided by a fitness function which evaluates how close an individual is to satisfy a given coverage goal. There exists several coverage criteria but the default criterion today is branch coverage. Nevertheless achieving high or full branch coverage does not imply that the generated test suite has good quality. In object oriented programs the state of the object affects its behavior. Thereupon, test cases that put the object under test, in new states are of interest in the testing context. In this article we propose a new fitness function which takes into consideration three factors for evaluation: the approach level, the branch distance and the new states reached by a test case. The coverage targets are still the branches, but during the search, the state of the object under test evolves with the scope to produce individuals that discover interesting features of the class and as a consequence can discover errors. We implemented this fitness function in the eToc tool. In our experiments the usage of the proposed fitness function towards the original fitness function results in a relative increase of 15.6% in the achieved average mutation score with the cost of a relative increase of 12.6% in the average test suite size.

Keywords: structural testing, test case generation, search based software testing, fitness function, object state, coverage criteria, mutation score.

I. Introduction

Due to the fact that the influence of software in all areas has grown rapidly in the past 40 years, the software has become very complex and also its reliability is fundamental. All the software development phases have been adapted to produce these complex software systems, but especially the testing phase is of critical importance and testing thoroughly today's software systems is still a challenge. According to a study [1] conducted by the National Institute of Standard & Technology, approximately 80% of the development cost is spent on identifying and correcting defects. It is a well-known fact that it is a lot more expensive to correct defects that are detected during later system operation. Considering past experiences, inadequate and ineffective testing can result in social problems and human/financial losses. In order to improve the testing infrastructure, several efforts have been made to automate this process.

In the unit testing level, there are three approaches towards automation: random testing, static analysis (Symbolic Execution [3]) and metaheuristic search. A considerable number of tools have been developed based on these approaches; eg. RANDOOP [4], EvoSuite [5], AgitarOne [6]. Nevertheless, the effectiveness of these tools is still not completely proved, because the results obtained from the experiments depend on the subjects under test. Usually, a coverage criteria is used to evaluate these tools, but achieving a high degree of code coverage does not imply that a test is actually effective at detecting faults [7]. According to [8], today there is no tool to find more than 40.6% of faults.

This article is focused on structural testing at the unit level of Java programs using Search-Based Software Testing (SBST) [9]. According to [10], SBST has been used to automate the testing process in several areas including the coverage of specific program structures, as part of a structural, or white-box testing strategy. Every unit (class) of the software must be tested before proceeding to the other stages of the development cycle. SBST is a branch of Search Based Software Engineering (SBSE). SBSE is an engineering approach in which optimal or near optimal solutions are sought in a search space of candidate solutions. The search is guided by a fitness function that distinguishes between better and worse solutions. SBSE is an optimization approach and it is suitable for software testing since test case generation is often seen as an optimization or search problem. Since SBST techniques are heuristic by nature, they must be empirically investigated in terms of how costly and effective they are at reaching their test objectives and whether they scale up to realistic development artifacts. However, approaches to empirically study SBST techniques have shown wide variation in the literature. There exist several search-based optimization methods used for test automation; e.g. genetic algorithms, hill climbing, ant colony optimization and simulated annealing, etc, but Genetic algorithms (GAs) are among the most frequently applied in test data generation.

GAs have several components which need to be defined in order for the GA to be implemented.

Author α: Department of Computer Engineering Polytechnic University of Tirana Tirana, Albania e-mail: ipapadhopulli@fti.edu.al
Author σ: Department of Computer Engineering Polytechnic University of Tirana Tirana, Albania. e-mail: ekajo@fti.edu.al

According to [10], the component that affects mostly the results obtained from the search is the fitness function. The fitness function is a mathematical representation of the coverage goal the search should achieve. There are different coverage goals each of them aims at covering certain parts of the unit under test. These different coverage criteria verify the quality of a test suite. The gold criterion is strong mutation, but today this criterion it is mainly used by the research community for evaluation of proposed techniques. The most used criterion is branch coverage [11]. However achieving high branch coverage (even 100%), for some classes is not sufficient.

In object oriented programs the state of the object is a factor that affects the execution of a method. This is why the state of the object of the Class Under Test (CUT), should evolve during the search in order to discover hidden features of the class [12]. A test case that puts the object in one or several new states is of interest in the testing context. The scope of this paper is to propose and evaluate a new fitness function, which rewards the test cases according to branch coverage and also according to the new states the object has taken during the execution of the test.

The rest of this paper is organized as follows: In the second section we explain in what unit testing of java programs consists and in the third section we present an overview of GAs. The fourth section is focused on branch coverage and the fifth section presents the proposed fitness function. The implementation of the proposed fitness function is described in section six. The seventh section gives details of the experimental setup and in the eighth section the results achieved are presented and discussed. We conclude finally with the conclusions we have come preparing and accomplishing this study.

II. Unit testing for Object Oriented Software

Software testing at the unit level (Java classes) consists of three steps:

1) The design of test cases
2) The execution of these test cases
3) The determination of whether the output produced is correct or not.

The second step is performed fully automatically using frameworks like JUnit [2]. Automatically generating the test oracle is still a challenge and there exists few research publications regarding this topic [13], therefore the third step is almost completely performed manually by the testers. Regarding the first step, there exist a lot of research effort for the generation of test cases automatically. Due to the complexity and the diversity of the programs under test this is still an open research topic. Moreover test cases in object oriented unit testing are not just a sequence of input values like in procedural languages. According to [14], a unit test of a Java class must accomplish the following four tasks:

1. Create an object of the class under test using one of the available constructors.
2. Invoke a sequence of zero or more methods on the created object.
3. Execute the method which is currently under test.
4. Examine the final state of the object to produce the pass/fail result

Some parameters in method calls are objects themselves, thus requiring further object constructions and as a consequence task 1 and 2 must be repeated for each parameter of object type.
The statements for Java unit test cases are:

1. *Primitive statements: declaration of variables e.g. int a = 15;*
2. *Constructor statements: construction objects of any given class e.g. String s = new String("Test");*
3. *Method statements: calling the methods of any given class e.g. char b = s.charAt(2);*
4. *Field statements: accessing the fields of any given class e.g. int c = ob.size;*
5. *Assignments statements: assign values to the fields of any given class e.g. ob.size = 17;*

Since objects have a state, the results are affected by the state of the object under test and of the object parameters.

III. Genetic Algorithms

Genetic Algorithms (GAs) are inspired by natural evolution. They were first introduced by Holland in 1975. Today GAs are used for optimization in testing real life applications. The most important components in GA are:

- *representation of individuals:* genotype (the encoded representation of variables) to phenotype (the set of variables themselves) mapping
- *fitness function:* a function that evaluates how close an individual is to satisfy a given coverage goal
- *population:* the set of all the individuals (chromosomes) at a given time during the search
- *parent selection mechanism:* selecting the best individuals to recombine in order to produce a better generation
- *crossover and mutation:* the two types of recombination used to produce new individuals
- *replacement mechanism:* a mechanism which replace the individuals with the lowest fitness function in order to produce a better population.

a) How does the GA work?

The space of potential solutions is searched in order to find the best possible solution. This process is started with a set of individuals (genotypes) which are generated randomly from the whole population space (phenotype space). New solutions are created by using the crossover and mutation operators. The replacement mechanism selects the individuals which will be removed so that the population size does not exceed a prescribed limit. The basis of selection is the fitness function which assigns a quality measure to each individual. According to the fitness function, the parent selection mechanism evaluates the best candidates to be parents in order to produce better individuals in the next generation. It is the fitness function which affects the search towards satisfying a given coverage criteria. Usually the fitness function provides guidance which leads to the satisfaction of the coverage criterion. For each individual the fitness is computed according to the mathematical formula which represents how close is a candidate to satisfy a coverage goal, e.g. covering a given branch in the unit under test. GAs are stochastic search methods that could in principle run for ever. The termination criterion is usually a search budget parameter which is defined at the beginning of the search and represents the maximum amount of time available for that particular search.

IV. Coverage Criteria

a) Types of Coverage Criteria

Automatic unit testing is guided by a structural coverage criterion. There exist many coverage criteria in literature, each of them aims at covering different components of a CUT. Nevertheless, not all the criteria have the same strength and can be fulfilled practically. Furthermore some criteria are subsumed by other criteria. Below is a list of coverage criteria for structural testing of Java programs.

1. Line Coverage
2. Branch Coverage
3. Modified Condition Decision Coverage [21]
4. Mutation
5. Weak Mutation
6. Method coverage
7. Top-level Method Coverage
8. No-Exception Top Level Method Coverage
9. Direct Branch Coverage
10. Output Coverage
11. Exception Coverage
12. Path Coverage
13. Condition Coverage
14. Multiple Condition Coverage
15. Condition/Decision Coverage

Mutation criterion is considered the gold criterion in research literature [15]. This criterion is difficult to apply and computationally expensive and it is practically only used for predicting suite quality by researchers. Another option to achieve high quality test cases with search based technique is to use a combination of multiple criteria. [16] performed an experiment to evaluate the effects of using multiple criteria and concluded that:

- *Given enough time the combination of all criteria achieves higher mutation score than each criterion separately (except Weak Mutation).*
- *Using all the criteria increases the test suite size by more than 50% that the average test suite size of each constituent criterion used separately.*
- *The next best criterion (after Weak Mutation) to achieve high mutation scores is branch coverage.*

The usage of multiple criteria increases the overall coverage and mutation score with the cost of a considerable increase in test suite length, so the usage of the combination in practice will be not feasible, because managing large test suites is difficult. A balance between mutation score and average test suite size is achieved with branch coverage criterion.

b) Branch Coverage

The most used criterion is branch coverage, but even though it is an established default criterion in the literature, it may produce weak test sets (mutation score less than 30% [17]). For example consider the Stack implementation in Figure 1.

```
public class Stack {
        private int size = 0;
        private int st [] = new int [4];
        void push (int x){
                if (size < st.length)
                        st[size++] = x;
        }
        int pop (){
                return st[size--];
        }
}
```

Figure1 : Example Stack implementation

The class Stack is very simple (8 LOC, 2 attributes, 2 methods). Suppose the test suite generated is the test suite given in Figure 2.

```
1.  @Test
2.  public void test0()  {
3.  Stack s0 =new Stack();
4.  s0.push(1);
5.  s0.push(0);
6.  int int0 = s.pop();
7.  assertEquals(0, int0);
8.  s.push(0);
9.  s.push(0);
10. s.push((-1916));
```

```
11. s.push((-1916));
12. }
```

Figure 2 : Test suite for class Stack

We used EclEmma [35] tool as a plugin in Eclipse to measure branch coverage. The branch coverage obtained by executing this test suite was 100%. There are 4 coverage goals in class Stack (2 methods and 2 branches from the predicate in line 5).

Even though class Stack is very simple, and the branch coverage obtained is 100%, the mutation score is relatively low (29%). We added an assertion in the test (line 7) and used the JUnit framework to run it in Eclipse. The test passed. The tester may assume the class is correct with 100% branch coverage and a passing test. *Is branch coverage sufficient for this class?*

Analyzing class Stack we notice the following errors:

- If method pop is called first and then is called method push, an uncaught exception is thrown (field size before calling push is -1).
- If method pop is called two times consequently an uncaught exception is thrown (field size before calling pop is -1).
- If method push is called four times consequently and then is called method pop an uncaught exception is thrown (field size before calling pop is 4).
- It is obvious that branch coverage is not sufficient for class Stack!

Is there any possibility to improve the fitness function for branch coverage in order to obtain a test suite with higher quality?

Both of the methods are covered by the test generated, but it is evident that the state of the object (the value of field size) before calling them affects the results of the tests. The same method called on different states of the object behaves differently. This is why, a possibility to improve the suite's ability to detect errors, is to evolve the state of the object during the search in order to put the object in new states that probably can discover interesting behaviors of the CUT. Since the search is guided by the fitness function, then this function should also consider the states reached by a test before evaluating it.

V. The Proposed Fitness Function

Fitness functions are a fundamental part of any search algorithm. They provide the means to evaluate individuals, thus allowing a search to move towards better individuals in the hope of finding a solution [18]. The approach considered here is to minimize the fitness function during the search. The fitness function proposed in this paper rewards the individuals based on how close they are at covering a target (branch) and the states they put the object under test. This function is a mathematical equation depending on the:

- Approach level
- Branch Distance
- New states achieved

a) Approach Level

For each target, the approach level show how many of the branch's control dependent nodes were not executed by a particular input [20]. The fewer control dependent nodes executed, the "further away" an input is from executing the branch in control flow terms. The approach level is the most used factor in the fitness function for structural criteria, but the fitness landscape contain plateaus because the search is unaware of how close a test case was to traversing the desired edge of a critical branching node.

b) Branch Distance

The branch distance is computed using the condition of the decision statement at which the flow of control diverted away from the current "target" branch. For every operator the branch distance is calculated using the formulas introduced by Tracey [19].

The approach level is more important that the branch distance and as a consequence the branch distance should be normalized at the fitness function formula. This distance will be normalized at a value between 0.0 and 1.0. Value 0.0 means "true"; the desired branch has been reached. Values close to 1.0 means that the condition is far from being fulfilled. Intermediate values guide slightly the search towards the accomplishment of the condition (in order to remove plateaus in the fitness landscape). The formula for branch distance in our proposed fitness function is the formula introduced by Arcuri [21].

$$BD(normalized) = \frac{BD}{BD + \beta}$$

BD is the branch distance before normalization and β is 1.

c) New States Achieved (NSA)

With the term state in this paper we refer to:

Definition 1. State: The set of the values of all the fields in the CUT before calling a method + the method called.

For example, for the class Stack the two states:

- *field size = 0 and filed st = !null, before calling method push*
- *field size = 0 and filed st = !null, before calling method pop*

are considered two different states and both of them are interesting in the testing context.

The total number of states in the CUT is computed as a product of all the possible combinations

of the class fields (declared non final) after abstraction (explained in the next section), with the number of public methods.

The approach level is more important that the number of new states achieved and as a consequence this factor should be normalized at the fitness function formula. The normalization formula is:

$$NSA = \frac{states\_total - states\_new}{states\_total}$$

The greater the number of the new states achieved by a test case the smaller this factor in the overall fitness.

The fitness function proposed considers the three factors described above and is computed with the formula:

$$itness = approach\_level + \frac{BD}{BD+1} + \frac{states\_total - states\_new}{states\_total}$$

d) Abstract States

If we use the real values of the fields, the number of states will be infinite. Moreover, not all the states are of equal relevance during testing. For example, from the testing prospective, calling method pop() of the class Stack with field size = 1, is the same as calling method pop with filed size = 2. On the other hand calling method pop() with filed size = 0 in not the same, since this state reveals an interesting behavior of the object under test. Therefore, we use *abstractions over the values of the fields* rather than the concrete values themselves. We use a state abstraction function provided by Dallmeier at al. [34]. The abstraction is performed based on the three rules below:

- If the type of the field is concrete (int, double, long etc), the value will be translated in three abstract values: $x_i < 0$, $x_i = 0$ and $x_i > 0$.
- If the type of the field is an object, the value will be translated in two abstract values: x_i = null dhe $x_i \neq$ null
- If the type of the field is Boolean, there is no need to do translation, since there are only two values.

For example the combinations of the field values of class Stack, after abstraction are those listed in Table 1.

Table 1 : Combination of Field Values for Class Stack

	size	st
state1	= 0	null
state2	> 0	null
state3	< 0	null

VI. Implementation of the Proposed Fitness Function

The proposed fitness function was implemented in the eToc [22] tool. eToc is a simple search based tool for unit testing of Java programs. Is uses GA and branch coverage criterion. This tool has been mentioned in many research works and has been used as the basis for the design of other tools. eToc is appropriate for the scope used in this work. In the high level architecture of this tool [22], the Branch Instrumentor module and the Test Case Generator module need to be differently implemented for the search to be guided by the proposed fitness function. The new implementation of these modules is described below.

a) The Intrumentor

The function of the instrumentor module is to transform the source code of the CUT in order to provide information about the executed branches, the branch distance and the states achieved during execution. The new statements added during instrumentation must not change the behavior of the CUT. In order to obtain information for the states reached by the object under test, for each of the attributes (except those declared final) of the CUT, a ***get*** method will be added. A static analysis can be used to provide information about the mutators and inspectors methods of a class [23][24], but in this case a static whole-program analysis is required, which is very expensive for this context used. Since it is not the purpose here to obtain a behavioral model of the CUT, the get methods are appropriate to be used as inspectors for obtaining the state of the object because these methods:

- Return the value of an attribute
- Do not take parameters
- Do not have any side effects in the execution of the program.

Based on the state definition given in section 5.C, the get methods should be called before the execution of each method of the CUT, so during instrumentation the statements calling the get methods are added before the existing statements of each method. The concrete values are translated in abstract values as described in section 5.D. Then the states reached by a test case are saved in a LinkedList and consequently during fitness evaluation the new states achieved by a test case can obtained.

```
public class Stack {
        private int size = 0;
        private int st [] = new int [4];
        void push (int x){
                returnState();
                if (size < st.length)
        }               st[size++] = x;
        int pop (){
                returnState();
                 return st[size --];
         }
         public int getsize1(){
                 return size;
```

```
        }
        public Object getst1(){
                return st;
        }
        public void returnState (){
          reachedStates.add(String.valueOf(getsize1()+"
"+getst1());
        }
        static java.util.List reachedStates;
        public static void newReachedStates()
        {
      reachedStates = new java.util.LinkedList();
        }
}
```

Figure3 : Class Stack after instrumentation for the new ststes achieved

b) The Test Case Generator

The instrumented version of the CUT is executed repeatedly with the scope to cover a specified target (branch of the CUT). The state lists resulting after each execution are compared with the state lists of the test cases that make up the population. The new states reached by an individual are used to compute part of its fitness.

This module is also responsible for the minimization of the generated test suite. Normally during minimization the tests that do not cover any target that is not covered by any other test are omitted from the test suite. Taking into consideration that a test case that reaches one or more new states is important in the testing context, before removing a test case because it does not cover any new target, it will be reconsidered regarding the states it puts the object under test in. The test cases which contain unreached states in their state lists, will be part of the final test suite. The proposed minimization has the advantage that it probably increases the number of tests in the generated test suite and as a consequence it also increases the length of the test suite. On the other side the usage of the proposed fitness function is expected to increase the capability of the test suite to detect errors. An experimental evaluation of the new fitness function is presented in the next section.

