



‘येथे बहुतांचे हित ।’

Marathwada Mitramandal's  
**COLLEGE OF ENGINEERING**

S.No.18, Plot No.5/3, Karvenagar, Pune-411 052

*Accredited with 'A' Grade by NAAC*

*Accredited by NBA (Electrical and Mechanical Engg. Department)*

*Recipient of 'Best College Award 2019' of SPPU*

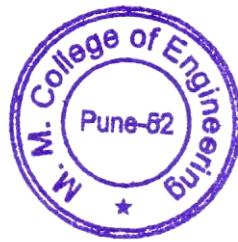
*Recognized under section 2(f) and 12B of UGC Act 1956*

## Criterion 3

### 3.3: Research Publication and Awards

3.3.2 Number of books and chapters in edited volumes/books published and papers published in national/ international conference proceedings per teacher during last five years

Sr. No.	Parameter	Academic Year	No. of Books and chapters in edited volumes / Books Published / Conference proceedings
1	Books and chapters in edited volumes / Books Published / Conference proceedings	2021-22	19
2		2020-21	20
3		2019-20	11
4		2018-19	24
5		2017-18	12





Marathwada Mitramandal's  
**COLLEGE OF ENGINEERING**

S.No.18, Plot No.5/3, Karvenagar, Pune-411 052

*Accredited with 'A' Grade by NAAC*

*Accredited by NBA (Electrical and Mechanical Engg. Department)*

*Recipient of 'Best College Award 2019' of SPPU*

*Recognized under section 2(f) and 12B of UGC Act 1956*

***Academic Year***  
**2018-19**

<b>Sr. No.</b>	<b>Title of Books and chapters in edited volumes / Books Published / Conference proceedings</b>	<b>Page No.</b>
1	<u>Detection of Suspicious Transaction with Database Forensics and Theory of Evidence</u>	4
2	<u>Parallel Crawling for detection and removal of Dust using Duster</u>	5
3	<u>Proposed Battlefield Simulator using GPU</u>	6
4	<u>De-duplication Approach with Enhance Security for Integrity</u>	7
5	<u>Hadoop in Action: Building a Generic Log Analyzing System</u>	8
6	<u>RAM Forensics: The Analysis and Extraction of Malicious Processes from Memory Image Using GUI Based Memory Forensic Toolkit</u>	9
7	<u>Tweet Classification with Convolutional Neural Network</u>	10
8	<u>Monitoring suspicious transactions in a financial transaction system through Database forensics audit</u>	11
9	<u>Content Centric Networks (CCN): A Survey</u>	12
10	<u>Ambiguity Resolution in English Language for Sentiment Analysis</u>	13
11	<u>Priority Based Sentiment Analysis for quick response to citizen complaints</u>	14
12	<u>Enhancement in Security Using Extended Security Techniques</u>	15
13	<u>Secure Dynamic Multi Owner System using Share generation scheme</u>	16
14	<u>Soil Toxicity Prediction and Recommendation System using Data Mining in Precision Agriculture</u>	17
15	<u>BMWA: A Novel Model for Behavior Mapping for Wormhole Adversary Node in MANET</u>	18
16	<u>Context Aware Computing: A survey</u>	19
17	<u>Security and Privacy Issues in Cloud Computing</u>	20
18	<u>Adulteration Detection in Petroleum Liquids using Stacked Multi Ring Resonator ( Received Best Paper Award IEEE Conference)</u>	21
19	<u>Review on Computer Aided Detection Systems of Breast Cancer</u>	22
20	<u>Discrete Structure(IT)</u>	23



Marathwada Mitramandal's  
**COLLEGE OF ENGINEERING**

S.No.18, Plot No.5/3, Karvenagar, Pune-411 052

*Accredited with 'A' Grade by NAAC*

*Accredited by NBA (Electrical and Mechanical Engg. Department)*

*Recipient of 'Best College Award 2019' of SPPU*

*Recognized under section 2(f) and 12B of UGC Act 1956*

'येथे बहुतांचे हित ।'

*Academic Year*  
**2018-19**

<b>Sr. No.</b>	<b>Title of Books and chapters in edited volumes / Books Published / Conference proceedings</b>	<b>Page No.</b>
21	<u>Software Testing Quality Assurance(IT)</u>	24
22	<u>Software Testing Quality Assurance(IT)</u>	25
23	<u>Software Testing Quality Assurance(Comp)</u>	26
24	<u>Implementing a hybrid of efficient algorithms for mining Top-K high utility item set</u>	27



[International Symposium on Security in Computing and Communication](#)

↳ SSCC 2018: [Security in Computing and Communications](#) pp 419–430 | [Cite as](#)

# Detection of Suspicious Transactions with Database Forensics and Theory of Evidence

[Harmeet Kaur Khanuja](#)  & [Dattatraya Adane](#) 

Conference paper | [First Online: 24 January 2019](#)

1231 Accesses

Part of the [Communications in Computer and Information Science](#) book series (CCIS, volume 969)

## Abstract

The aim of enabling the use of illegally obtained money for legal purposes, while hiding the true source of the funds from government authorities has given rise to suspicious transactions. Illegal transactions are detected using data mining and statistical techniques with the input data like various suspicious reports or the data set of all transactions within a financial institution. The output obtained is the set of highly suspicious transactions or highly suspicious entities (e.g., persons, organizations, or accounts). In this paper, we propose a database forensics methodology to monitor database transactions through audit logs. The Rule-based Bayesian Classification algorithm is applied to determine undetected illegal transactions and predicting initial belief of the transactions to be suspicious. Dempster-Shafer's theory of evidence is applied to combine different parameters of the transactions obtained through audit logs to verify the uncertainty and risk level of the suspected transactions. Thus a framework is designed and developed which can be used as a tool for the digital investigators.

## Keywords

Database forensics

Money laundering

Audit logs

Suspicious transactions

Outliers

Dempster Shafer theory



# Parallel Crawling for Detection and Removal of DUST using DUSTER

Jyoti G. Langhi

Department of Computer Engineering

Marathwada Mitra Mandal's

College of Engineering, Pune-52

jyotilanghi@mmcoe.edu.in

Prof Shailaja Jadhav

Department of Computer Engineering

Marathwada Mitra Mandal's

College of Engineering, Pune-52

jadhavsb@mmcoe.edu.in

**Abstract**—Web is commonly used medium to search information using Web Crawler. Web crawler fetches different pages related to given keyword but some of them contains duplicate content. Different URLs with similar text are DUST. To improve performance of search engine, DUSTER method is used. DUSTER detects and removes duplicate URLs without fetching their contents. Single crawler crawls single URL at a time. Multiple URLs are crawled parallelly by Parallel crawlers and the results of parallel crawlers are combined and given as an input to the DUSTER. Multiple sequence alignment is used to generate candidate rules and rules of validation. Then the candidate rules filtered out according to their performance in a validation set and finally removes the duplicate URLs. Using this method reduction of large number of duplicate URLs is achieved.

**Keywords**—Crawler, Parallel Crawling, DUSTER.

## I. INTRODUCTION

A Web crawler fetches data from various servers. Gathering data from various sources around the world takes huge amount of time. Such a single process faces problems on the processing power of a single machine and one network connection. If the workload of crawling Web pages is distributed, the job can be performed faster. Many search engines run multiple processes in parallel.

On the web there are different URLs that fetches the same page. These similar URLs are known as DUST. Duplicate URLs occur because of many reasons. DUST detection is important task for search engine because Crawling these duplicate URLs is a waste of resources. This results in the poor user experience. The existing system focused on document content to remove Duplicate URLs. Generation of Dynamic web pages leads to Duplication of contents. DUSTER converts duplicate URLs into same canonical form which can be used by web crawlers to avoid DUST. Instead of processing all URLs the existing system uses random sampling. In DUSTER framework, multiple sequence alignment is used to obtain a general and smaller set of rules and to avoid duplicate URLs. Multiple sequence alignment can be used to identify similar strings, so that normalization rules can be derived. More general rules can be generated using multiple sequence alignment algorithm to remove the duplicate URLs with similar text.

