

GLOBAL JOURNAL

OF COMPUTER SCIENCE AND TECHNOLOGY: C

Software & Data Engineering

Data Mining System

Technology in Gastroenterology

Highlights

Predictor Technology

Computing Technology

Discovering Thoughts, Inventing Future

VOLUME 13

ISSUE 12

VERSION 1.0



GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY: C
SOFTWARE & DATA ENGINEERING



GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY: C
SOFTWARE & DATA ENGINEERING

VOLUME 13 ISSUE 12 (VER. 1.0)

OPEN ASSOCIATION OF RESEARCH SOCIETY

© Global Journal of Computer Science and Technology. 2013.

All rights reserved.

This is a special issue published in version 1.0 of "Global Journal of Computer Science and Technology" By Global Journals Inc.

All articles are open access articles distributed under "Global Journal of Computer Science and Technology"

Reading License, which permits restricted use. Entire contents are copyright by of "Global Journal of Computer Science and Technology" unless otherwise noted on specific articles.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without written permission.

The opinions and statements made in this book are those of the authors concerned. Ultraculture has not verified and neither confirms nor denies any of the foregoing and no warranty or fitness is implied.

Engage with the contents herein at your own risk.

The use of this journal, and the terms and conditions for our providing information, is governed by our Disclaimer, Terms and Conditions and Privacy Policy given on our website <http://globaljournals.us/terms-and-condition/menu-id-1463/>

By referring / using / reading / any type of association / referencing this journal, this signifies and you acknowledge that you have read them and that you accept and will be bound by the terms thereof.

All information, journals, this journal, activities undertaken, materials, services and our website, terms and conditions, privacy policy, and this journal is subject to change anytime without any prior notice.

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
301st Edgewater Place Suite, 100 Edgewater Dr.-Pl,
Wakefield MASSACHUSETTS, Pin: 01880,
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: investers@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)

John A. Hamilton, "Drew" Jr.,
Ph.D., Professor, Management
Computer Science and Software
Engineering
Director, Information Assurance
Laboratory
Auburn University

Dr. Henry Hexmoor
IEEE senior member since 2004
Ph.D. Computer Science, University at
Buffalo
Department of Computer Science
Southern Illinois University at Carbondale

Dr. Osman Balci, Professor
Department of Computer Science
Virginia Tech, Virginia University
Ph.D. and M.S. Syracuse University,
Syracuse, New York
M.S. and B.S. Bogazici University,
Istanbul, Turkey

Yogita Bajpai
M.Sc. (Computer Science), FICCT
U.S.A. Email:
yogita@computerresearch.org

Dr. T. David A. Forbes
Associate Professor and Range
Nutritionist
Ph.D. Edinburgh University - Animal
Nutrition
M.S. Aberdeen University - Animal
Nutrition
B.A. University of Dublin- Zoology

Dr. Wenying Feng
Professor, Department of Computing &
Information Systems
Department of Mathematics
Trent University, Peterborough,
ON Canada K9J 7B8

Dr. Thomas Wischgoll
Computer Science and Engineering,
Wright State University, Dayton, Ohio
B.S., M.S., Ph.D.
(University of Kaiserslautern)

Dr. Abdurrahman Arslanyilmaz
Computer Science & Information Systems
Department
Youngstown State University
Ph.D., Texas A&M University
University of Missouri, Columbia
Gazi University, Turkey

Dr. Xiaohong He
Professor of International Business
University of Quinnipiac
BS, Jilin Institute of Technology; MA, MS,
PhD., (University of Texas-Dallas)

Burcin Becerik-Gerber
University of Southern California
Ph.D. in Civil Engineering
DDes from Harvard University
M.S. from University of California, Berkeley
& Istanbul University

Dr. Bart Lambrecht

Director of Research in Accounting and Finance
Professor 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 Regina
Ph.D., M.Sc. in Mathematics
B.A. (Honors) in Mathematics
University of Windsor

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., Etsv Lornd University
Postdoctoral Training,
New York University

Dr. Söhnke M. Bartram

Department of Accounting and Finance
Lancaster University Management School
Ph.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
Neuroscience
Northwestern 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

PRESIDENT EDITOR (HON.)

Dr. George Perry, (Neuroscientist)

Dean and Professor, College of Sciences

Denham Harman Research Award (American Aging Association)

ISI Highly Cited Researcher, Iberoamerican Molecular Biology Organization

AAAS Fellow, Correspondent Member of Spanish Royal Academy of Sciences

University of Texas at San Antonio

Postdoctoral Fellow (Department of Cell Biology)

Baylor College of Medicine

Houston, Texas, United States

CHIEF AUTHOR (HON.)

Dr. R.K. Dixit

M.Sc., Ph.D., FICCT

Chief Author, India

Email: authorind@computerresearch.org

DEAN & EDITOR-IN-CHIEF (HON.)

Vivek Dubey(HON.)

MS (Industrial Engineering),

MS (Mechanical Engineering)

University of Wisconsin, FICCT

Editor-in-Chief, USA

editorusa@computerresearch.org

Sangita Dixit

M.Sc., FICCT

Dean & Chancellor (Asia Pacific)

deanind@computerresearch.org

Suyash Dixit

(B.E., Computer Science Engineering), FICCTT

President, Web Administration and

Development , CEO at IOSRD

COO at GAOR & OSS

Er. Suyog Dixit

(M. Tech), BE (HONS. in CSE), FICCT

SAP Certified Consultant

CEO at IOSRD, GAOR & OSS

Technical Dean, Global Journals Inc. (US)

Website: www.suyogdixit.com

Email: suyog@suyogdixit.com

Pritesh Rajvaidya

(MS) Computer Science Department

California State University

BE (Computer Science), FICCT

Technical Dean, USA

Email: pritesh@computerresearch.org

Luis Galárraga

J!Research Project Leader

Saarbrücken, Germany

CONTENTS OF THE VOLUME

- i. Copyright Notice
 - ii. Editorial Board Members
 - iii. Chief Author and Dean
 - iv. Table of Contents
 - v. From the Chief Editor's Desk
 - vi. Research and Review Papers
-
- 1. An Aprori Algorithm in Distributed Data Mining System. **1-5**
 - 2. Automatic Segmentation of Punjabi Speech Signal using Group Delay. **7-10**
 - 3. A Parallel Computational Approach for String Matching- A Novel Structure with Omega Model. **11-29**
 - 4. Is Advanced Statistical Computing Technology a Clue in Applied Medicine? A Study using Data Mining as a Predictor Technology in Gastroenterology & Bariatric Surgery; Novel Elbanna Operations. **31-37**
-
- vii. Auxiliary Memberships
 - viii. Process of Submission of Research Paper
 - ix. Preferred Author Guidelines
 - x. Index



An Apriori Algorithm in Distributed Data Mining System

By Dr. C.Sunil Kumar, P.N.Santosh Kumar & Dr. C.Venugopal

Srrenidhi Institute of Science & Technology, India

Abstract- Many existing data mining (DM) tasks can be proficient effectively only in a distributed condition. The ground of distributed data mining (DDM) has therefore gained growing weightage in the preceding decades. The Apriori algorithm (AA) has appeared as one of the greatest Association Rule mining (ARM) algorithms. It also provides the foundation algorithm in majority of parallel algorithms (PAs). The size and elevated dimensionality of datasets characteristically existing as a key to difficulty of AR finding, makes it perfect difficulty for solving on numerous processors in parallel. The main causes are the computer memory and central processing unit pace constraints looked by single workstations. This paper is based on an Optimized Distributed AR mining algorithm for biologically distributed information is used in similar and distributed surroundings so that it decreases communication costs.

Keywords: *association rules (ars), apriori algorithm (aa), distributed data mining (ddm), xml data, parallel.*

GJCST-C Classification : *H.2.8*



Strictly as per the compliance and regulations of:



An Apriori Algorithm in Distributed Data Mining System

Dr. C.Sunil Kumar ^α, P.N.Santosh Kumar ^ο & Dr. C.Venugopal ^ρ

Abstract- Many existing data mining (DM) tasks can be proficient effectively only in a distributed condition. The ground of distributed data mining (DDM) has therefore gained growing weightage in the preceding decades. The Apriori algorithm (AA) has appeared as one of the greatest Association Rule mining (ARM) algorithms. It also provides the foundation algorithm in majority of parallel algorithms (PAs). The size and elevated dimensionality of datasets characteristically existing as a key to difficulty of AR finding, makes it perfect difficulty for solving on numerous processors in parallel. The main causes are the computer memory and central processing unit pace constraints looked by single workstations. This paper is based on an Optimized Distributed AR mining algorithm for biologically distributed information is used in similar and distributed surroundings so that it decreases communication costs.

Keywords: association rules (ars), apriori algorithm (aa), distributed data mining (ddm), xml data, parallel.

I. INTRODUCTION

ARM has turned out to be one of the hub DM tasks and has attracted marvelous interest among DM investigators. ARM is an unsupervised DM method which works on variable length data, and produces understandable results. There are two foremost approaches for utilizing numerous workstations that have appeared in distributed computer memory in which each CPU has a confidential memory; and public memory in which all CPUs access universal memory [4]. Collective memory planning has many gorgeous assets. Each CPU has straight and identical right to use memory in the computer system. Equivalent applications are easy to execute on such a distributed system. In allocated memory design each CPU has its own restricted memory that can simply right to use directly by that CPU [10]. For a CPU to have contact with facts in the restricted memory of another CPU a replica of the preferred data ingredient must be sent from one CPU to the other throughout message passing. XML information is used with the optimized distributed association rule mining (ODAM) algorithm. A similar application could be divided into numeral of jobs and implemented in parallel on different CPUs in the system [2, 9]. Though the performance of a similar function on a distributed system is mainly depe-

ndent on the distribution of the jobs contains the application onto the obtainable CPUs in the system.

Modern associations are biologically dispersed. Classically, each location locally stores its ever growing amount of every day data. Using centralized DM to find out useful patterns in such institutions' data isn't always practicable because integration of data sets from dissimilar locations into a centralized location earns enormous communication system costs. Information from these institutions' is not only spread to a variety of sites but also vertically incoherent, making it complex if not unfeasible to merge them in an essential site.

Distributed DM therefore emerged as vigorous subarea of DM investigation. In this paper an ODAM Algorithm is used for executing the mining procedure.

II. ASSOCIATION RULE MINING ALGORITHMS

An AR is a rule which implies certain association relationships among a set of objects in a database.

Given a set of transactions, where each transaction is a set of items, an AR is an expression of the form $X \rightarrow Y$, where X and Y are sets of items. The intuitive meaning of such a rule is that transactions of the database which contain X tend to contain Y [1].

a) Apriori Algorithm

Apriori has been urbanized for rule mining in large business databases by Quest project team of IBMs. An item set (IS) is a non-empty set of things.

They have decayed the difficulty of mining ARs into two (2) pieces

- Find all groupings of items that have business support above smallest support. Call those groupings frequent ISs [5].
- Use the recurrent IS to produce the preferred rules. The universal idea is that if, say, EFGH and EF are recurrent ISs, then we can determine if the rule EF \rightarrow GH holds by computing the ratio $R = \text{support}(EFGH)/\text{support}(EF)$. The rule holds only if $R \geq \text{min assurance}$. The algorithm is extremely scalable [7]. The AA used in Quest to find all recurrent ISs is specified below.

Procedure AprioriAlg () begin

```
M1 := {frequent 1-itemsets};
for ( k := 2; Mk-1 0; k++ ) do {Nk= Apriori-gen (Mk-1);
// new candidates for all transactions t in the dataset do}
{for all candidates c Nk contained in t do c: count++ }
```

Author α: Professor in ECM SNIST, Hyderabad, A.P, India.

e-mail: sunilkumarc@sreenidhi.edu.in

Author ο: Assistant Professor in ECM SNIST, Hyderabad, A.P, India.

e-mail: pnsk47@gmail.com

Author ρ: Professor in ECM, SNIST, Hyderabad, A.P, India.

e-mail: cherupalli_v@yahoo.com

```
Lk = {c Nk | c: count >= min-support}
Answer: = k Mk
end
```

It makes numerous passes over the database. In the first pass, the algorithm simply counts item occurrences to determine the frequent 1-itemsets. A succeeding pass, say pass k , consists of two phases. First, the recurrent ISs M_{k-1} found in the $(k-1)$ th pass are used to make the candidate ISs N_k , using the Apriori-gen () function. This task first joins M_{k-1} with M_{k-1} , the joining condition being that the lexicographically ordered first $k-2$ items are the same. Next, it deletes all those ISs from the join result that have some $(k-1)$ -subset that is not in M_{k-1} yielding N_k . The algorithm now examines the database. For each deal, it concludes which of the candidates in N_k are limited in the deal using a hash-tree data structure and increases the count of those candidates [8], at the end of the pass, C_k is observed to decide which of the candidates' frequent, yielding M_k . The algorithm quits when M_k becomes vacant.

III. OPTIMIZED DISTRIBUTED MINING ALGORITHM

The presentation of Apriori agent rule mining algorithms humiliates for diverse causes. It requires 'N' number of database examines to produce a common n-IS. In addition, it doesn't differentiate operations in the data set with identical ISs if that data set is not burdened into the memory.

Consequently, it needlessly occupies resources for frequently generating ISs from such matching transactions. For e.g., if a data set has 10 matching transactions, the AA not only indicates the same candidate ISs 10 times but also informs.

The support counts for those candidate ISs 10 times for every iteration. Furthermore, openly loading a unprocessed data set into the memory won't find an significant number of equal transactions because each business of a raw data set contains both recurrent and non-recurrent items. To overcome these difficulties, candidate support counts can't be supported from the raw data set following the first pass. This method not only decreases the average business length but also decreases the data set size considerably. The number of items in the data set might be huge, but only a few will gratify the support threshold (TH).

Consider the sample data set in Fig 1a. The data set is loaded in memory, and then only one indistinguishable transaction (EFGH) is found, as Fig 1b shows. However, if the data set is loaded into the main memory after eliminating rare items from every transaction, more identical transactions are found (as shown in Fig 1c). This technique not only reduces average transaction size but also finds more identical transactions. The following gives the pseudo code of O DAM algorithm.

```
NF= {Non-frequent global 1-itemset} for all transaction
t ∈ D {for all 2-subsets s of t
if (s ∈ C2) s.sup++;
t/=delete_nonfrequent_items(t);
Table.add (t/);}
Send_to_receiver (C2);
/* Global Frequent support counts from receiver*/
F2=receive_from_receiver (Fα);
C3= (Candidate item set);
T=Table.getTransactions (); k=3;
While (Ck ≠ {}) {For all transaction t ∈ T
For all k-subsets s of t
If (s ∈ Ck) s.sup++;
k++; Send_to_receiver (Ck);
/* Generating candidate item set of k+1 pass*/
Ck+1={Candidate item set}; }
```

Transactions	
T. No	Items
1.	EFGH
2.	FG
3.	EF
4.	EFGHI
5.	EGH
6.	EFI
7.	GHI
8.	EH
9.	FGI
10.	EFGH

Figure 1 : (a) An Example Dataset

Transactions	
T. No.	Items
1,10	EFGH
2.	FG
3.	EF
4.	EFGHI
5.	EGH
6.	EFI
7.	GHI
8.	EH
9.	FGI

Figure 1 : (b) Identical transactions

Transactions	
T. No.	Items
1,4,10	EFGH
2,9	FG
3,6	EF
5.	EGH
7.	GHI
8.	EH

Figure 1 : (c) Transactions after pruning infrequent items

ODAM eliminates all globally infrequent 1-itemsets from every transaction and inserts them into the main memory; it reduces the transaction size and finds more alike transactions. This is because the data set initially contains both frequent and rare items. However, total transactions could surpass the main memory limit.

ODAM removes rare items and inserts each transaction into the main memory. While inserting the transactions, it checks whether they are already in memory. If yes, it increases that transaction's counter by one. Otherwise, it inserts that transaction into the main memory with a count equal to one. Finally, it writes all main-memory entries for this separation into a temporary file. This process continues for all other divisions.

IV. PADA RULE WITH XML DATA

Parallelism is predictable to relieve current ARM methods from sequential blockages, providing the

ability to scale to enormous datasets and improving the response time. The parallel and Distributed Association rule (PADA) design space spans three main components including the hardware platform, the kind of parallelism broken and the load balancing strategy used [8]. Shared memory architecture has all the processors access common memory. Each processor has direct and equal access to all the memory in the system.

Parallel programs are easy to execute on such a system. The data warehouse (DW) is partitioned among 'P' processors logically. Each processor works on its local partition of the database but performs the same computation of counting support. Dynamic load balancing seeks to address this issue by balancing the load and reassigning the loads to the lighter ones. The development of distributed rule mining is a challenging and vital task, since it requires knowledge of all the data stored at different locations and the ability to combine

partial results from individual databases into a single result.

The AR from XML data with a sample XML document is considered. For example, the set of transactions are identified by the tag <transactions> and each transaction in the transactions set is identified by the tag <transaction>. The set of items in each transaction is: Transaction document (transactions.xml) is identified by the tag <items> and an item is identified by the tag <item>. Consider the problem of mining all ARs among items that emerge in the transactions document. With the understanding of traditional AR mining is expected to obtain the large item sets document and ARs document from the source document.

Let the minimum support (minsupp) = 35% and minimum confidence (minconfi) = 99%.

V. PERFORMANCE ASSESSMENT

The number of messages that ODAM exchanges among various locations to generate the globally frequent item sets in a distributed environment, the original data set is partitioned into five partitions. To decrease the dependency among dissimilar partitions, each one contains only 25 percent of the original data set's transactions. So, the number of identical transactions among different partitions is very low.

ODAM provides a proficient method for generating ARs from different datasets, distributed among various locations.

The datasets are generated arbitrarily depending on the number of different items, the maximum number of items in each transaction and the number of transactions. The performance of the XQuery implementation is dependent on the number of large item sets found and the size of the dataset as shown in the Fig 2.

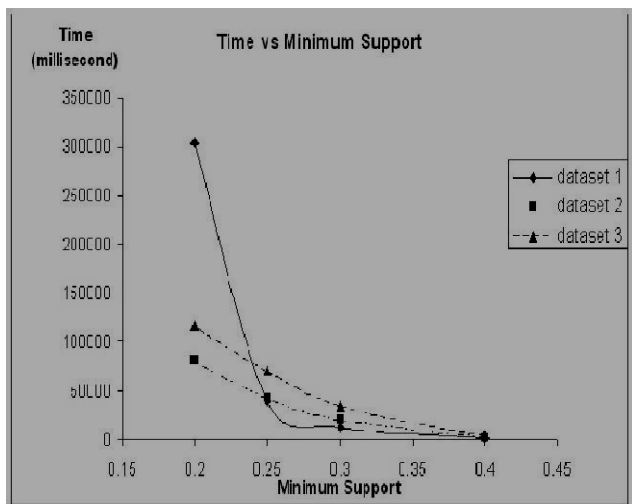


Figure 2: Time with Minimum support

The running time for dataset-1 with minimum support 20% is much higher than the running time of

dataset-2 and dataset-3, since the number of large item sets found for dataset-1 is about 2 times more than the other datasets. The Response time of the parallel and distributed data mining task on XML data is carried out by the time taken for communication, computation cost involved [6]. Communication time is largely dependent on the DDM operational model and the architecture of the DDM systems. The computation time is the time to perform the mining process on the distributed data sets.

VI. CONCLUSIONS

AR mining is a vital problem of DM. It's a new and challenging area to perform AR mining on XML data due to the difficulty of XML data. In our approach, numerous problems in XML data is handled suitably to assure the correctness of the result. The ODAM Algorithm is used for the mining process in a parallel and distributed setting. The response time with the communication and computation factors are measured to achieve an improved response time. The performance examination is done by increasing the number of processors in a distributed environment. As the mining process is done in parallel an optimal solution is obtained.