Table II : Characteristics Of The Classes Selected Fo The Experiments: Name Of The Project, Loc, Number Of Public Methods, Number Of Branches, Number Of Mutants, Number Of Non-Final Fields, Cyclomatic Complexity, Project

Project	Class	LOC	Branches	Mutants	Non-final Fields	Public Methods	Cyclomatic Complexity	URL e projektit
	Staku	12	4	22	2	2	1.5	
Commons CLI	Option	155	131	140	9	42	1.52	https://commons.apache.org
	TypeHandler	124	28	28	0	9	2.66	
	AlreadySelectedExcepti	26	4	1	2	2	1	
	OptionGroup	86	21	19	2	8	1.875	
Math4 J	Rational	61	36	161	2	19	1.526	https://sourceforge.net/projects/math4j
	ExponentialFunction	40	11	31	1	9	1	
	ArrayUtil	320	167	1769	0	36	3.48	
	PolyFunction	245	100	827	2	12	3.63	
	Complex	102	24	682	2	20	1.091	
jdk	StringTokenizer	313	78	434	7	6	3.12	
Genetic Algorit	GAAlgorithm	65	14	6	6	8	2	https://sourceforge.net/projects
	Genome	14	9	21	3	4	1.4	
	Population	62	13	44	4	11	1.08	
Object Explor	ExplorerFrame	158	26	74	8	9	1.44	https://sourceforge.n
	ObjectViewManager	114	41	41	8	17	1.571	
Newz Grabber	DirectoryDialog	177	47	155	16	13	2.235	https://sourceforge.net/projects/newsgrabber
	NewsFactory	121	45	88	4	7	4	
	SongInfo	55	12	59	3	4	2	
	BatchJob	28	11	29	8	10	1.27	
	StringSorter	63	12	47	1	4	2.2	
	OptionsPanel	363	75	214	15	4	9.8	
Jipa	Label	18	11	42	3	4	1.8	https://sourceforge.n
	Variable	40	23	87	3	4	2.1	
Total		*2762*	*943*	*5021*	*111*	*264*		

VII. Experimental Evaluation

In this work we aim to answer the following research questions:

- *RQ1: How does the usage of the proposed fitness function affect the branch coverage?*
- *RQ2: How does the usage of the proposed fitness function affect the mutation score of the suite?*
- *RQ3: How does the usage of the proposed fitness function affect the number of suite's test cases and their size?*

a) System Characteristics

For the experiments we used a desktop computer running Linux 32 bit Operating System, 1 GB of main memory and a Intel Core 2 Duo CPU E7400 2.8GHz x 2 Processor.

b) Subject Selection

Selecting the classes under test is very important since this selection affects the results of the experiments. We chose 7 open source projects and selected randomly 23 classes from them. Also, the class Stack discussed throughout this paper was used as a subject for the experiments. To obtain comprehensive results, the evaluation must be done to real and not simple subjects. Also these subjects should not have any common characteristics which affect the obtained results. The characteristics of the 24 classes are listed in Table 2. The information about LOC (without comments and empty lines) and cyclomatic complexity is obtained using Metrics 1.3.6 [25], as a plugin in Eclipse. As can be noted from Table 2, the classes have very different characteristics and complexity.

Five of the projects were downloaded from SourceForge [26] which is today the greatest open source repository (more than 300,000 projects and two million of users). One project was downloaded from the Apache Software Foundation [27] which exists from 1999 and has more than 350 projects (including Apache HTTP Server). Class StringTokenizer was taken from the java.util package which is part of jdk 1.8.0. This package has been used by several studies for evaluation of automatic test case generation techniques.

a) Parameters of GA

Defining the parameters of GAs to obtain the optimal results is difficult and a lot of research effort is dedicated to this topic [28][29]. Therefore we let the parameters of the GA to their default values [22]. The values of three of the most relevant parameters are listed in Table 3. Regarding the search budget, it was determined depending on the experiment and will be shown next for each experiment.

Table III : Parameters of Ga

Parameter	Value
Population Size	10
Search Budget	600s
Maximal number of generations/target	10

b) Experiment

For each of the classes we run eToc with the following configurations:

1. Original Fitness (OF) function with search budget of 2 min
2. Proposed Fitness (PF) function with search budget of 2 min
3. Original Fitness (OF) function with search budget of 10 min
4. Proposed Fitness (PF) function with search budget of 2 min

To overcome the randomness of the genetic algorithms each experiment was repeated 5 times.

The results of the experiments (average of all runs) are presented in Table 4.

Table IV : Branch Coverage, Mutation Score, Number Of Tests, Length Of Test Suite For Each Configuration, Average Of All Runs For Each Cut

Class	BC with OF (2 min)	BC with PF (2 min)	BC with OF (10	BC with PF (10	MS with OF (2	MS with PF (2	MS with OF (10	MS with PF (2	No. test with OF	Test length with OF	No. test with PF	Test length with PF
Staku	100	100	100	100	29	72	29	72	2	8	4	15
Option	69	69	69	69	41	49	41	49	62	147	71	166
TypeHandler	75	75	75	75	46	46	46	46	12	24	12	24
AlreadySelectedException	100	100	100	100	100	100	100	100	3	5	3	5
OptionGroup	100	100	100	100	84	89	84	89	8	27	7	35
Rational	94	94	94	94	75	79	75	79	12	24	12	31
ExponentialFunction	100	100	100	100	60	55	60	60	8	16	7	15

ExponentialFunction	100	100	100	100	60	55	60	60	8	16	7	15
ArrayUtil	100	99	100	100	9	9	9	9	64	141	64	141
PolyFunction	-	-	85	87	-	-	31	38	27	89	30	98
Complex	100	100	100	100	34	37	34	37	13	27	12	31
StringTokenizer	65	65	69	69	15	21	19	23	8	18	16	33
GAAlgorithm	93	93	93	93	33	33	33	50	10	21	8	19
Genome	44	44	55	55	0	4	0	4	3	6	4	10
Population	92	92	100	100	32	32	32	32	11	29	11	29
ExplorerFrame	8	15	8	15	0	3	0	3	2	2	2	3
ObjectViewManager	54	54	54	54	17	24	17	24	2	3	2	3
DirectoryDialog	6	6	6	6	0	0	0	0	5	11	5	11
NewsFactory	-	-	-	-	-	-	-	-	-	-	-	-
SongInfo	50	50	50	50	22	27	24	27	5	12	8	19
BatchJob	100	100	100	100	62	69	62	69	10	20	9	22
StringSorter	100	100	100	100	17	17	17	17	6	17	6	17
OptionPanel	-	-	37	37	-	-	3	9	7	21	8	19
Label	100	100	100	100	55	55	55	55	4	16	4	16
Variable	100	100	100	100	55	56	56	59	6	9	9	19
Average	60.5	69	74.8	75.2	37.9	42.7	35.9	41.5	-	-	-	-
Total	-	-	-	-	-	-	-	-	290	693	314	781

VIII. Results and Discussion

RQ1: How does the usage of the proposed fitness function affect the branch coverage?

The branch coverage was measured with EclEmma. For both functions the average branch coverage is greater when the search budget is 10 min. This result was expected since the individuals improve during the search and more time results in better solutions.

In order to do the best comparison of the approaches we focus on the case with search budget of 10 min in this section, since for the scope of the experiment, it is not appropriate to compare results affected by the limited search time.

The difference between the average branch coverage is inconsiderable (0.4%) when a search budget of 10 min is used. This difference may be due to the randomness of the results achieved by the search. Since the approach presented in this work does not change the targets to cover, the almost equal coverage was expected. For the class ExplorerFrame, there is an increase of 7% in the coverage achieved by the proposed approach. Even though the targets are identical, the proposed function rewards the individuals that reach more new states and therefore the test cases after minimization may be different and more complex. So, this increase probably is the effect of indirect coverage.

Only in the case of class ArrayUtil there was a decrease of 1% in the coverage achieved, with budget 2 min, but more likely it is due to the randomness of the search. For the class NewsFactory the search failed to produce results for both approaches. We changed the parameters of the GA, but even for a population of 20, or 30 individuals, no results were generated. It is not the scope of this work to investigate the reasons why this happened.

RQ1: In our experiments, there is no difference in the average branch coverage achieved between the usage of the original fitness function and the proposed fitness function.

- *RQ2: How does the usage of the proposed fitness function affect the mutation score of the suite?*

Since mutation score is the measure used in the strongest criterion (Mutation Coverage), here we have used it to measure the quality of the generated test suite. Computing the mutation score for a test suite requires determining, for every mutant, whether the test suite succeeds or fails when run on the mutant. In the worst case each test must be run on each mutant. For each of the classes the mutants were generated using as a plugin in Eclipse the tool MuClipse v1.3 [30]. Mu Clipse generates mutants using the traditional operators and the operators in the class level [31]. The number of generated mutants for each class is given in Table 2. Even classes with a small number of LOC can have many mutants (e.g. class Stack has 22 mutants). Assertions were inserted manually to the tests

generated, so that these cases can be used by MuClipse. Then, the generated tests were executed with JUnit against all the mutants and the presence of failures shows that the tests were able to kill the mutants.

The results of the mutation scores of each class for all the configurations are given in Table 4.

The mutation scores achieved by both of the fitness functions are far from the optimal value (100%). Almost this range of mutation scores is also obtained from other studies [32]. The main reasons of these low scores are:

- the targets to cover are the branches and not the mutants
- the presence of equivalent mutants (behave the same as the original program) which cannot be killed.

Nevertheless, despite the relatively low mutation scores, our interest is focused on the difference between the scores achieved by the original function against the proposed function.

For 6 classes (6/23 = 26%) there is an improvement in the mutation score achieved when using a search budget of 10 min against a search budget of 2 min.

For the same reasons mentioned in the discussion of RQ1, to answer RQ2 we are focusing mainly at the results achieved with a search budget of 10 min. The average mutation score reached by the original function is 35.9%, whereas the mutation score reached by the proposed function is 41.5%, thus a difference of 5.6%. The improvement is 5.6/35.9 = 15.6%. For 15 classes out of 23 (15/23 = 65%), there is an improvement in the mutation score achieved by the proposed function; for the remaining 8 classes (8/23 = 35%), the scores achieved are identical. There is no class where using the proposed function results in a lower mutation score. Even though we are aware that the results depend on the CUT (despite the fact that CUT chosen have different characteristics), the results obtained are very promising.

RQ2: In our experiments, the usage of the proposed fitness function results in a relative increase of 15.6% in the average mutation score achieved against the original fitness function.

- *RQ3: How does the usage of the proposed fitness function affect the number of suite's test cases and their size?*

Automatically generated JUnit tests need to be manually checked in order to detect faults because automatic oracle generation is not possible today. This is the reason why not only the achieved coverage of the generated test suite is important, but the size of the test suite is of the same importance [33].

Here we refer to the *size of a test suite* as the number of statements after the minimization phase (without assertions).

Only the results achieved with a search budget of 10 min, are shown in Table 4, because in answering RQ3 we are interested in the number of tests generated and their size in the "worst case". The minimization phase does not depend on the search budget, so the results with search budget of 10 min, subsume the scenario with a search budget of 2 min. The LOC of the generated suite was obtained with the tool Metrics 1.3.6.

There is an increase of 314 – 290 = 24 tests in the total number of test generated, or a relative increase of 24/290 = 8.2%. This increase is acceptable, although the number of tests in the test suite is not relevant in respect to the size of the test suite, because having many short size tests is not a problem for the tester who is detecting faults.

Regarding the size of the test suite, we can see from the results in Table 4, that using the proposed fitness function results in an average test suite size of 33.9 (781/23) statements. The relative increase is (33.9 – 30.1) / 30.1 = 12.6%. For 8 of the classes (34%), there is no change in the average test suite size. Regarding classes ExponentialFunction and GAAlgorithm (8.7% of the classes), there is a decrease in the average test suite size, although there is no decrease either in branch coverage or mutation score. These results are explained with the appearance of indirect coverage [36].

ArrayUtil is the class with the greatest test suite size because of the large number of branches (167). The average increase in test suite size with the usage of the proposed function is the consequence of two reasons:

- During the minimization phase the test cases that do not cover any target, but put the object under test in new states, are added in the minimized test suite (as explained in Section 6)
- Two different fitness functions probably will generate different test suites with different number of statements (not necessarily a larger number).

RQ3: In our experiments, the usage of the proposed fitness function results in a relative increase of 8.2% in the average number of test cases and 12.6% in the average test suite size achieved against the original fitness function.

IX. Conclusions

This paper concerns the fitness function used to guide the search during automatic unit test generation of Java classes. The branch coverage criterion is easy to implement but can produce weak test sets. Test cases that put the object under test in new states discover hidden behaviors and consequently are relevant in the testing context. Targeting all the states during the search is impossible due to the fact that some of them are infeasible. In this article we presented a new fitness function that takes into consideration the states reached during the execution of a test case. The implementation of this fitness function is very simple since the targets to cover remain the branches, but the state evolve during the search and the minimization phase the tests that reach one or more new states are not removed even though these tests does not reach any uncovered branches. The usage of the proposed fitness function does not decrease the branch coverage and results in a relative increase of 15.6% in the achieved average mutation score with the cost of a relative increase of 12.6% in the average test suite size. The results are promising but since the subjects under test are very different further evaluation of the proposed approach needs to be performed.

References Références Referencias

1. NIST (National Institute of Standards and Technology): The Economic Impacts of Inadequate Infrastructure for Software Testing, Report 7007.011,
2. Y. Cheon and G. T. Leavens. A simple and practical approach to unit testing: The JML and JUnit way. Technical Report 01-12, Department of Computer Science, Iowa State University, Nov. 2001.
3. C. Cadar, K. Sen, "Symbolic Execution for Software Testing: Three Decades Later". Communications of ACM, pages 82-90, 2013
4. C. Pacheco, S. Lahiri, M. Ernst, "Feedback-directed Random Test Generation". In Proceedings of International Conference in Software Engineering (ICSE) 2007
5. G. Fraser, A. Arcuri, "EvoSuite at the SBST 2015 Tool Competition". In Proceedings of International Conference in Software Engineering (ICSE) 2015
6. T. Tsuji, A. Akinyele, "Evaluation of AgitarOne". Analysis of Software Artifacts Final Project Report April 24, 2007
7. G. Fraser, P. McMinn, A. Arcuri, M. Staats, "Does Automated Unit Test Generation Really Help Software Testers? A Controlled Empirical Study". ACM Transactions on Software Engineering and Methodology, 2015
8. S. Shamshiri, R. Just, J. Rojas, G. Fraser, P. McMinn, A. Arcuri, "Do Automatically Generated Unit Tests Find Real Faults? An Empirical Study of Effectiveness and Challenges" In Proceedings of the 30th IEEE/ACM International Conference on Automated Software Engineering (ASE), 2015
9. F. Gross, G. Fraser, A. Zeller, "Search-based system testing: high coverage, no false alarms". In Proceedings of International Symposium on Software Testing and Analysis (ISSTA), 2012.
10. P. McMinn, "Search-based Software Test Data Generation: A Survey", Software Testing, Verification and Reliability, pp. 105-156, June 2004.
11. K. Lakhotia, P. McMinnb, M. Harman, "An empirical investigation into branch coverage for C programs using CUTE and AUSTIN". Journal of Systems and Software, 2010
12. M. Mirazz, "Evolutionary Testing of Stateful Systems: a Holistic Approach". PhD thesis, University of Torino, 2010
13. G. Fraser, A. Zeller, "Mutation-Driven Generation of Unit Tests and Oracles," IEEE Transactions on Software Engineering, 2012.
14. P. Tonella, "Search-Based Test Case Generation", TAROT Testing School Presenetation, 2013
15. G. Fraser, A. Arcuri, "Achieving Scalable Mutation-based Generation of Whole Test Suites". Empirical Software Engineering 2014.
16. Papadhopulli, E. Meçe "Coverage Criteria for Search Based Automatic Unit Testing of Java Programs", International Journal of Computer Science and Software Engineering (IJCSSE), Volume 4, Issue 10, October 2015
17. J. Miguel Rojas, J. Campos1, M. Vivanti, G. Fraser, A. Arcuri, "Combining Multiple Coverage Criteria in Search-Based Unit Test Generation" in Proceedings of the 26th IEEE/ACM International Conference on Automated Software Engineering (ASE), pp. 436-439, 2011
18. K. Lakhotia, M. Harman, H. Gross, "AUSTIN: A Tool for Search Based Software Testing for the C Language and Its Evaluation on Deployed Automotive Systems". International Symposium on SBSE, 2010
19. N. Tracey. "A Search-Based Automated Test-Data Generation Framework For Safety-Critical Software". PhD thesis, University of York, 2000
20. J. Wegener, A. Baresel, H. Sthamer. "Evolutionary Test Environment for Automatic Structural Testing". Information and Software Technology Special Issue on Software Engineering using Metaheuristic Innovative Algorithms, 43(14):841{854, December 2001.
21. Arcuri, "It Does Matter How You Normalise the Branch Distance in Search Based Software Testing". Third International Conference on Software Testing, Verification and Validation, 2010"
22. P. Tonella, "Evolutionary Testing of Classes". In Proceedings of International Symposium on Software Testing and Analysis (ISSTA) 2004

23. Rountev, "Precise identification of side-effect-free methods in java". 20th IEEE International Conference on Software Maintenance (ICSM '04), pages 82–91, 2004.
24. Salcianu, M. Rinard, "Purity and side effect analysis for java programs". In Proceedings of the 6th International Conference on Verification, Model Checking and Abstract Interpretation, pages 199–215, January 2005.
25. http://metrics.sourceforge.net/
26. http:/.sourceforge.net/
27. http://www.apache.org/
28. Aleti, L. Grunske, "Test Data Generation with a Kalman Filter-Based Adaptive Genetic Algorithm". Journal of Systems and Software, 2014.
29. E. Eiben, S. K. Smit, "Parameter tuning for configuring and analyzing evolutionary algorithms". Journal: Swarm and Evolutionary Cmputation, pages 19-31, 2011.
30. http://muclipse.sourceforge.net/
31. Y. Ma, J. Ouffut, "Description of Class Mutation Mutation Operators for Java", August 2014
32. D. Le, M. Alipour, R. Gopinath, and A. Groce, "MuCheck: An Extensible Tool for Mutation Testing of Haskell Programs". In Proc. of the International Symposium on Software Testing and Analysis, 2014.
33. G. Fraser, A. Arcuri, "Handling test length bloat". In Proceedings of ICST, 2013.
34. V. Dallmeier, C. Lindig, A. Vasilowski, "Mining Object Behaviour with ADABU". In Proceedings of the International Workshop on Dynamic Systems Analysis, 2006
35. http://www.eclemma.org/
36. Papadhopulli, N. Frasheri, "Today's Challenges of Symbolic Execution and Search-Based for Automated Structural Testing", In Proceedings of ICTIC, 2015.