To fetch the URLs from the web a crawler is used in an existing system. More than one crawler can be used in distributed web crawling. Each crawler in a system acts as

separate entity and does its own indexing. Distributed system can process a growing workload as we distribute the resources in the system. Data fetched by single crawler go by single physical link. If crawling process is distributed in several processes then it makes easy to build scalable system[3].

## II. LITERATURE SURVEY

DUST can be detected using two methods. First method is content based method and another one is URL based method. Content based method fetches the whole page and full content is inspected by comparing it using syntactic or semantic evidence. In URL based method, without examining the content of the page the duplicate URLs can be found out.

In the following paragraphs some URL-based methods are focused on.

In base paper [2] the DUSTER framework is proposed. DUSTER detects duplicate URLs and removes them. This method uses normalization rules that convert distinct URLs which refer to same content to a common canonical form. Normalization rules are generated to convert all duplicate URLs into same canonical form. This makes easy to detect them. The scalability and precision can be improved using other data sets.

S. Bal and G. Geetha [3] proposed a smart distributed web crawler. In this paper the authors suggested that use of distributed crawler is faster than that of single crawler. Distributed crawler is used to improve the scalability.

The work done by A. Agarwal and other authors [4] focuses that the basic and deep tokenization of URLs to extract all possible tokens from URLs which are mined by Rule generation techniques proposed by them for generating normalization Rules. Proposed system implements for giving output to the user efficiently and large-scale deduplication of documents. Short Web pages does not work well and does not find out noise ratio on web pages.

A new technique SizeSpotSigs [5] is used for effective near duplicate detection algorithm considering the size of page content in mining. Proposed system implements noise-content ratio to work better. The disadvantage of this technique is that the size of the core content of Web page does not automatically or approximately decided.



# Proposed Battlefield Simulator Using GPU

S. M. Chaware<sup>1</sup>

Department of Computer Engineering  
MMCOE, Pune, India  
+91 7738005902

sandeepchaware@mmcoe.edu.in

Kiran Joshi<sup>3</sup>

Computer Engineer  
MMCOE, Pune, India  
+91 7798681287

joshikiran2537@gmail.com

Omkar Udawant<sup>2</sup>

Computer Engineer  
MMCOE, Pune, India  
+91 8007791921

omkar80077@gmail.com

Tejas Deshpande<sup>4</sup>

Computer Engineer  
MMCOE, Pune, India  
+91 8412866983

tejasdeshpande619@gmail.com

## ABSTRACT

Battlefield is an area where you cannot predict the attacking situation from an opposition. The situation may become worse when the enemy tankers may attack from various position and we will not enough get chance to think about our security. If by any mean we can analysis the situation of battling, we can easily decide the attacking strategy against any attack. This entire environment may simulate through a simulator where we can decide to attack and defend ourselves.

In this paper, we had proposed a battlefield simulator which helps in eliminating manual efforts of artillery testing and the demonstration cost required for the same. This simulator takes parameters such as type of artillery to be tested, environmental conditions and strategic planning. Damage caused by the artillery is calculated using physics formulae designed for achieving actual results. We had compared the situation with CPU and GPU processor and found that GPU is must faster than CPU and gives more accuracy.

## CCS Concepts

Computing methodologies→Modeling and simulation→  
Simulation types and techniques→Massively parallel and  
high-performance simulations

## Keywords

Artillery shot; Concurrent computing; Military vehicles; Missiles;  
Optimization; Projectile; Simulation; Visual technologies

## 1. INTRODUCTION

The traditional way of artillery testing requires lot of human intervention through which it is only possible to test various artilleries in various conditions before actually buying it.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

ICBRA '18, December 27–29, 2018, Hong Kong, Hong Kong

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-6611-3/18/12...\$15.00

DOI: <https://doi.org/10.1145/3309129.3309131>

Comparing with the existing systems available in market there is no such simulator which provides projectile trajectory simulation along with actual damage calculation.

This simulator also provides actual damage calculation based on real time scenarios with actual intervening parameters. All these computing of mathematical formulation is done on GPU to achieve parallelism and using OpenCL we provide platform independence.

At the end of simulation, a report is generated which provides war strategies to counter assault enemy situations and preciseness of the projectile trajectory with some recommendations. In this system, we provide projectile trajectory which helps in analyzing the actual path being travelled by the missile after launching from barrel until it hits the target.

## 2. PROJECTILE TRAJECTORY

The trajectory of projectile motion is built in the starting coordinate system  $Oxyz$  related with the point of gun position and oriented to shot direction shown in Figure 1. The coordinates of projectile mass centre are defined by the following equations [4]:

$$\frac{dx}{dt} = V_k \cos\theta \cos\varphi \quad (1)$$

$$\frac{dy}{dt} = V_k \sin\theta \quad (2)$$

$$\frac{dz}{dt} = -V_k \cos\theta \sin\varphi \quad (3)$$

Equation (1), (2), (3) gives  $x$ ,  $y$ ,  $z$  coordinates respectively of the missile. These coordinates change with respect to time as the tank shell is in motion.

Here  $x$  is distance in shot plane;  $y$  is height of projectile flight;  $z$  is azimuth deviation;  $\theta$  is angle of trajectory inclination;  $\varphi$  is angle of shot direction;  $V_k$  is velocity of projectile mass centre.

Parameters of the projectile motion are defined in trajectory coordinate system  $Ox_k y_k z_k$  related with the projectile mass centre and oriented to velocity vector shown in Figure 1

$$\frac{dv_k}{dt} = -g \sin\theta - \frac{c_{xk} q S W}{m} \quad (4)$$

Equation (4) is used for calculating the velocity of the tank shell. The velocity varies with respect to time; therefore, we require this differential equation to calculate the velocity at different points in the projectile.



# De-duplication Approach with Enhance Security for Integrity

Minal Bharat Pokale  
Computer Engineering Department  
MMCOE, Karve Nagar  
Pune, India  
minalbpokale09@gmail.com

Dr. Sandeep M. Chaware  
Computer Engineering Department  
MMCOE, Karve Nagar  
Pune, India  
sandeepchaware@mmcoe.edu.in

**Abstract**—Cloud computing becomes one of the important part of the IT-enterprise. It delivers computing application, software, hardware and computing resources to the user. So many user gets connected to the cloud because of the facility it provided. In this, user pays only for the services it uses. Now a days, vast amount of information gets stored over cloud, which include personal file, images, pdf, text, multimedia data etc. So security becomes one of the important issue. Everyone just want to upload, store its data without worrying about the security. IT organization should look after to provide security to the document. Documents content get altered by some outsider attacker, so it is necessary to verify the integrity. This data get duplicated and unnecessary much of the space will get waste. So to avoid this, data must be de-duplicated and uploaded over the cloud server. For this secure hash algorithm is used. It calculates the hash value for chunk of data and store over cloud. Only reducing unnecessary wastage of space is not the solution administrator should look toward the security issue. TPA is used for verification of integrity of the document. This system is proposed for achieving this both issues.

**Keywords**—Cloud Server, Hash Table, Third Party Auditor, Secure Hash Algorithm, Chunk, Private Server, Encryption Algorithm.