REFERENCES RÉFÉRENCES REFERENCIAS

1. R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Database," Proc. 20th Int'l Conf. Very Large Databases (VLDB 94), Morgan Kaufmann, 1994, pp. 407-419.
2. R. Agrawal and J.C. Shafer, "Parallel Mining of Association Rules," IEEE Tran. Knowledge and Data Eng., vol. 8, no. 6, 1996, pp. 962-969.
3. D.W. Cheung, et al., "A Fast Distributed Algorithm for Mining Association Rules," Proc. Parallel and Distributed Information Systems, IEEE CS Press, 1996, pp. 31-42.
4. A. Savasere, E. Omiecinski, and S.B. Navathe, "An Efficient Algorithm for Mining Association Rules in Large Database," Proc. 21st Int'l Conf. Very Large Databases (VLDB 94), Morgan Kaufmann, 1995, pp. 432-444.
5. J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation," Proc. ACM SIGMOD Int'l. Conf. Management of Data, ACM Press, 2000, pp. 1-12.
6. M.J. Zaki and Y. Pin, "Introduction: Recent Developments in Parallel and Distributed Data Mining," J. Distributed and Parallel Databases, vol. 11, no. 2, 2002, pp. 123-127.
7. M.J. Zaki, "Scalable Algorithms for Association Mining," IEEE Trans. Knowledge and Data Eng., vol. 12, no. 2, 2000, pp. 372-390.
8. J.S. Park, M. Chen, and P.S. Yu, "An Effective Hash Based Algorithm for Mining Association Rules,

- "Proc. 1995 ACM SIGMOD Int'l Conf. Management of Data, ACM Press, 1995, pp. 175-186.
9. M.J. Zaki, et al., Parallel Data Mining for Association Rules on Shared-Memory Multiprocessors, tech. report TR 618, Computer Science Dept., Univ. of Rochester, 1996.
 10. D.W. Cheung, et al., "Efficient Mining of Association Rules in Distributed Databases, "IEEE Trans.Knowledge and Data Eng., vol. 8, no. 6, 1996, pp.911-922.



This page is intentionally left blank





Automatic Segmentation of Punjabi Speech Signal using Group Delay

By Anupriya Sharma & Amanpreet Kaur

RIMT, Gobindgrah, India

Abstract- This paper describes the concept of automatic segmentation of continuous speech signal. The language used for segmentation is the most widely spoken language i.e. Punjabi. Like all other Indian languages, Punjabi is a syllabic language, thus syllables are selected as the basic unit of segmentation. The traditional way of representing the speech signal is in terms of features derived from short-time Fourier analysis. It is difficult to compute the phase and processing the phase function from the FT phase. By processing the derivative of the FT phase, the information in the short-time FT phase function can be extracted. This paper describes the process of automatic segmentation of speech using group delay technique. This includes segmentation of continuous Punjabi speech into syllable like units by using the high resolution properties of group delay. This group delay function is found to be a better representative of the STE function for syllable boundary detection.

Keywords: *digital signal processing, speech signal, automatic speech segmentation, punjabi speech segmentation, asr, ste, syllables, units of speech.*

GJCST-C Classification : *H.5.5*



Strictly as per the compliance and regulations of:



Automatic Segmentation of Punjabi Speech Signal using Group Delay

Anupriya Sharma ^α & Amanpreet Kaur ^σ

Abstract- This paper describes the concept of automatic segmentation of continuous speech signal. The language used for segmentation is the most widely spoken language i.e. Punjabi. Like all other Indian languages, Punjabi is a syllabic language, thus syllables are selected as the basic unit of segmentation. The traditional way of representing the speech signal is in terms of features derived from short-time Fourier analysis. It is difficult to compute the phase and processing the phase function from the FT phase. By processing the derivative of the FT phase, the information in the short-time FT phase function can be extracted. This paper describes the process of automatic segmentation of speech using group delay technique. This includes segmentation of continuous Punjabi speech into syllable like units by using the high resolution properties of group delay. This group delay function is found to be a better representative of the STE function for syllable boundary detection.

Keywords: digital signal processing, speech signal, automatic speech segmentation, punjabi speech segmentation, asr, ste, syllables, units of speech.

I. INTRODUCTION

The Automatic speech recognizers (ASR) are used to facilitate communication between humans and machines. So it's a machine which understands human and the words spoken by them. The process of segmentation is one of the most important phases in the automatic recognition of speech. There are various units of speech into which it can be segmented, but syllables are found to be one of the most efficient units for automatic speech segmentation. The characteristics features of speech can be expressed by using STE and ZCR. The STE function also known as Short Term Energy function is known to be the better representative of speech segment boundaries. By computing the short-time Fourier analysis information in the speech signal can be extracted. But due to difficulty in computing the phase and also in processing the phase function over the past few decades the features of the FT phase were not exploited fully. By processing the derivative of the FT phase, the information in the short-time FT phase function can be extracted. There are various units of speech. The syllables are found to be the most suitable unit for automatic speech segmentation. A single component in the syllable is called as nucleus. The nucleus is found to be vowel while the onset and coda

are usually consonantal in form. The energy peak in the nucleus region can be viewed as the syllable; the consonants can be viewed as the valleys at both the ends. Many languages been spoken around the world posses a syllabic structure [10]. Mostly the syllable contains two phonetic segments of type CV such as in Japanese language. In contrast, English and German possess a more highly heterogeneous syllable structure [2].

II. RESEARCH BACKGROUND

a) Language Units of Speech in Punjabi

Punjabi is an Aryan language that is spoken by more than hundred million people those are inhabitants of the historical Punjab region (in north western India and Pakistan) and in the Diaspora, particularly Britain, Canada, North America, East Africa and Australasia [8].

Like other Indian languages the Punjabi language also contains segmental phonemes. The three basic units into which the speech can be segmented are: Words, Phonemes and Syllables. The syllable is the most important and widely used unit for automatic speech segmentation. Punjabi is a syllabic language thus syllables are selected as the basic units for segmentation.

b) Syllables as Basic unit of speech

Aksharas is the basic units of the writing system. An Akshara is an orthographic representation of a speech sound in an Indian language. Basically they are syllabic in nature; the typical forms of akshara are V, CV, CCV and CCCV type, where C and V are consonant vowel respectively [9]. There are thirty eight consonants in Punjabi language. Where ten are non-nasal and ten are nasal vowels. Vowels can appear alone but consonants can only appear with vowels. The number of nasal vowels is same as non-nasal vowels and is represented by Bindi or Tippi over the Non-Nasal Vowels. Following is the list of consonants in Punjabi language:

Author α: Department of computer science and engineering RIMTMandi Gobindgarh, India. e-mail: er.anupriya33@gmail.com

Author σ: Department of computer science and engineering BBSB Fatehgarh Sahib, India. e-mail: er.amanpreet.cse@gmail.com

CONSONANTS																																					
ਸ	ਹ	ਕ	ਖ	ਗ	ਘ	ਙ	ਚ	ਛ	ਜ	ਝ	ਞ	ਟ	ਠ	ਡ	ਢ	ਣ	ਤ	ਥ	ਦ	ਧ	ਨ	ਪ	ਫ	ਬ	ਭ	ਮ	ਯ	ਰ	ਲ	ਵ	ੜ	ਸ਼	ਖ਼	ਗ਼	ਜ਼	ਫ਼	ਲ਼
NON NASAL VOWELS																																					
ਅ	ਆ	ਇ	ਈ	ਏ	ਐ	ਉ	ਊ	ਓ	ਔ																												
NASAL VOWELS																																					
ਅੰ	ਆਂ	ਇੰ	ਈਂ	ਏਂ	ਐਂ	ਉਂ	ਊਂ	ਓਂ	ਔਂ																												

Figure 1 : Respresentation of 38 consonants and 20 vowels in Punjabi Language

As already mentioned syllables are the basic and most recommended used units of speech. Syllables are composed of vowel and consonants. Every syllable must have a vowel also known as its nucleus, where as presence of consonant is optional. Vowel (V) is always the nucleus part and the left part is onset and the right part is coda which is always a consonant. The seven types of syllables recognized in Punjabi language are represented in the following figure:

Syllable	Syllable Description	Word	Segments	Word	Segments
V	Vowel	ਉ	ਉ	ਈ	ਈ
VC	Vowel-consonant	ਉਡ	ਉ+ ਡ	ਇਸ	ਇ+ ਸ
CV	Consonant-Vowel	ਜਾ	ਜ+ ਆ	ਗਾ	ਗ+ ਆ
VCC	Vowel-consonant-consonant	ਉਤਰ	ਉ+ ਤ+ ਰ	ਅੰਦਰ	ਅੰ+ ਦ+ ਰ
CVC	Consonant-vowel-consonant	ਰਾਤ	ਰ+ ਆ+ ਤ	ਬਾਤ	ਬ+ ਆ+ ਤ
CVCC	Consonant-vowel-consonant-consonant	ਜੋਤਸ	ਜ+ ਐ+ ਤ+ ਸ	ਪੁਰਬ	ਪ+ ਊ+ ਰ+ ਬ
CCVC	Consonant-consonant-vowel-consonant	ਤਰੇਲ	ਤ+ ਰ+ ਏ+ ਲ	ਸਵੇਰ	ਸ+ ਵ+ ਏ+ ਰ

Figure 2 : Syllables in Punjabi language

III. THREE STATE REPRESENTATION OF SPEECH

The continuous speech signals composed of two elements one includes the speech information, and the other carries noise or silent sections. The verbal part of the speech can be further divided into two categories: voiced and unvoiced speech.

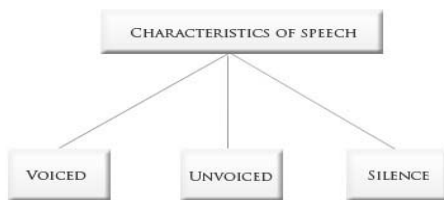


Figure 3 : Block diagram of characteristic features of voice

Moment the air from the lungs passes through the larynx voiced sound is produced. With the passage of air directly through the vocal tract formations the unvoiced speech sounds are produced. The speech production process is incomplete without the detection of voiced and unvoiced speech that is separated by a silence region. In case of silence region no excitation is supplied to the vocal tract and thus, no speech is produced. A regular speech is incomplete inaccurate without silence region. It helps to make the speech understandable [3].

IV. CHARACTERIZATION OF SPEECH

In order to segment continuous speech it is required to check its basic content, whether the signal is voiced or unvoiced. The two characteristics features of voice are the zero crossing rate (ZCR) and short term energy (STE) [13].

a) Zero Crossing Rate

The rate at which the signal crosses zero provides the information regarding its (source of creation) i.e. zero crossing rate. Unvoiced speech has higher zero crossing rate. Whereas in case of voiced speech the zero crossing rate is low. Thus, the amplitude of unvoiced segments is lower than that of the voiced segments.

ZCR can be defined as:

$$Z_n = \sum_{m=-\infty}^{\infty} |\text{sgn}[x(m)] - \text{sgn}[x(m-1)]| w(n-m) \quad (1)$$

Where

$$\text{sgn}[x(n)] = \begin{cases} 1, & x(n) \geq 0 \\ -1, & x(n) < 0 \end{cases}$$

b) Short Term Energy (STE)

Short-time energy of speech signals reflects the amplitude variation. By processing STE function the speech can be segmented. STE shows the voiced content of the signal [13].

The STE can be defined as follows:

$$E_n = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2 \quad (2)$$

The STE of voiced signal is always much greater than that of unvoiced signals. In a speech signal where there are voiced signal its STE will be high, the peaks in the signal represents nucleus that is denoted as vowel where as the valleys at both the ends represents the coda.

V. SEGMENTATION OF SPEECH

The syllable is composed of three parts, the onset, rime (nucleus) and coda. The rime also known as

nuclei, where as the onset and coda consist of consonants. The high energy regions are represented by the nuclei where as the valleys at both ends corresponds to syllable boundaries. The vowel region corresponds to much higher energy region compared to that of a consonant region [9]. In case of spontaneous speech, the definition of a syllable in terms of short-term energy function is suitable for almost all the languages.

Due to local energy fluctuations the STE function alone cannot be directly used to perform segmentation. Techniques such as fixed or even adaptive threshold will not work when the energy variation across the signal is quite high [1].

To overcome the problems of local energy fluctuations, the STE function should be smoothed. The information in speech signals can be represented in terms of features derived from short-time Fourier analysis. The information in the short-time FT phase function can be extracted by computing the group delay function [9].

$$H(\omega) = H1(\omega) \cdot H2(\omega), \tag{3}$$

group delay function can be represented as

$$\begin{aligned} \tau_h(\omega) &= -\frac{\partial(\arg(H(\omega)))}{\partial\omega} \\ &= \tau_{h1}(\omega) + \tau_{h2}(\omega). \end{aligned} \tag{4}$$

The equation (1) shows the multiplicative property of magnitude spectra where as equation (2) is in group delay domain it shows an addition. The group delay spectrum has been found better due to its additive. It was observed that in case of the magnitude

spectra the peaks are clearly visible, but when the two poles are combined together the peaks are not resolved. The research shows the disadvantage of multiplicative property of magnitude spectra. In case of group delay spectra the peaks and valleys are better resolved when the signal is in minimum phase [2].

For any syllable, the STE function of the voiced region, the energy is quite high and diminishes at the ends, representing the consonants, due to which local energy fluctuations. If these local variations are smoothed, then the minima at both ends of a voiced region correspond to syllable boundaries [9].

The algorithm for group delay based segmentation

Step 1 - Let $x[n]$ be continuous speech signal.

Step 2 - Compute N , the length(x) of the input signal.

Step 3 - Calculate the STE function $E[m]$, where $m=1,2,\dots,M$ is the number of frames.

Step 4 - Inverse the STE i.e $E(i) = 1/E(m)$

Step 5 - Compute the IFFT of $E(i)$, It gives the magnitude of the input signal in form of complex function i.e. $a+ib$.

Step 6 - The phase angle is computed from the above values, i.e. $\phi = \tan^{-1}(b/a)$.

Step 7 - Compute the negative derivative of Fourier transformation i.e. the group delay function.

Step 8 - Compute the minimum phase of group delay, i.e. $\text{phase}(n) - \text{phase}(n - 1)$, let the signal be of length n . Locate the positive peaks in the minimum phase group delay function, $(E_i \text{ gd}[f])$. If $E_i \text{ gd}[f]$ is positive, and $E_i \text{ gd}[f-1] < E_i \text{ gd}[f] < E_i \text{ gd}[f+1]$ then $E_i \text{ gd}[f]$ is considered as a peak. These peaks represent the syllable boundaries.

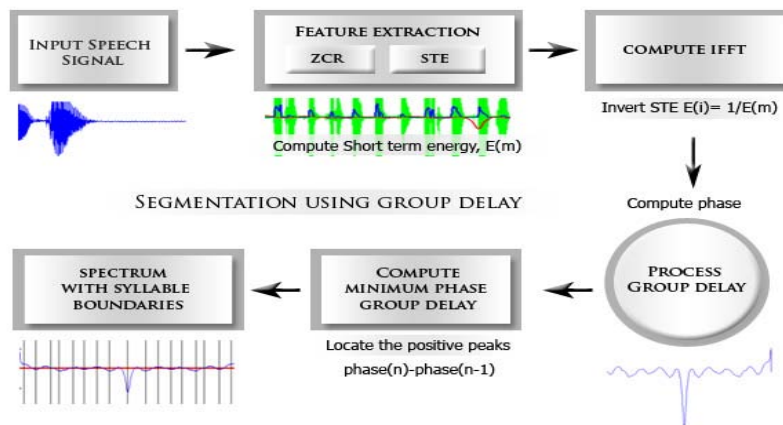


Figure 4 : Steps involved in finding syllable boundaries

VI. RESULTS AND DISCUSSIONS

The technique of automatic segmentation is applied on the continuous Punjabi speech. The method was implemented in Matlab. The group delay algorithm is applied to segment the continuous Punjabi speech waveform. The following sentence is given as an input to the system.

ਅੰਮ੍ਰਿਤਸਰ ਸਿੱਖਾਂ ਦਾ ਸਭ ਤੋਂ ਉੱਚਾ ਧਾਰਮਿਕ ਸਥਾਨ ਹੈ

The system has efficiently marked the syllable boundaries. The onset and offset values are shown in the following table.

Table 1 : Results of Segmentation Obtained with Onset and Offset Syllable Values

Sentence	Onset	Offset	Duration
AMimR	0.576	1.664	1.088
qsr	1.664	2.88	1.216
isW	2.88	3.904	1.024
KF	3.904	5.056	1.152

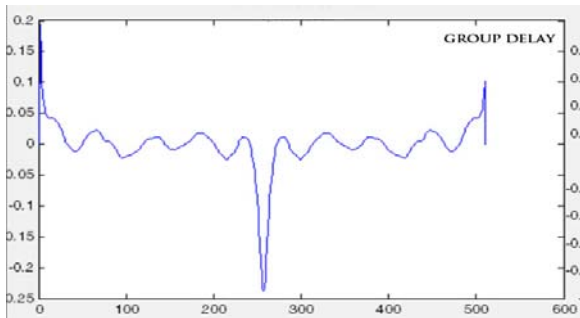


Figure 5 : Signal Representing the Group delay

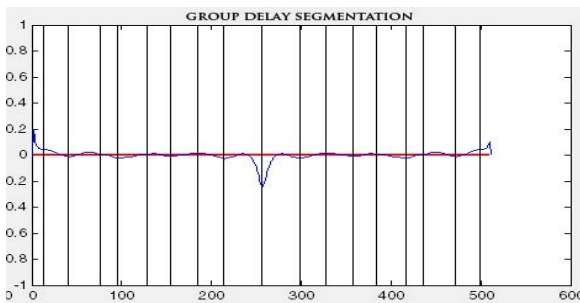


Figure 6 : Group delay based segmentation with marked syllable boundaries

REFERENCES RÉFÉRENCES REFERENCIAS

1. T.Nagarajan et al. "Segmentation of speech into syllable-like units," in Eurospeech Sixth biennial conference of signal processing, Geneva, 2003.
2. T. Nagarajan and H. A. Murthy, "Subband-Based Group Delay Segmentation of Spontaneous Speech into Syllable-Like Units," in Eurasip Journal on Applied Signal Processing , Hindawi Publishing Corporation 2004:17, pp. 2614–2625.
3. Mikael, E. Marcus, "Speech Recognition using Hidden Markov Model, Performance evaluation in noisy environment", Degree of master of science in Electrical Engineering, Department of telecommunications and engineering, Blekinge Institute of Technology, March 2002.
4. G. Pradeep "Text-to-Speech Synthesis for Punjabi Language", Thesis degree of Master of Engineering in Software Engineering submitted in Computer Science and Engineering Department of Thapar Institute of Engineering and Technology (Deemed University), Patiala, May 2006.
5. V. Kamakshi Prasad, T. Nagarajan, Hema A. Murthy, "Automatic segmentation of continuous speech,

- using minimum phase group delay functions," in the proceedings of science direct, Speech Communication 42, 2004, pp. 429–446.
6. G Lakshmi Sar ada, et al. "Automatic transcription of continuous speech into syllable-like units for Indian languages," in Sadhana, Vol. 34, Part 2, April 2009, pp. 221–233.
7. K. Amanpreet, and S. Tarandeep, "Segmentation of Continuous Punjabi Speech Signal into Syllables," in the Proceedings of the World Congress on Engineering and Computer Science 2010 Vol I, WCECS 2010, San Francisco, USA, October 20-22, 2010.
8. S. Parminder, L. Gurpreet, "Corpus Based Statistical Analysis of Punjabi Syllables for Preparation of Punjabi Speech Database," in International Journal of Intelligent Computing Research (IJICR), Volume 1, Issue 3, June 2010.
9. A.Hema, and B.Yegnanarayan, "Group delay functions and its applications in speech technology," in Sadhana, Vol. 36, Part 5, October 2011, pp. 745–782.
10. S. Nishi, and S. Parminder, "Automatic Segmentation of Wave File," in International Journal of Computer Science & Communication Vol. 1, No. 2, July-December 2010, pp. 267-270.
11. A. Hema, B. Ashwin, et al., IIT-Madras, IIT-Kharagpur, CDAC-Trivandrum, CDAC- mumbai, IIIT-Hyderabad, "Building Unit Selection Speech Synthesis in Indian Languages," An Initiative by an Indian Consortium, 2009.
12. Zhihong Hu, Johan Schalkwyk, Etienne Barnard, Ronald Cole, "Speech Recognition Using Syllable-Like Units," Center for Spoken Language Understanding, Oregon Graduate Institute of Science and Technology, September 2008, pp. 218-222.
13. R.G. Bachu, S. Kopparthi, B. Adapa, B.D. Barkana, " Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal," Electrical –Engineering Department School of Engineering, University of Bridgeport, March 2010, volume 7340, 2012, pp. 539-546.
14. B. Shneiderman, "The Limits of Speech Recognition," Communications of the ACM, vol. 43, no. 9, pp. 63–65, September 2000.
15. O. Fujimura, "Syllable as a unit of speech recognition," in IEEE Trans.Acoust., Speech, Signal Processing, vol. 23, February 1975, pp. 82–87.
16. J. L. Gauvian, "A syllable-based isolated word recognition experiment," in IEEE International Conf. on ICASSP'86, vol. 11, April 1986, pp. 57–60.