This page is intentionally left blank

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 16 Issue 2 Version 1.0 Year 2016
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

A Systematic Review of Learning based Notion Change Acceptance Strategies for Incremental Mining

By D. S. S K. Dhanalakshmi & Dr. Ch.Suneetha

CMR college of Engineering & Technology, India

Abstract- The data generated contemporarily from different communication environments is dynamic in content different from the earlier static data environments. The high speed streams have huge digital data transmitted with rapid context changes unlike static environments where the data is mostly stationery. The process of extracting, classifying, and exploring relevant information from enormous flowing and high speed varying streaming data has several inapplicable issues when static data based strategies are applied. The learning strategies of static data are based on observable and established notion changes for exploring the data whereas in high speed data streams there are no fixed rules or drift strategies existing beforehand and the classification mechanisms have to develop their own learning schemes in terms of the notion changes and Notion Change Acceptance by changing the existing notion, or substituting the existing notion, or creating new notions with evaluation in the classification process in terms of the previous, existing, and the newer incoming notions. The research in this field has devised numerous data stream mining strategies for determining, predicting, and establishing the notion changes in the process of exploring and accurately predicting the next notion change occurrences in Notion Change.

Keywords: notion change, defencing notion change, conventional learning, supervised notion change acceptance, unsupervised notion change acceptance, data stream mining and concept evolution.

GJCST-C Classification : H.2.8, D.3.4, D.2.3

Strictly as per the compliance and regulations of:

A Systematic Review of Learning based Notion Change Acceptance Strategies for Incremental Mining

D. S. S K. Dhanalakshmi [α] & Dr. Ch.Suneetha [σ]

Abstract- The data generated contemporarily from different communication environments is dynamic in content different from the earlier static data environments. The high speed streams have huge digital data transmitted with rapid context changes unlike static environments where the data is mostly stationery. The process of extracting, classifying, and exploring relevant information from enormous flowing and high speed varying streaming data has several inapplicable issues when static data based strategies are applied. The learning strategies of static data are based on observable and established notion changes for exploring the data whereas in high speed data streams there are no fixed rules or drift strategies existing beforehand and the classification mechanisms have to develop their own learning schemes in terms of the notion changes and Notion Change Acceptance by changing the existing notion, or substituting the existing notion, or creating new notions with evaluation in the classification process in terms of the previous, existing, and the newer incoming notions. The research in this field has devised numerous data stream mining strategies for determining, predicting, and establishing the notion changes in the process of exploring and accurately predicting the next notion change occurrences in Notion Change. In this context of feasible relevant better knowledge discovery in this paper we have given an illustration with nomenclature of various contemporarily affirmed models of benchmark in data stream mining for adapting the Notion Change.

Keywords: notion change, defencing notion change, conventional learning, supervised notion change acceptance, unsupervised notion change acceptance, data stream mining and concept evolution.

I. Introduction

The data streams generated in real time are dynamic in content unlike the contemporary static data environments and involve huge volumes of data transmitted at great speeds. These dynamic data communication environments are used in various fields such as, real time surveillance and monitoring, web traffic internet networks, applications producing huge HTTP data requests, weather monitoring or environment systems, RFID and wireless sensor networks, retail transactions, real time media streaming processes, cloud automations systems, telephone networks, etc.

The applications of data streams mining are many such as, financial analysis of stock market drifts, customer data transactions analysis, predicting customer preferences in retail or online shopping, telephone call records analysis, fraud prevention, social networks user content generation, internet networks traffic mining for knowledge exploration, in spam and intrusion detection, etc. The data generated by the World Wide Web in the year 2013 is stated to be around 4 zettabytes of data [1] and this is growing with magnified volumes and speed continuously. In this context and the many application areas of mining streaming data the research of real world data classification has acquired great importance both for researchers as well as for the business community.

The incremental mining process has to be associated with an efficient strategy to handle the huge volume of dynamically changing data streams that do not have any established notions in Notion Change [2] [3] and the learning algorithm applied over the drifting data streams has to find context changes in terms of the drift. The process of Notion Change Acceptance in data streams involves the objects of the streaming data categorized in terms of a concept individually either as positive or as negative case concepts. The newer concepts are mined and analyzed with visible case concepts variables based on recent topology information with a learning algorithm for predicting each incoming streaming objects case concept. The learning process strategy of adapting to the notion change is either with a set of fixed notions called supervised learning, or uses the dynamic notion which is called as unsupervised learning strategy. The learning model in exploration and extraction of the information incorporates successful predictions and if the prediction is in contrast to the objects real case concept, it changes the existing model with deletion of the old concepts.

The research in recent years has generated several strategies of benchmark Notion Change Acceptance. To further widen the scope of understanding and aiding future studies in this field we in this paper review existing strategies available in recent literature in terms of their, advantages, weaknesses, applicability, and compatibility with various domain streams and the wide scope of Notion Change

Author α : Assistant Professor, Dept., of MCA, CMR college of Engineering & Technology, Hyderabad-501401, Telangana State.
e-mail : dasaridhana123@gmail.com
Author ρ: Associate Professor, Dept., of Computer Applications, RVR & JC College of Engineering, Chowdavaram-522019, Andhra Pradesh.
e-mail : suneethachittineni@gmail.com

influenced streaming data under different contexts of notion change. Also in this paper we review the notion progress in the process of adapting Notion Change and a description of the nomenclature of data stream mining.

The remaining sections of the paper is structured as follows; Section 2 assesses the streaming data based mining nomenclature and the Notion Change impact; Section 3 reviews the models of benchmark devised and affirmed in contemporary literature. Section 4 gives conclusion of this paper with future research scope.

II. The Streaming data and Notion Change Acceptance

a) Incremental mining

The data streams mining process involves processing, classifying the dynamic data where the concepts might change, appear or not appear again requiring constant adaptation according to the notion change and in accordance to the data influx speed for an efficient exploration and retrieval of hidden relevant contexts.

b) Notion Change and the objective

The objective of learning over streaming data must be for noticing the change of notion or Notion Change by applying efficient mining strategies with the learning mechanism. If we consider the content streaming sites a considerable notion change of viewer's preferences alters the data streams in terms of the drift of concepts. Hence the principal aim in learning and mining of user preferences must be to recognize the notions changes.

c) Notion Change Acceptance

The notion change due to Notion Change based on a learning mechanism is adapted with appropriate learning model changes for mining efficiency and for significant data retrieval.

i. *Notion changeover frequency*

The frequency of notion change or "speed of the Notion Change" in streaming data is the average recorded time for every Notion Change Acceptance occurrence. The learning mechanism of drift Acceptance is of 2 different kinds, regular Notion Change, and impulsive Notion Change. In regular Notion Change the Notion Change event is reflected in fixed time intervals for the probable prediction of the time of next Notion Change occurrence which is usually a recursive Notion Change that converts the data to earlier state. The impulsive Notion Change learning mechanism tries to replicate unexpected and irregular Notion Change occurrences.

d) The taxonomy of Notion Change Acceptance strategies

The strategies for adapting the Notion Change implement the learning mechanism in two stages; stage 1 involves determining the significant notion change which has importance towards the Notion Change, and stage 2 adapts the data streams newest state into the learning process. The Notion Change adaptation into the existing learning mechanism may be ordered and explained as below,

e) Mutable learners

An easy technique for Notion Changes Acceptance into the learning mechanism are the mutable learners which use supervised or semi supervised learning approaches for the data streams visible scope to be dynamically expanded or restricted in terms of the newer state of the data streamed for updating the learning model with deletion of obsolete data instances.

i. *Mutable Training set based Learners*

A type of Notion Change class which use a mutable training set in the learning process these learners use a strategy of unsupervised learning based on a window of a set records grouped by considering comparable notion, or comparable instance weights. The classifiers based on the windows newest notion or newest instances having comparable weights mute the existing training set and update the learning set applied over the data streams for Notion Change Acceptance.

ii. *Colle ctive Learners*

A strategy that applies multiple learners' collectively is a well known standard data engineering approach for its realistically achievable efficiency. A learning mechanism based on diverse classifiers may be applied over streaming data with either unassigned notion changes or imbalanced data classes for achieving substantial variation in the learning process. The collective learner strategy may be applied over identical data to expand the achievable accuracy with the inclusion of a predictive classifier. A learning mechanism not based on multiple learners' experiences over fitting and decrease of performance.

Another extensively important concept for leaning over Notion Changes prone data streams in the progression of new classes. This notion progression maybe further defined using two types of networks, internet networks using the learning mechanism for intrusion prevention, and social networks use for identifying initiation of new trends. In Internet networks the notion progression is visible in case we associate a class label to every kind of attack and when the traffic is under an entirely new type of attack it results in notion progression. The social networks data streams have class labels associated with trends and the origin of a new trend whose posts are unlike the earlier posts

enables notion progression. However very less importance has been given to notion progression in the contemporary research.

III. Notion Change Acceptance Strategies

The decision trees is a type mutable learner's model C4.5 [4] to be specific. The model Very Fast Decision Tree (VFDT) [5] is known as one of the initial models designed based on decision trees. The essentials of decision-tree learning are used in the design of the VFDT algorithm by which accurate and very fast decision trees are generated. These decision trees formed for data streams with the VFDT model are based on the application of the Hoeffding tree algorithm where a comparable notion subset is produced using Hoeffding bounds [6], [7]. This model mines real time data streams which are imperfect or having uncertainty characteristics depicting the entire streams as a unique advanced model in which with the arrival of newer data the decision tree's existing database is constantly updated making it more efficient in the prediction of drift in new incoming data.

The problem with this model in incremental mining is mostly due to the noise in the data used for training which forms unnecessary tree branches causing over-fitting that is further complicated because of the run-time memory inadequacy in the total decision tree accommodation declining the prediction accuracy frequently. This affects the main reason for implementing this model which is the achievable reasonable accuracy. This need for rapidly adapting the Notion Change with precision has made the model undergo several revisions.

The rule k Nearest Neighbor (kNN) is one of the earliest classification rules and also the easiest one that has been researched widely for numerous different objectives in various fields especially Notion Change Acceptance. The kNN algorithm is a type of incremental classifier which does not include any previous conventions of the data distribution and with the rapidly changing streaming data performs the learning and training continuously updating the classification model.
The major difficulty with the approach however is it initiates the learning only during the time of prediction which increases the overhead in terms of time and cost especially in case of instances related to multi-label data. Also the approach is involved with computational complications when used with non-incremental type of base classifiers. A revision of the kNN based algorithm for streaming data by Alippi and Roveri [8][9] in case of a streaming data not under Notion Change is based on choosing k samples from the data using the theory based outcomes by Fukunga [10] where these newer examples are added to the knowledge base and the kNN based classifier is updated. In case of Notion Change the approach retains all the newer examples and eliminates all the old examples from its knowledge base. For data streams drifting regularly at a lesser pace a revised model is presented later by the authors, called adaptive weighted kNN which uses the strategy of assigning weights to the examples based on their nearness to the present concept where older instances comparatively still have considerably higher weights associated.

A Notion Change learning process for a streaming data is an algorithms capability of learning incrementally the newer incoming streaming information while maintaining the earlier data in the classification process. The research of Carpenter, G., Grossberg, S., Markuzon, et. al., of this difficult problem has led to the design of the Adaptive Resonance Theory (ART) model for effective classification and prediction of concept change with a model called ARTMAP (Adaptive Resonance Theory Map) [11] which is a strategy of unsupervised learning that recognizes from the data set all the different patterns incrementally with cluster formation and applies supervised learning over these clusters in the classification. The attributes of every cluster is used to map the cluster to a class considering class compatibility in terms of their labels. Carpenter, G., Grossberg, et. al., devised a Fuzzy ARTMAP [12] which applies fuzzy principals in the ARTMAP model with 2 variants of the ART model called ARTa and ARTb which are connected with an inter Adaptive Resonance module. The strategy of assessing the patterns recognized by the model is implemented with an unsupervised ART's model and the prediction process is implemented with a supervised ART model in an incremental order. In this prediction process a class in ARTa is linked to a class in ARTb and this mapped field is used to form predictive classes in learning the class associations. In case of an incompatibility scenario with existing classes the search is either repeated or a new clusters is created for properly including the newer input patterns that are dissimilar to the earlier observed examples. The incremental rule pruning strategy for fuzzy ARTMAP by Andres-Andres, A., Gomez-Sanchez, E., BoteLorenzo, M., in [13] extends the fuzzy ARTMAP models devised earlier. The model is based on updating fuzzy rules frequently with dynamic pruning the inactive and or obsolete fuzzy rules based on a pruning strategy in the paper [14] which prunes the rules set in terms of their attributes, rule confidence, rule usage frequency and rule significance.

These models of ARTMAP and Fuzzy ARTMAP are used widely in the process of incremental learning. However the problem with these models is with noisy training data where the performance becomes ineffective. The fuzzy ARTMAP model constructs maximum possible classes for learning the entire static training set and in the statistical assessment, which due to overfit leads to pending of parameter selection and

the resulting generalization is ineffective. These problems are overcome with the strategy proposed in [13]. However an assessment of all the 3 models [11][12][13], considering recursive concept impacted Notion Change of high frequency shows the rule update process to be computationally complex and redundant in comparison to other similar methods.

The model AO-DCS (Attribute-oriented Dynamic Classifier) devised by Xingquan Zhu, Xindong Wu et.al., [15] is a supervised learning approach based on a single best learning algorithm whose performance efficiency is far advanced in contrast to other collective learners currently existing. This approach instead of a CC (Classifier Combination) method uses a CS (Classifier Selection) technique called DCS (Dynamic Classifier Selection) to overcome the inefficiency of the Classifier Combination method in mining highly noisy data streams under dramatic Notion Changes. The dynamic classifier selection scheme uses attribute values of instances to partition the evaluation set into subsets. This approach uses instead of the clustering technique the attribute values of the evaluation set to classify the data set into a number of small sets and the new examples are used to find the final subsets. The existing base classifiers are applied on these subsets with new examples to evaluate the performance effectiveness of a base classifier in terms of a specific domain and determine the choice of a best classifier.

The experiments are executed with 8 datasets of benchmark data streams of the UCI database repository comprising of synthetic and real time data. The experiments using a real time simulated scenario evaluates the systems performance for incremental mining under dramatic Notion Changes applying different DCS approaches with different factors like scalability, robustness and accuracy. In this process prior to data partitioning several levels of class noise or manual errors are fed into the data stream. The execution of every experiment is assessed with 10-fold cross-validation and the obtained average accuracy is used as the final result. The test outcomes show an enhanced performance with the devised DCS method compared to most other CC or CS based approaches like SAM, CVM, DCS_LA and Referee in mining real-time data streams. However a major problem with this approach is in case of high frequency Notion Changes and since the accuracy is inversely proportional to the Notion Change frequency, the learning factors of accuracy, scalability and robustness decrease the performance.

A Notion Change Rule mining Tree (CDR-Tree), for exploring, finding and precisely assessing Notion Change rules by Chien-I Lee, Cheng-Jung Tsai, Jhe-Hao Wu, and Wei-Pang Yang [16] is a unique and different approach of determining the Notion Change causes. The previous approaches devised were based on the strategy of modifying the current database for accurately classifying the incoming data and not for finding the reasons for the drift occurrences. The authors of the CDR-Tree model represent the reasons of drift occurrences in terms of categorically ordered rules set based on which the examples of old data and the incoming newer data related to different time periods are coupled to create a CDR-Tree where IG (information Gain) is used to find the node's split point in the process of forming the CDR-Tree structure. The defined Notion Change rules set is further defined by CDR-Tree with RS (Rule Support) and a RC (Rule Confidence) to screen less important instances with user specific threshold values that can be set for notable rules.

The experiments were performed with Microsoft VC++ 6.0 to depict the CDR-Tree and IBM Data Generator is used for the generating the experimental data comprising of 1 Boolean target class and 9 basic attributes given by 4 random classification functions. Here 20 integrated data sets with 6 dissimilar drift levels are tested and the results show the proposed approach achieves high accuracy in all the 20 data sets considered. The devised approach overcomes the limitations of the earlier strategies which are unable to continue the node split process in case of real time streaming data. The model is able to correctly compute the drift in case of data streams truly under Notion Change. However the concept-drifting rules in case of higher cNotion Change levels make the CDR-Tree more complex affecting the accuracy of mining. To reduce the complexity of the CDR-Tree discretization algorithm are proposed which however fail in achieving the desired accuracy. Also the chances of tree construction are highly reduced in case of streaming data under recursive concepts based Notion Change.

A stacking style ensemble-based strategy by Yang Zhang, Xue Li, [17] is devised to address the problem of single class classification of Notion Change influenced high speed and constantly changing noisy data streams with limited memory space. The objective is to use few class labels during training. A stacking strategy based ensemble learning approach is used for classifying the Notion Change exposed texts. The approach presumes the data to be coming in batches streamed in varying lengths. The classification uses only a single class and every batch streamed is classified using very few training samples. The training data selected from every batch is a positive training data set of k number of documents selected initially from 2 scenarios with the remaining data used as unlabeled data. For a more reliable sample extraction subsequently the negative samples are included along with positive samples in the data used in training. The ensemble of classifiers created from the different batches of streaming data are used to find class labels of new incoming data and inefficient classifiers in the classification process are removed from the ensemble to control the ensemble capacity dependent on

limitations of memory. The algorithms devised are used to build the classifier where linear SVM classifier is used as base classifier. In the training of the base classifiers every batch is trained on 2 base classifiers, one with positive samples and another including positive samples on earlier batches as well. The learning mechanism based on ensemble stacking uses the concept descriptions preserved in its database in the learning with prediction for voting and selecting the best base classifier.

The experiments of this approach are implemented in Java simulated in WEKA with 1G memory with a dataset comprising of 20 newsgroup classes where each class has 1000 texts documents. The documents after preprocessing are vector represented with weights using the TFIDF algorithm. The simulations are done with 15 different scenarios where each scenario has 10 batches of text data. Each batch has 100 documents from each of the 20 different classes equaling to 2000 text documents. The simulations test outcomes show that the classifier achieves good performance in classification and predicting the different types of Notion Changes occurring in every batch due to variations in the user interests and distribution of data. The stacking approach is a successfully strategy for managing data streams with recursive concepts based Notion Changes. The classification efficiency achieved is higher with the devised EN methodology compared to similar window-based methods like single window (SW), fixed window (FW), and full memory (FM). However the problems with this approach are, in case of a high frequency of Notion Change the approach is unable to regulate the usage of memory where more number of stacks are required, and in case of noisy streaming data the complications associated with the process also increase.

A collective learner approach by Stephen H. Bach, Marcus A. Maloof in [18] adapts a learner pair for streaming data classification with better performance compared to other contemporary approaches. In an online learning task the Notion Change learners have to be reactive and stable for detecting the frequently occurring concept changes and this aspect is used in the devised PL-NB approach's learning mechanism where a stable learner is paired with a reactive learner in the process of finding the Notion Change and securing the newly incoming target concept. The approach focuses on the most recent time period during which concept change has occurred in the streaming data. In this window of concept change the reactive learner has better accuracy for determining the Notion Change occurrence compared to the stable online learner which has better accuracy over the reactive learner in acquiring the target concept. The approach compares the performance of the two learners in a data stream under concept change occurrences for updating the existing stable learner based learning model with the newer instances gained from the reactive learner. The better performance of the reactive learner over the stable learner in predicting the Notion Change is because the stable learner strategy is based on using all the information learned in the classification process, while the reactive learner predicts considering only the information learned in training over a recent window of time during which the concept change occurs.