## I. INTRODUCTION

Cloud Computing is one of the fastest growing technology. Previously IT enterprise, organization kept their data over physical disk, hard drive. They can kept their data over own servers. Now a days, the cloud and virtualization makes it possible that data reside in physical infrastructure but it is under the logical control of organization. As outsider use, owns and control the resources, services, how can be the organization gives the assurance about the security, privacy of data. Following factors needs to be consider, in order to provide the security to the cloud:

- Government, organization protect the confidentiality of data.
- Consider the recycling of data on disk and erase the unused data.
- Auditing for giving assurance about the content of data. To prevent data from the insider attacker.

Such factor makes organization, IT infrastructure to look after the security issue seriously. Everyone wish to use the cloud because of its low cost, pay per use, location independence facility. When we consider the cloud security, different threat needs to be consider in order to provide the security to the data. We know that cloud offers SaaS, PaaS,

IaaS. In SaaS, the running of application is done e.g. salesforce.com. PaaS, it provides environment for development e.g. Microsoft Windows Azure. IaaS, it provides the basic computing resources, storage, network e.g. Amazon Elastic Compute Cloud (EC2). Security risk is come when consider about facility.

As virtualization makes user to use, upload or download his data anywhere anytime that store over cloud. Data duplication increases that's why much of the space will get waste. De-duplication makes it possible to reduce the space and make the space available to store another document, file, images etc. Only one unique copy get store over cloud, then the next task is security. Organization, IT infrastructure, work continuously to solve the problem for providing security to the data store over cloud. Organization, It infrastructure place some security policy according which, it can provide access to the authorize user only. Cloud provide so many techniques for data protection in the cloud.

Auditing is one of the techniques that admin used for checking integrity of the data. In Third Party Auditing include the third party but which is trusted server. This audit the file, data and document as data gets stored over the cloud dynamically. In providing security, first is to make the data non-readable, second comes the access policies so that the data can be access by the authorize user only, third comes auditing of data to check for the integrity.

### A. De-duplication

It becomes one of the interesting term in IT infrastructure. De-duplication is used to reduce the space so that more space gets available for storing another documents. Cloud provides many techniques for data de-duplication. Various version of message digest are used. It generates the hash value for identifying the data uniquely and for de-duplication of the data. This algorithm have some vulnerabilities that not resistant to collision. That are overcome by the hash algorithm. Hash algorithm have various version. In this, if data gets changed then the respective hash value for the file will also get change.



# Hadoop in Action: Building a Generic Log Analyzing System

Vinayak Bhosale, Anuja Thakar, Chinmaya Pandit, Anirudha Deshpande, Harmeet Khanuja

MMCOE, Computer Dept.  
Pune, India

vinayakbhosale.comp@mmcoe.edu.in, anujathakar.comp@mmcoe.edu.in, chinmayapandit.comp@mmcoe.edu.in,  
anirudhadeshpande.comp@mmcoe.edu.in

**Abstract**—The massive amount of structured, semi-structured and unstructured data can be attributed as Big Data. Traditional data processing applications and on-hand database system tools are impotent to evaluate and process such data. When the traditional systems were invented in the beginning, we never anticipated that we would have to deal with such monstrous amount of data, which would be difficult to process due to its characteristics viz. high volume, velocity, variety and veracity. The count of internet users has increased tremendously by virtue of which gargantuan amount of multifarious data gets generated. Moreover, the advent of IoT, has resulted in a boom in data generation. Thus, there was a pressing need to build a platform/framework which could process such huge, multifaceted data efficiently. That is where Hadoop came into picture. Hadoop facilitates scalable and distributed processing of Big Data in a proficient manner, saving a lot of time. Our work sheds some light on processing gigantic log data using Apache Hadoop open source framework. The data warehousing package built over Hadoop, Apache Hive, is used to summarize, analyze and query the different types of logs. Lastly, Apache Zeppelin, a powerful multipurpose notebook environment, is used for analysis, visualization and collaboration of log data in the form of bar graphs, pie charts, scatter charts, line graphs and area graphs.

**Keywords**—Big Data, Hadoop, Log Analysis, Network Security.

## I. INTRODUCTION

The events occurring in a system or a network are documented in the log files. For example, the requests made on a server are listed in the corresponding log files. Log data is in the unstructured format. This data if analyzed efficiently can help organizations in making decisions, gaining business insights, mitigating different risks, forensics and intrusion detection. However, we need to pre-treat this diverse and voluminous log data [1] to remove all noisy and irrelevant data and to resolve the inconsistencies. Log data preprocessing is the most significant step in the cleaning stage of the log analysis process, without which one cannot expect to mine knowledge from it. We then try to provide some structure to it, by storing it into Hive tables and then perform analysis using HQL (Hive Query Language) statements [5]. Fig. 1 shows the log analysis process:



Fig. 1. Log analysis process.

## II. TOOLS INVOLVED

1) *Hadoop*: Apache Hadoop is a collection of services that opens the door for using a combination of computers [4] to solve problems involving large amounts of data. It is used for processing gigantic data in a parallel and distributed manner.

2) *Hive*: Apache Hive provides a higher layer of abstraction over MapReduce. It facilitates reading and writing large datasets in a distributed environment using HQL which is similar to the SQL (Structured Query Language).

3) *Derby*: Apache Derby is the default database of Hive.

4) *Wireshark*: Wireshark is a cross platform, open source packet analyzer which is used for network troubleshooting.

5) *Nmap*: It is an open source tool that sends specially formulated packets to the target machine and then examine the responses [10].

6) *Zeppelin*: Apache Zeppelin is a multi-purpose web-based notebook which endows collective data analysis and visualization [1].

## III. ALGORITHM

Fig. 2 shows the sequence of tasks for the proposed methodology and the algorithm for the proposed method is explained in this section.



# RAM Forensics: The Analysis and Extraction of Malicious processes from memory Image using GUI based Memory Forensic Toolkit

Mr. Vivek Ravindra Sali  
Dept. of Computer Engineering  
Marathwada Mitra Mandal's  
College of Engineering  
Karvenagar, Pune  
vivekravindrasali@gmail.com

Mrs. H.K.Khanuja  
Dept. of Computer Engineering  
Marathwada Mitra Mandal's  
College of Engineering  
Karvenagar, Pune  
harmeetkaurkhanuja@mmcoe.edu.i

n

*Abstract*— In today's world the use of internet and information technology has grown up very rapidly. Due to increasing use of Internet the amount of cyber crimes have been increased. Hence it's become a very challenging task for the cyber crime investigator to not only find out the root cause of the crime but also to prove it correctly in the court of law. Computer Forensics is the science of investigating the computer system to obtain the digital evidences to find out the root cause of cyber crimes. Memory forensics is one of the branches of the Computer Forensics. The present techniques of memory forensics like Live Response and Memory Imaging, used by investigators during analysis and seizure operations involves either carrying the live analysis of volatile memory(RAM) of victimized computer system or by making the image of the RAM of suspect as machine and performing post analysis on different machine. In this paper Memory imaging approach of RAM analysis is used to find out the malicious processes using the GUI based tool that can analyze the volatile memory artifacts those are affected by malwares. The architecture of extracting the malicious processes is mentioned.

*Index Terms*— Digital investigation, digital evidence, GUI Framework, computer forensics, volatile memory dump, Live Response, YARA Scanner.