A Parallel Computational Approach for String Matching- A Novel Structure with Omega Model

By K Butchi Raju & Dr. S. Viswanadha Raju

Griet Aff to Jntuh, India

Abstract- In recent day's parallel string matching problem catch the attention of so many researchers because of the importance in different applications like IRS, Genome sequence, data cleaning etc.,. While it is very easily stated and many of the simple algorithms perform very well in practice, numerous works have been published on the subject and research is still very active. In this paper we propose a omega parallel computing model for parallel string matching. The algorithm is designed to work on omega model parallel architecture where text is divided for parallel processing and special searching at division point is required for consistent and complete searching. This algorithm reduces the number of comparisons and parallelization improves the time efficiency. Experimental results show that, on a multi-processor system, the omega model implementation of the proposed parallel string matching algorithm can reduce string matching time.

Keywords: *string matching; parallel string matching; computing model; omega model.*

GJCST-C Classification : *C.1.4*



Strictly as per the compliance and regulations of:



A Parallel Computational Approach for String Matching- A Novel Structure with Omega Model

K Butchi Raju ^α & Dr. S. Viswanadha Raju ^σ

Abstract- In recent day's parallel string matching problem catch the attention of so many researchers because of the importance in different applications like IRS, Genome sequence, data cleaning etc.,. While it is very easily stated and many of the simple algorithms perform very well in practice, numerous works have been published on the subject and research is still very active. In this paper we propose a omega parallel computing model for parallel string matching. The algorithm is designed to work on omega model parallel architecture where text is divided for parallel processing and special searching at division point is required for consistent and complete searching. This algorithm reduces the number of comparisons and parallelization improves the time efficiency. Experimental results show that, on a multi-processor system, the omega model implementation of the proposed parallel string matching algorithm can reduce string matching time.

Keywords: string matching; parallel string matching; computing model; omega model.

I. INTRODUCTION

String matching has been extensively studied in the past 30 years. A string C of length n is a sequence of characters $C_1C_2\dots C_n$. Let $\Sigma = \{Y_1, Y_2, \dots, Y_N\}$ be a finite set of strings called patterns, and let I be an arbitrary string. The string matching problem is to identify and locate all substrings of I which are patterns in Σ . It performs important tasks in many applications including information retrieval; library systems, artificial intelligence, pattern recognition, molecular biology, and text search and edit systems. The challenge is that for the string matching to be accurate, it needs to be able to search every byte of every input data streaming for a potential match from a large set of strings. So, normal software solutions are not enough, for this we need the hardware computing models [1-7].

The main contributions of this work are summarized as follows. This work offers a comprehensive study as well as the results of typical parallel string matching algorithms at various aspects and their application on computing models. This work suggests the most efficient algorithmic models and demonstrates the performance gain for both synthetic and real data. The rest of this work is organized as, review typical algorithms, algorithmic models and finally conclude the study.

Author α : Associate Professor, Department of CSE, GRIET, Hyderabad, AP, India.

Author σ : Professor & HOD, Department of CSE, JNTU College of Engineering, JNTUH, Jagithyal, A.P., India.

II. PARALLEL COMPUTING MODELS

Parallelism takes many forms and appears in many areas. It is exhibited at the CPU level when microinstructions are executed simultaneously. It is also present when an arithmetic or logic operation is realized by a circuit of small depth, as with carry-save addition. And it is present when multiple computers are connected together in a network. Parallelism can be available but go unused, either because an application was not designed to exploit parallelism or because a problem is inherently serial [8-11].

A parallel computer is any computer that can perform more than one operation at time. By this definition almost every computer is a parallel computer. For example, in the pursuit of speed, computer architects regularly perform multiple operations in each CPU cycle: they execute several microinstructions per cycle and overlap input and output operations (I/O) with arithmetic and logical operations. Architects also design parallel computers that are either several CPU and memory units attached to a common bus or a collection of computers connected together via a network. Clearly parallelism is common in computer science today [8-11].

A parallel programming model is a concept that enables the expression of parallel programs which can be compiled and executed. The value of a programming model is usually judged on its generality: how well a range of different problems can be expressed and how well they execute on a range of different architectures. The implementation of a programming model can take several forms such as libraries invoked from traditional sequential languages, language extensions, or complete new execution models [8-11].

III. LITERATURE

Hundreds of articles, literally, have been published about string matching with computing models, exploring the multitude of theoretical and practical facets of this fascinating fundamental problem. For an n -character text T and an m -character pattern x , the classical algorithm by Knuth, Morris and Pratt takes $O(n+m)$ time and uses $O(m)$ auxiliary space to find all pattern occurrences in the text, namely, all text positions i , such that $x = T[i \dots i + m - 1]$. Many other algorithms have been published; some are faster on the average, use only constant auxiliary space, operate in real-time,

or have other interesting benefits. This work categorizes the algorithms into some categories to emphasize the data structure that drives the matching. These categories are discussed here.

a) *Intrusion Detection Systems (Ids)*

Tuck et al. proposed modifications to the well-known Aho-Corasick string matching algorithm to reduce the amount of memory required to store known malicious strings and improve worst-case timing [12]. They achieve results in both of these areas, while slightly degrading average-case performance. The proposed data storage methods for string matching are bitmap compression and path compression. Their experiments consider both an ASIC and a programmable router design. The ASIC design is tailored to only string matching, while the programmable design assumes an implementation that can be used for many different types of router applications. Experiment results show that the proposed compression optimizations resulted in a 50 times reduction in database size over the Aho-Corasick implementation.

Dharmapurikar et al.[13] proposed a hardware architecture based on parallel Bloom filters for network packet inspection. A Bloom filter is a space-efficient probabilistic data structure that is used to test whether or not an element is a member of a set. It stores a set of signatures compactly by computing multiple hash functions on each member of the set. The answer to querying a database of strings to check for the membership of a particular string can be “false positive”, but never “false negative”. False positive means a condition exists when in fact it does not. False negative means a condition does not exist when in fact it does. The computation time involved in performing the query is independent of the number of strings in the database, provided the memory used by the data structure scales linearly with the number of strings stored in it.

J.Nandhini et al., [14] provide a systematic virus detection software solution for network security for computer systems. Instead of placing entire matching patterns on a chip, proposed solution is based on an antivirus processor that works as much of the filtering information as possible onto a reference memory. The infrequently accessing off-reference data to make the matching mechanism scalable to large pattern sets. Dual port BITCAM processing program is used along with the Exact Matching Engine and Bloom Filter process. This Dual port BITCAM processes next to the exact matching engine and bloom filter process. This Dual port BITCAM process is placed exclusively for obtaining higher throughput.

Safaa O. Al-Mamory et al.,[15] suggest distributed environment in order to enhance the problems of Snort IDS, one of the problems is the efficiency problem. They achieved this goal by enhancing the Snort's string matching engine through

using a LAN of computers, where each computer in the LAN matching a subset of the monitored attacks. Snort is an open source IDS which enable us to detect the previously known intrusion.

Performance Evaluation : From the above papers, it is possible to improve Snort's efficiency using distributed environment and testability of Snort has been enhanced.

b) *Dfa Based Approaches*

Dharmapurikar et al. presented a scheme[16] that can process multiple characters per clock cycle and attain average throughput up to multi-gigabit with moderate memory consumption. But in the worst case they must access the relatively slow off-chip SRAMs frequently for exact string comparisons. Nan et al [17]. Introduced a variable-stride method to deal with string matching ruleset without inciting the byte alignment problem. While enhancing the throughput, this method is sensitive to both ruleset and input string causing greatly reduced throughput in worst cases. Lu et al. [18] introduced parallel DFAs with overlapping input windows to achieve the goal of processing multiple characters in each clock cycle. By slight modification to the straightforward representation of the transition rules, the complexity of each DFA is distinctively reduced. Brodie et al.[19] increased the throughput of regex matching by expanding the alphabet set, resulting in an exponentially increased memory requirement in the worst case. A recent method D. Ficara et.al [20] introduced the sampling techniques to accelerate regex matching, but not all kinds of regex are supported.

WANG Xiaofei et. al[21] proposed a parallel Length-based matching (LBM) architecture to increase the throughput without extra memory cost. The basic idea is to process multiple characters between some specific tags in parallel. For this they use multiple hash functions solution to reduce the possibility of false positive. The evaluation shows that parallel architecture can reduce nearly 55% processing time with less memory consumption than the traditional DFA. According to proposed statistics, there are 99.41% of the input stream has been filtered by architecture which means only 0.58% (263.5M) need to be sent to StriD2FA to match instead of sending to the matching engine byte by byte. Besides the different type of traces, in this paper the other kind of trace that was collected from the World Wide Web should also be used to test the performance and stability.

HyunJin Kim and Seung-Woo Lee [22] proposed a memory-based parallel string matching engine using the compressed state transitions. In the finite-state machines of each string matcher, the pointers for representing the existence of state transitions are compressed. In addition, the bit fields for storing state transitions can be shared. Therefore, the total memory requirement can be minimized by reducing the memory size for storing state transitions. In this, four

large rule sets were extracted from Snort v2.8 rules. Several parameters were swept to find their optimal values. The maximum number of states S was 128 after considering the maximum length of Snort rules. When S was 128, the number of bits in a PMV P was either 16, 24, 32, or 40; when the number of states S was 256, P was either 32, 48, 64, or 80; the maximum number of entries in the LSB transition existence table, K , and the maximum number of shared state transitions, T , were double S , respectively. In Table 1, the total memory requirements were shown by varying P . For the rule sets, the total memory requirements were minimized when S and P were 128 and 24, respectively.

Performance Evolution : The memory requirements were not minimized by only increasing P over 24; therefore, there was a threshold point of P for minimizing memory requirements. In the evaluations in which S was 256, because the required K and P increased rapidly, the required number of string matchers was not decreased. Then, by fixing S and P as 128 and 24, K was varied to find the optimal value; K was either 192, 256, 320, or 384 because K should be greater than S . Each state can have one or more than one state transition. In these evaluations, T was fixed as 256. As a result, when K was 256, the total memory requirements were greatly reduced for all rule sets. This was mainly due to the limited number of state transitions toward noninitial states in each string matcher. With the optimal K of 256, T was varied, that is, T was either 128, 192, 256, or 320. The obtained optimal T was 256. In these evaluations, because state pointers can be shared in the transition table, the required T was small. Considering the optimal values of K and T , it is concluded that many state pointers can be shared in the transition table in the proposed string matcher.

c) *Parallel Processing Based Approaches*

Akhtar Rasool and Nilay Khare[23] proposed an approach, which is designed to work on SIMD parallel architecture where text is divided for parallel processing and special searching at division point is required for consistent and complete searching. This algorithm reduces the number of comparisons and parallelization improves the time efficiency. This algorithm achieves a better result as compared to the multithreaded version of the algorithm where again by text dividing, the parallelization is achieved.

cheng zhong and guo-liang chen[24] presented a perfect hash function for processing string is constructed by applying the Chinese Remainder Theorem, and a fast string matching algorithm, which is suited to process the successive sequences like the network traffic data. The determinate match results and fast execution for the string matching algorithm are very important to the network intrusion detection systems. This paper constructs a perfect hash function for processing string by applying the Chinese Remainder Theorem, transforms uniquely a pattern of length m and

each substring of text of length m into a pair of integer values respectively, and presents a fast string matching algorithm which is suited to process the successive sequences like the network traffic data. The presented algorithm not only obtains the determinate match results, but also holds a linear time complexity in the worst case. The experiment results for matching a sequence database in the network intrusion detection systems also shows that the presented algorithm is efficient.

Panwei Cao and Suping Wu[25] proposed a Parallel KMP algorithm based on MPI to get higher efficiency. The tradition pattern matching algorithm need backtrack and compare repeatedly, so that affects efficiency of algorithm. Knuth and others put forward KMP algorithm in order to promote efficiency of the pattern matching. They combine MPI and KMP algorithm using MPI's Multi process to parallel KMP algorithm. By reducing the time waiting for matching, improve the string matching efficiency.

d) *Aho Corasick Based Approaches*

Wei Lin, Bin Liu[26] presented a pipelined parallel approach for hardware implementation of Aho-Corasick (AC) algorithm for multiple strings matching called P2-AC. P2-AC organizes the transition rules in multiple stages and processes in pipeline manner, which significantly simplifies the DFA state transition graph into a character tree that only contains forwarding edges. In each stage, parallel SRAMs are used to store and access transition rules of DFA in memory. Transition rules can be efficiently stored and accessed in one cycle. The memory cost is less than 47% of the best known AC-based methods. P2-AC supports incremental update and scales well with the increasing number of strings. By employing two-port SRAMs, the throughput of P2-AC is doubled with little control overhead.

Chuanpeng Chen and Zhongping Qin[27] proposed a high throughput configurable string matching architecture based on Aho-Corasick algorithm. The architecture can be realized by random-access memory (RAM) and basic logic elements instead of designing new dedicated chips. The bit-split technique is used to reduce the RAM size, and the byte-parallel technique is used to boost the throughput of the architecture. By the particular design and comprehensive experiments with 100MHz RAM chips, one piece of the architecture can achieve a throughput of up to 1.6Gbps by 2-byte-parallel input, and we can further boost the throughput by using multiple parallel architectures.

Hyun Jin Kim et.al [28] proposed an Aho-Corasick algorithm based parallel string matching. In order to balance memory usage between homogeneous finite-state machine (FSM) tiles for each string matcher, an optimal set of bit position groups is determined. Target patterns are sorted by binary-reflected gray code (BRGC), which reduces bit transitions in patterns

mapped onto a string matcher. In the evaluations of Snort rules, the proposed string matching outperforms the existing bit-split string matching.

Alicherry et al. [29] proposed an architecture consisting of TCAM and SRAM to implement the AC algorithm that utilizes the property of ternary matching of TCAM to achieve the matching of characters expressed in negation expressions. As a result, the space required for the transitions can be reduced. Pao et al. [30] and W. Lin and B. Liu [31] proposed pipeline architectures to implement the partial trie that only contains goto functions of the AC-trie so that it can reduce the space induced by failure functions. N. Hua et al. [32] proposed another approach based on a block-oriented scheme instead of usually byte-oriented processing of patterns to reduce the memory usage. D. P. Scarpazza et al. [33] proposed an optimized software approach for a multi-core processor that splits keywords to fit in the local memories of the processing cores such that it can reach very high overall throughput. Y. Sugawara et al. [34] proposed a string matching method called suffix based traversing (SBT) that is an extension of the AC-algorithm to process multiple input characters in parallel and to reduce the size of the lookup table.

e) *Finite Automata Based Approaches*

HyunJin Kim et al.,[35] proposes a memory-efficient parallel string matching scheme. In order to reduce the number of state transitions, the finite state machine tiles in a string matcher adopt bit-level input symbols. Long target patterns are divided into subpatterns with a fixed length and deterministic finite automata are built with the sub patterns. Using the pattern dividing, the variety of target pattern lengths can be mitigated, so that memory usage in homogeneous string matchers can be efficient. In order to identify each original long pattern being divided, a two stage sequential matching scheme is proposed for the successive matches with sub patterns. Experimental results show that total memory requirements decrease on average by 47.8%-62.8% for Snort and ClamAV rule sets, in comparison with several existing bit-split string matching methods.

Yi-Hua E. Yang and Viktor K. Prasanna [36] proposed a head-body finite automaton (HBFA) which implements SPM in two parts: a head DFA (H-DFA) and a body NFA (B-NFA). The H-DFA matches the dictionary up to a predefined prefix length in the same way as AC-DFA, but with a much smaller memory footprint. The B-NFA extends the matching to full dictionary lengths in a compact variable-stride branch data structure, accelerated by single-instruction multiple-data (SIMD) operations. A branch grafting mechanism is proposed to opportunistically advance the state of the H-DFA with the matching progress in the BNFA. Compared with a fully-populated AC-DFA, proposed HBFA prototype has < 1/5 construction time, requires < 1/20 run-time memory, and achieves 3x to 8x throughput when

matching real-life large dictionaries against inputs with high match ratios. The throughput scales up 27x to over 34 Gbps on a 32-core Intel Manycore Testing Lab machine based on the Intel Xeon X7560 processors.

Yi Tang et al., [37] proposed a paper that extends the classic longest prefix principle from single-character to multi-character string matching and proposes a multi-string matching acceleration scheme named Independent Parallel Compact Finite Automata (PC-FA). In this scheme, DFA is divided into k PC-FAs, each of which can process one character from the input stream, achieving a speedup up to k with reduced memory occupation. They introduce their observation against the prefix based automata algorithms and propose a new conception of inclusion-equivalence principle. Compared with traditional DFA approach and other improved work, PC-FA achieves a high speed-up with a lower memory cost.

Performance Evolution : Experimental evaluations show that seven times of speedup can be practically achieved with a reduced memory size than up-to-date DFA-based compression approaches. They further propose a memory-efficient multi-string matching acceleration scheme named PC-FA Match Engine.

f) *Hardware Related Based Approaches*

KSMV Kumar et.al [38] compared string matching on single processor with multi-processors in parallel environment on hypercube network. The total time taken by search pattern is going to reduce as the No. of processors increases in network. This application developed for text documents of size only MB. It may extend to any size i.e GB to TB also and any other format like image and video files etc. There is lot of scope to develop new trends in this area by evolving modern methods and models for increasing search speed and accuracy. To fulfill here considered both KV-KMP and KV-boyer-moore string matching algorithms for pattern matching in large text data bases using three data sets and graph's drawn for different patterns. Actual test is conducted separately for single processor, two processors, three processors and four processors. Every time, while the test is conducted the program gives elapse time for each processor separately. Therefore the average time is calculated from output result based on the maximum time taken by the individual processor among the processors involved for the particular test. The results shows that the search time taken by single processor is more when compared with multiple processors. It is also observed that as the pattern size increases the search time decreases further. For bigger pattern sizes string matching is more easier for Boyer moore algorithm because of less number of mismatches.

Yao Xin et al [39] presented a hardware architecture for the BWT-based inexact sequence mapping algorithm using the Field Programmable Gate

Array(FPGA). The proposed design can handle up to two errors, including mismatches and gaps. The original recursive algorithm implementation is dealt with using hierarchical tables, and is then parallelized to a large extension through a dual-base extension method. Extensive performance evaluations for the proposed architecture have been conducted using both Virtex6 and Virtex7 FPGAs. This design is considerably faster than a direct implementation. When compared with the popular software evaluation tool BWA, their architecture can achieve the same match quality tolerating up to two errors.