The simulations experiments with WEKA of the proposed PL-NB algorithm is done by combining the paired learner with the base learner using the naïve Bayes online algorithm. The execution is done with 2 variations of the PL-NB algorithm using a similar online NB algorithm as stable learner and with dissimilar reactive learners. The scheme is assessed by comparing it with 4 different schemes, NB (single base learner), DWM (dynamic weighted majority), AWE (accuracy weighted ensemble), and streaming ensemble algorithm (SEA) with 2 synthetic problem concepts, the Stagger concepts and the SEA concepts, and with 3 data sets of real time, a meeting scheduling data set, a electricity prediction data set, and a malware detection data set. The tests outcomes indicate for the above problems, the approach of paired learners has an equivalent or an enhanced performance over other schemes as it uses only an ensemble of 2 learners where the other methods use an ensemble with a higher number of learners. The approach uses lesser space, time, and cost in contrast to the high overhead incurred with the other schemes. The efficiency achieved in mining unfamiliar type of class labels with paired learner classifier is also comparatively very high. However the problem with the devised paired learner scheme is that both the classifiers are inefficient in tracking the noise in the data streams which affects their accuracy of predicting the Notion Change.

A unique framework of an ensemble classifier called WEAP-I by Zhenzheng Ouyang, Min Zhou, et. al., [19] is an approach developed based on the collective learning strategy. This design strategy combines the models of WE [31] and AP [32] for addressing the existing PL-NB approach constraints [18] in the enhancement of the performance of classifying noisy data streams. The averaging ensemble classifier AE has lesser probable occurrences of errors comparatively though in classification the accuracy is low as it is not based on future instances led alterations and evolution of concept in noisy data, and has a low stability as in training it doesn't consider older data portions. The model weighted ensemble classifier WE is capable of handling noisy data though incapable of handling concept evolution constantly. These two issues in incremental mining are effectively handled by the WEAP-1 devised by integrating the structure of an online learner WE trained on the highest possible portions of data with a reactive learner AE trained on the most recently available portion of data. In the completion of

this process all the base classifiers selected are joined to create the WEAP-I ensemble classifier.

The experiments are executed in Weka with real time instruction detection data set KDDCUP'99 comprising of a series of TCP connection records which are of 2 types, one a normal connection, and the second is an instruction connection of 4 different attack types, DOS, U2R, R2L, and Probing. The tests are executed with 100 data portions where each portion has 2000 sample data, first with a normal connection and second with an attack connection where the data is not replaced in between them. Next noise is added to approximately 30% of the data and the performance evaluated and then the tests are repeated by adding noise to each selected data portion. The basic classification algorithms DT (Weka J4.8 implementation) and SVM (Weka SMO implementation) are applied over these data sets and evaluated with the parameters of classified Algorithm L, Average Accuracy (Aacc), Average Ranking (AR) and Standard Deviation. The results of the WEAP-I model shows it is more robust and efficient in solving the learning and classification problems of real time data stream irrespective of the levels of noise in training data compared to the performance of averaging probability ensemble. The difficulty associated with this model is its inefficiency in the classification considering the context of the Notion Change and its incapability of handling recursive concepts based Notion Change.

A unique E-Tree Indexing structure by Peng Zhang, Chuan Zhou et. al., in [20] is a collective learner based ensemble classifier. The approach is devised for handling cost and time impaired high speed real time data streams where the incurred constrains related overhead including process complexity increases with the data dimensionality. These problems deter a feasible ensemble learning and mining classification solution to be devised in terms of response time and overall performance efficiency. This distinct ensemble-tree or simply E-tree solution models or indexes the base classifiers to form an ensemble in an orderly way for fast decision making in the predictive process of classifying the newer instances with minimal complexity associated with the factors of time and related overhead. The strategy of this E-trees approach considers an ensemble of base classifiers as spatial databases by modeling every base classifier as a set of spatial data objects. The ensemble model E-tree is mapped to the spatial database that creates a spatial index supporting the search process of the spatial database and thus the predictive complexity associated with the new instance classification is effectively minimized. In this classification approach the E-tree is searched for every new instance and from the leaf node(s) the decision rules related to the new instance are determined and merged for predicting its class label. A new classifier thus formed is merged with the E-tree structure and a new entry associated to this new classifier is created in the database and the retrieved decision rules are sequentially inserted and further connected in the tree structure. The classifiers that are old and inapplicable in terms of the newer instances in the classification due to overcapacity are removed from the E-tree ensemble which might otherwise lead to increase in the process cost. The E-trees ensemble model evolves with constant and automatic updating process which adds the incoming new classifiers and deletes the old inapplicable classifiers and adapts to the streaming data's latest patterns and trends. The E-trees are devised for binary classification only whereas to a certain extent the multi-class problems are solved with an E-forests model that merges several E-trees.

The experiments for assessed the E-trees performance is done in terms of prediction time, memory usage, and prediction accuracy with 3 real-time and synthetic data streams intrusion detection, spam detection, malicious url detection collected from the UCI repository. F-Score is used for feature selection and the devised approach is compared with 4 benchmark models Global E-tree (GE-tree), Local E-tree (LE-tree), Global Ensemble (G-Ensemble), and Local Ensemble (L-Ensemble) where the decision trees algorithms C4.5 is used for training and retrieving the data rules. The assessment of the online query traversal in the devised E-tree methods is analyzed and compared with 4 methods, the TS model, the fractal model, selectivity method, and the ERF model and is done with 3 measures, time-cost, memory cost, and accuracy with a decision rules set of total 200 rules used to quantify the average relative error. These benchmark approaches are compared with each other with varied ensemble size, node size and target indexing class and 10 data sets. The performance of our approach demonstrate that LE-tree outperforms all other methods, is faster with lesser prediction time, and occupies lesser memory with the exception compared to L-Ensemble approach where the proposed approach consumes more memory significantly. The method effectively contributes towards achieving accuracy of prediction comparatively and the approach may also be implemented with different other types of classifications not related to ensemble learning and for data analysis of spatial or temporal databases also. The model does not effectively describe the Notion Change supervision and prediction and lacks proper assessment of Notion Change and of the class labels temporal validity.

An approach devised for solving the data stream classification problems is proposed in the paper [21] by Kapil Wankhade, Snehlata Dongre, et,al, is a supervised learning based strategy devised for achieving high accuracy in the classification performance of high speed, huge size, and noisy Notion Change influenced data streams. The devised models strategy is based on using two different methods the

weighted majority method together with the method of adaptive sliding window for achieving the objective of achieving better and high classification accuracy over other models. In this model the approach polls a new example by all the ensembles algorithms considered experts. The predictions polled and the weights linked to the algorithms are combined, and in terms of the maximum accumulated weights it determines the global prediction of the labels of the class. The prediction accuracy is improved by incremental learning where incorrect predications by an expert has the related algorithms weight being reduced and the process repeated where experts with below the threshold values are deleted and new experts created. The performance is further improved by normalization of the weights where each expert is scaled according to the maximum weight so that the decision and prediction process is not totally influenced by the recently created experts. The weighted majority technique thus accurately classifies the Notion Changeing data streams mostly with noise. The accuracy in processing the fast streaming data is achieved with the sliding window concept which monitors the existing learning model and if the pace of change is greater than a set threshold value the windows obsolete sections are automatically removed from the strategy and the model gets updated by the base learners according to statistically determined distribution changes. This learning and classification is very fast in pace with the speed of the Notion Changing streaming data using sub linear memory

The experiments are performed with existing models Oz a Bag, Oz a Boost, OC Boost, Oz a Bag ADWIN, AEBC, and the devised model. The datasets used in the experiments are synthetic datasets of two types' hype plane and RBF where the Notion Change is synthetically applied and with real datasets of the UCIML repository. These approaches are tested with factors of accuracy, time, and memory. The devised model aims for better accuracy so in terms of classification accuracy it shows performance improvement compared to the other models.

The study by G. R. Marrs • M. M. Black et.al., of the streaming and Notion Change influenced data classification devise an approach [22] based on the latency of new instances arriving and the importance of the time stamp of the instances in the life cycle of the learning process. The authors apply a time stamp based learning strategy with latency applied arbitrarily on the data resulting in new rules of classification. The proposed model has 2 algorithms CDTC 1 and CDTC 2 which use the time stamp protocol or time of classification protocol for a latency impacted data classification with a proper definition given for the ordering of the instances selected in a temporal environment.

The experiments with 4 online learners, the contemporary CD3 and CD5 algorithms and the time stamp based proposed meta data tagging protocol approaches CDTC version 1 and CDTC version 2 are implemented with different scenarios of latency based Notion Change influenced streaming data. The tests with a normal latency shows, the CDTC algorithms ver 1 and ver 2 are immediately affected by the drift and the recovery is much faster and the rate of classification achieved is much greater before occurrence of another drift. The approach shows equivalent performance with other domains such as binary class value, airplane arrival data and real protein data which validates the time stamp protocol performance overcoming the constraints of memory and time for different classification scenarios.

A new approach for data stream classification devised by Zohre Karimi, Hassan Abolhassani et. al·[23] handles batch data with discrete and continuous variables, the data streams of huge volume for reduced overhead incurrence. The devised approach is a batch classifier based on the harmony search algorithm called harmony-based classifier (HC) in which the every classifier is a potential solution determined by user specified parameter based rules for the selection of a class. A Harmony is defined by the user parameters set in terms of variables sourced from memory which can be changed as per user requirements and the fitness of a harmony is determined by its accuracy. The performance of an incoming classifier if is efficient compared to a least performing classifier in the memory it is substituted and the obtained classification model is used for class label prediction. The HC approach is not capable for handling streaming data where there is no pre-determined training data available and so is combined with the Stream Miner framework for a new classification model called IHC (Incremental Harmony-based Classifier). The evaluation of the fitness by the IHC is done by a detecting and incrementally learning mechanism over the Notion Change influenced data streams with n-time cross validation towards determining the classifiers accuracy and selecting the final classifier with maximum accuracy. The IHC approach is further improved for the method called IIHC (Improved incremental harmony-based classifier) for handling the overhead incurred due to computation of learning stable and recurring concepts and learning data with noise for increasing the robustness of the model.

The experiments of the IIHC model are performed with 8 benchmark data sets of real world and synthetic datasets known for their accurateness in prediction The outcomes of the performed experiments prove that compared to other classifiers available for streaming data classification the speed and accuracy achieved with the IIHC classifier is improved for predicting the drift and is also robust in performance in data impacted by noise. However the issues of lesser

important Notion Change and the recursive concepts based Notion Change are not properly assessed.

An approach devised by Mayank Pal Singh in the paper [24] is a novel approach that uses a strategy of supervised adaptive learning with fixed window that identifies the Notion Change, trains, updates, and evolves the model continually in the classification process of the data. The devised model performs data classification using a classifier based on the Naïve Bayes theorem. The incoming traffic is separated into ingress and egress traffic and the related attributes like Source IP, Destination IP, Source Port, Destination Port, Flags, Protocol are extracted. The training dataset termed as base class is used to classify the current class data set collected from the incoming streaming data. The examples of the base class are linked to the current class examples using the NB classifier and the resulting ROC curves is used to determine and quantify the Notion Change occurring. The devised model finds the drift occurrence using the ROC curve and identifies the flow specific data attributes responsible for the drifting concept.

The experiments are performed with the WEKA simulation tool on lab collected real time dataset and on the KDD datasets. The classification is implemented with the complete dataset and also using the flow specific attributes with a training window ranging from a few hours to a couple of days depending on the data under drift. The drift is generated in the traffic by using a packet generator tool that injects in normal traffic a protocol based traffic which causes drift to occur. The analysis of the results show for a KDD dataset the model is able to correctly distinguish normal and anomalous traffic. The model may be used with other classifiers as a pre-processing tool for better classification. The models classification performance in terms of the cost incurred and the accuracy achieved may be further enhanced. However the model does not totally validate the importance of data streams characterized by capricious data.

An unsupervised clustering framework that is an on-demand resources aware classification strategy defined by conditional rules called SRASTREAM is proposed in the paper [25] by Gansen Zhao, Ziliu Li, Fujiao Liu, et.al,. The methodologies available now focus on the accuracy or on the speed whereas the devised approach based on the resource available classifies the data streams. If there is no drifting of the concept the approach does not perform the clustering and if the Notion Change occurs then the cluster refining is done in terms of the drift detected which greatly reduces the time and cost overhead and makes possible the mining of huge streaming data in real-time. The devised framework combines different tasks such as clustering, computing, evolution detection and resource monitoring.

The experiments performed are 3 comparison tests with the devised approach and existing approach CluStream. The datasets used are the KDDCUP99 data and synthetic dataset. The results of the tests show clustering performance with the proposed approach is capable of specifically clustering data of huge data size. The proposed results of the approach do not specifically validate the approach and the model is unable to completely address the issue of recursive concept based Notion Change.

A new ensemble classifier called Rot-SiLA by Muhammad Shaheryar, Mehrosh Khalid and Ali Mustafa Qamar [26] is a collective learner approach which has Rotation Forest algorithm [30] integrated with the Similarity Learning Algorithm (SiLA) ([29]. The classification strategy of the approach is devised based on similarity where relevant similarity metrics are used instead of the distance measure. The Rotation Forest classifier can be used with different selections of base classifiers and is a feature extraction based strategy which uses the PCA (Principal Component Analysis) technique to divide the feature set into K subsets and maintains all the principal components information in the process of classification.The Similarity Learning Algorithm (SiLA approach strategy is built by integrating kNN (k nearest neighbor) algorithm with Voted Perceptron technique and the learning strategy for classifying any kind of data uses the related similarity metrics instead of the distances. The assigning of an example by the Rot-SiLA algorithm to a specific nearest class has the similarity associated to a class equal to the total all the similarities existing among an example being classified and all the k nearest neighbors in the class.

The experiments are done with a fourteen UCI benchmark datasets of different domains such as medical, biology, and materials classified first with SiLA using kNN-A and SkNN-A and then with the ensemble learner Rot-SiLA kNNA and Rot-SiLA SkNN-A algorithms. The learning schemes classification accuracies gained with the Rot-SiLA ensemble learners are compared with the SiLA kNNA and SiLA SkNN-A and also with the Rotation Forest ensemble which has various integrations with dissimilar base classifiers. The test outcomes show the devised models is optimal compared to the other existing approaches. However as the extracted feature set is first separated into subsets with the PCA technique the devised models accuracy is defined by the accuracy of the variance matrix formation in the principal component analysis process.

The SA-Miner strategy proposed by Chao-Wei Li , Kuen-Fang Jea in the paper [27] for incremental mining models the frequently occurring item sets by their frequency relationships with a support approximation strategy for definitively characterizing the data streams in terms of concepts. The algorithm SA-MINER collects the examples defining a concept with

the support approximation strategy which generates the concepts for the document. The techniques of other types could be used for monitoring variations of the support relationships to find the new trends and for capturing gradual drifts.

This devised model is tested and evaluated with a number of experiments and performance compared with many approximate algorithmic methods such as Stream Mining, Loss-Counting, DSCA, and SWCA. The test data used in the experiments uses synthetic as well as real-life datasets with 3 type's metrics, space efficiency, time efficiency, and mining quality. The criteria of the tests performed are set as maximum or satisfactory in terms of efficiency in achieving accuracy in mining with least memory usage. The approach achieves better classification accuracy compared to the other streaming data classification strategies.

The density-based unsupervised learning approaches reviewed in the paper [28] are capable of learning data comprising of undefined cluster shapes as well with noise. This density based model for robustness and scalability combines 2 algorithms, one called micro-cluster formation algorithm and second the grid formation algorithm. The model does not use any previous clusters number information explores the Notion Changes influenced data streams. The paper reviews the important density based clustering algorithms for streaming data classification and the issues faced with these algorithms. The algorithms are classified into two type's micro-cluster and grid algorithms by the authors.

The simulation experiments of the different algorithms are done to evaluate their performance using real life data sets and with different metrics for cluster quality. The density based algorithms are able to mine data with different clusters those without any particular shape in terms of robust and scalable performance factors. However the performance of the density based algorithms is dependent on a large number of parameters and only a few algorithms are able to handle high dimensional data streams or complex clustering processes, or different other types of data streams.

IV. Conclusion

The objective of the paper is taxonomy and systematic review over incremental mining under the influence of Notion Change. The information retrieval and knowledge discovery progression from the strategies based on static data volumes has moved to the streaming data scenario where the notion change is not available, the established concept is not static but due to changes in the environment drifts with time, where the existing static data classification approaches are not applicable. The growth in the research in the data stream mining field has been propelled with the rapid developments in computing and communications where numerous organizations have varied interests in information exploration, extraction and knowledge discovery. The focus of these research activities in recent years has been for devising Notion Change Acceptance strategies for high speed and noisy data streams considering factors of higher accuracy, lower time complexity, scalability and robustness in the mining process and among these devised strategies a considerable number of them have materialized as benchmark strategies. These models of benchmark have been reviewed in this paper with their merits and demerits giving a better perception of these models for Notion Change and their algorithms for assessing their performance. The domain of research which is reviewed in this paper offers many new and superior strategies for mining streaming data under the influence of Notion Change. The research scope in this field is still huge as these existing models are not comprehensive and also not totally compatible with the many different types and domain contexts of streaming data influenced by diverse scenarios of Notion Change and notion changes. The factors like Notion Change context, temporal validity of Notion Change, and recursive concepts based Notion Change are not given the needed importance. Based on these factors the research for devising newer strategies and models for Notion Change Acceptance in data stream mining has wide opportunities. These opportunities will be the focus of our future research and design of newer models and strategies for Notion Change Acceptance.