## I. INTRODUCTION

The computing resources and Internet play a significant role as vital business tool to provide the necessary information to an individual. Due to massive use of the Internet, cyber crimes have been increased. Cyber crime is any illegal activity which involves a computer system or its related systems or their applications. Today solving any cyber crime put up new challenges for a digital forensics investigator [5]. Digital forensics is the process of uncovering and interpreting an electronic data. The goal of investigation is to preserve the evidence that is obtained during an investigation process. This evidence is termed as digital evidence which must be preserved to reconstruct the past events. The analysis of volatile memory plays a very

significant role in a process of digital investigation process. The volatile memory contains many important artifacts which can be used in forensic investigation process. The information may contain passwords, event logs, cryptographic keys, process information and other vital data related to number of processes running in a system[2][8]. The collection of volatile data from a victimized computer system under investigation can be done using a conventional approach known as Live Response approach. In this approach the investigator first establishes a trusted command shell to acquire the data for investigation process. Volatile memory analysis using a Live Response method helps to collect all relevant evidences from a system. These evidences can be used to prove any incident occurred that might have compromised a system resulting into a cyber crime[2]. Another method to analyze a volatile memory is to perform memory image analysis. The analysis of a volatile memory is performed by capturing an image of RAM known as memory dump. Digital forensics contains the collection, validation, analysis, interpretation, documentation and presentation of the digital evidences[15]. Digital Forensics investigator make use of forensics tools in an investigation process, which are present in commercial and open domains. Depending upon the requirement of analysis, forensic toolkits are categorized like file system and data analysis tools, memory analysis tools, disk analysis tools, registry analysis tools, Internet analysis tools and many more analysis tools. The commonly used toolkits for analyzing file systems are Encase, FTK, X-Ways, Nuix, Sleuthkit, DFF, Snorkel and LibForensics. Of these tools, Encase, FTK and X-Ways are commercial toolkits while Sleuthkit, DFF and LibForensics are in open domain.

To extract the malicious processes from the processes of memory image dump, the file signature scanner tool known as YARA tool can be used. The YARA is an open source tool designed to help malware researcher to identify and classify malware samples. It uses the efficient pattern-matching rule. YARA supports the use of three different types of strings for pattern-matching:

(a) Hexadecimal Strings



# Tweet Classification with Convolutional Neural Network

Santosh Shivaji Kolekar  
Computer Engineering Department  
MMCOE Karvenagar  
Pune, India  
santoshkolekar.pune1@gmail.com

H.K.Khanuja  
Computer Engineering Department  
MMCOE Karvenagar  
Pune, India  
harmeetkaurkhanuja@mmcoe.edu.in

**Abstract**— Tweet classification task is used to determine opinion of tweets. Such opinion is useful for making new strategy and taking right decision as per situation. Due to high speed and high availability of internet, large numbers of people are involving in social media to share their opinion towards any happening event like sport. It needs to analyze the behavior of people whether they are happy or unhappy towards the event.

We consider convolutional neural network model, one of the deep learning approach for tweet classification. We used word embedding technique like word vector for text representation. We used APNews corpus as word embedding technique to give word vector called pre-trained word vector. On top of pre-trained word vector, we apply convolutional neural network to know the polarity of tweet. Here, we map each word of tweet to already pre-trained word vector of APNews corpus.

We fetched the tweets from social website and performed pre-process for training and testing purpose. During training phase, we found 87.25% accuracy. During testing phase, we found 79% accuracy of proposed CNN model.

**Keywords**— Convolutional Neural Network, Deep Learning, Word Embedding, Word Vector.

## I. INTRODUCTION

Solving opinion analysis problem is one of the challenging tasks in text classification. Because, people openly discuss and share their thoughts towards any new products or any event which is text based communication. It is very essential to know about the sentiments and thinking of the people and their opinion on it.

Today, internet users are increasing day by day. As per the survey it is observed that traditional machine learning techniques are not capable of handling such big data on social media. They give the good results only on small datasets and require feature engineering. We found that there are two approaches to know the opinion of people on social media like Twitter. One is traditional method using machine learning and other is deep learning approach.

In this paper, we presented the deep learning approach with review of literature, system architecture, system analysis and conclusion for tweet classification.

### A. Opinion Mining

In opinion mining, opinion of tweet either positive or negative is analyzed. It is carried out in two phase like training and testing phase. The process fetches tweet from social website like data.world. Pre-process technique is applied before training and testing phase. A training dataset

contain tweet with corresponding class label either positive or negative. Deep learning approach like proposed CNN model learns through training dataset and validated against test dataset. In such way opinion mining is performed on tweet to know the opinion of people with deep learning approach.

We use binary classification in which the process classifies the tweet either positive or negative opinion.

### B. Deep Learning

Deep learning approach like proposed CNN automatically handles feature extraction task. It is more robust and adaptable model.

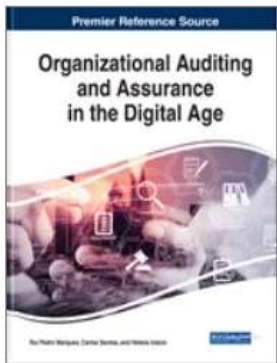
Deep learning approach is used for solving the sentiment analysis or opinion mining [4]. In sentiment analysis task, it can be used for distributed representation of text. Distributed representation is learned from large amount of training dataset. Due to distributed representation of single word it is less dependent on task specific feature. Neural network is statistical learning model used for the problem of classification. It is used in natural language processing. There are different complex network architectures having different neural cell model. It is used for several tasks achieving remarkable result. It reduces the gap between the machine learning and classical artificial intelligent. It reduces feature engineering costs [6].

## II. REVIEW OF LITERATURE

Aliaksei Severyn et al. [1] 2015, they had discussed phrase and message level sentiment analysis. Text is represented by word embedding and CNN is used as classifier. Network architecture is consist of sentence matrix, convolution feature map, activation function, pooling and softmax function. Shiyang Liao et al.[2] 2017,they had discussed CNN model which has better result compare to Support vector Machine and Naive Bayes model. It considers sentence matrix is equal to image matrix. Benchmark dataset like Movie Review (MR) and Stanford Twitter Sentiment (STS) Gold are used for conducting experiment.

Jonatas Wehrmann et al. [3] 2017, they had discussed a language agnostic translation-free method for twitter sentiment analysis. They make use of deep convolutional neural networks with character level embedding. Model is capable of learning latent feature from languages (employed during the training process). They had used Conv-Char-S character based neural network for sentiment classification





## To Monitor and Detect Suspicious Transactions in a Financial Transaction System Through Database Forensic Audit and Rule-Based Outlier Detection Model

Harmeet Kaur Khanuja (Marathwada Mitra Mandal's College of Engineering, India) and Dattatraya Adane (Shri Ramdeobaba College of Engineering and Management, India)

Source Title: [Organizational Auditing and Assurance in the Digital Age](#)

Copyright: © 2019 | Pages: 32

DOI: 10.4018/978-1-5225-7356-2.ch012

**OnDemand PDF**    **\$29.50**

**Download:**

Available

[Current Special Offers](#) ▾

### Abstract

The objective of this chapter is to monitor database transactions and provide information accountability to databases. It provides a methodology to retrieve and standardize different audit logs in a uniform XML format which are extracted from different databases. The financial transactions obtained through audit logs are then analyzed with database forensic audit. The transactions are examined, detected, and classified as per regulations and well-defined RBI antimoney laundering rules to obtain outliers and suspicious transactions within audit logs. Bayesian network is used in this research to represent rule-based outlier detection model which identifies the risk level of the suspicious transactions.

### Chapter Preview

[Top](#)

### Introduction



# Content Centric Networks (CCN): A Survey

Anita Shinde Ph.D. Scholar  
SKN COE SPPU Pune  
anitavshinde8@gmail.com

Prof. S. M. Chaware Ph.D. Guide  
SKN COE SPPU Pune  
sandeepchaware@mmcoe.edu.in

**Abstract**— In recent years, there is tremendous growth in the internet applications and communication in the field of Information Technology. They are continuously producing large size, variety of data called big data. By year 2020, several billion smart devices will be connected to Internet. With the speedy growth of the social network, information overload has become a significant problem. Service providers send a lot of additional contents and advertisements to consumers on daily basis. Because of this, consumers' concern and likelihood of reading them have decreased significantly. Thus, the network load is wasted. Precise content recommendation is needed to handle this problem. The main concern here is to provide accurate user information and handling the big data nature of users and content. This tremendous growth in devices is a challenge to the current internet architecture where connectivity is based on host-to-host communication. Content-centric networking (CCN) has come forward as novel network architecture to meet today's need for content access and delivery. CCN uses data names instead of host names and has in-network caching. CCN helps to find real time interest of consumer. In this paper, content centric network, its working, and various catching strategies are discussed in detail.