Performance Evaluation : Their major contributions include: (1) improving the original inefficient recursive algorithm using hierarchical tables, (2)parallelizing the inexact search process and constructing a parallel architecture by using the consecutive dual-base extension method (3)evaluating the architectures with a different number of stack arrays and processing elements(PEs). Extensive evaluation experiments are performed using both simulation datasets and real datasets. With the same inexact search options with in two errors for their architecture and BWA software, the hardware architecture can realize the same search quality as BWA. Compared with different CPU platforms running the BWA aln process, their architecture is also capable of better performance in execution speed: the Virtex6 FPGA with 2PEs implemented exceeds all software platforms except for the multithread Xeon CPU; the Virtex7 FPGA implementing 6PEs, however, can reach up to 2 times faster than the Xeon CPU with 6 threads[39].

Hoang Le, and Viktor K. Prasanna [40] proposed an algorithm called "leaf-attaching" to preprocess a given dictionary without increasing the number of patterns. The resulting set of post processed patterns can be searched using any tree search data structure. It also present a scalable, high-throughput, Memory-efficient Architecture for large-scale String Matching (MASM) based on a pipelined binary search tree. The proposed algorithm and architecture achieve a memory efficiency of 0.56 (for the Rogets dictionary) and 1.32 (for the Snort dictionary). As a result, our design scales well to support larger dictionaries. Implementations on 45 nm ASIC and a state-of-the-art FPGA device (for latest Rogets and Snort dictionaries) show that proposed architecture achieves 24 and 3.2 Gbps, respectively. The MASM module can simply be duplicated to accept multiple characters per cycle, leading to scalable throughput with respect to the number of characters processed in each cycle. Dictionary update involves simply rewriting the content of the memory, which can be done quickly without reconfiguring the chip.

TAN Jianlong et al., [41] proposes a novel method to extract the partial strings from each pattern which maximizes search speed. More specifically, with

this method they can compute all the corresponding searching time cost by theoretical derivation, and choose the location which yields an approximately minimal search time.. String matching plays a key role in web content monitoring systems. Suffix matching algorithms have good time efficiency, and thus are widely used. These algorithms require that all patterns in a set have the same length. When the patterns cannot satisfy this requirement, the leftmost m -characters, m being the length of the shortest pattern, are extracted to construct the data structure. They call such m -character strings as partial strings. However, a simple extraction from the left does not address the impact of partial string locations on search speed. They evaluate their method on two rule sets: Snort and ClamAV. Experiments show that in most cases, their method achieves the fastest searching speed in all possible locations of partial string extraction, and is about 5%-20% faster than the alternative methods.

Prasad et.al.,[42] propose two new bit-parallel algorithms to solve the same problem. These new algorithms requires no verification and can handle patterns of length $> w$. These two techniques also use the same BPA of approximate matching and concatenation to form a single pattern from the set of r patterns. It compares the performance of new algorithms with existing algorithms and found that proposed algorithms MASM1 and MASM2 have better running time than the existing algorithms: MASM and BPA (running r times). Mosleh M. Abu-Alhaj et al.,[43] proposes a general platform for improving the existing Exact String-Matching algorithms executing time, called the PXSMAlg platform. The function of this platform is to parallelize the Exact String-Matching algorithms using the MPI model over the Master/Slaves paradigms. The PXSMAlg platform parallelization process is done by dividing the Text into several parts and working on these parts simultaneously. This improves the executing time of the Exact-String- Matching algorithms. They have simulated the PXSMAlg platform in order to show its competence, through applying the Quick Search algorithm on the PXSMAlg platform. The simulation result showed significant improvement in the Quick Search executing time, and therefore extreme competence in the PXSMAlg platform.

PXSMAlg Platform Performance Analysis: They have built a simulation to demonstrate the feasibility of the PXSMAlg platform and its compatibility with the Exact-String-Matching algorithms. The simulation is done to compare the performance of the PXSMAlg platform with the conventional method, that is, the sequential method. The simulation built is based on three main factors: executing time, speedup, and efficiency. Their simulation runs under the Aurora server, which consists of 14 nodes, with each node having 2 CPUs, a speed of 1300MHz and a 1GB memory; all nodes run the Linux OS. The results showed high

performance of the PXSMAlg platform over the sequential methods

g) *Gpu's Based Approaches*

Benedikt Forchhammer et al [44] presented a complete duplicate detection workflow that utilizes the capabilities of modern graphics processing units (GPUs) to increase the efficiency of finding duplicates in very large datasets. Proposed solution covers several well-known algorithms for pair selection, attribute-wise similarity comparison, record-wise similarity aggregation, and clustering. Here redesigned these algorithms to run memory-efficiently and in parallel on the GPU. Proposed experiments demonstrate that the GPU-based workflow is able to outperform a CPU-based implementation on large, real-world datasets. For instance, the GPU-based algorithm deduplicates a dataset with 1.8m entities 10 times faster than a common CPU-based algorithm using comparably priced hardware.

Antonino Tumeo et al.,[45] focus on the matching of unknown inputs streamed from a single source, typical of security applications and difficult to manage since the input cannot be preprocessed to obtain locality. They consider shared-memory architectures (Niagara 2, x86 multiprocessors and Cray XMT) and distributed memory architectures with homogeneous (InfiniBand cluster of x86 multicores) or heterogeneous processing elements (InfiniBand cluster of x86 multicores with NVIDIA Tesla C1060 GPUs).

Performance Evolution: They have presented several software implementations of the Aho-Corasick pattern matching algorithm for high performance systems, and carefully analyzed their performance. They presented optimized designs for the various architectures, discussing several algorithmic strategies, for shared memory solutions, GPU-accelerated systems and distributed memory systems. They describe how each solution achieves the objectives of supporting large dictionaries, sustaining high performance, and enabling customization and flexibility using various data sets. They found that the absolute performance obtained on the Cray XMT is one of the highest reported in literature, at ≈ 28 Gbps (using 128 processors) for a software solution with very large dictionaries. This work compares several software-based implementations of the Aho-Corasick algorithm for high performance systems.

Antonino Tumeo et al [46] presented several software implementations of the Aho-Corasick pattern matching algorithm for high performance systems, and carefully analyzed their performance. It considered the various tradeoffs in terms of peak performance, performance variability, and data set size. It presented optimized designs for the various architectures, discussing several algorithmic strategies, for shared-memory solutions, GPU-accelerated systems, and distributed-memory systems. Finally from this paper

found that the absolute performance obtained on the Cray XMT is one of the highest reported in the literature, at 28 Gbps (using 128 processors) for a software solution with very large dictionaries. Through multithreading and memory hashing the XMT is able to maintain stable performance across very different sets of dictionaries and input streams. A dual Niagara 2 obtains stable performance only in low and medium matching conditions, while a dual Xeon 5560 has more varied results, obtaining high peak rates for light matching conditions, but progressively reducing its performance as the number of matches' increases.

h) *Ram Based Approaches*

Vinod.O et al [47] proposes an alternative algorithm using a Hash Function which uses a SRAM that creates fingerprints of the packet payload which are then compared with the patterns signatures. The proposed hash based system consumes around 0.56 times or 56 percent less memory than the memory consumed by the RTCAM method. It can also be observed from the results that as the TCAM width doubles the initial width the memory consumption increases around 1000kb the initial memory consumption value in RTCAM method. But in the case of hash based method as the block size is doubled the memory consumption increases by a small value around 200kb only from the initial memory consumption value. Hence the proposed hash based method is efficient than the RTCAM method in terms of memory consumption.

Oren Ben-Kiki et al.,[48] proposed macro-level algorithm only uses the standard AC instructions of the word-RAM model (i.e. no integer multiplication) plus two specialized micro-level AC 0 word-size packed-string instructions. The main word-size string matching instruction wssm is available in contemporary commodity processors. The other word-size maximum-suffix instruction wslm is only required during the pattern pre-processing.

Performance evolution: They demonstrated how to employ word-size string matching instructions to design optimal packed string matching algorithms in the word-RAM, which are fast both in theory and in practice. They also consider the complexity of the packed string matching problem in the classical word-RAM model in the absence of the specialized micro-level instructions wssm and wslm. They propose micro-level algorithms for the theoretically efficient emulation using parallel algorithms techniques to emulate wssm and using the Four-Russians technique to emulate wslm. Surprisingly, bit-parallel emulation of wssm also leads to a new simplified parallel random access machine string matching algorithm. As a byproduct to facilitate their results they develop a new algorithm for finding the leftmost (most significant) 1 bit in consecutive non-overlapping blocks of uniform size inside a word.

j) Approaches For Genome Sequences

More approaches have been developed for the k differences problem in fields including molecular biology. Some of them are briefly mentioned in this part. Cheng and Fu [49] proposed VLSI architecture of two dimensional arrangements of $n \times m$ processing elements. P-NAC (Princeton Nucleic Acid Comparator) [50] was built using linear systolic array architecture for comparing DNA sequences.

Similarly, Sastry et al [51] presented a VLSI chip for computing similarity between two strings. Two generations of the Splash processors, which are based on systolic arrays of FPGAs (field-programmable gate arrays) have been designed [52]. As Foster and Kung [53] mentioned, the good algorithms for VLSI implementation are not necessarily those requiring minimal computation. Computation is cheap in VLSI and the communication determines the performance. This matter is also applicable to the dataflow environment. As used in most high performance software string matching algorithms, trial of skipping operations will degrade the overall performance of the dataflow algorithm. Thus we start from the naïve algorithm which requires $n-m+1$ attempts and each attempt takes m comparisons.

Carla Correa Tavares dos Reis and Oswaldo Cruz[54] presented the development of algorithms for approximate string matching using parallel methods. It intends to do the maximum of molecular sequences comparisons per unity of time. The parallel program implementation has carried out in C on an available twenty processing nodes clustering architecture using a model of parallel programming systems, the MPI (Message-Passing Interface), which is as library of subroutines. In this paper also concerned with reporting the speedup and efficiency measures. More precisely, present a parallel algorithm for approximate sequence matching, showing its implementation and reporting its measures in comparison to its sequential version. It use one of the possible approaches to reduce the time spent on comparisons of molecular database sequences by distributing the data among processors, which achieves a linear speedup (time) and requires constant space memory per processor. It also compares between the serial processing and the parallel processing (under the operation conditions offered by MPI ambient), the parallel version always gave the best results (execution and data distribution times).

Muhammad Zubair et al.,[55] propose a new concept to solve the problem of exact string matching by scanning text string for the rightmost character of the pattern in preprocessing phase. In matching phase TSPRC (Test Scanning for Pattern Rightmost Character) compares the pattern with text window from both directions simultaneously. They proposed a new algorithm TSPRC, in addition to proposed algorithm Naive, Not So Naive, quick Search, Boyer Moor Bad

Character and Berry Ravindran algorithms are experimented with TSPRC. In the experiment they took a text string T of the size of sixty thousand characters and pattern P of lengths $\{6, 12, 18, 24, 30, 36, 42, 48, 54, 60\}$. Text String is consisted of the four characters $L = \{A, C, G, T\}$ these are the characters occurred in DNA pattern. Pattern is also of $L = \{A, C, G, \text{ and } T\}$. Several experiments have been conducted and the obtained results are compared with Naive, Not So Naive, Quick Search and Berry Ravindr algorithms. Comparison made on two bases; total number of characters compared by each algorithm and the number attempts taken by each algorithm for finding all possible occurrence of the pattern in the text.

Performance Evolution : Comparison of proposed algorithm is made with existing algorithms on the bases of the number of characters compared and the attempts made by experimented algorithms to complete the task. In preprocessing phase of TSPRC; rightmost character of pattern is searched in the text string. Index of the character in the text string is used to calculate the shift's length and align the pattern with next text window. The analysis and the experimental results illustrate that the TSPRC algorithm is better than the number of existing algorithms.

Tomohiro I et al.,[56] presented a linear-time algorithm to solve the palindrome pattern matching problem. The first algorithm is a Morris–Pratt type algorithm, and the second one is a suffix-tree type algorithm. The palindrome pattern matching problem is to compute all positions i of t such that $\text{Pals}(p) = \text{Pals}(t[i : i + m - 1])$, given a text t of length n and a pattern p of length m . Palindromes in strings have widely been studied both in theoretical and practical contexts, such as in word combinatorics and in bioinformatics. In practical applications such as DNA and RNA sequence analysis, it is desired to cope with gapped palindromes which have a space between the left and right arms of the palindromes.

IV. SCATTERED COMPUTING AND SCATTERED ALGORITHMS

The new technologies of networking and the dramatic evolution of the internet and intranet impacts the way that we use computers and changes the way we create applications for them. Distributed applications are becoming the natural way to build software. "Distributed computing" is all about designing and building applications as a set of processes that are distributed across a network of machines and work together as an ensemble to solve a common problem[8-11]. Distributed algorithms are algorithms designed to run on a distributed system; where many processes cooperate by solving parts of a given problem in parallel. For this purpose, the processes have to exchange data and synchronize their actions. In contrast to so called parallel algorithms, communication and

synchronization is solely done by message passing - there are no shared variables- and usually the processes do not even have access to a common clock. Since message transmission time cannot be ignored, no process has immediate access to the global state. Hence, control decisions must be made on a partial and often outdated view of the global state which is assembled from information gathered gradually from other processes. Distributing computing requires a tool by which the distributed machines can communicate. Many tools are available such as Remote Method Invocation (RMI), CORBA and Java Space. Each tool has its own specifications; the application designer chooses the appropriate one for his application requirement[8-11].

In this, paper we examine a number of explicitly parallel models of computation for string matching,

including shared and distributed memory models proposed by different researchers and proposed a computing model with omega architecture.

V. PROPOSED SYSTEM ARCHITECTURE

a) System Architecture

System Architecture describes “the overall structure of the system and the ways in which the structure provides conceptual integrity”. Architecture is the hierarchical structure of a program components (modules), the manner in which these components interact and the structure of data that are used by that components. The existing string matching system architecture is as shown in Fig 1 and in this the efficiency is not good.

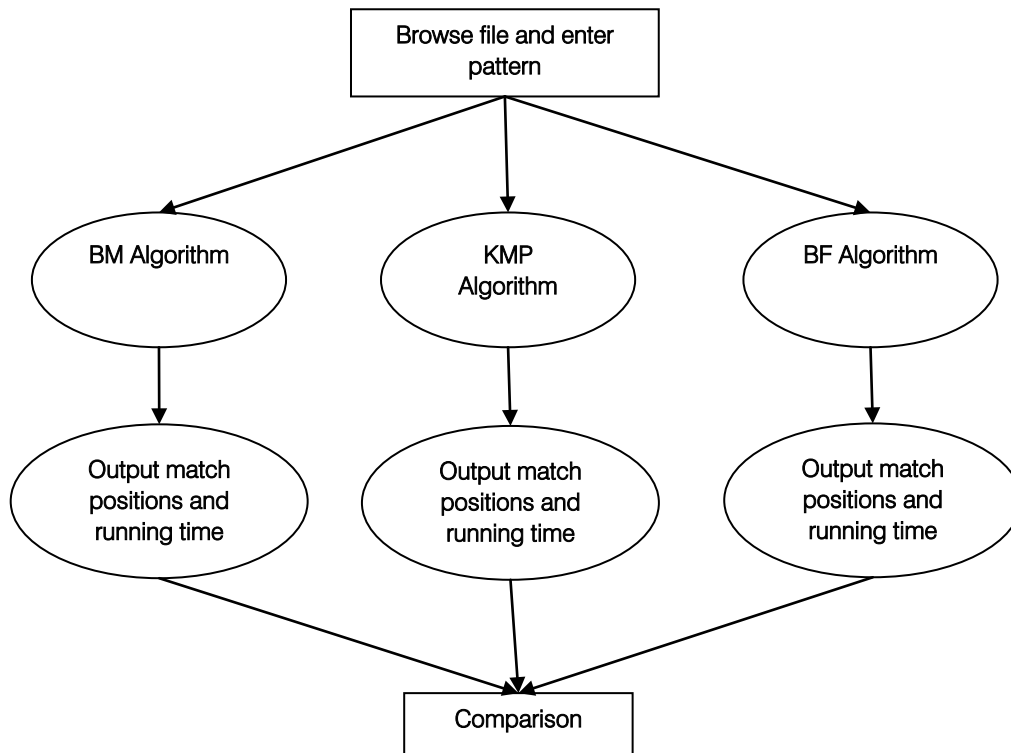


Figure 1 : Existing System

In the existing string matching architecture we search the required pattern sequentially at first we pass the required that is to be searched and this pattern is searched by using the three algorithms Brute force, KMP, Boyer Moore the entire string is passed through all the algorithms and the output match and the running time is calculate for the required pattern from all the algorithms and the algorithm with the least running time is selected, all this is done sequentially which takes more time to execute to improve the efficiency and the performance in this we use the parallel string matching algorithms with multicores processors as shown in Fig2.

The proposed system Architecture of Comparison of parallel String Matching Algorithms is as

follows in the below diagram. In this search the pattern parallel. in this at first we take the input as a string or text. The required text that is to be searched is further divided into further small patterns and all this patterns are passed on the different parallel algorithms like KMP boyar Moore, brute force and at all the output position match and running time of all the patterns is calculated and the all the patterns of same algorithm are added and all the resulted running time are compared with other algorithms resulting time and from them the best one is taken as the efficient algorithm for the string matching.



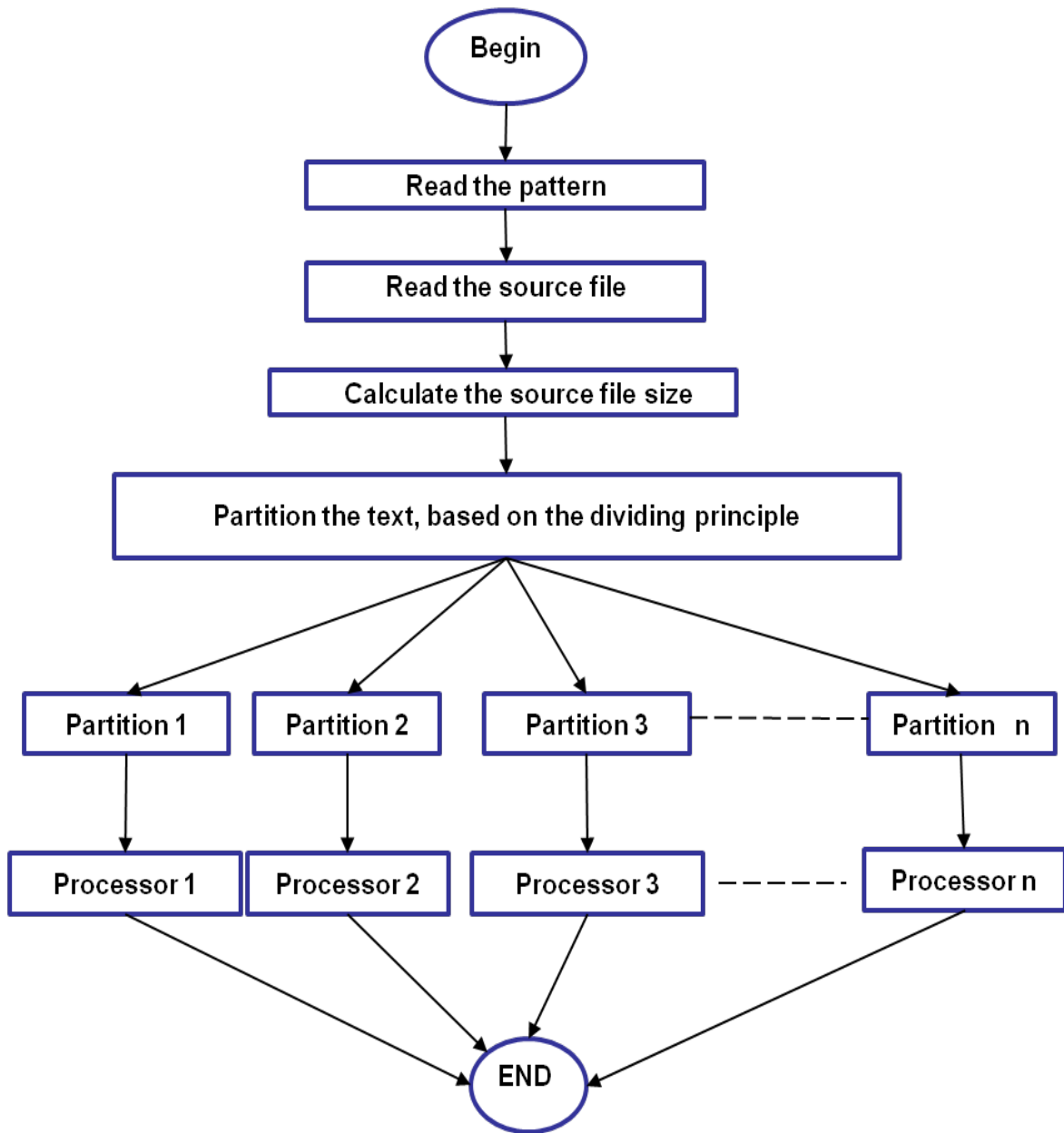


Figure 2 : Proposed Systems

b) Proposed Approach

In now a day as the current free textual database is growing vast there is a problem of finding the pattern by string matching the efficiency is decreased and takes more time. In our paper, we use parallel algorithms to increase the efficiency on multicore processor we pass the same string to all the three algorithms and we select the best based on the running time. Here we have to implement the proposed system with JAVA 1.7 multi threading; initially we have to implement the proposed algorithm with threading on Multicore processor. Here we discuss some of them.