References Références Referencias

1. https://en.wikipedia.org/wiki/Zettabyte.
2. SCHLIMMER, J. AND GRANGER, R. 1986. Incremental learning from noisy data. Mach. Learn. 1, 3, 317–354.
3. WIDMER, G. AND KUBAT, M. 1996. Learning in the presence of Notion Change and hidden contexts. Mach. Learn. 23, 1, 69–101.
4. Quinlan, J.: C4.5: programs for machine learning. Morgan Kaufmann (1993)
5. Domingos, P., Hulten, G.: Mining high-speed datastreams. In: KDD, pp. 71–80. ACM (2000)
6. Hoeffding, W.: Probability inequalities for sums of bounded random variables. JASA 58(301), 13–30 (1963)
7. Maron, O., Moore, A.W.: Hoeffding races: Accelerating model selection search for classification and function approximation. In: NIPS, pp. 59–66 (1993)
8. Alippi, C., Roveri, M.: Just-in-time adaptive classifiers in non-stationary conditions. In: IJCNN, pp. 1014–1019. IEEE (2007)
9. Alippi, C., Roveri, M.: Just-in-time adaptive classifierspart ii: Designing the classifier. TNN 19(12), 2053–2064 (2008)

10. Fukunaga, K., Hostetler, L.: Optimization of k nearest neighbor density estimates. Information Theory 19(3), 320–326 (2002)
11. Carpenter, G., Grossberg, S., Markuzon, N.,Reynolds, J., Rosen, D.: Fuzzy artmap: A neural network architecture for incremental supervised learning of analog multidimensional maps. TNN 3(5), 698–713 (1992)
12. Carpenter, G., Grossberg, S., Reynolds, J.: Artmap:Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. Neural Networks 4(5), 565–588 (1991)
13. Andres-Andres, A., Gomez-Sanchez, E., BoteLorenzo, M.: Incremental rule pruning for fuzzy artmap neural network. ICANN pp. 655–660 (2005)
14. Carpenter, G., Tan, A.: Rule extraction: From neural architecture to symbolic representation. Connection Science 7(1), 3–27 (1995)
15. Xingquan Zhu, Xindong Wu, and Ying Yang, "Dynamic Classifier Selection for Effective Mining from Noisy Data Streams", Proceedings of the Fourth IEEE International Conference on Data Mining (ICDM'04)
16. Chien-I Lee, Cheng-Jung Tsai, Jhe-Hao Wu, Wei-Pang Yang "A Decision Tree-Based Approach to Mining the Rules of Notion Change", Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007) IEEE.
17. Yang Zhang, Xue Li, "One-class Classification of Text Streams with Notion Change" 2008 IEEE International Conference on Data Mining Workshops.
18. Stephen H. Bach, Marcus A. Maloof, "Paired Learners for Notion Change", Eighth IEEE International Conference on Data Mining, 2008.
19. Zhenzheng Ouyang, Min Zhou, Tao Wang, Quanyuan Wu,"Mining Concept-Drifting and Noisy Data Streams using Ensemble Classifiers"International Conference on Artificial Intelligence and Computational Intelligence, 2009.
20. Peng Zhang, Chuan Zhou, Peng Wang, Byron J. Gao, Xingquan Zhu, Li Guo, "E-Tree: An Efficient Indexing Structure for Ensemble Models on Data Streams", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, JUNE 2011 1.
21. Kapil Wankhade, Snehlata Dongre, Ravindra Thool, "New Evolving Ensemble Classifier for Handling Notion Changeing Data Streams", 2nd IEEE International Conference on Parallel, Distributed and Grid Computing, pp. 657-662, 2012
22. G. R. Marrs • M. M. Black • R. J. Hickey, "The use of time stamps in handling latency and Notion Change in online learning", Springer-Verlag, 2012.
23. Zohre Karimi · Hassan Abolhassani · Hamid Beigy, "A new method of incremental mining using harmony search", Springer Science+Business Media, LLC February 2012.
24. Mayank Pal Singh, "Quantifying Notion Changeing in Network Traffic using ROC Curves from Naive Bayes Classifiers", Nirma University International Conference on Engineering (NUiCONE), 2013
25. Gansen Zhao, Ziliu Li, Fujiao Liu and Yong Tang, "A Notion Changeing based Clustering Framework for Data Streams", Fourth International Conference on Emerging Intelligent Data and Web Technologies, 2013.
26. Muhammad Shaheryar, Mehrosh Khalid and Ali Mustafa Qamar, "Rot-SiLA: A Novel Ensemble Classification approach based on Rotation Forest and Similarity Learning using Nearest Neighbor Algorithm", 12th International Conference on Machine Learning and Applications, pp. 46-51, 2013.
27. Chao-Wei Li , Kuen-Fang Jea, "An approach of support approximation to discover frequent patterns from concept-drifting data streams based on concept learning", Springer-Verlag London, 2013.
28. Amineh Amini, Member, IEEE, Teh Ying Wah, and Hadi Saboohi, Member, ACM, IEEE, "On Density-Based Data Streams Clustering Algorithms: A Survey", JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 29(1): 116–141 Jan. 2014.
29. M. Qamar, E. Gaussier, J.-P. Chevallet, and J. H. Lim, "Similarity learning for nearest neighbor classification," in Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on. IEEE, 2008, pp. 983–988.
30. J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 28, no. 10, pp. 1619–1630, 2006.
31. M. Scholz and R. Klinkenberg. An ensemble classifier for drifting concepts. In Proc. ECML/PKDD'05 Workshop on Knowledge Discovery in Data Streams.
32. Jin Gao, Wei Fan, and Jiawei Han, On appropriate assumptions to mine data streams: Analysis and Practice, Proc. of ICDM'07, pp. 143-152

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 16 Issue 2 Version 1.0 Year 2016
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Guiding Software Developers using Automated Adaptation of Object Ensembles Plug-in

By Ziaur Rahman, Md. Abir Hosain, Md. Badrul Alam Miah & Md. Hadifur Rahman

Mawlana Bhashani Science and Technology University, Bangladesh

Abstract - Software developing process has been improving day by day. The development process can be affected through different ways like changing the development environment, strategies and upcoming technologies. In order to save valuable times and to speed up the process, we can guide programmer during the development time through providing relevant recommendations. There are some strategies that suggest related code snippets and API-items to the software programmers. There are some techniques that apply general code searching approaches and some techniques that employ online based repository mining process. But it is kind of difficult tasks to guide programmers when they need specific type conversion like adapting existing interfaces from the previously used types as per their demands. One of the familiar approaches to guide developers in such a situation is to adapt collections and arrays through automated adaptation of object ensembles. But how does it help a novice developer in real time software development that is not explicitly exemplified. In this paper, we have tried to introduce a system that works as a plug-in tool incorporated with a data mining integrated environment to recommend the relevant interfaces while they look for a type conversion. We have a mined repository of respective adapter classes and related APIs from where programmers search their query and get their result using the relevant transformer classes.

Keywords: adaptation of object ensembles (aoe); repository mining; development process; data mining integrated environment.

GJCST-C Classification : D.2, D.2.1, D.2.3

Strictly as per the compliance and regulations of:

Guiding Software Developers using Automated Adaptation of Object Ensembles Plug-in

Ziaur Rahman [α], Md. Abir Hosain [σ], Md. Badrul Alam Miah [ρ] & Md. Hadifur Rahman [ω]

Abstract- Software developing process has been improving day by day. The development process can be affected through different ways like changing the development environment, strategies and upcoming technologies. In order to save valuable times and to speed up the process, we can guide programmer during the development time through providing relevant recommendations. There are some strategies that suggest related code snippets and API-items to the software programmers. There are some techniques that apply general code searching approaches and some techniques that employ online based repository mining process. But it is kind of difficult tasks to guide programmers when they need specific type conversion like adapting existing interfaces from the previously used types as per their demands. One of the familiar approaches to guide developers in such a situation is to adapt collections and arrays through automated adaptation of object ensembles. But how does it help a novice developer in real time software development that is not explicitly exemplified. In this paper, we have tried to introduce a system that works as a plug-in tool incorporated with a data mining integrated environment to recommend the relevant interfaces while they look for a type conversion. We have a mined repository of respective adapter classes and related APIs from where programmers search their query and get their result using the relevant transformer classes. The system that recommends developers entitled automated objective ensembles (AOE plug-in). From the investigation that we have done, we can see that our approach works much better than some of the existing approaches.

Keywords: adaptation of object ensembles (aoe); repository mining; development process; data mining integrated environment.

I. Introduction

Software development process improving rapidly. A lot of guidelines are suggested which influence software development process, especially in the coding stage. Reusing previously completed software repository to enhance the development process is a common phenomenon in the field of Mining Software Repository (MSR). If developers get suggestions in e.g. API recommendations, object usage pattern, class structure or code snippets from the existing projects they might be benefited a lot what they eventually expect while coding. Some of the approaches have integrated web based code, searching in their customized tool like MAC (Hsu & Lin, 2011) and MAPO (Xie & Pei, 2006), before mining the code source abstractions. Although MAC and MAPO are server dependent or online based, by which they are not flexible for a developer. Automated Adaptation of Object Ensembles (AOE) shows a process of adapting the collection frameworks and Arrays, but it is not clear view how to use it as a plug-in tool.

In software development there are different ways available to guide software developers during the development period. A programmer can use programming code in a software system easily by using an automated adaptation of object ensembles. By this process user can find out required data easily. We have used an AOE Plug-in by which a software developer can complete a code in a short time. It takes less time compared to existing approaches. The existing approaches like simply using the Integrated Development Environment (IDE) like Net Beans (NetBeans plug-in, 2015), IntellijIdea (Genuitec, 2015), and Eclipse (Seiffert & Hummel, 2015) is vulnerable to flaws and it is unable to provide us the required interfaces. In essence, it consumes developers valuable times. We have investigated between the conventional ways and our AOE plug-in approach to evaluate the efficiency of our proposed tool. The study shows that it is able to accelerate the developer's performance and facilitate less time consuming with decreasing code flaws and errors.

Although in MAC (Hsu & Lin, 2011), MAPO (Xie & Pei, 2006), AOE (Shahnewaz et. al, 2014), and many others code repository had several limitations such as server dependent searching, either database based or Internet based where data is not preprocessed according to rules of data mining so these are not friendly for a developer. In our methodology we try to show our repository plug-in as an offline repository that is filed based instead of a database and it has special source abstraction technique. Adapter classes are the key point in our approach. In AOE the result depends on the resource of the repository of adapter classes. Recently, a number of works are available to guide developers in the field of software engineering. All we have seen is that our approach is comparatively easier to handle than other existing approaches.

Author α σ ρ: Department of Information and Communication Technology.
Author ω: Department of Computer Science and Engineering Mawlana Bhashani Science and Technology University Santosh, Tangail-1902,Bangladesh. e-mails: zia@iutdhaka.edu, abirh500@gmail.com, badrul_ict05@yahoo.com, hadifur@gmail.com

The paper is so far structured in the followings: Section II background and related works of this study. Section III shows that full design and proposed approach. The detailed results and evaluation of this paper is presented in Section IV. Section V concludes with the set of observation and future work of this research.

II. Backgrounds and Related Works

As the software development process can be affected by using different strategies. Researchers from parts of the world have been trying to provide ways that can speed up the development process. We can say that previously completed software repository technique to enhance the development process is a common framework in the field of mining software repository. The developers can be benefited by following the provided suggestions from various recommendations like API recommendations, object usage pattern, class structure or code snippets from the existing projects. Some of the approaches have integrated web based code, searching in their customized tool like MAC (Hsu & Lin, 2011) and MAPO (Xie & Pei, 2006) before mining the code source abstractions.

There are some approaches by which a software system is established by code reusing. But in our approach we use a repository of adapter classes and a tool which adapt this code. By using this AOE plug-in which integrates with the IDE, the user can search the required data type by investing less effort. There are some existing efforts such as adapting collection and array by using Automated Adaption of Object (AOE) (Shahnewaz et. al, 2014). Some approaches like Code reusing in MAPO (Xie & Pei, 2006), better user recommendation using enhancing software development repository, Scenario Based API Recommendation System (Seiffert & Hummel, 2015), and others are also used to speed up the software development process. As we have proposed that if there have been adapter tools (Kabir,Rahman & Islam, 2015), which adapt the given interfaces it will be more helpful for the programmers to find the required interfaces. Mining API Usages from the Open Source Repositories (MAPO) (Xie & Pei, 2006),(NetBeans plug-in,2015), was one of the first and MAC (Hsu & Lin, 2011), was one of the updated efforts to mine API usage pattern. Other recent works called Enhancing Software Development Process (ESDP) (Reiss et. al, 2009), where the developers are highly guided by recommendations from a mined repository is also one of our referral works.

One of the popular concepts of test-driven reuse showed by Reiss (Reiss et. al, 2009), common test-cases issued as input for a component search engine in (Hummel,&Janjic,2013). Nevertheless, there exist some difficulties such as license problem and dependency issues. When the user changes the parameter types, then it might need an even more propagated deep adjustment of type changes. The formal and rule-based language is proposed by Kell (Kell et. al,2010), that was named Cake for automated wrapper generations. The designing used to define interface relations; transformation object structures are possible by applying these rules and strategies. We introduced that transformation should happen automatically but it is most overhead for a developer writing mapping rules, basically for an unknown object instance. It overcame Nita and Notkin (Nita & Notkin, 2010) by providing an approach which concerned with adapting programs to alternative APIs. When the variations among the APIs are small its schema considers not-straightforward structural respect as out of scope, which is the main challenge. Another Challenge is providing transformation. The work showed by Hummel (Hummel & Atkinson, 2010), is depending on the Identity Map Pattern from Fowler (Janjic & Atkinson, 2012) and identifies the answer about the problems of the Gang of Four adapter pattern. The approach is integrated into another work by Hummel and Atkinson (Hummel & Atkinson, 2010), that supplies relaxed-signature matching for primitive data types.

Recently there are different works available to enhance the Software Development Process (SDP). Some of them have applied many tools such as adapter generation tool (Seiffert & Hummel, 2015). In this way a user can find the required method easily. As a result a user can save time and solve any problem easily. The software development process is an easy task for a developer.

In the approach we have tried to overcome the limitations of existing repository tools. We have tried to provide as an offline repository tool that are file based instead of a database that has special source abstraction. Providing recommendations using the respective Adapter Classes and the Transformer Classes are the key point in our approach.

III. Proposed Approach

In our approach we have developed a plug-in tool that is able to guide software development through suggesting interfaces by using the respective adapter classes. The tool is completely written in java and is executable as a standalone application. It can work with IDE like Net beans (NetBeans plug-in, 2015), and Eclipse (Seiffert & Hummel, 2015) as they have the software extensibility.

Open Source Repository (OSR) is an online software code repository. In this repository many projects problem solving codes are stored. When a programmer stays in online and gets any programming problem, then the programmer can search in this repository for required code. OSP is an Internet based repository. In many software companies, there are stored many projects. It is called the Enterprise Repository.

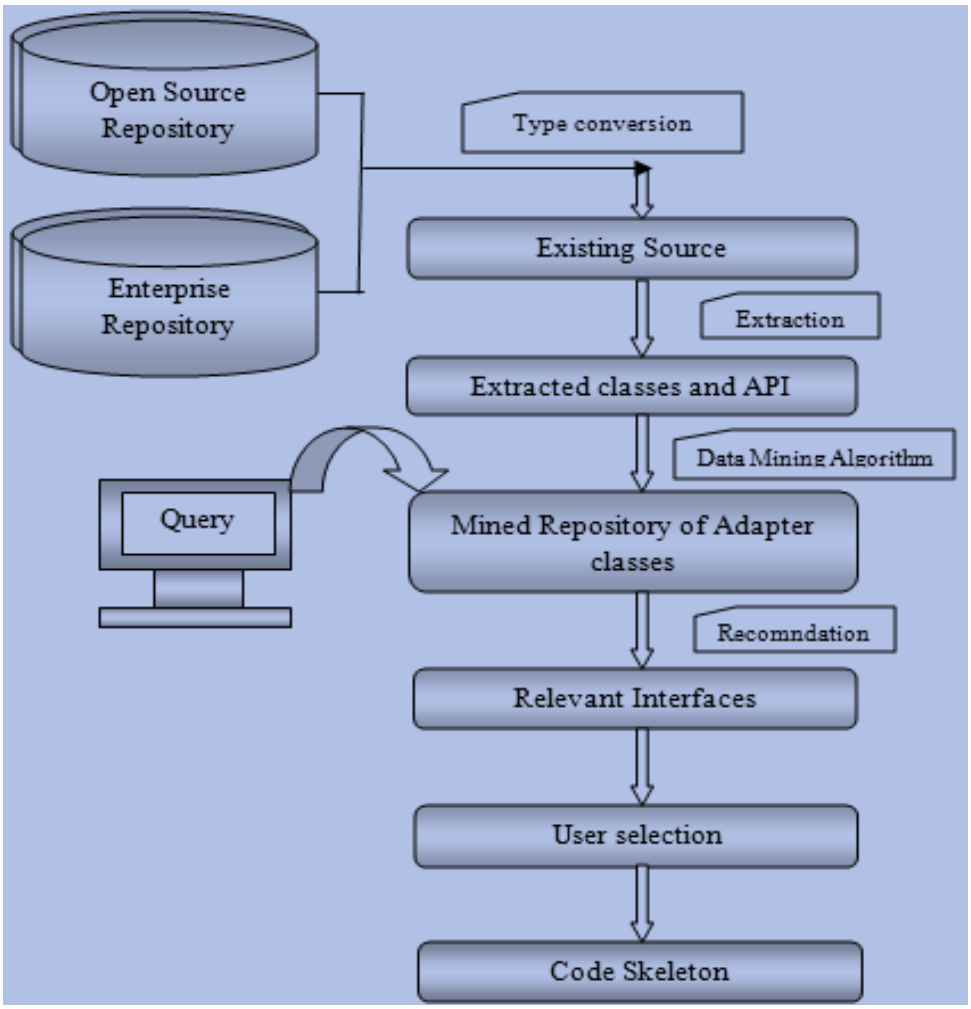

Figure 1 : Automated Object Ensembles (AOE) plugin Framework

In repository, there are stored a large number of programs. Many unusual code lines are staying in programs. We take the only important line for a program. By using type conversion we build up Existing Source where projects are stored without unusual code lines. Programming code is stored in Existing Sources. We extract classes of API from these projects which exist in Existing Sources.

Data Mining Algorithm is used to build up a mined repository of Adapter classes. When a programmer searches any classes, the required classes are shown at first. Then we will find related classes. Programmers, search their needed query in the mined repository of Adapter Classes. Then we get relevant interfaces with the help of a transformer and recommendation interfaces. The user selects the required interfaces an14014d gets the code skeleton.