**Keywords**—: Future internet, Content-Centric Networking CCN, Big Data, Catching Strategy.

## I. INTRODUCTION

The speedily growth of social networks such as Facebook, Youtube, Twitter, Google + has increased human's social interaction. It has provided end users easily access to information as there is fast spread of information which causes the issue of information overload. Service providers suggest different types of contents and recommend related advertisements to all users [1]. Maximum delivered contents are not used by customer so the network load is wasted. At the same time, the profits of providers are decreased because advertisements are main source of income for social network service providers. The novel approach for accurate recommendation is needed with the goal of mapping context space to content space. User's real-time context is used to make personalized recommendation. The main hindrance here is that both users and data are big data. Both of them are large in volume, in diversity and are growing rapidly. Jacobson proposed Content-centric networking (CCN) paradigm as a new network paradigm for content push. It focuses on the fact that customers observe what contents the Internet holds rather than where they communicate to.

Content centric networking (CCN) makes content directly addressable and routable [2]. Communication between endpoints is based on named data instead of conventional IP addresses. It has content request messages and content return messages. It is also called as an information-centric networking (ICN) architecture. The main objective of CCN is to give flexible, protected and

scalable network to meet requirement of secure data delivery on large scale set of end devices. It has protected and named content in distributed caches which are automatically used on demand. When there is named request, CCN provides named content to consumer from nearby cache with less network hops thus avoids redundant requests and overall consumes less resources. In CCN, named content is used as packet address instead of host identifier. It mainly helps to save the network load and energy efficient than IP based network [3]. The idea of CCN is absolutely suitable in the online social network activities which include one-to-many or many-to-many dissemination and retrieval of content. CCN permits service providers to fetch accurate context of user which helps to provide accurate content push to user [4].

The performance of previous work about social media recommendation has not given better results in big data area [5]. When the data set is comparatively small, it computes recommendations for every individual data, giving high accuracy results. But as the scale of content increases, it causes high computational cost and makes it difficult for algorithm to perform [6]. All contents are organized in a tree structure to accomplish high accuracy recommendations. It separates tree construction process away from recommendations which creates difficulty in performing very well [7].

## II. RELATED WORK

Various future Internet architectures such as data-oriented networking architecture, content-centric networking, named-data networking, publish/ subscribe, and network of information focus on data delivery instead of host centric approach [8]. The main goal is to become familiar with the revolution of these architectures. All of these architectures differ in terms of implementation, but they have the similar objective i.e. to improve the performance and customer experience of the Internet by presenting access to content and services by name instead of original location. All architectures are compared in terms of naming scheme used, Security, Name resolution, Caching, Transport etc.

CCN is presented as a different broadcasting protocol for sensor data. CCN is a complete architecture which executes all pieces essential for a network layer. Message broker servers or proxies are not needed in CCN because clients can demand for any data through issuing an Interest message depicting the data. As CCN does not deal with data location, the data can be present at source node or on the path towards the source [9]. Benefits of using Content Centric Networking in Internet of Things are :i) host-to-host



# Ambiguity Resolution in English Language for Sentiment Analysis

Kamalakshi V. Deshmukh

Department of Computer science  
Marathwada Mitra Mandal's college of Engineering  
Pune, India  
kamalakshideshmukh.comp@mmcoe.edu.in

Sankirti S. Shiravale

Asst. Prof., Department of Computer Engineering  
Marathwada Mitra Mandal's College of Engineering  
Pune, India  
sankirtishiravale@mmcoe.edu.in

**Abstract**—Today's citizens need a platform where they can register their complaints about municipal corporation. Citizens need to submit their daily complaints for municipal corporation. In a traditional system like Telephonic system for registration of complaints is a very time-consuming method. The customer has to wait until call is received by service executive. The proposed system is for the Pune Municipal Corporation (PMC) where citizens can insert query and intelligent reply is given to the citizen by short text understanding and machine learning algorithms. In the post query system analyses the sentiment of citizen by sentiment analysis. Accordingly, priorities to the given citizens on the basis of the intensity of the citizen complaint. Understanding the short text is the major challenge in the system like short texts do not follow the syntax of written language, short text does not have sufficient statistics to support for approaches like text mining, short text is ambiguous and noisy. In this system to understand natural language semantic knowledge is provided by the knowledgebase. This system will help many organizations to ensure quality service provision and customer satisfaction with less human efforts.

**Keywords**—Natural Language Processing, Short text understanding, Sentence Similarity, Sentiment Analysis, text segmentation, type detection, Sentiment Analysis, WordNet.

## I. INTRODUCTION

NLP is the one of the important field in the computer science where the language of the human is understood by the computer. This can be achieved with the help of the computational linguistics. The modish study of the linguistics using computer science tools is called as the computational linguistics. To understand the natural language requires a vast amount of knowledge, syntax, semantics and knowledge about the real world.

In the era of the big data, we all are surrounded by data like E-mail, SMS, web pages, transactions etc. This data need to be analyzed to mine the meaningful information. In this project we specifically concentrate on the short text. Short text is nothing but text with little context like tweet, microblog and web search. This short text is more difficult to handle as it is generated in vast amount and it does not contain more statistics. So, understanding short text can give us more value in our knowledge.

In the NLP one of the important domain is the sentiment analysis where the analysis and summarization of the opinions is done on the data. Sentiment analysis is also called as the opinion mining and it is emerging field for research.

In the ultramodern age no one has a time to rely on the telephonic system where the long waiting period is expected. The proposed system is based on the NLP, machine learning and sentiment analysis. In the system citizen will enter the query or complaint in natural language and they will get quick response in the natural language. The intensity of the user's complaint is also calculated with the help of the sentiment analysis so it will help us to give a proper level of priority amongst the users.

## II. LITERATURE SURVEY

For the correct interpretation of the short text has many challenges like short text is ambiguous and noisy, short text does not have sufficient statistical support for approach text mining, short text not always follow the syntax of written language and generated in vast amount. Wen Hua [1] has proposed a framework for understanding the short text considering the semantic knowledge from well-known knowledgebase like Probase. There are two phases offline and online phase. In the offline phase, they have constructed the index on vocabulary and also grasped knowledge from web corpus and knowledgebase.

Precomputed the semantic score between the related terms. In the online part they have used the three important steps text segmentation, type detection and concept labelling for correct interpretation of the short text.

Taesung Lee [18] has proposed approach for the extraction of the attribute and its scoring. Attribute extraction is very important to understand the meaning of the concept. Knowledgebase consists of the concept, entities, instance, attribute and relations. They have used the concept and instance-based approach for the attribute extraction from different data sources by using Probase knowledgebase.

One of the significant concepts in natural language processing is sentiment analysis. Analysis of the opinion and view expressed in the text to be classified with text classifying algorithms as positive, negative or neutral is called as sentiment analysis. Sentiment analysis can be done at the sentence level where each sentence is given polarity as positive or negative. SA on the document level is done where whole document is given polarity positive, negative or objective. SA on the Aspect level is the deeper analysis of the text where the different features are extracted and on the basis of these different features classify view as positive or negative [14] [21].