1. Choose the TEXT file and Open the file in read mode
2. Read a line from TEXTFILE into variable ProcessLine (text of size n)
3. Read the pattern of size m and number of processors
4. Distribute the pattern to all the processors in the omega model structure
5. Each Processor searches the pattern in the ProcessLine with proposed approach and returns the result

Proposed approach

Begin

- 5.1. If ProcessLine equals to NULL GOTO Step 5.7
- 5.2. It compares the last character of the pattern with the rightmost text character of the ProcessLine, then
- 5.3. if found it compares the first character of the pattern with the leftmost text character of the window, then
- 5.4. if found it compares the middle character of the pattern with the middle text character of the window. And finally
- 5.5. if found it actually compares the other characters from the second to the last but one.
 - else
- 5.6. If a mismatch occurs, it will shift according to the value in pre-processing stage.
- 5.7. Increment the counter in TEXTFILE and GOTO Step 5.1. Repeat the steps from 5.1 to 5.6 until the end of the file.

End

Pre-Processing Stage

Computing model consists of the good suffix and bad-character shift function. The bad character shift means to shift the pattern so that the text character of the mismatch is aligned to the last occurrence of that character in the initial part of the pattern (pattern minus last pattern character), if there is such an occurrence, or one position before the pattern if the mismatched character doesn't appear in the initial part of the pattern at all.

The other shift, the good suffix shift, aligns the matched part of the text, m, with the rightmost occurrence of that character sequence in the pattern that is preceded by a different character (including none, if the matched suffix is also a prefix of the pattern) than the matched suffix m of the pattern.

6. Each processor stores the MATCH results and RETURNS to the main program.
7. The main program collect the return results from all the processors and summing them
8. STOP.

c) Claims

Implementation is the stage where the theoretical design is turned into a working system. The most crucial stage in achieving a new successful system and in giving confidence on the system for the users that will work efficiently and effectively. The system will be implemented only after thorough testing and if it is found to work according to the specification. For testing our proposed system we will take the gene sequence data set, consists of the four nucleotides a, c, g and t (standing for adenine, cytosine, guanine, and thymine, respectively) used to encode DNA. Therefore, the alphabet is $O = \{A, C, G, T\}$. The text is consisted of 7,50,000 records. Our test tested with different processors like i3, i5 etc., here we put some achievements what we develop and observe, finally our system shows that parallel approach is much better than sequential approach with multi core processor. The Fig 3 shows (Graph) Execution time vs File size on sequential search with intel i3 processor using Brute force algorithm. From the graph we clearly observe that sequential is better compared to parallel approach. The Fig 4 shows (Graph) Execution time vs File size on

sequential search with intel i5 processor using Brute force algorithm. From the graph we clearly observe that sequential is better compared to parallel approach.

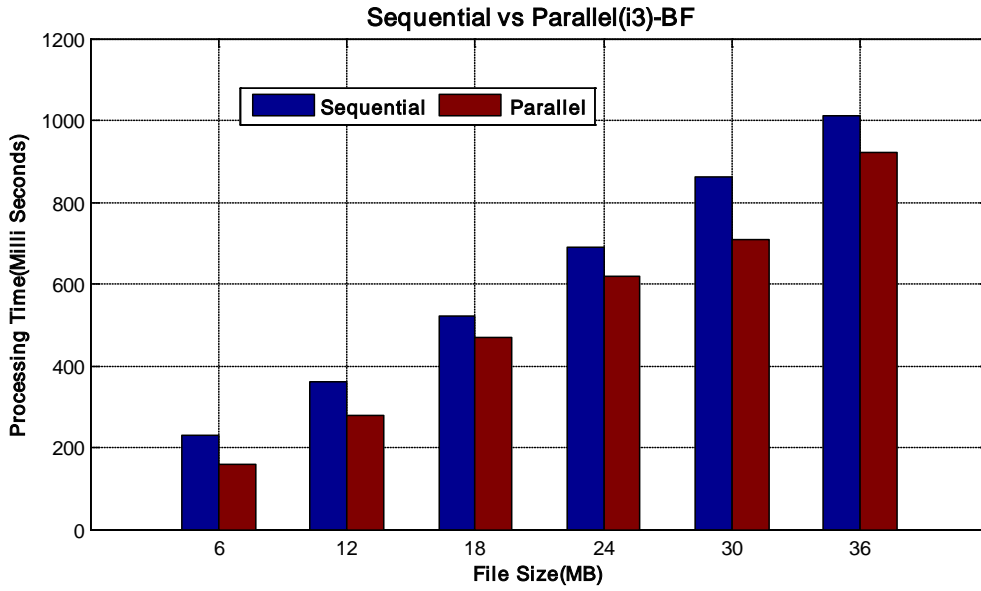


Figure 3 : Sequential VS Parallel(i3)-BF

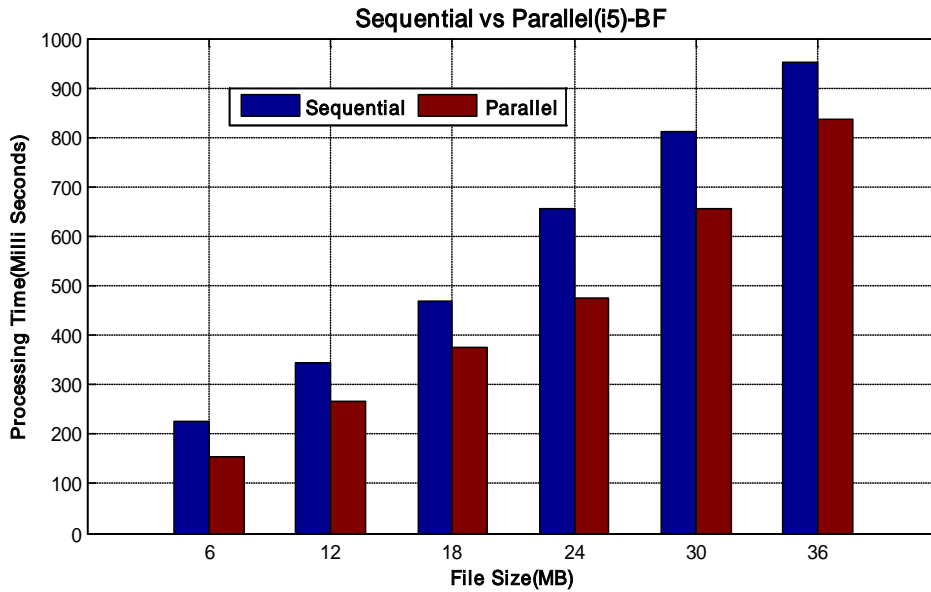


Figure 4 : Sequential VS Parallel(i5)-BF

The Fig 5 shows(Graph) Execution time vs File size on sequential search with intel i3 processor using KMP algorithm. From the graph we clearly observe that sequential is better compared to parallel approach. The Fig 6 shows(Graph) Execution time vs File size on sequential search with intel i5 processor using KMP algorithm. From the graph we clearly observe that sequential is better compared to parallel approach.

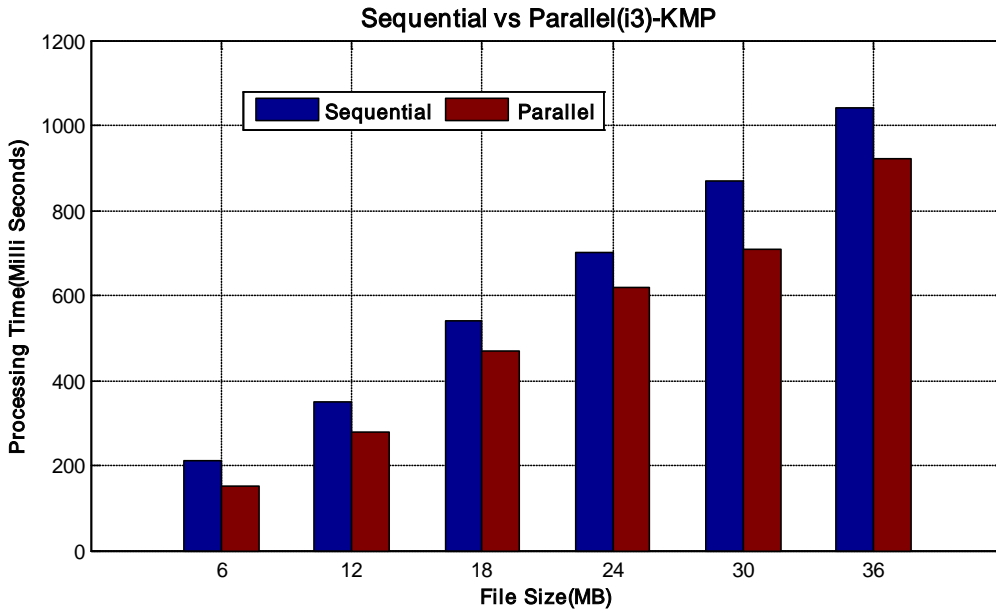


Figure 5 : Sequential VS Parallel(i3)-KMP

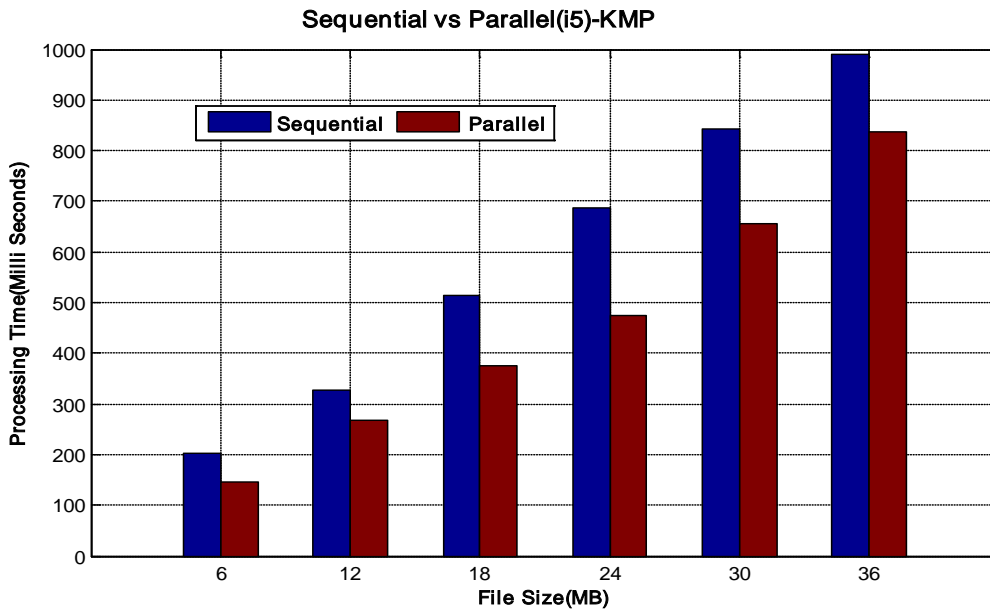


Figure 6 : Sequential VS Parallel(i5)-KMP

The Fig 7 shows (Graph) Execution time vs File size on sequential search with intel i3 processor using BM algorithm. From the graph we clearly observe that sequential is better compared to parallel approach. The Fig 8 shows (Graph) Execution time vs File size on sequential search with intel i5 processor using BM algorithm. From the graph we clearly observe that sequential is better compared to parallel approach.

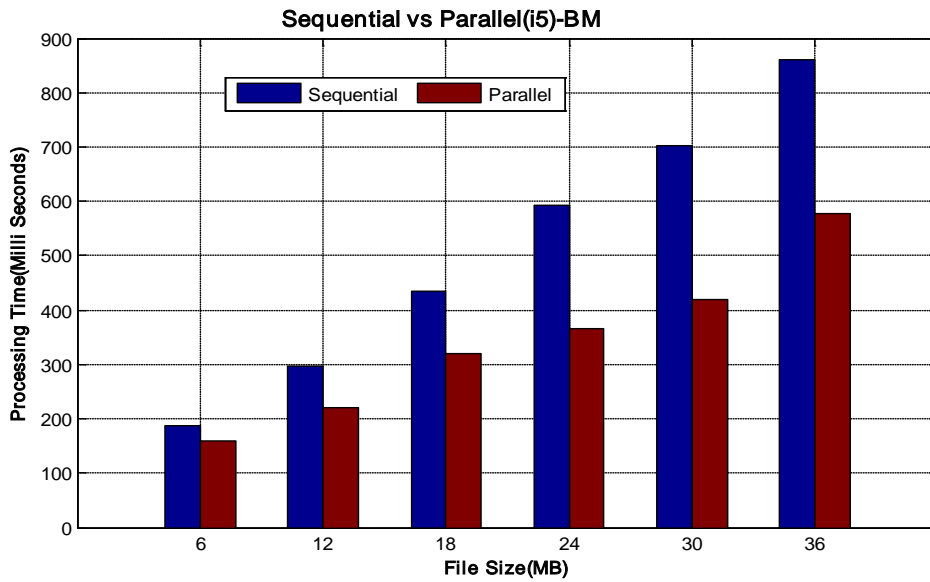


Figure 7 : Sequential VS Parallel(i5)-BM

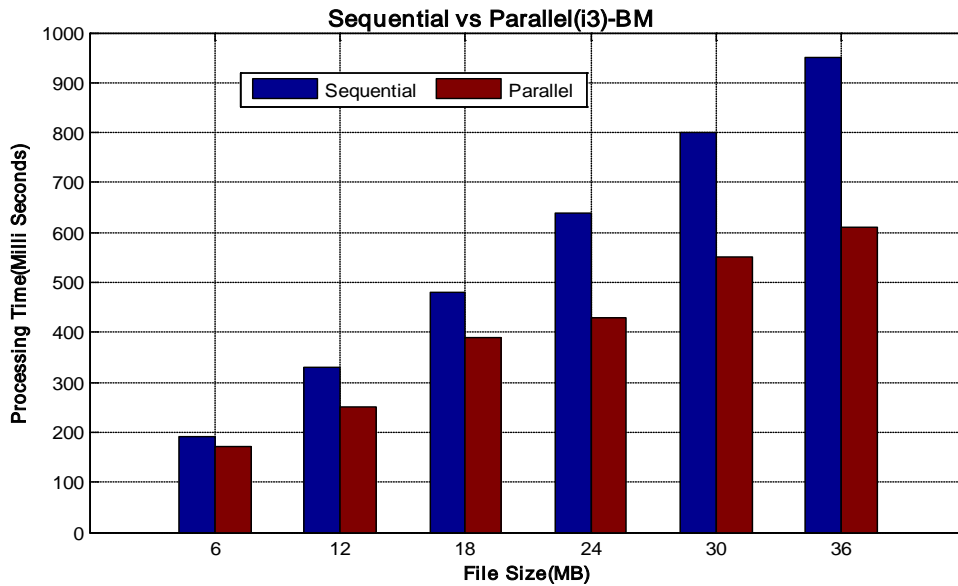


Figure 8 : Sequential VS Parallel(i3)-BM

The Fig 9 shows (Graph) Execution time vs File size on sequential search with intel i3 processor using Proposed algorithm. From the graph we clearly observe that sequential is better compared to parallel approach. The Fig 10 shows (Graph) Execution time vs File size on sequential search with intel i5 processor using Proposed algorithm. From the graph we clearly observe that sequential is better compared to parallel approach.

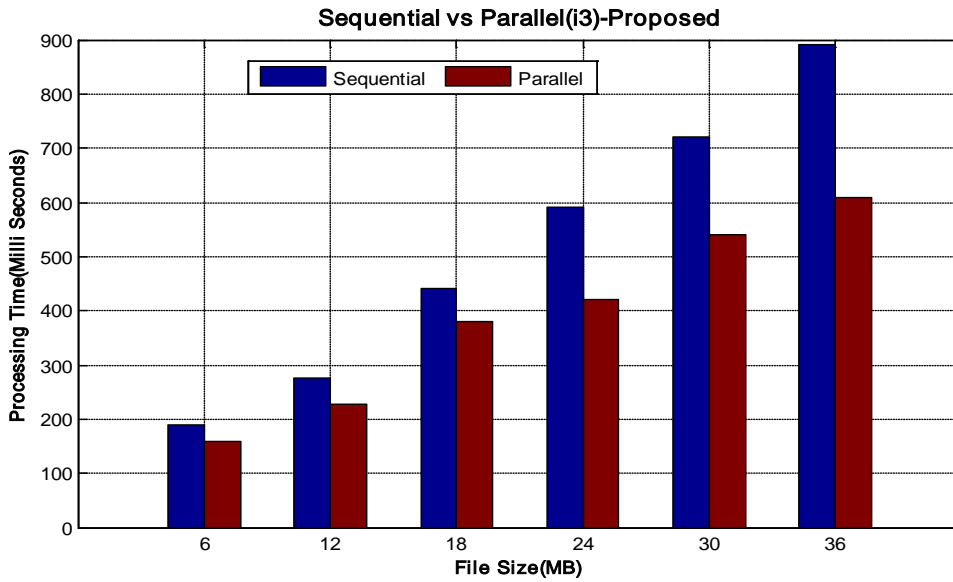


Figure 9 : Sequential VS Parallel(i3)-Proposed

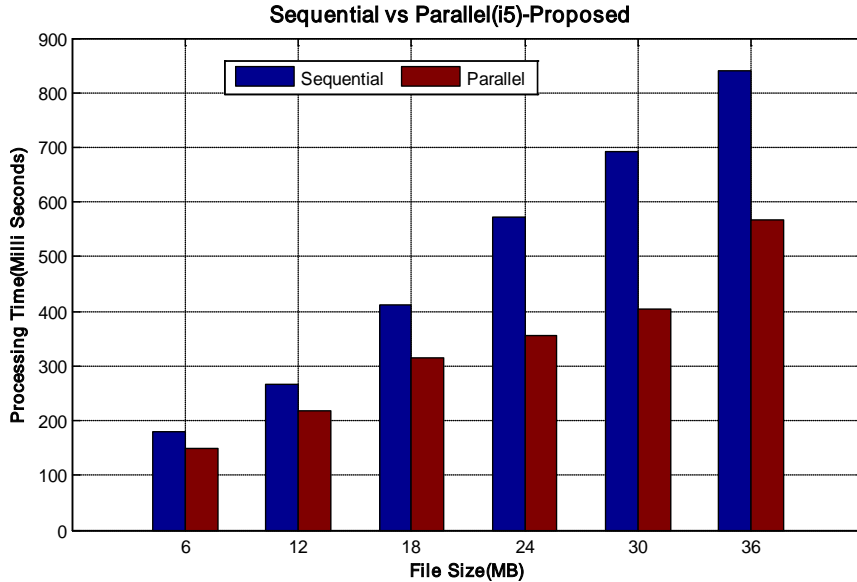


Figure 10 : Sequential VS Parallel(i5)-Proposed

The Fig 11 shows (Graph) Execution time vs File size on sequential search with intel i3 processor using Boyer Moore, Brute force, KMP and proposed Algorithm. This graph shows the performance difference between Boyer Moore, Knuth Morris Pratt and Brute force algorithms. From the graph clearly observe that proposed is better compared to other approaches.