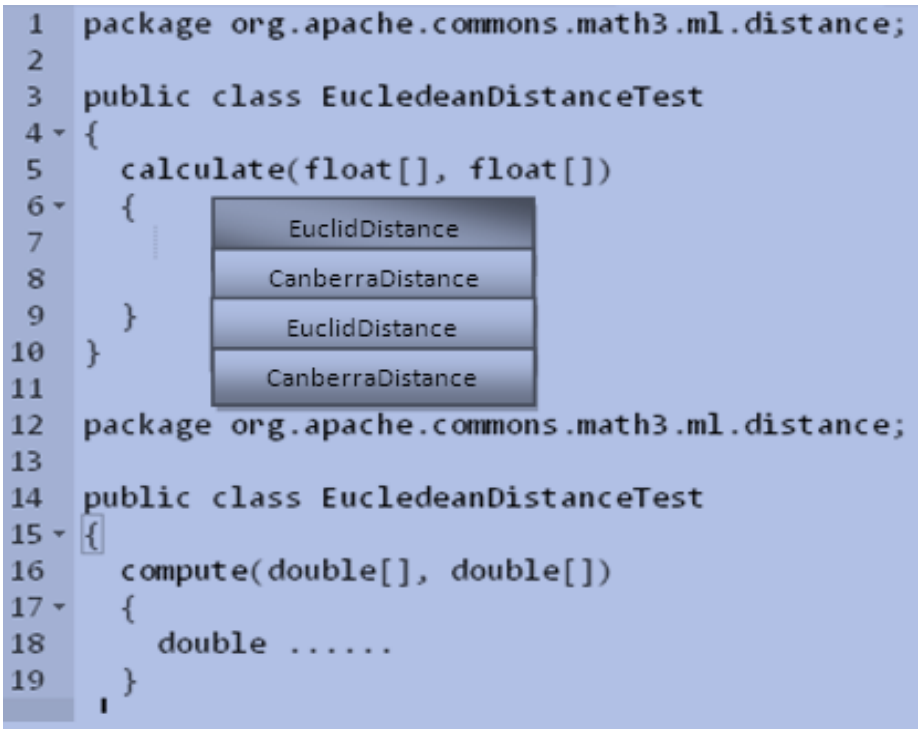

Figure 2 : Adaptation testing overview

Now a day, there has various tools to increase the optimizing capability in software development field; adapter generation tool is one of them. To overcome mismatches on the signature level a good way is to add an adapter that controls message forwarding from one interface to the other. The adapters allow classes to work together that could not otherwise because of incompatible interfaces. It can also be familiar as a "Wrapper" which wraps the incompatible class into the adapter class, where adapter adapt any types of object, method and interface that helps in the programming fields. There exist projects available the newly integrated ability to transform arrays and collections, which can be executed for verifying the adaptation capabilities. This idea is to describe by Hummel (Hummel & Atkinson,2010), where able to generate a random Array List and a sorted Vector instance the helper class Generator is used. When a client uses an automated adapter class that depends on the interfaces that are provided by collection frameworks then the client needs help to use any plug-in.

For example, when a programmer writes their code in the IDE like as Net Beans they must write the full code to reach the required result, But when they use AOE plug-in by pressing the right button, then the programmer will be guided by the several adapter classes that are shown in the Fig. 2, which is remarked by red color box is chosen. Suppose coder write ,calculate (float [], float []) then AOE plaguing suggest Euclid Distance adoptee adapt that generate in Seffert and Hummel (Seiffert & Hummel, 2015) where adapter depends on transformer that transform provided instance to require instance and vice versa. Suppose a client requires a specific data types, then its search in the adapter fields after matching needed data types client can choose any of them. By finishing the process of adaptation successfully the modified test case was executed here using the final adapter instead of the adapter directly. Nevertheless, test case executed full filly the adapter's transformation capability was verified. For example, in the test case the compute method of the selected class from *org. apache. commons. math3.ml.distance* package is tested. That takes as an input two vectors, showed by an array of type double each, and calculates the distance between them. The distance should be zero, if the same vector is provided as the first and second parameter such as in this example. The public double computes (double [], double []) is the interface of the compute method.

The array of type double was replaced with a Link List<Float> after the verification of the original test case executed successfully, where the expected name of the method was changed from compute to calculate. The test case is changed to public void double calculate (Linked List<Float>) for the require interfacing. The adapter generator overcome a parameter type and the method name mismatch, namely from Linked List<Float> on double [] and calculate in a computer.

Imagine a client requires one method within two parameters, but there exist more than two parameters in a similar method in the tool then adaptation process solves this complexity. In this paper, we want to

describe how easily use collection frameworks as a plug-in tool in software development fields. The collection framework refers the way of implementing interfaces with the help of several classes that are considered as a supported plaguing tool.

We have seen in the approaches (Xie & Pei, 2006), [MAPO] that they have used a code search engine to find the desired item following a search query given by a particular client. Like MAPO this paper also enhances the automated adaptation of object ensembles as a repository tool. In this approach the require source may online or offline repository that contain various kinds of tools. Eventually, the approaches are not only suggesting an adapter class, but also provide its related code implementation to reach the desired goal of a software developer. It works automatically when a client type any code during programming according to their require data types or interfaces. In order to consider the existing paper (Kabir, Rahman, & Islam ,2015) that shows how to generate an adapter with the help of transforms that can able to solve the matching problems of complex data types. In Fig. 2 shows the snapshot of the adapter generation tool. On other existing paper there have no idea about the plug-in. But in our paper, we want to use plug-in for a user.

A plug-in is a software component that adds a specific feature to an existing once. When an application supports plug-in, it enables customization. In our paper we want to provide an adapter plug-in by which any client can complete a task more easily than existing processes. When a client wants to find classes or interfaces, there have given some adapter class options. From these options which is chosen then it finds out the required data by using adapter plug-in.

```
1  package org.apache.commons.math3.ml.distance;
2
3  public class EucledeanDistanceTest
4  {
5      final DistanceMeasure distance = new EucliadeanDistance();
6      public void testZero()
7      {
8          final double a [] = {0, 1, -2, 3.4, 5, -6.7, 8, 9 };
9          Assert.assertEquals(0, distance.compute(a,a), 0);
10     }
11 }
```

Figure 3 : Snapshot 1 of Adapter generation tool (AOE)

```
1  package result;
2  import java.util.WeakHashMap;
3
4  public class Matrix{
5      private adaptees Rechteck adaptee;
6      private static WeakHashMap<adaptees.Rechteck, Matrix> map;
7
8      public Matrix (int param0, int param1, String param2){
9          adaptee = new adapatees.Rechteck(param0, param1, param2);
10         map.put(adaptee, this);
11     }
12
13     public Matrix (adapatees.Rechteck adaptee){
14         this.adaptee = adaptee;
15         map.put(adaptee, this);
16     }
17
18    public adaptees Rechteck getAdaptee(){
19        return adaptee;
20     }
21 }
```

Figure 4 : Snapshot 2 of Adapter generation tool (AOE)

IV. Experimental Evaluation

There have given a guideline for a programmer to enhance the software development process. An automated adaptation of object ensembles is a process by which any software developer can find the required data easily. There have given some comparison by which we will understand which process is better than another.

a) Environmental setup

In this development process, we have used a repository of adapter classes. How much healthier this repository the development process is more easy. If our adapter class repository is enriched, a user can find the required interfaces in a short time and easily. There have given same adapter classes as shown in the Table I.

Table 1 : List of Adapter Classes

SL	Adapter Classes	Line of Code (LOC)
A	EuclidDistance	17
B	CanberraDistance	11
C	ChebyshevDistance	15
D	ManhattanDistance	14
E	OnewayAnova	16

There used an adapter generation tool. Some object oriented language is used. There it needs an adapter by which adaptation code is plagued with user. There have given an adapter generation tool such as adaptation tool.zip (Seiffert & Hummel, 2015).

b) Time complexity

Time is an important thing when a program is solved. By which a program is solved quickly this process is better than other. When a programmer solves a program in IDE (Net Beans, eclipse) there have needed some times such as 40sec.

Table 2 : Time Complexity

Task	IDE with AOE plug-in(no of solved problem)	NetBeans(no of solved problem)
T1	60 sec	120 sec
T2	130 sec	200 sec
T3	150 sec	310 sec
T4	120 sec	360 sec
T5	140 sec	400 sec

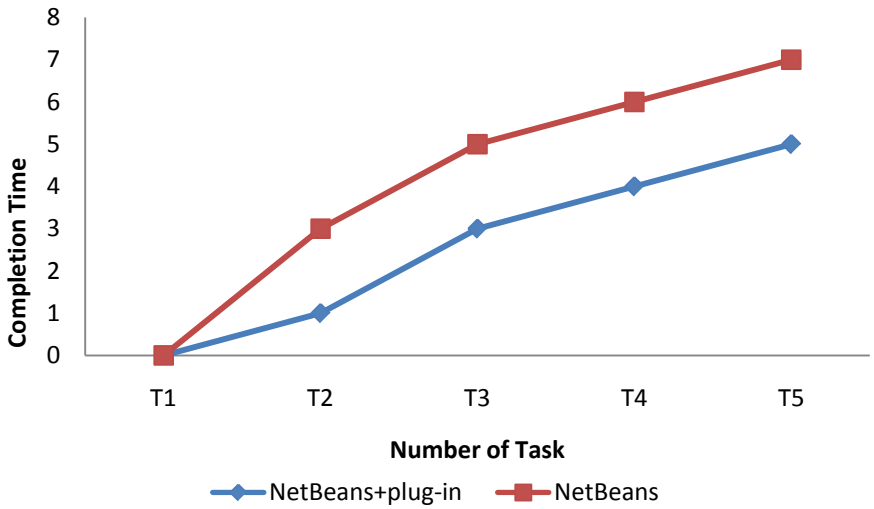

Figure 5 : Time complexity between NetBeans and NetBeans+plug-in

But the same programmer solves this same problem in net beans, but there have plugged-in AOE. As a result, we see that at this time the programmer can solve this problem in 0.20sec. When AOE is plugged-in, they have saved 20sec. The Table II shows the comparison that how much time it takes to solve a problem without AOE plugged-in and within the AOE plugged-in.

c) Error Vulnerability

In TABLE III we see that after the total time in IDE with plague-in there have found 13 errors, but at the same time problem solving only in IDE there have found out 21 errors.

Table 3 : Error Vulnerability

Time	NetBeans +plug-in	NeatBeans
T1	0	0
T2	1	3
T3	3	5
T4	4	6
T5	5	7
Total error	=13	=21

As a result, we understand that when the plug-in is used error rate is low. So this process is better than another.

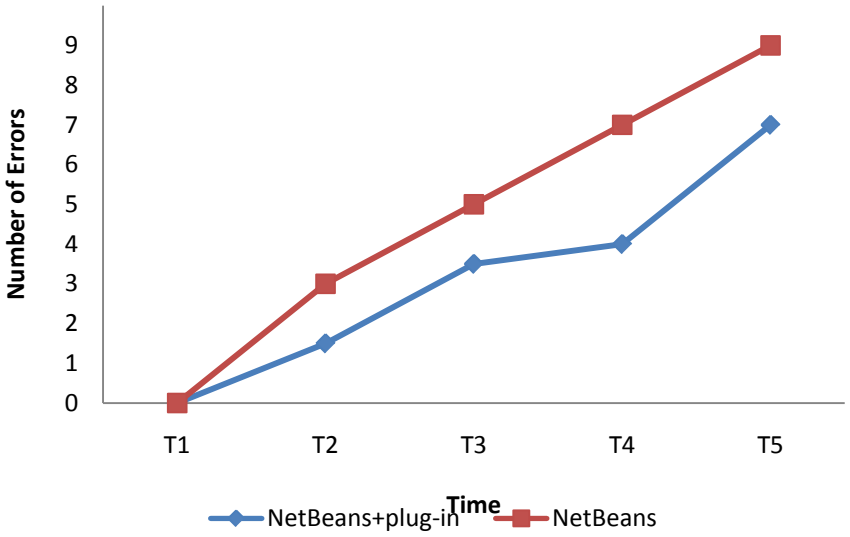

Figure 6 : Error comparison between net beans and NetBeans+plug-in

d) Thread of the evaluation

Everything has a limitation. There has some limitation of evaluation. This guideline for software development process is more effective. The evaluation is observed at the same time and same experiment. Such as

a. By having more AOE adapter classes in the repository a user finds more accurate data.
b. A program is evaluated by the same user.
c. It is a lengthy process to plug in a user in an adapter class repository.
d. A user cannot find the mining data.

The results observed in the empirical study may not be applicable to the programming tasks other than those considered in the study, being a threat to the external validity. If the tasks mentioned out there in the study change the results may also be changed. Before we start our evaluation the team members are well trained. The receiving capacity of team members may vary. So the learning curve of these numbers may affect the results.

Within many problems this guideline for a programmer is more effective to develop software process. By this process a user can find any data very quickly. It will keep an important role to enhance the software development process.

V. Conclusion

The AOE plug-in approach is more enriched than any other Existing approaches. The approach is able to find out any required data easily and there have needed less time than any other approach. A user finds a data in adapter classes than the AOE plug-in give the required data to the client. The process which we have provided in this paper is different from any other existing process. That is able to enhance the SDP recently.

This paper provided a guideline by which a user can get the required data easily and it is comfortable to use. But these data are not mined. In Future there have used data mining algorithm and find out mined data to enhance software development process.

References Références Referencias

1. Hsu, S. K., & Lin, S. J. (2011). MACs: Mining API code snippets for code reuse. *Expert Systems with Applications, 38*(6), 7291-730
2. Xie, T., & Pei, J. (2006, May). MAPO: Mining API usages from open source repositories. In Proceedings of the 2006 i*nternational workshop on Mining software repositories* (pp. 54-57). ACM.
3. NetBeans plug-in(2015),Retrieved from: https://www.plugins.netbeans.org/
4. Jet Brain, (2015), Retrieved from: https://www.jetbrains.com/idea/
5. Genuitec,(2015),Retrievedfrom:http://www.genuitec.com/myeclips
6. Seiffert, D., & Hummel, O. (2015). Adapting Collections and Arrays: Another Step towards the Automated Adaptation of Object Ensembles. In *Software Reuse for Dynamic Systems in the Cloud and Beyond* (pp. 348-363).Springer International Publishing.
7. Shahnewaz, "A Scenerio based API recommendation system using syntax and semantics of client source code". Master's Thesis, Department of CSE, Islamic University of Technology, OIC, Gazipur-1704, Bangladesh. April, 2014.
8. Kabir, H., Rahman, Z., & Islam, N. (2015, December).Enhancing software development process using automated adaptation of object

ensembles. In *Electrical Information and Communication Technology (EICT), 2015 2nd International Conference on* (pp. 560-565). IEEE.

9. Reiss, S. P. (2009, May). Semantics-based code search. In *Proceedings of the 31st International Conference on Software Engineering* (pp. 243-253).IEEE Computer Society.
10. Hummel, O., & Janjic, W. (2013). Test-driven reuse: Key to improving precision of search engines for software reuse. In *Finding Source Code on the Web for Remix and Reuse* (pp. 227-250).Springer New York.
11. Kell, S. (2010, October). Component adaptation and assembly using interface relations. In *ACM Sigplan Notices* (Vol. 45, No. 10, pp. 322-340).ACM.
12. Nita, M., & Notkin, D. (2010, May).Using twinning to adapt programs to alternative APIs. In *Proceedings of the 32nd ACM/IEEE International* Conference on Software Engineering-Volume 1 (pp. 205-214). ACM.
13. Hummel, O., & Atkinson, C. (2010).Automated creation and assessment of component adapters with test cases. In *Component-Based Software Engineering* (pp. 166-181).Springer Berlin Heidelberg.
14. 14. Janjic, W., & Atkinson, C. (2012, June). Leveraging software search and reuse with automated software adaptation. In S*earch-Driven Development-Users, Infrastructure, Tools and Evaluation (SUITE), 2012 ICSE Workshop on* (pp. 23-26).IEEE.

Global Journals Inc. (US) Guidelines Handbook 2016

www.GlobalJournals.org

FELLOW OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (FARSC)

Global Journals Incorporate (USA) is accredited by Open Association of Research Society (OARS), U.S.A and in turn, awards "FARSC" title to individuals. The 'FARSC' title is accorded to a selected professional after the approval of the Editor-in-Chief/Editorial Board Members/Dean.

- The "FARSC" is a dignified title which is accorded to a person's name viz. Dr. John E. Hall, Ph.D., FARSC or William Walldroff, M.S., FARSC.

FARSC accrediting is an honor. It authenticates your research activities. After recognition as FARSC, you can add 'FARSC' title with your name as you use this recognition as additional suffix to your status. This will definitely enhance and add more value and repute to your name. You may use it on your professional Counseling Materials such as CV, Resume, and Visiting Card etc.

The following benefits can be availed by you only for next three years from the date of certification:

FARSC designated members are entitled to avail a 40% discount while publishing their research papers (of a single author) with Global Journals Incorporation (USA), if the same is accepted by Editorial Board/Peer Reviewers. If you are a main author or co-author in case of multiple authors, you will be entitled to avail discount of 10%.

Once FARSC title is accorded, the Fellow is authorized to organize a symposium/seminar/conference on behalf of Global Journal Incorporation (USA).The Fellow can also participate in conference/seminar/symposium organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent.

You may join as member of the Editorial Board of Global Journals Incorporation (USA) after successful completion of three years as Fellow and as Peer Reviewer. In addition, it is also desirable that you should organize seminar/symposium/conference at least once.

We shall provide you intimation regarding launching of e-version of journal of your stream time to time. This may be utilized in your library for the enrichment of knowledge of your students as well as it can also be helpful for the concerned faculty members.

The FARSC can go through standards of OARS. You can also play vital role if you have any suggestions so that proper amendment can take place to improve the same for the benefit of entire research community.

As FARSC, you will be given a renowned, secure and free professional email address with 100 GB of space e.g. johnhall@globaljournals.org. This will include Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.

The FARSC will be eligible for a free application of standardization of their researches. Standardization of research will be subject to acceptability within stipulated norms as the next step after publishing in a journal. We shall depute a team of specialized research professionals who will render their services for elevating your researches to next higher level, which is worldwide open standardization.

The FARSC member can apply for grading and certification of standards of their educational and Institutional Degrees to Open Association of Research, Society U.S.A. Once you are designated as FARSC, you may send us a scanned copy of all of your credentials. OARS will verify, grade and certify them. This will be based on your academic records, quality of research papers published by you, and some more criteria. After certification of all your credentials by OARS, they will be published on your Fellow Profile link on website https://associationofresearch.org which will be helpful to upgrade the dignity.

The FARSC members can avail the benefits of free research podcasting in Global Research Radio with their research documents. After publishing the work, (including published elsewhere worldwide with proper authorization) you can upload your research paper with your recorded voice or you can utilize chargeable services of our professional RJs to record your paper in their voice on request.

The FARSC member also entitled to get the benefits of free research podcasting of their research documents through video clips. We can also streamline your conference videos and display your slides/ online slides and online research video clips at reasonable charges, on request.

The FARSC is eligible to earn from sales proceeds of his/her researches/reference/review Books or literature, while publishing with Global Journals. The FARSC can decide whether he/she would like to publish his/her research in a closed manner. In this case, whenever readers purchase that individual research paper for reading, maximum 60% of its profit earned as royalty by Global Journals, will be credited to his/her bank account. The entire entitled amount will be credited to his/her bank account exceeding limit of minimum fixed balance. There is no minimum time limit for collection. The FARSC member can decide its price and we can help in making the right decision.

The FARSC member is eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get remuneration of 15% of author fees, taken from the author of a respective paper. After reviewing 5 or more papers you can request to transfer the amount to your bank account.

MEMBER OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (MARSC)

The ' MARSC ' title is accorded to a selected professional after the approval of the Editor-in-Chief / Editorial Board Members/Dean.
The “MARSC” is a dignified ornament which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., MARSC or William Walldroff, M.S., MARSC.