# Priority Based Sentiment Analysis for Quick Response to Citizen Complaints

Kamalakshi V. Deshmukh

Department of Computer science  
Marathwada Mitra Mandal's college of Engineering  
Pune, India  
kamalakshideshmukh.comp@mmcoe.edu.in

Prof. Sankirti S. Shiravale

Department of Computer Engineering  
Marathwada Mitra Mandal's College of Engineering  
Pune, India  
sankirtishiravale@mmcoe.edu.in

**Abstract**—Today metropolitan citizens need a common platform to register their complaints. In the traditional telephonic system task of complaint registration is very time-consuming process so, citizen have to wait until call is received by the service executive. The proposed framework is developed for Pune Municipal corporation (PMC) will be very helpful for registering queries in natural language and get immediate response.

Understanding the short text is the main challenge of the system as short text do not follow syntax of the written language, short text does not have sufficient statistics to support approach of text mining, short text is noisy and ambiguous. So, traditional Part-of-Speech (POS) tagging tools cannot be easily applied. In proposed framework for understanding natural language semantic knowledge provided by well-known knowledgebase WordNet is used. In prequery citizens inserts complaint to system and get immediate response to query with the help of knowledgebase and machine learning algorithm. In postquery system analyses the citizen sentiment to handle grievance level and accordingly prioritize the citizens by sentiment analysis. The proposed framework will help many organizations to ensure quality service provision and customer satisfaction with less human efforts.

**Index Terms**— Short text understanding, WordNet, Sentence Similarity, text segmentation, NLP, type detection, concept labelling, Sentiment Analysis.

## I. INTRODUCTION

In Computer Science Natural language processing (NLP) is emerging field where interactions between human and computers takes place. There are many challenges in natural language processing like speech recognition, natural language understanding, to understand human expressions or emotions and question answering systems or some combinations of above.

Information burst pinpoints the requirement for machines to well understand natural language texts. In this project, we emphasis on short texts which denote texts with limited context [1]. Short text is like tweet, microblogging services, user entered queries or web search etc. are generated in huge amount so, there is necessity to tackle short texts. As a result, improved understanding of short texts will bring great value. The sentiment analysis is typical problem in the NLP. The process of analyzing and summarizing the opinions expressed

in these huge user generated data is usually called Sentiment Analysis or Opinion Mining which is a very interesting and popular domain for researchers nowadays [20].

In the today's fast world no one has time to rely on the telephonic system where one has to wait long time to register their complaint. The proposed system is based on the concepts of NLP, sentiment analysis and machine learning. This system is developed for the resolving queries of the citizens in natural language. Also giving the priority to the citizens request based on the intensity calculated by sentiment analysis. Citizens enter the complaints or queries are considered to be short text. Correct interpretation of short text semantically and sentiment analysis of the user entered text are two important tasks of this framework.

## II. RELATED WORK

To correct interpret the short text has many challenges like short texts does not follow syntax of written language, short text does not have sufficient statistics to support for approaches for text mining, short text is ambiguous and noisy. Therefore, traditional natural language tools such as POS tagging cannot be easily applied. In this system for understanding natural language processing semantic knowledge present in the recognized knowledgebase and sentence similarity measure can be used [1]. Wen Hua et al proposed framework for considering semantic knowledge in order to understand the short text. Short text is noisy, ambiguous and are produced in massive amount thus it adds another trouble to tackle them. In this paper they have used knowledge intensive approach for tasks like text segmentation, NER, concept labelling and type detection for understanding short text [1].

Study of views and sentiments expressed in texts by means of text classifying algorithms is called as sentiment analysis [20]. we will consider the basic definition of sentiment as giving positive or negative view. "I liked the music concert" will be treated as positive mood however "Concert was having very poor performances" conveys a negative comment or mood. "I am going to my college" can be considered as neutral comment as it does not convey any moods. In the sentiment analysis many times we need to classify the text according to the sentiment polarity positive, negative or



# Enhancement in Security using Extended Security Techniques

Aishwarya Shahaji Mane  
Department of Computer Engineering  
M.M.C.O.E., Karvenagar  
Pune, India  
aish301194@gmail.com

Prof. Rupali Dalvi  
Department of Computer Engineering  
M.M.C.O.E., Karvenagar  
Pune, India  
rupalidalvi@mmcoe.edu.in

**Abstract**— Security has become an inseparable issue as internet is ruling the world. Every day we are doing various activities such as e-banking, e-shopping or information transfer through internet so there exists a need for safe and secure transactions. Extended visual cryptography is very advanced technique for providing security to information which will be going to transfer through network. In this paper, we exploit extended visual cryptography technique for military application security. Without sending password directly as it is from sender to receiver we are going to create shares of password and those shares will be transfer through secure server to receiver. When receiver will combine this shares then only complete password will be revealed. In this paper we provide overview of extended visual cryptography and Algorithms implementation is presented in this paper for share generation of password.

**Index Terms**— Extended visual cryptography, Extended share generation scheme, Secure server, XOR operation, Cover image, Pixel.

## I. INTRODUCTION

Now we are in 21st century so we are dealing with huge amount data and for transaction of such data we are using internet. Information transfer through network is not secure because of attackers attacking strategies. In today's world, cybercrime is rapidly gaining momentum. Data which we transfer through internet is heterogeneous. If that is very confidential then it requires more security. There are various types of attacks which takes place over network like passive attacks and active attacks. In passive attacks there is no modification of data while in case of active attacks there is modification of data.

For providing security to confidential data various data hiding techniques are being introduced. Steganography, Watermarking, Cryptography are some of the data hiding techniques [3]. These techniques have various drawbacks. In case of watermarking receiver is not able to easily extract the original information. Steganography is hiding data under another data. Problem with steganography is that there are overcomes drawback of steganography by using key for security of data. But in this case security of key is main concern.

Visual cryptography is technique introduced by Adi and Shamir to overcome the drawbacks of data hiding techniques

[3]. The main aim of hiding technique is not only to hide the data but also to mislead attackers in wrong direction. In visual cryptography for strong security "sharing scheme" and "steganography" is used. As we said above misleading an attacker is also important. In first phase of visual cryptography that is share generation misleading is takes place. Figure 1. Represent general idea of visual cryptography.

There are mainly three phases i.e. confidential data, share generation phase and recovery of original data. Second phase plays an important role in system. It considers black and white image in the form of matrix say  $M$ . This matrix is filled with 0's and 1's. Black pixel is represented by 0 and white pixel is represented by 1. Now for each pixel from  $M$  shares will be generated. By only overlapping those shares original data will be reveal.

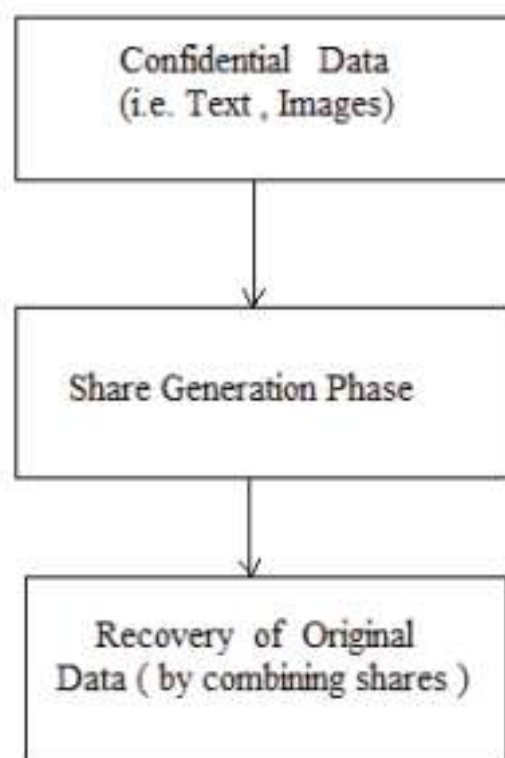


Figure. 1. General idea of extended visual cryptography

Institutional Sign In

Browse

My Settings

Get Help

Subscribe

Advertisement

Advertisement

Conferences > 2018 Fourth International Con...