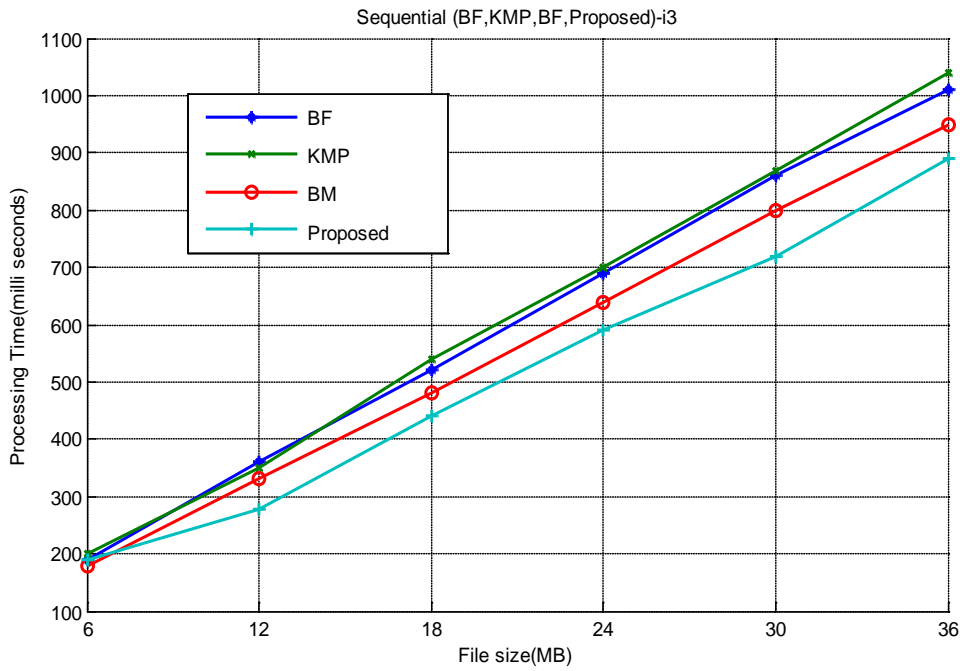


Figure 11 : Sequential approaches BF,BM,KMP and Proposed (i3)

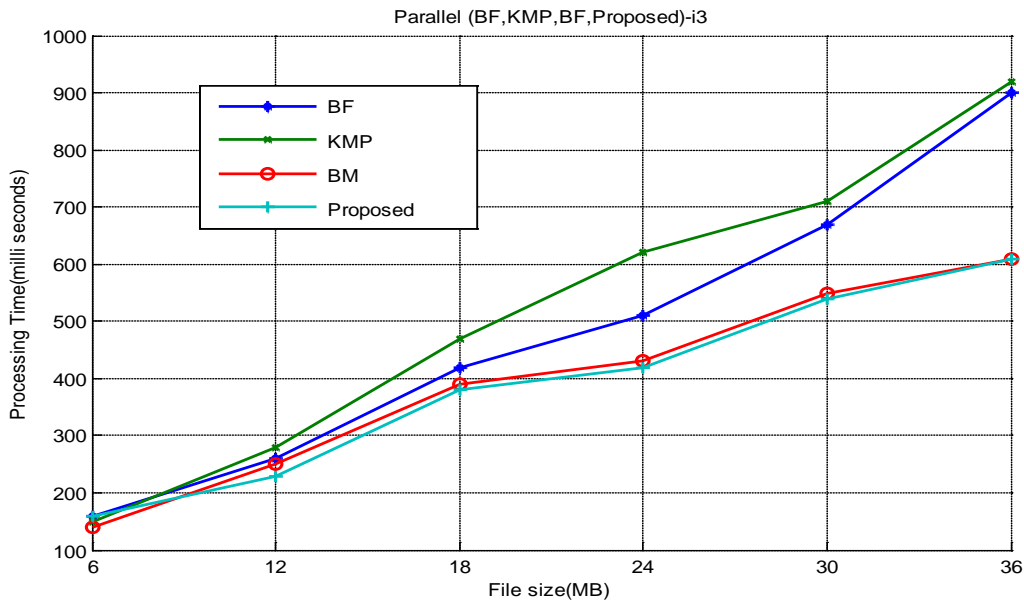


Figure 12 : parallel approaches BF,BM,KMP and Proposed (i3)

The Fig 12 shows (Graph) Execution time vs File size on parallel search with intel i3 processor using Boyer Moore, Brute force, KMP and proposed Algorithm. This graph shows the performance difference between Boyer Moore, Knuth Morris Pratt and Brute force algorithms. From the graph clearly observe that proposed is better compared to other approaches, as well as this parallel approach is much better compared to sequential approaches.

Algorithm. This graph shows the performance difference between Boyer Moore, Knuth Morris Pratt and Brute force algorithms. From the graph clearly observe that proposed is better compared to other approaches.

The Fig 13 shows (Graph) Execution time vs File size on sequential search with intel i5 processor using Boyer Moore, Brute force, KMP and proposed

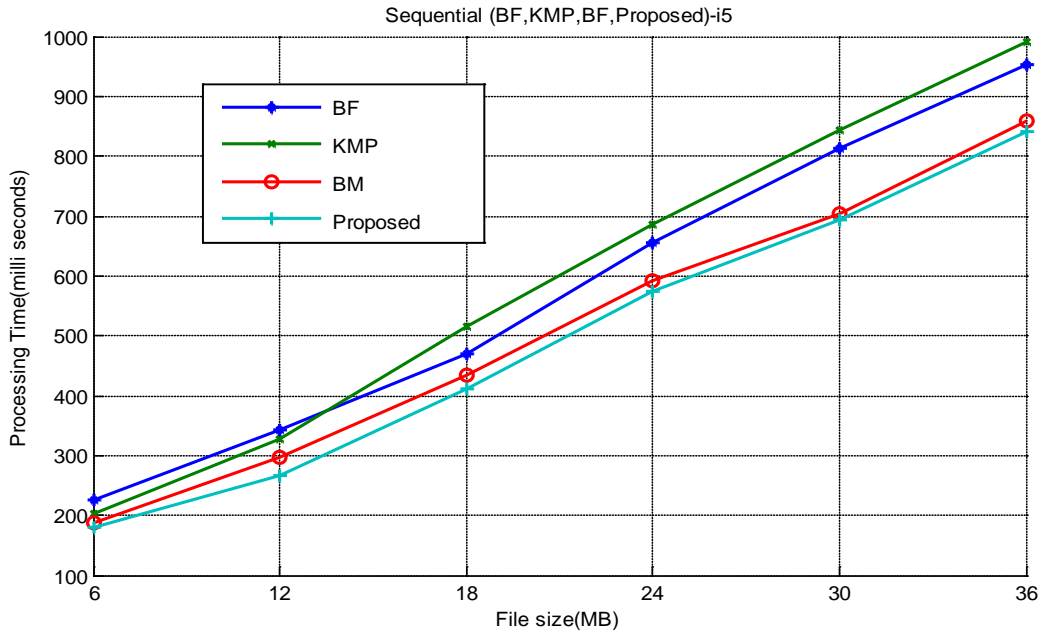


Figure 13: sequential approaches BF, BM, KMP and Proposed (i5)

The Fig 14 shows (Graph) Execution time vs File size on parallel search with intel i5 processor using Boyer Moore, Brute force, KMP and proposed Algorithm. This graph shows the performance difference between Boyer Moore, Knuth Morris Pratt and Brute

force algorithms. From the graph clearly observe that proposed is better compared to other approaches, as well as this parallel approach is much better compared to sequential approaches.

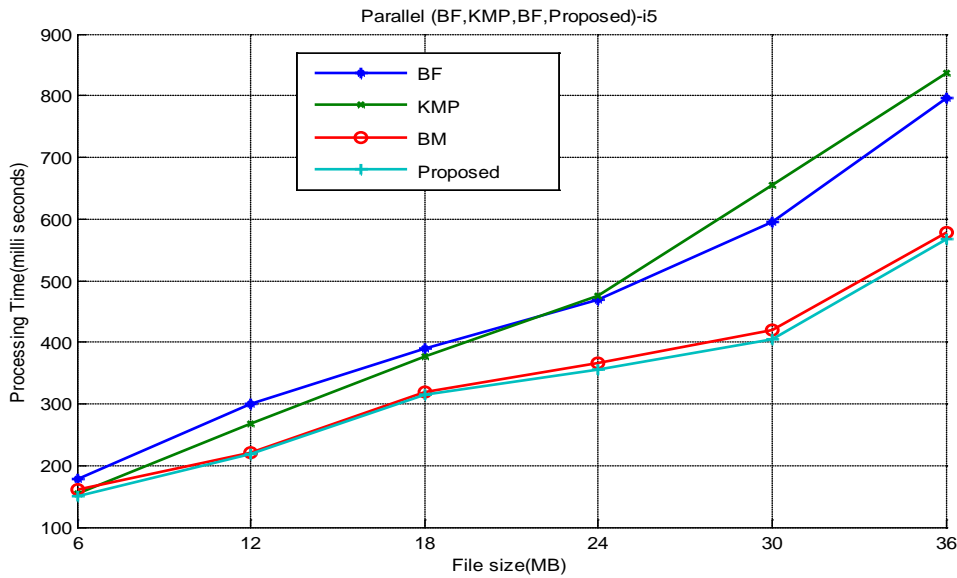


Figure 13: Parallel approaches BF, BM, KMP and Proposed (i5)

VI. CONCLUSIONS

In this study, we widely investigate the problem of parallel string matching in the context of Parallel Computing with omega model. This Parallelization greatly improves the matching efficiency if the text size is very large and a sufficient numbers of processors are available. The most important characteristic of the

proposed algorithm is that reduces the number of comparisons by making better use of next bit characters. Further parallelization of proposed algorithm provides parallel computing of pattern searching on the text in omega architecture. Experimental results shows that proposed algorithm much more better than other approaches both sequential and parallel.

REFERENCES RÉFÉRENCES REFERENCIAS

1. Chinta Someswararao, K Butchiraju, S ViswanadhaRaju, "Recent Advancement is Parallel Algorithms for String matching on computing models - A survey and experimental results", LNCS, Springer, pp.270-278, ISBN: 978-3-642-29279-8, 2011.
2. Chinta Someswararao, K Butchiraju, S ViswanadhaRaju, "PDM data classification from STEP- an object oriented String matching approach", IEEE conference on Application of Information and Communication Technologies, pp.1-9, ISBN: 978-1-61284-831-0, 2011.
3. Chinta Someswararao, K Butchiraju, S ViswanadhaRaju, "Recent Advancement is Parallel Algorithms for String matching - A survey and experimental results", IJAC, Vol 4 issue 4, pp-91-97, 2012.
4. Simon Y. and Inayatullah M., "Improving Approximate Matching Capabilities for Meta Map Transfer Applications," Proceedings of Symposium on Principles and Practice of Programming in Java, pp.143-147, 2004.
5. Chinta Someswararao, K Butchiraju, S ViswanadhaRaju, "Parallel Algorithms for String Matching Problem based on Butterfly Model", pp.41-56, IJCST, Vol. 3, Issue 3, July – Sept, ISSN 2229-4333, 2012.
6. Chinta Someswararao, K Butchiraju, S ViswanadhaRaju, "Recent Advancement is String matching algorithms- A survey and experimental results", IJCIS, Vol 6 No 3, pp.56-61, 2013.
7. S. viswanadha raju, "parallel string matching algorithm using grid", "international journal of distributed and parallel systems" (ijdps) vol.3, no.3, 2012.
8. Leslie G. Valiant, A bridging model for parallel computation, Commun. ACM, volume 33, issue 8, August, 1990, pages 103—111
9. I. Foster. Designing and Building Parallel Programs. Addison Wesley, 1996.
10. I. Foster and S. Tuecke. Parallel Programming with PCN. Technical Report ANL-91/32, Argonne National Laboratory, Argonne, December 1991.
11. J. Darlinton, M. Ghanem, H. W. To (1993), "Structured Parallel Programming", *In Programming Models for Massively Parallel Computers. IEEE Computer Society Press. 1993*
12. N. Tuck, T. Sherwood, B. Calder, and G. Varghese. Deterministic Memory-Efficient String Matching Algorithms for Intrusion Detection. In *Proceedings of IEEE Infocom*, Hong Kong, March 2004.
13. S. Dharmapurikar, P. Krishnamurthy, T. S. Sproull, and J. W. Lockwood. Deep Packet Inspection using Parallel Bloom Filters. *IEEE Micro*, 24(1):52–61, 2004.
14. J.Nandhini, Dr.M. Nithya, Dr.S.Prabhakaran "Advance virus detection using combined techniques of pattern matching and dynamic instruction sequences", International Journal of Communication and Computer Technologies, Volume 01 – No.45, pp.156-161 2013
15. Safaa O. Al-Mamory et al., "String Matching Enhancement for Snort IDS", pp.1020-1023.
16. S. Dharmapurikar and J.W. Lockwood, "Fast and scalable pattern matching for network intrusion detection system", *IEEE JSAC*, Vol.24, No.10, pp.1781–1792, 2006.
17. N. Hua, H. Song and T.V. Lakshman, "Variable-stride multipattern matching for scalable deep packet inspection", in *Proc. of INFOCOM*, Rio de Janeiro, Brazil, pp.415–423, 2009.
18. H. Lu, K. Zheng, B. Liu, X. Zhang and Y. Liu, "A memoryefficient parallel string matching architecture for high-speed intrusion detection", *IEEE JSAC*, Vol.24, No.10, pp.1793–1804, 2006.
19. B.C. Brodie, D.E. Taylor and R.K. Cytron, "A scalable architecture for high-throughput regular-expression pattern matching", in *Proc. Of ACM/IEEE International Symposium on Computer Architecture (ISCA)*, Boston, MA, USA, pp.191–202, 2006.
20. D. Ficara, G. Antichi, A.D. Pietro, S. Giordano, G. Procissi and F. Vitucci, "Sampling techniques to accelerate pattern matching in network intrusion detection systems", in *Proc. of IEEE ICC*, Cape Town, South Africa, pp.1–5, 2010.
21. WANG Xiaofei, HU Chengchen, TANG Yi, ZHANG Ting, WU Chunming, LIU Bin and WANG Xiaojun, "Parallel Length-based Matching Architecture for High Throughput Multi-Pattern Matching", Chinese Journal of Electronics, Vol.21, No.3, pp.489-494, 2012.
22. HyunJin Kim and Seung-Woo Lee, "A Hardware-Based String Matching Using State Transition Compression for Deep Packet Inspection", *ETRI Journal*, Volume 35, Number 1, pp.154-157, 2013.
23. Akhtar Rasool and Nilay Khare, "Parallelization of KMP String Matching Algorithm on Different SIMD architectures-Multi-Core and GPGPU", *International Journal of Computer Applications*, pp.26-28, 2012.
24. cheng zhong and guo-liang chen, " a fast determinate string matching algorithm for the network intrusion detection systems", *Proceedings of the Sixth International Conference on Machine Learning and Cybernetics*, Hong Kong, pp.,3173-3177, 2007.
25. Panwei Cao and Suping Wu, "Parallel Research on KMP Algorithm", pp.4252-4255, IEEE, 2011.
26. Wei Lin, Bin Liu, "Pipelined Parallel AC-based Approach for Multi-String Matching", 14th IEEE International Conference on Parallel and Distributed Systems, pp. 665- 672, 2008.

27. Chuanpeng Chen and Zhongping Qin A Bit-split Byte-parallel String Matching Architecture, *IEEE*, pp.214- 217, 2009.
28. HyunJin Kim, Hyejeong Hong, Hong-Sik Kim, and Sungho Kang, "A Memory-Efficient Parallel String Matching for Intrusion Detection Systems", *IEEE COMMUNICATIONS LETTERS*, VOL. 13, NO. 12, pp. 1004-1006, 2009.
29. M. Alicherry, M. Muthuprasanna and V. Kumar, High speed pattern matching for network IDS/IPS, *IEEE ICNP*, pp.187-196, 2006.
30. D. Pao, W. Lin and B. Liu, A memory-efficient pipelined implementation of the Aho-Corasick string- matching algorithm, *ACM Trans. on Archit. Code Optim.*, vol.7, pp.1-27, 2010.
31. W. Lin and B. Liu, Pipelined parallel AC-based approach for multi-string matching, *IEEE ICPADS*, pp.665-672, 2008.
32. N. Hua, H. Song and T. V. Lakshman, Variable-stride multi-pattern matching for scalable deep packet inspection, *IEEE INFOCOM*, pp.415-423, 2009.
33. D. P. Scarpazza, O. Villa and F. Petrini, Exact multi-pattern string matching on the cell/b.e. processor, *ACM CF*, 2008.
34. Y. Sugawara, M. Inaba and K. Hiraki, Over 10Gbps string matching mechanism for multi-stream packet scanning systems, *Field Programmable Logic and Application*, vol.3203, pp.484-493, 2004.
35. HyunJin Kim et al., "A Memory-Efficient Bit-Split Parallel String Matching using Pattern Dividing for Intrusion Detection Systems", *IEEE Transactions On Parallel And Distributed Systems*, Third Draft, September 2010, pp.1-8,2011.
36. Yi-Hua E. Yang and Viktor K. Prasanna, "Robust and Scalable String Pattern Matching for Deep Packet Inspection on Multi-core Processors", *IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS*, pp.1-11,2012.
37. Yi Tang et al., "Independent Parallel Compact Finite Automata for Accelerating Multi-String Matching", *IEEE Globecom 2010 proceedings*, 2010.
38. KSMV Kumar, S. Viswanadha Raju and A. Govardhan, "Overlapped Text Partition Algorithm for Pattern Matching on Hypercube Networked Model", *GJCST*, pp. 1-8,2013.
39. Yao Xin , Benben Liu, Biao Min, WillX.Y.Li, Ray C.C. Cheung, Anthony S.Fong, Ting Fung Chan" Parallel architecture for DNA sequence inexact matching with Burrows-Wheeler Transform", *Microelectronics Journal*, pp. 670-682,Elsevier,2013.
40. Hoang Le, and Viktor K. Prasanna, "A Memory-Efficient and Modular Approach for Large-Scale String Pattern Matching", *IEEE Transactions on Computers*, VOL. 62, NO. 5, pp. 844-857, 2013.
41. TAN Jianlong et al., "Speeding up Pattern Matching by Optimal Partial String Extraction", the first international workshop on security in computers, networking and communications, pp.1030-1035, IEEE, 2011.
42. Rajesh Prasad, Anuj Kumar Sharma, Alok Singh, Suneeta Agarwal and Sanjay Misra, "Efficient bit-parallel multi-patterns approximate string matching algorithms", *Scientific Research and Essays Vol. 6(4)*, pp. 876-881, 18 February, 2011.
43. Mosleh M. Abu-Alhaj et al., "An Innovative Platform To Improve The Performance Of Exact-String Matching ALGORITHMS",2010, (IJCSIS) International Journal of Computer Science and Information Security,pp.280-283, Vol. 7, No. 1, 2010
44. Benedikt Forchhammer, Thorsten Papenbrock, Thomas Stening, Sven Viehmeier, Uwe Draisbach, Felix Naumann, "Duplicate Detection on GPUs", pp.165-188,2013
45. Antonino Tumeo and G et al., "Aho-Corasick String Matching on Shared and Sistributed Memory Parallel Architectures", *IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS*, pp. 1-9, IEEE, 2011.
46. Antonino Tumeo, Oreste Villa, and Daniel G. Chavarría-Miranda, "Aho-Corasick String Matching on Shared and Distributed-Memory Parallel Architectures", *IEEE Transactions on Parallel and Distributed Systems*, VOL. 23, NO. 3, PP.436-443, 2012.
47. Vinod.O, B.M.Sagar, "Hash-Based String Matching Algorithm For Network Intrusion Prevention systems (NIPS)", *International Journal of Advanced Computer Theory and Engineering*, Volume-2, Issue-2, pp.31-35, 2013.
48. Oren Ben-Kiki, Philip Bille, Dany Breslauer, Leszek Gasieniec, Roberto Grossi, Oren Weimann "Towards Optimal Packed String Matching", pp. 1-33, ---July 10, 2013.
49. H. D. Cheng and K. S. Fu, "VLSI Architecture for String Matching and Pattern Matching," *Pattern Recognition*, Vol.20, pp. 125-141, 1987.
50. Daniel P. Lopresti, "P-NAC : A Systolic Array for Comparing Nucleic Acid Sequences," *Computer*, pp. 98- 99, July 1987.
51. Amar Mukhopadhyay, "Hardware Algorithms for Nonnumeric Computation," *IEEE trans. on Computers*, Vol. C-28, No. 6, pp. 384-394, June 1979.
52. Bradly Fawcett, "Reconfiguring a Computing Pattern," *Electronic Engineering Times*, Manhasset, pp. 64- , April. 1995.
53. M. J. Foster and H. T. Kung, "The Design of Special-Purpose VLSI Chips," *Computer*, pp. 26-38, Jan. 1980.
54. Carla Correa Tavares dos Reis and Oswaldo Cruz, "Approximate String Matching Algorithm Using Parallel Methods for Molecular Sequence Comparisons", *IEEE*, pp.140-143,2005.