MARSC accrediting is an honor. It authenticates your research activities. After becoming MARSC, you can add 'MARSC' title with your name as you use this recognition as additional suffix to your status. This will definitely enhance and add more value and repute to your name. You may use it on your professional Counseling Materials such as CV, Resume, Visiting Card and Name Plate etc.

The following benefitscan be availed by you only for next three years from the date of certification.

MARSC designated members are entitled to avail a 25% discount while publishing their research papers (of a single author) in Global Journals Inc., if the same is accepted by our Editorial Board and Peer Reviewers. If you are a main author or co-author of a group of authors, you will get discount of 10%.

As MARSC, you will be given a renowned, secure and free professional email address with 30 GB of space e.g. johnhall@globaljournals.org. This will include Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.

We shall provide you intimation regarding launching of e-version of journal of your stream time to time. This may be utilized in your library for the enrichment of knowledge of your students as well as it can also be helpful for the concerned faculty members.

The MARSC member can apply for approval, grading and certification of standards of their educational and Institutional Degrees to Open Association of Research, Society U.S.A.

Once you are designated as MARSC, you may send us a scanned copy of all of your credentials. OARS will verify, grade and certify them. This will be based on your academic records, quality of research papers published by you, and some more criteria.

It is mandatory to read all terms and conditions carefully.

AUXILIARY MEMBERSHIPS

Institutional Fellow of Open Association of Research Society (USA)-OARS (USA)

Global Journals Incorporation (USA) is accredited by Open Association of Research Society, U.S.A (OARS) and in turn, affiliates research institutions as "Institutional Fellow of Open Association of Research Society" (IFOARS).

The "FARSC" is a dignified title which is accorded to a person's name viz. Dr. John E. Hall, Ph.D., FARSC or William Walldroff, M.S., FARSC.

The IFOARS institution is entitled to form a Board comprised of one Chairperson and three to five board members preferably from different streams. The Board will be recognized as "Institutional Board of Open Association of Research Society"-(IBOARS).

The Institute will be entitled to following benefits:

The IBOARS can initially review research papers of their institute and recommend them to publish with respective journal of Global Journals. It can also review the papers of other institutions after obtaining our consent. The second review will be done by peer reviewer of Global Journals Incorporation (USA) The Board is at liberty to appoint a peer reviewer with the approval of chairperson after consulting us.

The author fees of such paper may be waived off up to 40%.

The Global Journals Incorporation (USA) at its discretion can also refer double blind peer reviewed paper at their end to the board for the verification and to get recommendation for final stage of acceptance of publication.

The IBOARS can organize symposium/seminar/conference in their country on behalf of Global Journals Incorporation (USA)-OARS (USA). The terms and conditions can be discussed separately.

The Board can also play vital role by exploring and giving valuable suggestions regarding the Standards of "Open Association of Research Society, U.S.A (OARS)" so that proper amendment can take place for the benefit of entire research community. We shall provide details of particular standard only on receipt of request from the Board.

The board members can also join us as Individual Fellow with 40% discount on total fees applicable to Individual Fellow. They will be entitled to avail all the benefits as declared. Please visit Individual Fellow-sub menu of GlobalJournals.org to have more relevant details.

We shall provide you intimation regarding launching of e-version of journal of your stream time to time. This may be utilized in your library for the enrichment of knowledge of your students as well as it can also be helpful for the concerned faculty members.

After nomination of your institution as "Institutional Fellow" and constantly functioning successfully for one year, we can consider giving recognition to your institute to function as Regional/Zonal office on our behalf.
The board can also take up the additional allied activities for betterment after our consultation.

The following entitlements are applicable to individual Fellows:

Open Association of Research Society, U.S.A (OARS) By-laws states that an individual Fellow may use the designations as applicable, or the corresponding initials. The Credentials of individual Fellow and Associate designations signify that the individual has gained knowledge of the fundamental concepts. One is magnanimous and proficient in an expertise course covering the professional code of conduct, and follows recognized standards of practice.

Open Association of Research Society (US)/ Global Journals Incorporation (USA), as described in Corporate Statements, are educational, research publishing and professional membership organizations. Achieving our individual Fellow or Associate status is based mainly on meeting stated educational research requirements.

Disbursement of 40% Royalty earned through Global Journals : Researcher = 50%, Peer Reviewer = 37.50%, Institution = 12.50% E.g. Out of 40%, the 20% benefit should be passed on to researcher, 15 % benefit towards remuneration should be given to a reviewer and remaining 5% is to be retained by the institution.

We shall provide print version of 12 issues of any three journals [as per your requirement] out of our 38 journals worth $ 2376 USD.

Other:

The individual Fellow and Associate designations accredited by Open Association of Research Society (US) credentials signify guarantees following achievements:

- The professional accredited with Fellow honor, is entitled to various benefits viz. name, fame, honor, regular flow of income, secured bright future, social status etc.

- In addition to above, if one is single author, then entitled to 40% discount on publishing research paper and can get 10%discount if one is co-author or main author among group of authors.
- The Fellow can organize symposium/seminar/conference on behalf of Global Journals Incorporation (USA) and he/she can also attend the same organized by other institutes on behalf of Global Journals.
- The Fellow can become member of Editorial Board Member after completing 3yrs.
- The Fellow can earn 60% of sales proceeds from the sale of reference/review books/literature/publishing of research paper.
- Fellow can also join as paid peer reviewer and earn 15% remuneration of author charges and can also get an opportunity to join as member of the Editorial Board of Global Journals Incorporation (USA)
- • This individual has learned the basic methods of applying those concepts and techniques to common challenging situations. This individual has further demonstrated an in–depth understanding of the application of suitable techniques to a particular area of research practice.

Note :

"

- In future, if the board feels the necessity to change any board member, the same can be done with the consent of the chairperson along with anyone board member without our approval.
- In case, the chairperson needs to be replaced then consent of 2/3rd board members are required and they are also required to jointly pass the resolution copy of which should be sent to us. In such case, it will be compulsory to obtain our approval before replacement.
- In case of "Difference of Opinion [if any]" among the Board members, our decision will be final and binding to everyone.

"

Process of submission of Research Paper

The Area or field of specialization may or may not be of any category as mentioned in 'Scope of Journal' menu of the GlobalJournals.org website. There are 37 Research Journal categorized with Six parental Journals GJCST, GJMR, GJRE, GJMBR, GJSFR, GJHSS. For Authors should prefer the mentioned categories. There are three widely used systems UDC, DDC and LCC. The details are available as 'Knowledge Abstract' at Home page. The major advantage of this coding is that, the research work will be exposed to and shared with all over the world as we are being abstracted and indexed worldwide.

The paper should be in proper format. The format can be downloaded from first page of 'Author Guideline' Menu. The Author is expected to follow the general rules as mentioned in this menu. The paper should be written in MS-Word Format (*.DOC,*.DOCX).

The Author can submit the paper either online or offline. The authors should prefer online submission.Online Submission: There are three ways to submit your paper:

(A) (I) First, register yourself using top right corner of Home page then Login. If you are already registered, then login using your username and password.

(II) Choose corresponding Journal.

(III) Click 'Submit Manuscript'. Fill required information and Upload the paper.

(B) If you are using Internet Explorer, then Direct Submission through Homepage is also available.

(C) If these two are not convenient, and then email the paper directly to dean@globaljournals.org.

Offline Submission: Author can send the typed form of paper by Post. However, online submission should be preferred.

Preferred Author Guidelines

MANUSCRIPT STYLE INSTRUCTION (Must be strictly followed)

Page Size: 8.27" X 11'"

- Left Margin: 0.65
- Right Margin: 0.65
- Top Margin: 0.75
- Bottom Margin: 0.75
- Font type of all text should be Swis 721 Lt BT.
- Paper Title should be of Font Size 24 with one Column section.
- Author Name in Font Size of 11 with one column as of Title.
- Abstract Font size of 9 Bold, "Abstract" word in Italic Bold.
- Main Text: Font size 10 with justified two columns section
- Two Column with Equal Column with of 3.38 and Gaping of .2
- First Character must be three lines Drop capped.
- Paragraph before Spacing of 1 pt and After of 0 pt.
- Line Spacing of 1 pt
- Large Images must be in One Column
- Numbering of First Main Headings (Heading 1) must be in Roman Letters, Capital Letter, and Font Size of 10.
- Numbering of Second Main Headings (Heading 2) must be in Alphabets, Italic, and Font Size of 10.

You can use your own standard format also.
Author Guidelines:

1. General,

2. Ethical Guidelines,

3. Submission of Manuscripts,

4. Manuscript's Category,

5. Structure and Format of Manuscript,

6. After Acceptance.

1. GENERAL

Before submitting your research paper, one is advised to go through the details as mentioned in following heads. It will be beneficial, while peer reviewer justify your paper for publication.

Scope

The Global Journals Inc. (US) welcome the submission of original paper, review paper, survey article relevant to the all the streams of Philosophy and knowledge. The Global Journals Inc. (US) is parental platform for Global Journal of Computer Science and Technology, Researches in Engineering, Medical Research, Science Frontier Research, Human Social Science, Management, and Business organization. The choice of specific field can be done otherwise as following in Abstracting and Indexing Page on this Website. As the all Global

Journals Inc. (US) are being abstracted and indexed (in process) by most of the reputed organizations. Topics of only narrow interest will not be accepted unless they have wider potential or consequences.

2. ETHICAL GUIDELINES

Authors should follow the ethical guidelines as mentioned below for publication of research paper and research activities.

Papers are accepted on strict understanding that the material in whole or in part has not been, nor is being, considered for publication elsewhere. If the paper once accepted by Global Journals Inc. (US) and Editorial Board, will become the copyright of the Global Journals Inc. (US).

Authorship: The authors and coauthors should have active contribution to conception design, analysis and interpretation of findings. They should critically review the contents and drafting of the paper. All should approve the final version of the paper before submission

The Global Journals Inc. (US) follows the definition of authorship set up by the Global Academy of Research and Development. According to the Global Academy of R&D authorship, criteria must be based on:

1) Substantial contributions to conception and acquisition of data, analysis and interpretation of the findings.

2) Drafting the paper and revising it critically regarding important academic content.

3) Final approval of the version of the paper to be published.

All authors should have been credited according to their appropriate contribution in research activity and preparing paper. Contributors who do not match the criteria as authors may be mentioned under Acknowledgement.

Acknowledgements: Contributors to the research other than authors credited should be mentioned under acknowledgement. The specifications of the source of funding for the research if appropriate can be included. Suppliers of resources may be mentioned along with address.

Appeal of Decision: The Editorial Board's decision on publication of the paper is final and cannot be appealed elsewhere.

Permissions: It is the author's responsibility to have prior permission if all or parts of earlier published illustrations are used in this paper.

Please mention proper reference and appropriate acknowledgements wherever expected.

If all or parts of previously published illustrations are used, permission must be taken from the copyright holder concerned. It is the author's responsibility to take these in writing.

Approval for reproduction/modification of any information (including figures and tables) published elsewhere must be obtained by the authors/copyright holders before submission of the manuscript. Contributors (Authors) are responsible for any copyright fee involved.

3. SUBMISSION OF MANUSCRIPTS

Manuscripts should be uploaded via this online submission page. The online submission is most efficient method for submission of papers, as it enables rapid distribution of manuscripts and consequently speeds up the review procedure. It also enables authors to know the status of their own manuscripts by emailing us. Complete instructions for submitting a paper is available below.

Manuscript submission is a systematic procedure and little preparation is required beyond having all parts of your manuscript in a given format and a computer with an Internet connection and a Web browser. Full help and instructions are provided on-screen. As an author, you will be prompted for login and manuscript details as Field of Paper and then to upload your manuscript file(s) according to the instructions.

To avoid postal delays, all transaction is preferred by e-mail. A finished manuscript submission is confirmed by e-mail immediately and your paper enters the editorial process with no postal delays. When a conclusion is made about the publication of your paper by our Editorial Board, revisions can be submitted online with the same procedure, with an occasion to view and respond to all comments.

Complete support for both authors and co-author is provided.

4. MANUSCRIPT'S CATEGORY

Based on potential and nature, the manuscript can be categorized under the following heads:

Original research paper: Such papers are reports of high-level significant original research work.

Review papers: These are concise, significant but helpful and decisive topics for young researchers.

Research articles: These are handled with small investigation and applications.

Research letters: The letters are small and concise comments on previously published matters.

5. STRUCTURE AND FORMAT OF MANUSCRIPT

The recommended size of original research paper is less than seven thousand words, review papers fewer than seven thousands words also.Preparation of research paper or how to write research paper, are major hurdle, while writing manuscript. The research articles and research letters should be fewer than three thousand words, the structure original research paper; sometime review paper should be as follows:

Papers: These are reports of significant research (typically less than 7000 words equivalent, including tables, figures, references), and comprise:

(a)Title should be relevant and commensurate with the theme of the paper.

(b) A brief Summary, "Abstract" (less than 150 words) containing the major results and conclusions.

(c) Up to ten keywords, that precisely identifies the paper's subject, purpose, and focus.

(d) An Introduction, giving necessary background excluding subheadings; objectives must be clearly declared.

(e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition; sources of information must be given and numerical methods must be specified by reference, unless non-standard.

(f) Results should be presented concisely, by well-designed tables and/or figures; the same data may not be used in both; suitable statistical data should be given. All data must be obtained with attention to numerical detail in the planning stage. As reproduced design has been recognized to be important to experiments for a considerable time, the Editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned un-refereed;

(g) Discussion should cover the implications and consequences, not just recapitulating the results; conclusions should be summarizing.

(h) Brief Acknowledgements.

(i) References in the proper form.

Authors should very cautiously consider the preparation of papers to ensure that they communicate efficiently. Papers are much more likely to be accepted, if they are cautiously designed and laid out, contain few or no errors, are summarizing, and be conventional to the approach and instructions. They will in addition, be published with much less delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and to make suggestions to improve briefness.

It is vital, that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

Format

Language: The language of publication is UK English. Authors, for whom English is a second language, must have their manuscript efficiently edited by an English-speaking person before submission to make sure that, the English is of high excellence. It is preferable, that manuscripts should be professionally edited.

Standard Usage, Abbreviations, and Units: Spelling and hyphenation should be conventional to The Concise Oxford English Dictionary. Statistics and measurements should at all times be given in figures, e.g. 16 min, except for when the number begins a sentence. When the number does not refer to a unit of measurement it should be spelt in full unless, it is 160 or greater.

Abbreviations supposed to be used carefully. The abbreviated name or expression is supposed to be cited in full at first usage, followed by the conventional abbreviation in parentheses.

Metric SI units are supposed to generally be used excluding where they conflict with current practice or are confusing. For illustration, 1.4 l rather than 1.4 × 10-3 m3, or 4 mm somewhat than 4 × 10-3 m. Chemical formula and solutions must identify the form used, e.g. anhydrous or hydrated, and the concentration must be in clearly defined units. Common species names should be followed by underlines at the first mention. For following use the generic name should be constricted to a single letter, if it is clear.

Structure

All manuscripts submitted to Global Journals Inc. (US), ought to include:

Title: The title page must carry an instructive title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) wherever the work was carried out. The full postal address in addition with the e-mail address of related author must be given. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining and indexing.

Abstract, used in Original Papers and Reviews:

Optimizing Abstract for Search Engines

Many researchers searching for information online will use search engines such as Google, Yahoo or similar. By optimizing your paper for search engines, you will amplify the chance of someone finding it. This in turn will make it more likely to be viewed and/or cited in a further work. Global Journals Inc. (US) have compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Key Words

A major linchpin in research work for the writing research paper is the keyword search, which one will employ to find both library and Internet resources.

One must be persistent and creative in using keywords. An effective keyword search requires a strategy and planning a list of possible keywords and phrases to try.

Search engines for most searches, use Boolean searching, which is somewhat different from Internet searches. The Boolean search uses "operators," words (and, or, not, and near) that enable you to expand or narrow your affords. Tips for research paper while preparing research paper are very helpful guideline of research paper.

Choice of key words is first tool of tips to write research paper. Research paper writing is an art.A few tips for deciding as strategically as possible about keyword search:

- One should start brainstorming lists of possible keywords before even begin searching. Think about the most important concepts related to research work. Ask, "What words would a source have to include to be truly valuable in research paper?" Then consider synonyms for the important words.
- It may take the discovery of only one relevant paper to let steer in the right keyword direction because in most databases, the keywords under which a research paper is abstracted are listed with the paper.
- One should avoid outdated words.

Keywords are the key that opens a door to research work sources. Keyword searching is an art in which researcher's skills are bound to improve with experience and time.

Numerical Methods: Numerical methods used should be clear and, where appropriate, supported by references.

Acknowledgements: Please make these as concise as possible.

References

References follow the Harvard scheme of referencing. References in the text should cite the authors' names followed by the time of their publication, unless there are three or more authors when simply the first author's name is quoted followed by et al. unpublished work has to only be cited where necessary, and only in the text. Copies of references in press in other journals have to be supplied with submitted typescripts. It is necessary that all citations and references be carefully checked before submission, as mistakes or omissions will cause delays.

References to information on the World Wide Web can be given, but only if the information is available without charge to readers on an official site. Wikipedia and Similar websites are not allowed where anyone can change the information. Authors will be asked to make available electronic copies of the cited information for inclusion on the Global Journals Inc. (US) homepage at the judgment of the Editorial Board.

The Editorial Board and Global Journals Inc. (US) recommend that, citation of online-published papers and other material should be done via a DOI (digital object identifier). If an author cites anything, which does not have a DOI, they run the risk of the cited material not being noticeable.

The Editorial Board and Global Journals Inc. (US) recommend the use of a tool such as Reference Manager for reference management and formatting.

Tables, Figures and Figure Legends

Tables: Tables should be few in number, cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g. Table 4, a self-explanatory caption and be on a separate sheet. Vertical lines should not be used.

Figures: Figures are supposed to be submitted as separate files. Always take in a citation in the text for each figure using Arabic numbers, e.g. Fig. 4. Artwork must be submitted online in electronic form by e-mailing them.

Preparation of Electronic Figures for Publication

Even though low quality images are sufficient for review purposes, print publication requires high quality images to prevent the final product being blurred or fuzzy. Submit (or e-mail) EPS (line art) or TIFF (halftone/photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Do not use pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings) in relation to the imitation size. Please give the data for figures in black and white or submit a Color Work Agreement Form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution (at final image size) ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs) : >350 dpi; figures containing both halftone and line images: >650 dpi.

Color Charges: It is the rule of the Global Journals Inc. (US) for authors to pay the full cost for the reproduction of their color artwork. Hence, please note that, if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a color work agreement form before your paper can be published.