# Secure Dynamic Multiowner System Using Share Generation Scheme

Publisher: IEEE

2 Author(s) Aishwarya Shahaji Mane ; Rupali Dalvi View All Authors

10 Full Text Views

## Abstract

**Abstract:** This Nowaday's internet is an part of our life. Various activities are performed using internet. While transferring confidential private data through network we need to t... [View more](#)

Document Sections

- I. Introduction
- II. Review of Literature
- III. Limitation of Existing System
- IV. An Illustration
- V. System Architecture

### Metadata

**Abstract:** This Nowaday's internet is an part of our life. Various activities are performed using internet. While transferring confidential private data through network we need to think about security. In various critical multiowner applications like military, banking, medical security is main issue. In proposed system we are going to provide security for military application while transferring password to multiple number of receiver. Password will be forwarded in the form of encrypted shares. Share generation will be at dynamic basis as per number of receiver present. For more security algorithms will be applied at secure server. Only by overlapping of shares by authorized receivers original data will be reveal. This paper provides an overview of experimentation and result analysis of proposed system.

**Published in:** 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)

**Date of Conference:** 16-18 Aug. 2018 **INSPEC Accession Number:** 18618049

**Date Added to IEEE Xplore:** 25 April 2019

**DOI:** 10.1109/ICCUBEA.2018.8697527

**ISBN Information:**

**Publisher:** IEEE

**Conference Location:** Pune, India, India

Authors

Figures

References

Keywords

Metrics

More Like This

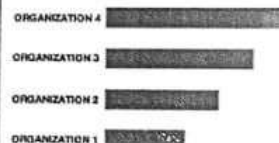
### More Like This

Deployment of secure sharing: Authenticity and authorization using cryptography in cloud environment  
2015 International Conference on Advances in Computer Engineering and Applications  
Published: 2015

Shared authorization in industrial automation systems using threshold cryptography  
2005 IEEE Conference on Emerging Technologies and Factory Automation  
Published: 2005

[View More](#)

### Top Organizations with Patents on Technologies Mentioned in This Article



Advertisement

Advertisement

## Contents

### I. Introduction

Today's world relies on internet for performing various day to day activities. Now a days confidential data can easily accessible via internet (e.g. Posting personal information on facebook). In every field security is always a main concern. According to Symantec

IEEE websites place cookies on your device to give you the best user experience. By using our websites, you agree to the placement of these cookies. To learn more, read our Privacy Policy.

Accept & Close



# Soil Toxicity Prediction and Recommendation System Using Data Mining In Precision Agriculture

Mayuri Pawar

Department of Computer Engineering  
MMCOE  
Pune, India  
mayuripawar.comp@mmcoe.edu.in

Geetha Chillarge

Department of Computer Engineering  
MMCOE  
Pune, India  
geethasb@mmcoe.edu.in

**Abstract**— India is agricultural land. India ranks second worldwide in agriculture output, but GDP share is declining. There are many factors contribute for declining agriculture GDP which are inadequate irrigation, inadequate power supply, changing environmental conditions, conventional agricultural method etc. In this paper, the proposed system can help farmers by making them aware about their soil conditions. Farmers can maximize crops yield by knowing proportion of nutrients present in the soil. Soil toxicity affects the soil nutrients which indirectly affects crops health.

The proposed system predicts the level of toxicity present in the soil and makes farmer aware about it. Many farmers are depending on rainfall which is the one of the factor for poor growth and decreases crops yield. Thus the proposed system recommends the farmer about the crop, fertility of soil, level of toxicity and water supply. For this recommendation system, sensor's accuracy is very important as well as classification algorithm. For classification, decision tree J48 algorithm is used which is simple to implement and having more accuracy as compared with other classification algorithms. Issue of power supply can be overcome by using solar panel system.

**Index Terms**— prediction, precision agriculture, soil nutrients, sensors, decision tree, J48, toxicity, soil pH.

## I. INTRODUCTION

Soil is the soul of agricultural land. All the nutrients are present in soil in different proportions based on soil type. At every step, in our country the properties of soil are different. So it is important to maintain soil nutrients with existing diversity. Conventional methods are time consuming and cannot find the cause of poor plant growth. With enhancing technology now we are able to continuously monitor field. Sensors are playing major role in precision agriculture.

Number of work has been done with sensor network in this area of application under IoT. Sensor network is the connecting real world to cyberspace which is key concept of IoT [1]. There are various applications of IoT due to its various characteristics like interoperability, heterogeneity with increasing automation with less effort by employing internet standards etc. [2]

With machine learning algorithms prediction systems are developed like crop selection and crop yield prediction, disease prediction [7][15], weather forecasting, deciding minimum support price, irrigation system etc.[3][4][5] As well as data mining techniques are used for classification of soil and for prediction along with machine learning to increase accuracy [6].

Data mining is important technique to extract information from huge data which supports decision making system. Classification, clustering, association rule mining and regression these techniques are used to discover hidden information. Classification is the technique [18] of classifying data new data from known one. It includes decision tree, naïve bayes, nearest neighbor, neural network [15], support vector machine (SVM) [13]. Proposed system uses decision tree J48 algorithm as accuracy is more than other classification algorithms [16] [19].

The proposed system uses J48 algorithm for classifying fertility and based on it predicting soil toxicity level. The objectives of proposed system is 1) monitoring crop health based on soil fertility and to recommend required fertilizers 2) to predict toxicity of soil so farmers can take action it 3) alerting about irrigation requirement 4) recommendation of crop.

## II. LITERATURE SURVEY

Heamin Lee et al. [7] designed IoT system for disease and pest prediction in orchard. All the weather conditions affect the growth of plants. Amount and frequency of pesticide can be reduced by monitoring weather data. They developed four modules: weather forecast, pest prediction, user application (web application, mobile phone), server. They are attached sensors to weather station near orchard as well as video camera is placed for monitoring purpose. The weather data is stored in database. Pest prediction module reads weather data and real-time data and calculates the probability of infection and their occurrence date. This estimated date is stored in database and server sends notification to user priority so that user can manage the things. This helps farmer economically as it prevents from huge loss.



[Intelligent Computing and Information and Communication](#) pp 505–513 | [Cite as](#)

# BMWA: A Novel Model for Behavior Mapping for Wormhole Adversary Node in MANET

[S. B. Geetha](#)  & [Venkanagouda C. Patil](#)

Conference paper | [First Online: 20 January 2018](#)

**1318** Accesses | **1** Citations

Part of the [Advances in Intelligent Systems and Computing](#) book series (AISC, volume 673)

## Abstract

Wormhole attack has received very less attention in the research community with respect to mobile ad hoc network (MANET). Majority of the security techniques are toward different forms of wireless network and less in MANET. Therefore, we introduce a model for behavior mapping for Wormhole Attacker considering the unknown and uncertain behavior of a wormhole node. The core idea is to find out the malicious node and statistically confirm if their communication behavior is very discrete from the normal node by formulating a novel strategic approach to construct effective decision. Our study outcome shows enhanced throughput and minimal overhead-latency with increasing number of wormhole node.