55. Muhammad Zubair et al., "Text Scanning approach for Exact String Matching", International Conference on Networking and Information Technology, pp.118-121,2010.
56. Tomohiro I , Shunsuke Inenaga, Masayuki Takeda "Palindrome pattern matching" Theoretical Computer Science, pp. 162-170,Elsevier,2013.



This page is intentionally left blank





Is Advanced Statistical Computing Technology a Clue in Applied Medicine? A Study using Data Mining as a Predictor Technology in Gastroenterology & Bariatric Surgery; Novel Elbanna Operations

By Abduh Elbanna & Abd Elrazek M. Aly Abd Elrazek

Al-Azhar University, Egypt

Abstract- Advanced computer science is very important in applied medicine, we have several publications regarding use of data mining computing technology in many clinical fields, here we will present to our experience in gastroenterology and surgery using data mining in obesity related gastrointestinal motility disorder; (IBS) and surgical bariatric approaches. Obesity is a chronic disease that is increasing in prevalence worldwide, data from United States and those from other countries are indicative of a major international epidemic, a steady and distressing increase worldwide, whatever numbers of large epidemiologic studies have evaluated the relationship between obesity and several Co- morbidities. Furthermore; obesity – related - morbidities are common diseases affecting population.

Keywords: *IBS, BMI, obesity, Elbanna, data mining.*

GJCST-C Classification : *H.2.8*



Strictly as per the compliance and regulations of:



Is Advanced Statistical Computing Technology a Clue in Applied Medicine? A Study using Data Mining as a Predictor Technology in Gastroenterology & Bariatric Surgery; Novel Elbanna Operations

Abduh Elbanna ^α & Abd Elrazek M. Aly Abd Elrazek ^σ

Abstract- Advanced computer science is very important in applied medicine, we have several publications regarding use of data mining computing technology in many clinical fields, here we will present to our experience in gastroenterology and surgery using data mining in obesity related gastrointestinal motility disorder; (IBS) and surgical bariatric approaches. Obesity is a chronic disease that is increasing in prevalence worldwide, data from United States and those from other countries are indicative of a major international epidemic, a steady and distressing increase worldwide, whatever numbers of large epidemiologic studies have evaluated the relationship between obesity and several co - morbidities. Furthermore; obesity - related - morbidities are common diseases affecting population. Irritable bowel syndrome (IBS) is a common gastrointestinal disorder with special concern for those with comorbid or morbid obesity, characterized by chronic or recurrent abdominal pain with constipation, diarrhea and/or an alternation of the two, and often bloating. Patients often use manipulation of diet as a mean of controlling symptoms. Adopting a healthy lifestyle, one that includes not smoking or drinking excess alcohol, eating right, daily physical activity, and a healthy weight, is associated with improving IBS-related morbidities. Bariatric surgery would be the clue therapy if diet and medical therapies fail. Novel Elbanna techniques are bariatric maneuvers successfully performed to those with morbid or comorbid obesity in more than 200 patients. Although data mining is very common applicable method in economy and global trade, it is not fully known by most of physicians and surgeons worldwide, for its limited use in applicable medicine. To our knowledge, there is no previous studies can evaluated such a relationship between IBS and BMI using Data mining computing analysis, also we used data mining to compare both Elbanna bariatric techniques, to look for the best leading factors in both Elbanna bariatric operations.

Keywords: IBS, BMI, obesity, Elbanna, data mining.

I. INTRODUCTION

Obesity impairs health related quality of life (HRQL) in adolescents. Obesity is increasing in prevalence worldwide, unfortunately many of

Author α: Bariatric and Laparoscopic surgery Department – Al-Azhar University – School of Medicine – AlHusain University hospital – Cairo – Egypt.

Author σ: Department of Hepatology & Gastroenterology, Al Azhar Faculty of Medicine, Al Azhar Asuit University Hospital, 712-572- Al-Azhar University, Asuit, Egypt. e-mail: ahmadrazek@gmail.com

obese and morbid obese individuals suffering irritable bowel syndrome (IBS) dominant constipation. Prevalence of obesity was 35.5 and 35.8 percent among adult American men and women respectively, interestingly many of those with high BMI may complaining of IBS- dominant constipation. Irritable bowel syndrome (IBS) is a serious health problem that affects an estimated 10–15% of people worldwide and has economic consequences in the United States of over \$30 billion annually [1]. IBS is characterized by abdominal pain or discomfort and is associated with changes in stool frequency and/or consistency. The etiopathogenesis of IBS may be multifactorial, as is the pathophysiology, which is attributed to alterations in gastrointestinal motility, visceral hypersensitivity, intestinal microbiota, gut epithelium and immune function, dysfunction of the brain-gut axis or certain psychosocial factors. Current therapeutic strategies are often unsatisfactory. There is now increasing evidence linking alterations in the gastrointestinal microbiota and IBS [2-7]. Furthermore weight reduction may improve the psycho-organic symptoms of chronic IBS [8-11]. Data mining is applicable in economics, business, global trade and biology with limitation use in clinical medicine. In their graphical representations, The advanced neural network is a data mining trying to mimic the human brain connecting attributes to each other aiming to compare these information-related attributes to one another, finally looking for the strongest connections, the neural mining nets are drawn using nodes (ganglia) and neurons (nerves). Using the routine graphical view, may be very difficult to be read also it may need a long time to be understood, which can often happen when there are plenty of numbers of examined attributes, the computer can read the network and apply the suitable required model to score the applicable data in order to make useful predictions and publishable results [12].

II. PATIENTS AND METHODS

We retrospectively reviewed the files of 2 groups of patients; first group (Gastrointestinal group);

was 35 Patients (16 male and 19 female) aged 13 to 83 years old, presented in the period from 1/9/2011 to 30/8/2012 with psychosomatic manifestations of Irritable Bowel Syndrome (IBS) in the form of Intermittent colicky abdominal pain, frequent changes in the bowel habit; constipation on top of chronic diarrhea or diarrhea on the top of chronic constipation, back ache, anxiety disorders and infrequent attacks of nausea and vomiting. Investigations were done to exclude other comorbidities e.g.: U/S, CT, Gastroscopy or even, full colonoscopy when IBD; (Crohn's, ulcerative), malignancy or other diseases suspected. Diagnosis of (IBS) was done by exclusion. All patients received both pharmacotherapy and psychotherapy; One by One psychoanalysis. BMI was calculated for each patient before dedicated therapy.

Second group (Obesity group) was 20 patients underwent bariatric surgery aged 22 to 39 years old, 13 Male and 7 female in the period from 20/11/2012 to 21/11/2013, using Elbanna operations; Elbanna intestinal bypass (for 10 patients) and Modified Elbanna; Intestinal bypass with fundal resection (for the other 10 patients).

III. STATISTICAL-ANALYSIS

10 folds cross validation using naive Bayes application was helpful to demonstrate our results specifically, a descriptive model was generated using a decision tree algorithms. The decision tree decided the most significant independent variable in each stage of predicting dependent variables. (Using the Rapid Miner, Rapid I, version 4.6, Berlin, Germany), in both gastrointestinal and the bariatric groups was a successfully application in clinical medicine.

IV. RESULTS

Following the 35 patients for 12 months duration, 18 individuals (12 male and 7 female) improved, whatever 17(8 male and 9 female) did not improve.

17 out of 18 patients, showed clinical improvement (94.4%), had normal BMI or slightly overweight, whatever 6 out of 17, did not improve (35.2%) were overweight, obese or morbid obese.

Regarding who underwent novel Elbanna operations, there was significant Excessive Weight Loss (EWL) in those underwent Modified Elbanna technique, than those underwent classic Elbanna operation (90 ± 10) versus (71 ± 9.2).

V. DISCUSSION

Data mining as a discipline is largely transparent to the world Most of the time, we never even notice it is happening. But whenever we sign up for a grocery store shopping card, place a purchase using a credit card, or surf a web site, we are creating data.

These data are stored in large sets on powerful computer systems owned by the companies we deal with every day. Lying within those data sets are patterns-indicators of our interests, our habits and our behaviour, indeed companies will know about customers distribution areas. In another point of view application of data mining in medicine could discover many of obscured factors led to several morbidities and mortalities according to our previous publications, furthermore using of data mining in medical fields is encouraged [13]. There are several well-established health hazards associated with obesity, including liver disease, Diabetes Mellitus, heart disease, cerebrovascular stroke, certain cancers, osteoarthritis, gastrointestinal motility disorders, sexual disorders and depression. Weight loss is encouraged in any mean to overcome morbidity and diseases – affecting survival [14,15]. IBS related-obesity is not uncommon problem especially in those with morbid obesity associated intestinal microbiota. The percentage of Americans with a BMI above 25 kg/m² or 30 kg/m² has been determined in several government surveys, beginning in 1960. In 2009 to 2010 the prevalence of obesity was 35.5 and 35.8 percent among adult American men and women, respectively. In Canada more than 27 percent of men and 23 percent of women are obese. Reported prevalence rates of obesity include 11 percent of men and 10 percent of women in Belgium (2002 to 2004), 23 percent of men and women in the UK (2009), 24 percent of men and 34 percent of women in Mexico (2006), 9 percent of men and 27 percent of women in South Africa (2003), 50.5% of boys and 46.5% of girls in Kuwait and 8 percent of men and 13 percent of women in Pakistan (1994) [16-19]. Diagram (1)

Diagram (1)

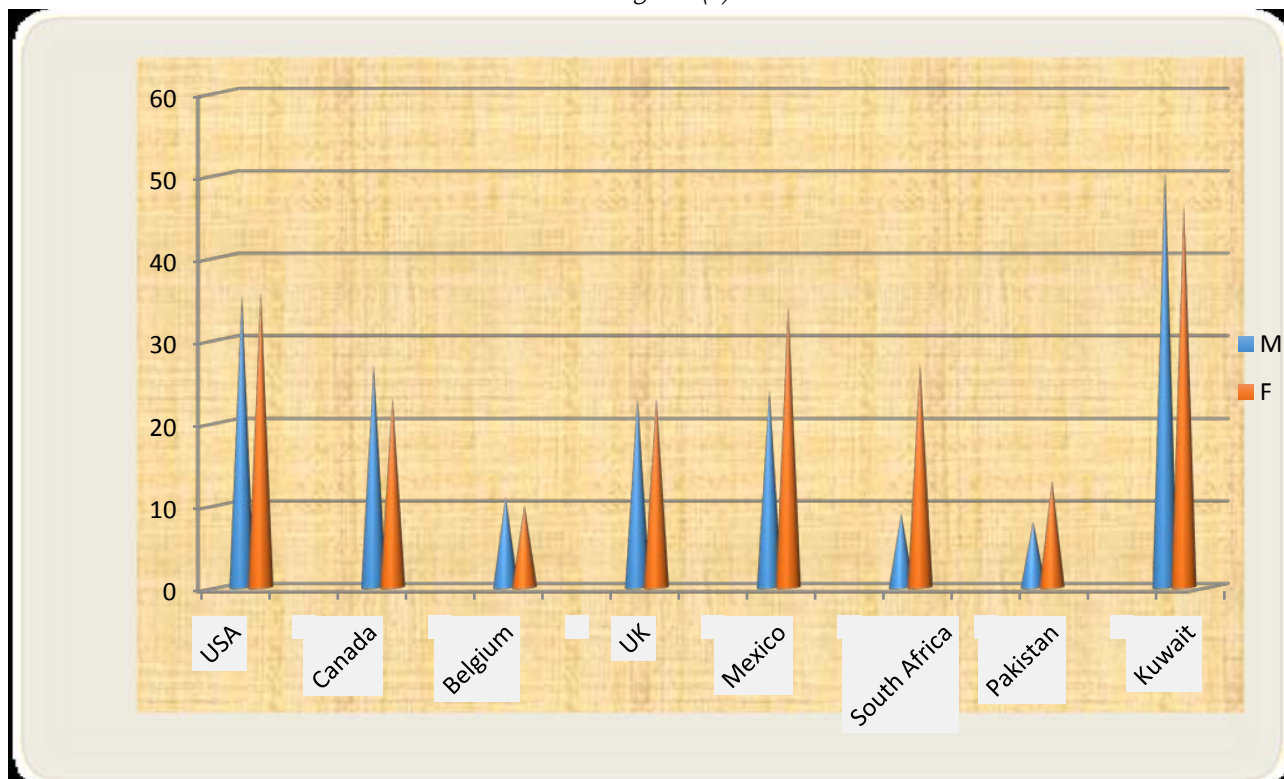


Diagram (1) showing the percentage of obesity in different populations worldwide, USA; Canada; & Mexico; are examples of North and South America; UK; & Belgium; are examples of Europe; Pakistan; & Kuwait; are examples of Asia; whatever South Africa is the only example of Africa due to lack of information-related obesity in different African countries. Note the average obesity of Female to Male in all countries 53% versus 47%.

For patients with BMI ≥ 40 kg/m² who have failed to lose weight with diet, exercise and drug therapy, and those with BMI >35 kg/m² with obesity-related comorbidities, bariatric surgery become the definitive clue, whatever the laparoscopic bariatric approach is preferred over the open approach, because of improved recovery and lower morbidity and is the most common approach used worldwide [21].

Irritable bowel syndrome (IBS) is a common gastrointestinal disorder worldwide with special concern to obese patients, characterized by chronic or recurrent abdominal pain with constipation, diarrhea and/or an alternation of the two, and often bloating. It is well recognized in many with high BMI, that ingestion of food is a trigger for functional bowel symptoms, particularly those with chronic irritable bowel syndrome (IBS). Occasionally chemicals fall to control the associated IBS-morbidities, whatever nutritional elements could improve the course of such chronic functional bowel disorder. Patients often use manipulation of diet as a means of controlling symptoms. Despite description of multiple dietary methods, few have scientific backing or quality evidence of efficacy [21,22].

Accordingly, obesity - related IBS morbidities should be considered. To our knowledge no previous studies could evaluate such a relationship of IBS and High BMI using Data mining computing analysis.

Patients who responded to the IBS-therapy were on controlled medications, some of them withdrew the medications completely. The gold standard for helping patients with IBS to lose weight is behavioral counseling providing advice on how to increase physical activity and reduce calorie intake in a high-intensity program with a qualified healthcare provider. According to our experience, patients presented with IBS, have shown sustain improvement in both their organic and psychogenic symptoms related-chronic IBS, when weight reduction is obtained. Our current study dedicating that psychogenic programs together with gaining ideal BMI significantly improve quality of life. Such as this can reliably produce and sustain IBS-free Symptoms.

Rapid I showed the association between overall outcome therapy and BMI, it was evaluated the positive and negative groups for the therapy, according to different related factors, examining those with BMI >25 (Figure.1).

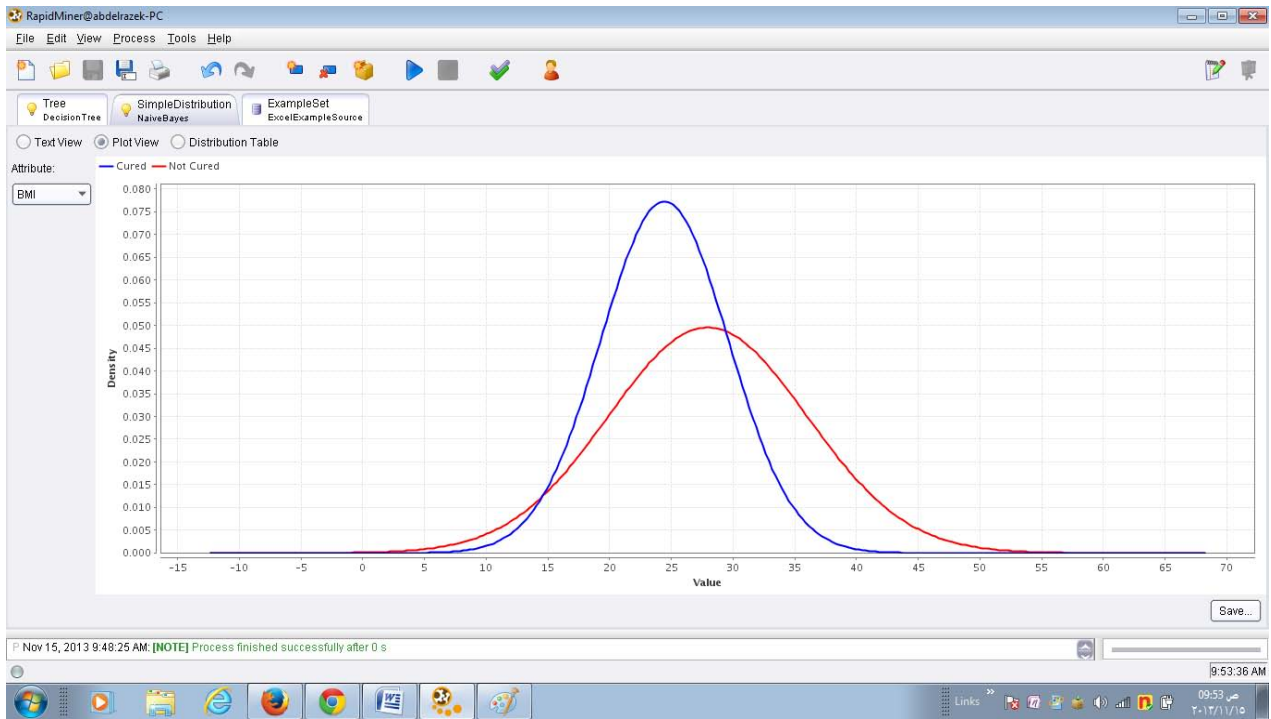


Figure 1 : showing improved group(Blue) and unimproved (Red), patients with IBS, in related to BMI. BMI; (Mean \pm SD) for improved group (24.44 ± 5.17), for unimproved (27.94 ± 8.05).