Figure Legends: Self-explanatory legends of all figures should be incorporated separately under the heading 'Legends to Figures'. In the full-text online edition of the journal, figure legends may possibly be truncated in abbreviated links to the full screen version. Therefore, the first 100 characters of any legend should notify the reader, about the key aspects of the figure.

6. AFTER ACCEPTANCE

Upon approval of a paper for publication, the manuscript will be forwarded to the dean, who is responsible for the publication of the Global Journals Inc. (US).

6.1 Proof Corrections

The corresponding author will receive an e-mail alert containing a link to a website or will be attached. A working e-mail address must therefore be provided for the related author.

Acrobat Reader will be required in order to read this file. This software can be downloaded

(Free of charge) from the following website:

www.adobe.com/products/acrobat/readstep2.html. This will facilitate the file to be opened, read on screen, and printed out in order for any corrections to be added. Further instructions will be sent with the proof.

Proofs must be returned to the dean at dean@globaljournals.org within three days of receipt.

As changes to proofs are costly, we inquire that you only correct typesetting errors. All illustrations are retained by the publisher. Please note that the authors are responsible for all statements made in their work, including changes made by the copy editor.

6.2 Early View of Global Journals Inc. (US) (Publication Prior to Print)

The Global Journals Inc. (US) are enclosed by our publishing's Early View service. Early View articles are complete full-text articles sent in advance of their publication. Early View articles are absolute and final. They have been completely reviewed, revised and edited for publication, and the authors' final corrections have been incorporated. Because they are in final form, no changes can be made after sending them. The nature of Early View articles means that they do not yet have volume, issue or page numbers, so Early View articles cannot be cited in the conventional way.

6.3 Author Services

Online production tracking is available for your article through Author Services. Author Services enables authors to track their article - once it has been accepted - through the production process to publication online and in print. Authors can check the status of their articles online and choose to receive automated e-mails at key stages of production. The authors will receive an e-mail with a unique link that enables them to register and have their article automatically added to the system. Please ensure that a complete e-mail address is provided when submitting the manuscript.

6.4 Author Material Archive Policy

Please note that if not specifically requested, publisher will dispose off hardcopy & electronic information submitted, after the two months of publication. If you require the return of any information submitted, please inform the Editorial Board or dean as soon as possible.

6.5 Offprint and Extra Copies

A PDF offprint of the online-published article will be provided free of charge to the related author, and may be distributed according to the Publisher's terms and conditions. Additional paper offprint may be ordered by emailing us at: editor@globaljournals.org .

You must strictly follow above Author Guidelines before submitting your paper or else we will not at all be responsible for any corrections in future in any of the way.

Before start writing a good quality Computer Science Research Paper, let us first understand what is Computer Science Research Paper? So, Computer Science Research Paper is the paper which is written by professionals or scientists who are associated to Computer Science and Information Technology, or doing research study in these areas. If you are novel to this field then you can consult about this field from your supervisor or guide.

TECHNIQUES FOR WRITING A GOOD QUALITY RESEARCH PAPER:

1. Choosing the topic: In most cases, the topic is searched by the interest of author but it can be also suggested by the guides. You can have several topics and then you can judge that in which topic or subject you are finding yourself most comfortable. This can be done by asking several questions to yourself, like Will I be able to carry our search in this area? Will I find all necessary recourses to accomplish the search? Will I be able to find all information in this field area? If the answer of these types of questions will be "Yes" then you can choose that topic. In most of the cases, you may have to conduct the surveys and have to visit several places because this field is related to Computer Science and Information Technology. Also, you may have to do a lot of work to find all rise and falls regarding the various data of that subject. Sometimes, detailed information plays a vital role, instead of short information.

2. Evaluators are human: First thing to remember that evaluators are also human being. They are not only meant for rejecting a paper. They are here to evaluate your paper. So, present your Best.

3. Think Like Evaluators: If you are in a confusion or getting demotivated that your paper will be accepted by evaluators or not, then think and try to evaluate your paper like an Evaluator. Try to understand that what an evaluator wants in your research paper and automatically you will have your answer.

4. Make blueprints of paper: The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

5. Ask your Guides: If you are having any difficulty in your research, then do not hesitate to share your difficulty to your guide (if you have any). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work then ask the supervisor to help you with the alternative. He might also provide you the list of essential readings.

6. Use of computer is recommended: As you are doing research in the field of Computer Science, then this point is quite obvious.

7. Use right software: Always use good quality software packages. If you are not capable to judge good software then you can lose quality of your paper unknowingly. There are various software programs available to help you, which you can get through Internet.

8. Use the Internet for help: An excellent start for your paper can be by using the Google. It is an excellent search engine, where you can have your doubts resolved. You may also read some answers for the frequent question how to write my research paper or find model research paper. From the internet library you can download books. If you have all required books make important reading selecting and analyzing the specified information. Then put together research paper sketch out.

9. Use and get big pictures: Always use encyclopedias, Wikipedia to get pictures so that you can go into the depth.

10. Bookmarks are useful: When you read any book or magazine, you generally use bookmarks, right! It is a good habit, which helps to not to lose your continuity. You should always use bookmarks while searching on Internet also, which will make your search easier.

11. Revise what you wrote: When you write anything, always read it, summarize it and then finalize it.

12. Make all efforts: Make all efforts to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in introduction, that what is the need of a particular research paper. Polish your work by good skill of writing and always give an evaluator, what he wants.

13. Have backups: When you are going to do any important thing like making research paper, you should always have backup copies of it either in your computer or in paper. This will help you to not to lose any of your important.

14. Produce good diagrams of your own: Always try to include good charts or diagrams in your paper to improve quality. Using several and unnecessary diagrams will degrade the quality of your paper by creating "hotchpotch." So always, try to make and include those diagrams, which are made by your own to improve readability and understandability of your paper.

15. Use of direct quotes: When you do research relevant to literature, history or current affairs then use of quotes become essential but if study is relevant to science then use of quotes is not preferable.

16. Use proper verb tense: Use proper verb tenses in your paper. Use past tense, to present those events that happened. Use present tense to indicate events that are going on. Use future tense to indicate future happening events. Use of improper and wrong tenses will confuse the evaluator. Avoid the sentences that are incomplete.

17. Never use online paper: If you are getting any paper on Internet, then never use it as your research paper because it might be possible that evaluator has already seen it or maybe it is outdated version.

18. Pick a good study spot: To do your research studies always try to pick a spot, which is quiet. Every spot is not for studies. Spot that suits you choose it and proceed further.

19. Know what you know: Always try to know, what you know by making objectives. Else, you will be confused and cannot achieve your target.

20. Use good quality grammar: Always use a good quality grammar and use words that will throw positive impact on evaluator. Use of good quality grammar does not mean to use tough words, that for each word the evaluator has to go through dictionary. Do not start sentence with a conjunction. Do not fragment sentences. Eliminate one-word sentences. Ignore passive voice. Do not ever use a big word when a diminutive one would suffice. Verbs have to be in agreement with their subjects. Prepositions are not expressions to finish sentences with. It is incorrect to ever divide an infinitive. Avoid clichés like the disease. Also, always shun irritating alliteration. Use language that is simple and straight forward. put together a neat summary.

21. Arrangement of information: Each section of the main body should start with an opening sentence and there should be a changeover at the end of the section. Give only valid and powerful arguments to your topic. You may also maintain your arguments with records.

22. Never start in last minute: Always start at right time and give enough time to research work. Leaving everything to the last minute will degrade your paper and spoil your work.

23. Multitasking in research is not good: Doing several things at the same time proves bad habit in case of research activity. Research is an area, where everything has a particular time slot. Divide your research work in parts and do particular part in particular time slot.

24. Never copy others' work: Never copy others' work and give it your name because if evaluator has seen it anywhere you will be in trouble.

25. Take proper rest and food: No matter how many hours you spend for your research activity, if you are not taking care of your health then all your efforts will be in vain. For a quality research, study is must, and this can be done by taking proper rest and food.

26. Go for seminars: Attend seminars if the topic is relevant to your research area. Utilize all your resources.

27. Refresh your mind after intervals: Try to give rest to your mind by listening to soft music or by sleeping in intervals. This will also improve your memory.

28. Make colleagues: Always try to make colleagues. No matter how sharper or intelligent you are, if you make colleagues you can have several ideas, which will be helpful for your research.

29. Think technically: Always think technically. If anything happens, then search its reasons, its benefits, and demerits.

30. Think and then print: When you will go to print your paper, notice that tables are not be split, headings are not detached from their descriptions, and page sequence is maintained.

31. Adding unnecessary information: Do not add unnecessary information, like, I have used MS Excel to draw graph. Do not add irrelevant and inappropriate material. These all will create superfluous. Foreign terminology and phrases are not apropos. One should NEVER take a broad view. Analogy in script is like feathers on a snake. Not at all use a large word when a very small one would be sufficient. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Amplification is a billion times of inferior quality than sarcasm.

32. Never oversimplify everything: To add material in your research paper, never go for oversimplification. This will definitely irritate the evaluator. Be more or less specific. Also too, by no means, ever use rhythmic redundancies. Contractions aren't essential and shouldn't be there used. Comparisons are as terrible as clichés. Give up ampersands and abbreviations, and so on. Remove commas, that are, not necessary. Parenthetical words however should be together with this in commas. Understatement is all the time the complete best way to put onward earth-shaking thoughts. Give a detailed literary review.

33. Report concluded results: Use concluded results. From raw data, filter the results and then conclude your studies based on measurements and observations taken. Significant figures and appropriate number of decimal places should be used. Parenthetical remarks are prohibitive. Proofread carefully at final stage. In the end give outline to your arguments. Spot out perspectives of further study of this subject. Justify your conclusion by at the bottom of them with sufficient justifications and examples.

34. After conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium though which your research is going to be in print to the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects in your research.

Informal Guidelines of Research Paper Writing

Key points to remember:

- Submit all work in its final form.
- Write your paper in the form, which is presented in the guidelines using the template.
- Please note the criterion for grading the final paper by peer-reviewers.

Final Points:

A purpose of organizing a research paper is to let people to interpret your effort selectively. The journal requires the following sections, submitted in the order listed, each section to start on a new page.

The introduction will be compiled from reference matter and will reflect the design processes or outline of basis that direct you to make study. As you will carry out the process of study, the method and process section will be constructed as like that. The result segment will show related statistics in nearly sequential order and will direct the reviewers next to the similar intellectual paths throughout the data that you took to carry out your study. The discussion section will provide understanding of the data and projections as to the implication of the results. The use of good quality references all through the paper will give the effort trustworthiness by representing an alertness of prior workings.

Writing a research paper is not an easy job no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record keeping are the only means to make straightforward the progression.

General style:

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear

· Adhere to recommended page limits

Mistakes to evade

- Insertion a title at the foot of a page with the subsequent text on the next page
- Separating a table/chart or figure - impound each figure/table to a single page
- Submitting a manuscript with pages out of sequence

In every sections of your document

· Use standard writing style including articles ("a", "the," etc.)

· Keep on paying attention on the research topic of the paper

· Use paragraphs to split each significant point (excluding for the abstract)

· Align the primary line of each section

· Present your points in sound order

· Use present tense to report well accepted

· Use past tense to describe specific results

· Shun familiar wording, don't address the reviewer directly, and don't use slang, slang language, or superlatives

· Shun use of extra pictures - include only those figures essential to presenting results

Title Page:

Choose a revealing title. It should be short. It should not have non-standard acronyms or abbreviations. It should not exceed two printed lines. It should include the name(s) and address (es) of all authors.

Abstract:

The summary should be two hundred words or less. It should briefly and clearly explain the key findings reported in the manuscript--must have precise statistics. It should not have abnormal acronyms or abbreviations. It should be logical in itself. Shun citing references at this point.

An abstract is a brief distinct paragraph summary of finished work or work in development. In a minute or less a reviewer can be taught the foundation behind the study, common approach to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Yet, use comprehensive sentences and do not let go readability for briefness. You can maintain it succinct by phrasing sentences so that they provide more than lone rationale. The author can at this moment go straight to shortening the outcome. Sum up the study, with the subsequent elements in any summary. Try to maintain the initial two items to no more than one ruling each.

- Reason of the study - theory, overall issue, purpose
- Fundamental goal
- To the point depiction of the research
- Consequences, including definite statistics - if the consequences are quantitative in nature, account quantitative data; results of any numerical analysis should be reported
- Significant conclusions or questions that track from the research(es)

Approach:

- Single section, and succinct
- As a outline of job done, it is always written in past tense
- A conceptual should situate on its own, and not submit to any other part of the paper such as a form or table
- Center on shortening results - bound background information to a verdict or two, if completely necessary
- What you account in an conceptual must be regular with what you reported in the manuscript
- Exact spelling, clearness of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else

Introduction:

The **Introduction** should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be capable to comprehend and calculate the purpose of your study without having to submit to other works. The basis for the study should be offered. Give most important references but shun difficult to make a comprehensive appraisal of the topic. In the introduction, describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will have no attention in your result. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here. Following approach can create a valuable beginning:

- Explain the value (significance) of the study
- Shield the model - why did you employ this particular system or method? What is its compensation? You strength remark on its appropriateness from a abstract point of vision as well as point out sensible reasons for using it.
- Present a justification. Status your particular theory (es) or aim(s), and describe the logic that led you to choose them.
- Very for a short time explain the tentative propose and how it skilled the declared objectives.

Approach:

- Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done.
- Sort out your thoughts; manufacture one key point with every section. If you make the four points listed above, you will need a least of four paragraphs.

- Present surroundings information only as desirable in order hold up a situation. The reviewer does not desire to read the whole thing you know about a topic.
- Shape the theory/purpose specifically - do not take a broad view.
- As always, give awareness to spelling, simplicity and correctness of sentences and phrases.

Procedures (Methods and Materials):

This part is supposed to be the easiest to carve if you have good skills. A sound written Procedures segment allows a capable scientist to replacement your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt for the least amount of information that would permit another capable scientist to spare your outcome but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section. When a technique is used that has been well described in another object, mention the specific item describing a way but draw the basic principle while stating the situation. The purpose is to text all particular resources and broad procedures, so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step by step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

- Explain materials individually only if the study is so complex that it saves liberty this way.
- Embrace particular materials, and any tools or provisions that are not frequently found in laboratories.
- Do not take in frequently found.
- If use of a definite type of tools.
- Materials may be reported in a part section or else they may be recognized along with your measures.

Methods:

- Report the method (not particulars of each process that engaged the same methodology)
- Describe the method entirely
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures
- Simplify - details how procedures were completed not how they were exclusively performed on a particular day.
- If well known procedures were used, account the procedure by name, possibly with reference, and that's all.

Approach:

- It is embarrassed or not possible to use vigorous voice when documenting methods with no using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result when script up the methods most authors use third person passive voice.
- Use standard style in this and in every other part of the paper - avoid familiar lists, and use full sentences.

What to keep away from

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings - save it for the argument.
- Leave out information that is immaterial to a third party.

Results:

The principle of a results segment is to present and demonstrate your conclusion. Create this part a entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Carry on to be to the point, by means of statistics and tables, if suitable, to present consequences most efficiently.You must obviously differentiate material that would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matter should not be submitted at all except requested by the instructor.

Content

- Sum up your conclusion in text and demonstrate them, if suitable, with figures and tables.
- In manuscript, explain each of your consequences, point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation an exacting study.
- Explain results of control experiments and comprise remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or in manuscript form.

What to stay away from

- Do not discuss or infer your outcome, report surroundings information, or try to explain anything.
- Not at all, take in raw data or intermediate calculations in a research manuscript.
- Do not present the similar data more than once.
- Manuscript should complement any figures or tables, not duplicate the identical information.
- Never confuse figures with tables - there is a difference.

Approach

- As forever, use past tense when you submit to your results, and put the whole thing in a reasonable order.
- Put figures and tables, appropriately numbered, in order at the end of the report
- If you desire, you may place your figures and tables properly within the text of your results part.

Figures and tables

- If you put figures and tables at the end of the details, make certain that they are visibly distinguished from any attach appendix materials, such as raw facts
- Despite of position, each figure must be numbered one after the other and complete with subtitle
- In spite of position, each table must be titled, numbered one after the other and complete with heading
- All figure and table must be adequately complete that it could situate on its own, divide from text

Discussion:

The Discussion is expected the trickiest segment to write and describe. A lot of papers submitted for journal are discarded based on problems with the Discussion. There is no head of state for how long a argument should be. Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implication of the study. The purpose here is to offer an understanding of your results and hold up for all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of result should be visibly described. Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved with prospect, and let it drop at that.

- Make a decision if each premise is supported, discarded, or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."
- Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work
- You may propose future guidelines, such as how the experiment might be personalized to accomplish a new idea.
- Give details all of your remarks as much as possible, focus on mechanisms.
- Make a decision if the tentative design sufficiently addressed the theory, and whether or not it was correctly restricted.
- Try to present substitute explanations if sensible alternatives be present.
- One research will not counter an overall question, so maintain the large picture in mind, where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

- When you refer to information, differentiate data generated by your own studies from available information
- Submit to work done by specific persons (including you) in past tense.
- Submit to generally acknowledged facts and main beliefs in present tense.

Please carefully note down following rules and regulation before submitting your Research Paper to Global Journals Inc. (US):

Segment Draft and Final Research Paper: You have to strictly follow the template of research paper. If it is not done your paper may get rejected.

- The **major constraint** is that you must independently make all content, tables, graphs, and facts that are offered in the paper. You must write each part of the paper wholly on your own. The Peer-reviewers need to identify your own perceptive of the concepts in your own terms. NEVER extract straight from any foundation, and never rephrase someone else's analysis.

- Do not give permission to anyone else to "PROOFREAD" your manuscript.

- Methods to avoid Plagiarism is applied by us on every paper, if found guilty, you will be blacklisted by all of our collaborated research groups, your institution will be informed for this and strict legal actions will be taken immediately.)
- To guard yourself and others from possible illegal use please do not permit anyone right to use to your paper and files.

Criterion for Grading a Research Paper (Compilation) by Global Journals Inc. (US)

Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals Inc. (US).

Topics	Grades		
	A-B	**C-D**	**E-F**
Abstract	Clear and concise with appropriate content, Correct format. 200 words or below	Unclear summary and no specific data, Incorrect form Above 200 words	No specific data with ambiguous information Above 250 words
Introduction	Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited	Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter	Out of place depth and content, hazy format
Methods and Procedures	Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads	Difficult to comprehend with embarrassed text, too much explanation but completed	Incorrect and unorganized structure with hazy meaning
Result	Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake	Complete and embarrassed text, difficult to comprehend	Irregular format with wrong facts and figures
Discussion	Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited	Wordy, unclear conclusion, spurious	Conclusion is not cited, unorganized, difficult to comprehend
References	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring

Index

save our planet

Global Journal of Computer Science and Technology

Visit us on the Web at www.GlobalJournals.org | www.ComputerResearch.org
or email us at helpdesk@globaljournals.org

ISSN 9754350