Using Rapid Miner version 4.6.Berlin, revealed a significant relation in those who improved to overall chemical, nutritional and psychotherapy; (94.4%), whatever (35.2%) of those did not improve due to overweight, obesity or morbid obesity and (23.5%) due

to obesity or morbid obesity, explaining insignificant relationship in non responding group.

The decision tree algorithm showed such a preliminary idea about IBS associated BMI (Figure 2).

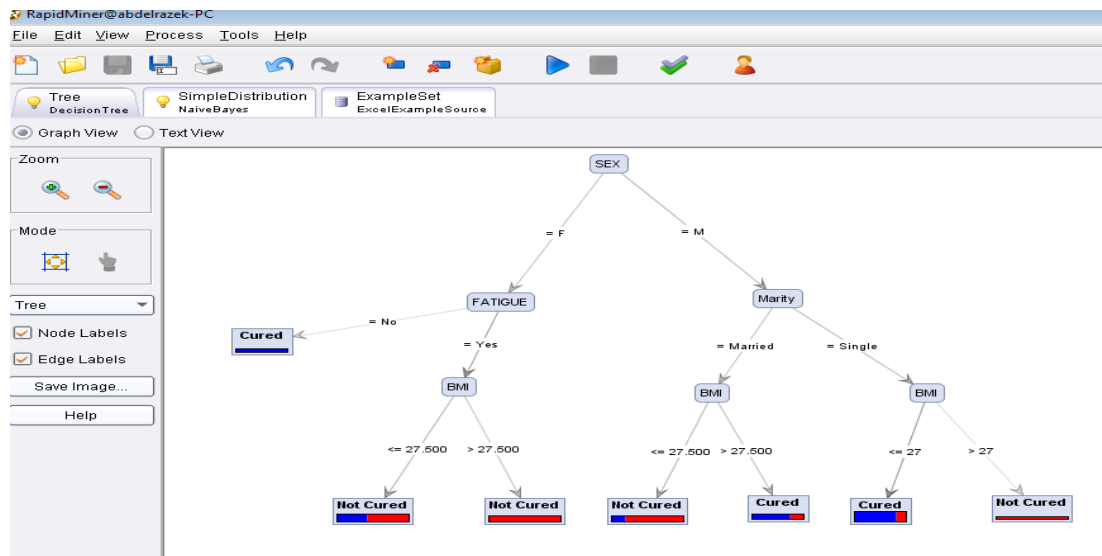


Figure 2 : Showing the Decision tree Algorithm Created by (Rapid I, Rapidminer Ver.4.6, Berlin, Germany). Note the Correlation Between IBS Outcome Therapy and Corresponding BMI, Marriage and Fatigue are Considerable Factors.

Unfortunately obesity became a worldwide stigma, microbiota that increase with obesity have a strong retaliation with IBS, currently obese subjects are often exposed to public disapproval because of their fatness affecting significantly their psychosocial

behavior, affecting the progressive course of chronic IBS in a psychosomatic way of explanation. According to our point of view, if diet, exercise or drugs failed to control morbid obesity or co- morbid obesity disorders, bariatric surgery should be considered aiming to

overcome obesity related morbidities including functional bowel disorder. A new trend of bariatric operation; Novel strategy using a good digestive & selective absorption technique has been evaluated; Modified intestinal bypass (Elbanna operation) or with fundal resection (Modified Elbanna), informed as a brief poster presentation or a talk discussion presentation in congresses of the international federation for the surgery

of obesity and metabolic disorders (IFSO) 2009 and 2013.

Elbanna techniques were experienced in more than 200 patients, in the period from 1998 till 2013, showed significant improvement in BMI related morbidities by avoiding vitamins and trace elements deficiency obtained followed other surgical diversion techniques (Figure 3).

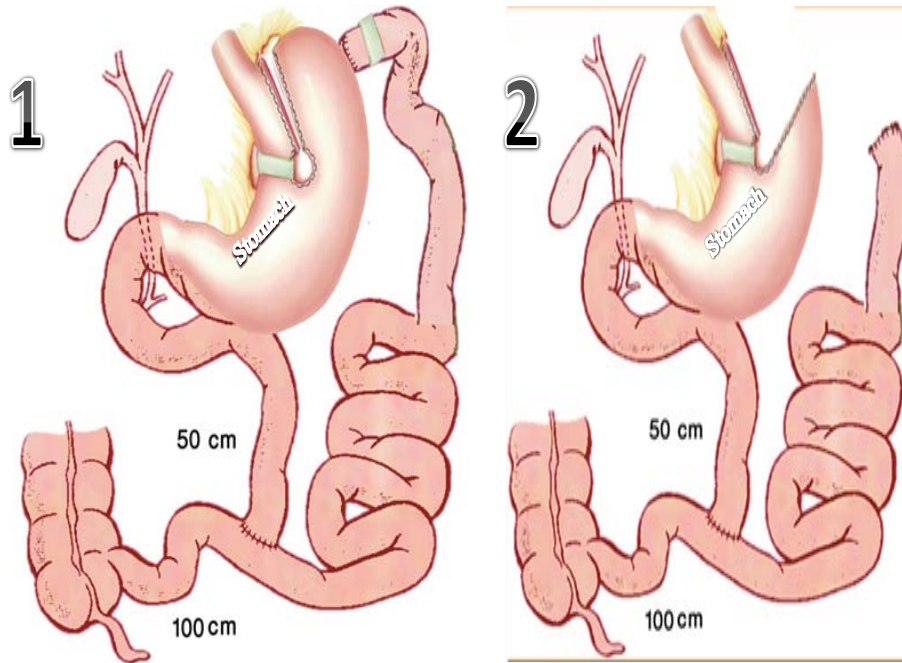


Figure 3 : 1-Elbanna operation; Modified Intestinal Bypass. 2- Modified Elbanna; with Fundal Resection. In Both Operations No nutritional, Vitamins or Minerals deficiencies.

Figure(3).Elbanna

Comparing two surgical techniques using the data mining computing analysis could find a significant EWL in the modified technique with significant overall success comparing to Elbanna operation, whatever

each operation has its indication according to each patient's condition (Figure 4). In another point of view, bariatric Elbanna techniques are better in young age, that EWL is significantly higher in young group (Figure5).

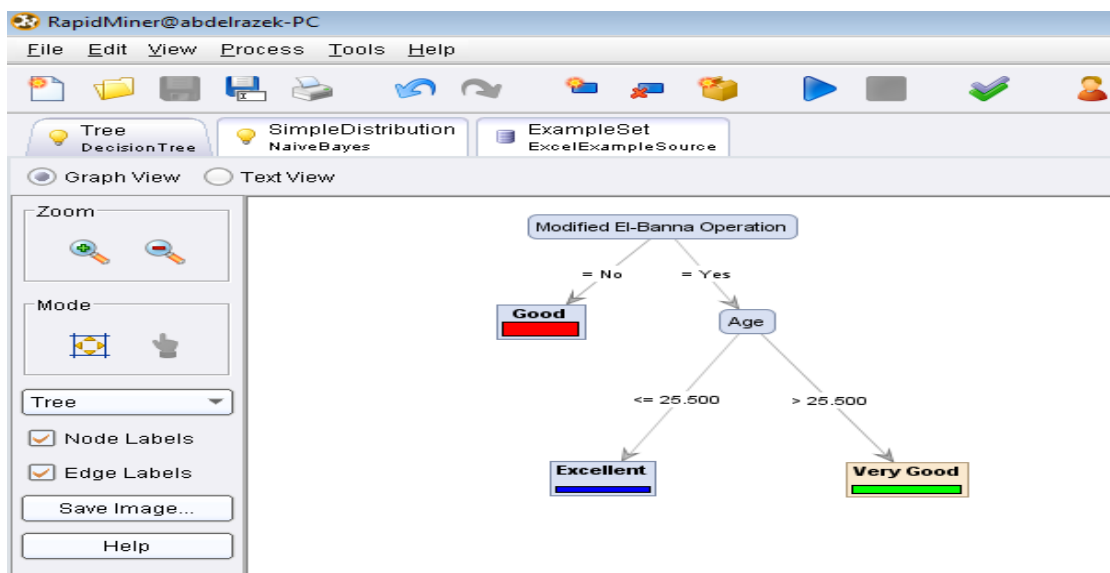


Figure 4 : Showing the decision tree algorithm for EWL in both El Banna techniques.

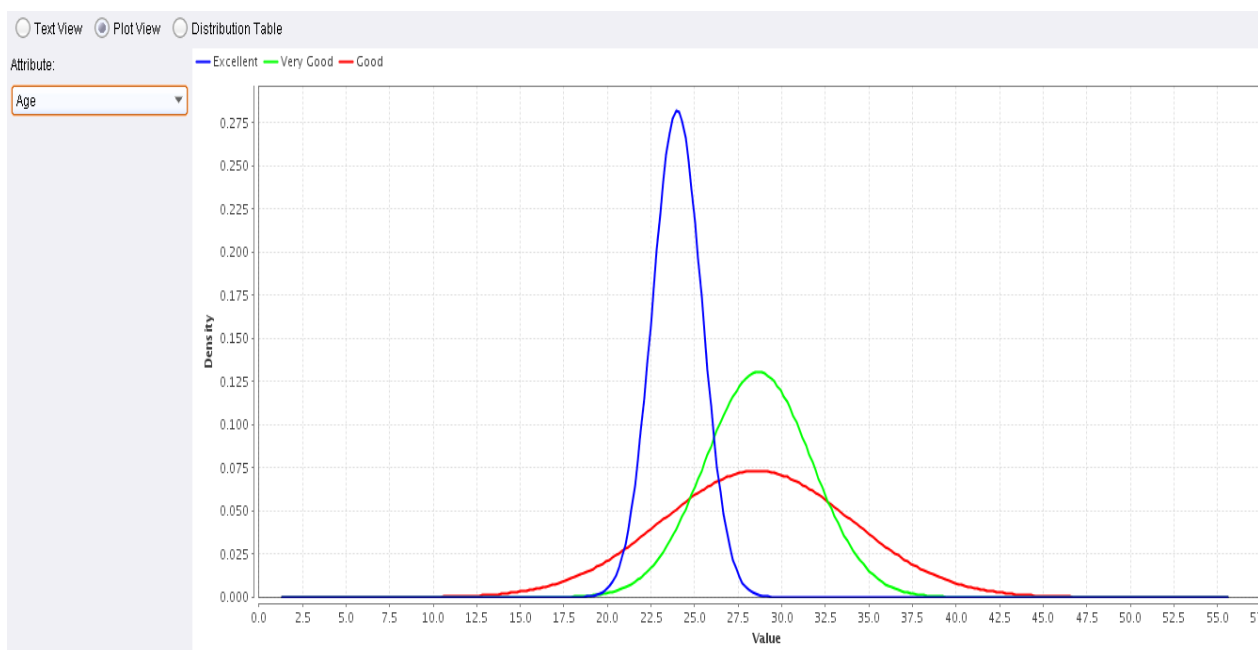


Figure 5 : Showing age group corresponding EWL inpatients underwent Elbanna operations; Young are better.

The message we have to sail for readers that to enjoy a healthy lifestyle, one that includes not smoking or drinking excess alcohol, eating right, daily physical activity, and a healthy weight and psychogenic stability is associated with improving quality of life. Every Effort should be gained to prevent obesity. Bariatric surgery would be the clue therapy if diet and medical therapies fail. Data mining computing analysis should be considered more and more in applied medicine, sure it is a clue in many obscured diseases we do not know. Using Data mining in applied medicine is of great important consideration, to predict factors leading to disease progression and even which may lead to associated morbidities such as high BMI and IBS-associated morbidities.

VI. STUDY LIMITATIONS AND FUTURE RECOMMENDATIONS

In the current research we just present to a preliminary study, whatever our clinical experience played a major role in assessing the information mentioned in our current study. Given a small sample size we recommend a second study involving more patients with multidisciplinary disorders, our results must be confirmed using more evidence-based criteria using more data mining applicable opportunities.

Our results might be changed according further innovations in Elbanna techniques, furthermore expected more data with large sample groups may significantly changed our results variations.

According to our point of view, Data mining is still unknown in many medical fields; whatever it has a great importance in business, economy, global trade negotiations, national labour relations and even a break

through in evolutionary biology and software engineering.

REFERENCES RÉFÉRENCES REFERENCIAS

1. Sethi S, Wadhwa V, Leclair J, Mikami S, Park R, Jones M In-et, al. patient discharge rates for the irritable bowel syndrome - an analysis of national trends in the United States from 1997 to 2010.
2. Dai C, Zheng CQ, Jiang M, Ma XY, Jiang LJ. Probiotics and irritable bowel syndrome. World J Gastroenterol. 2013 Sep 28.
3. Gilkin RJ. The spectrum of irritable bowel syndrome: A clinical review. Clin Ther. 2005; 27:1696-1709.
4. Longstreth GF, Thompson WG, Chey WD, Houghton LA, Mearin F, Spiller RC. Functional bowel disorders. Gastroenterology. 2006; 130:1480-1491.
5. Hammer J, Eslick GD, Howell SC, Altiparmak E, Talley NJ. Diagnostic yield of alarm features in irritable bowel syndrome and functional dyspepsia. Gut. 2004; 53:666-672.
6. Hussain Z, Quigley EM. Systematic review: Complementary and alternative medicine in the irritable bowel syndrome. Aliment Pharmacol Ther. 2006; 23:465-471.
7. Peckham EJ, Nelson EA, Greenhalgh J, Cooper K, Roberts ER, A Homeopathy for treatment of irritable bowel syndrome. Cochrane Database Syst Rev. 2013 Nov 13; 11:CD009710. [Epub ahead of print].
8. American Society for Metabolic and Bariatric Surgery (2009) Fact Sheet: Metabolic & Bariatric Surgery. Available online at: www.asbs.org/Newsite07/media/asbs_presskit.htm

9. Nguyen NT, Masoomi H, Magno CP, et al. Trends in use of bariatric surgery, 2003-2008. *J Am Coll Surg* 2011; 213:261.
10. Quartarone G. Role of PHGG as a dietary fiber: a review article. *Minerva Gastroenterol Dietol.* 2013 Dec;59(4):329-40.
11. Magge SS, Wolf JL. Complementary and alternative medicine and mind-body therapies for treatment of irritable bowel syndrome in women. *Womens Health (Lond Engl).* 2013 Nov; 9(6):557-67.
12. Zhang BH, Gao R, Li ZH, Li BS, Wang FY, Tang XD. [Treatment of irritable bowel syndrome by Chinese medicine and pharmacy: an analysis of data mining on experiences of experts]. *Zhongguo Zhong Xi Yi Jie He Za Zhi.* 2013 Jun; 33(6):757-60. [Article in Chinese].
13. Mathew North. *Data Mining for the masses, USA, 2008.*
14. Abd El Razek, Hamdy M Mahfouz, M. Afifi, Abd Elwahhab fathy, Khaled Abd elazeem, Ahmed El-Shamy, et al., Detection of Risky Esophageal varices by 2D U/S: when to perform Endoscopy. *Am J Med Sci.* 2012 Dec 21. [Epub ahead of print]. PMID: 23267234.
15. Abd Elrazek M. Aly Abd Elrazek, Hamdy M Mahfouz, Amro Metwally and Ahmed El-Shamy. Mortality Prediction of Non -Alcoholic patients presenting with Upper gastrointestinal bleeding using data mining. *European journal of Gastroenterology & Hepatology*, 1st of October 2013, [Epub ahead of print].
16. Abd Elrazek M Aly abd Elrazek & Hamdy M. Mahfouz. Prediction analysis of esophageal variceal degrees using data mining: Is validated in Clinical medicine? *Global Journal of computer science and technology*, 1-5 volume 13, ver.10, 2013.
17. Flegal KM, Carroll MD, Kit BK, Ogden CL. Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. *JAMA* 2012; 307:491.
18. Flegal KM, Carroll MD, Ogden CL, Curtin LR. Prevalence and trends in obesity among US adults, 1999-2008. *JAMA* 2010; 303:235.
19. Finucane MM, Stevens GA, Cowan MJ, et al. National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9•1 million participants. *Lancet* 2011; 377:557.
20. Flegal KM, Graubard BI, Williamson DF, Gail MH. Cause-specific excess deaths associated with underweight, overweight, and obesity. *JAMA* 2007; 298:2028.
21. Ambrosini GL, Oddy WH, Robinson M, et al. Adolescent dietary patterns are associated with lifestyle and family psycho-social factors. *Public Health Nutr* 2009; 12:1807.
22. Al-Haifi AR, Al-Fayez MA, Al-Athari BI, Al-Ajmi FA, Allafi AR, Al-Hazzaa HM, Musaiger AO. Relative contribution of physical activity, sedentary behaviors, and dietary habits to the prevalence of obesity among Kuwaiti adolescents. *Food Nutr Bull.* 2013 Mar; 34(1):6-13.
23. Snow V, Barry P, Fitterman N, et al. Pharmacologic and surgical management of obesity in primary care: a clinical practice guideline from the American College of Physicians. *Ann Intern Med* 2005; 142:525.
24. Eaton DK, Kann L, Kinchen S, et al. Youth risk behavior surveillance - United States, 2011. *MMWR Surveill Summ* 2012; 61:1.
25. Stang J, Story MT, Harnack L, Neumark-Sztainer D. Relationships between vitamin and mineral supplement use, dietary intake, and dietary adequacy among adolescents. *J Am Diet Assoc*
26. <http://link.springer.com/article/10.1007/s11695-013-0986-z>.

GLOBAL JOURNALS INC. (US) GUIDELINES HANDBOOK 2013

WWW.GLOBALJOURNALS.ORG

FELLOWS

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.

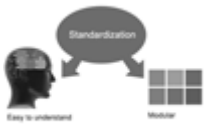




Journals Research
inducing researches

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 benefits can 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.

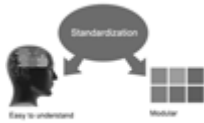
The Board can also play vital role by exploring and giving valuable suggestions regarding the Standards of “Open Association of Research Society, U.S.A (OARS)” so that proper amendment can take place for the benefit of entire research community. We shall provide details of particular standard only on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



Journals Research
inducing researches

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 brevity.

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 $1.4 \times 10^{-3} \text{ m}^3$, or 4 mm somewhat than $4 \times 10^{-3} \text{ m}$. 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 through 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.



ADMINISTRATION RULES LISTED BEFORE
SUBMITTING YOUR RESEARCH PAPER TO GLOBAL JOURNALS INC. (US)

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 perceptives 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
<i>Abstract</i>	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
<i>Introduction</i>	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
<i>Methods and Procedures</i>	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
<i>Result</i>	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
<i>Discussion</i>	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
<i>References</i>	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring



INDEX

A

Adenine · 23
Amplitude · 8
Apriori · 1, 3

C

Cerebrovascular · 35
Chuanpeng · 13, 31
Comorbidities · 34, 36
Constipation · 32, 34, 36
Corasick · 12, 13, 17, 31
Cytosine · 23

E

Enormous · 1, 5
Etiopathogenesis · 32

H

Hierarchical · 16, 21

L

Laparoscopic · 36
Larynx · 8
Lexicographically · 3

M

Morbidities · 32, 35, 36, 37, 38, 39, 40

O

Osteoarthritis · 35

P

Palindrome · 19
Pharmacotherapy · 34
Psychotherapy · 34, 38

T

Tomohiro · 19

Z

Zhongping · 13, 31



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