## Keywords

Mobile ad hoc network

Wormhole attack

Adversary

Malicious node behavior

Attack pattern

This is a preview of subscription content, [access via your institution.](#)



# Context Aware Computing Systems: A survey

<sup>1</sup>Ms. Pooja S. Gandodhar, Research Scholar, SKNCOE, SPPU, Pune

<sup>2</sup>Dr. S.M. Chaware, PhD Guide, SPPU, Pune

**Abstract**— The term context has been studied in different areas of Computer Science. Context-aware systems have high demand in areas like Intelligent Environments, Pervasive & Ubiquitous Computing and Ambient Intelligence. Such systems gather data and adapt system behavior accordingly using context information like physical context, computational context, and user context/tasks. Developing such context-aware applications is inherently complex and hence should be supported by adequate context information modeling and reasoning techniques. In this paper the concept of context awareness is discussed along with requirements for context identification and formulation, procedure for converting context modeling to intelligent actions and context recognition techniques & algorithms are discussed. The applications of context aware are reviewed.

**Index Terms**—Context, Context Awareness, Context modeling, Context Aware Computing, Pervasive Computing

## I. INTRODUCTION

One of the areas of pervasive (ubiquitous) computing is context-aware computing. Mobility of devices and use of services make context-aware systems a popular research field. In real physical context means to use the data gathered from sensors to the context-aware computing platform, which includes: what to sense, how to acquire the information and reasoning to that information to infer the context of a user. It is highly needed that programs and services should respond according to the user's situation and behave the way she wants these to be, i.e. services and systems should be more dynamic. To identify the context the information which we require can be captured in a number of ways, for example from user profile information, network (to sense location, time, nearby objects etc.), sensors (for activity) and other sources.

Many researchers have defined context as per their understanding in an effort to consider a more general concept of context. Schilit and Theimer [6] used the term context-aware in 1994 and described as identities, location, objects and nearby people [6]. In 1996 Brown defined context as the elements that surround a user which a computer can identify [5]. An often cited and quite generic definition of context is that by Dey and Abowd: "Context is any information that can be used to characterize the situation of an entity. An entity is a

person, place or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves." The context is useful and people have worked on it, focusing on location mostly, although the situation of an entity might consist time, location, activity and the surrounding factors that might affect the activity of an entity[6].

The term context has been classified into two categories (physical and logical). Physical context can be determined by hardware sensors and logical context is either given by the user's input or by capturing her interactions with the services available [1][2]. For example through observing or analyzing the user's profile, activities, working routines, typing activity etc. Most research in this area makes use of physical sensors for light, movement, sound, temperature, touch, and of course location. The logical sensors however provide related information by reading user's information from opened web pages and other documents and also analyses user's data (interactions) and based on those interactions targets advertising.

## II. LITERATURE REVIEW

Following section includes discussion on concepts of context, context awareness, and context aware systems.

*Context* The term Context is been studied by many researchers. The most popular definition is information that can be used to relate or characterize an entity [2]. Using various definitions and interpretations "Context is any information that can be used to characterize the situation of an entity. A person, place, or object can become entity which is considered applicable to the interaction between a user and an application, including the user and applications themselves."[3][6]

Context Awareness was introduced Schilit and Theimer as sentient [7]. "A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task.[4] [9]"

Context-aware systems are able to adapt their activities based on current context. This also increases effectiveness by taking environmental context into account. Context aware systems watch the environment all the time and propose suitable suggestions to users so they can take necessary actions. For example publishing a user's location to appropriate members of a social network, and allowing retailers to publish special offers to potential customers who are near to the retailers.





# Security and Privacy Issues in Cloud Computing

Asma A. Shaikh<sup>1(✉)</sup> and Kamatchi Iyer<sup>2(✉)</sup>

<sup>1</sup> Amity School of Engineering and Technology, Mumbai, India  
asmamokashi@mmcoe.edu.in

<sup>2</sup> CSE, Amity School of Engineering and Technology, Mumbai, India  
kiyer@mum.amity.edu

**Abstract.** IT Organization are moving for Cloud Computing from last few years and now days Cloud Computing becomes their first choice because it provides flexible and measurable provision as per their use. IT Organization is having major concern about Security and Privacy. It is difficult to provide safety and privacy in Cloud Computing because of its open and distributed architecture. Most intruders are trying to attack Cloud to get private information from IT Organization. This paper gives overview of different security issues in Cloud Computing.

**Keywords:** Cloud computing · Cloud security · Security controls  
Intrusion detection

## 1 Introduction

As per the NICT's definition of Cloud Computing mentioned five important features:

**On-demand self-service.** A customer of Cloud can get the access to computing capabilities like Server time and network storage, as needed automatically without requiring human intervention with each service provider.

**Broad network access.** A Cloud computing provides capabilities to accessed the network through standard technique that help use by different types of thin or thick client platforms (e.g. tablets, laptops, and workstations, mobile phones).

**Resource pooling.** The Cloud Computing service providers provides pool of resources to handle many customer using a multi-tenant model, with many physical and virtual resources dynamically allocated and reallocated according to customer demand.

**Rapid elasticity.** It provides capabilities to elastically provisioned and released as per the need access in some cases automatically, to scale rapidly upward and inward proportionate with demand. It provides features to give of unlimited access to the Cloud Resources as per the need.

**Measured service.** Cloud System provide service of Pay per Use. Cloud systems automatically maintain and optimize resource utilization by investing a measuring system capability at some level of abstraction needed to the type of service (e.g., active user accounts, storage, processing, bandwidth). Resource usage can be observed, maintained, and reported, providing clarity for both the Cloud provider and customer of the utilized service.



# Adulteration Detection in Petroleum Liquids using Stacked Multi Ring Resonator

Anaght Kunte

Electronics & Telecommunication Department  
Marathwada Mitra Mandal's College of Engineering,  
Research Scholar, Shrihgad College of Engineering  
(Savitribai Phule Pune University)  
Pune, India  
anaghtkunte7@gmail.com

Arun Gaikwad  
Electronics & Telecommunication Department  
Zeal College of Engineering and Research  
(Savitribai Phule Pune University)  
Pune, India  
arungkwd47@gmail.com

**Abstract**—Adulteration detection and estimation are essential for quality control of fuel. The microstrip ring resonator sensor can be used for permittivity measurement of liquid petroleum. The novel Stacked Multi Ring Resonator (SMRR) is specially designed for permittivity measurement with ease of handling liquids. Lumped model development of ring resonator structure is done with the help of ADS (Advanced Digital System). CST MWS (Computer Simulation Technology Microwave Studio) software is used for simulation of 3 D model of SMRR. Resonating frequency, insertion loss, quality factor are main parameters in the determination of permittivity of adulterated liquids. SMRR is low cost and simple microwave sensor for adulteration detection in liquids.

**Keywords**—Adulteration; Permittivity; Stacked Multi Ring Resonator; Resonating frequency; Quality Factor.

## I. INTRODUCTION

Quality monitoring in Indian gasoline like petrol and diesel is essential due to hike in prices. Petroleum properties such as viscosity, density, boiling point, and color of petroleum may vary widely, for a large number of petroleum samples over a narrow range. The carbon content is relatively constant, while the hydrogen and heteroatom contents are responsible for the significant differences between petroleum samples. Indian gasoline is adulterated by mixing ethanol, kerosene, naphtha. This is because if adulteration is limited to small volume, it is difficult to detect by the automobile user. To check adulteration and to monitor fuel quality, some sensor is needed at the distribution point.

Dielectric constant measurement of liquid petroleum like N-Hexane, Petrol, Diesel, Bio-diesel and kerosene is the important parameter for adulteration prediction. Adulteration detection and estimation can be done using many techniques like Evaporation test (ASTM D3810), Distillation Test (ASTM D86), Gas Chromatography, Optical Fiber, Ultrasound, etc. [1]. But these tests require costly infrastructure, substantial computational time and laborious efforts. Microwave sensor like microstrip resonator sensor is a nondestructive technique and chemical composition of the material is retained.

## II. SMRR DESIGN

Two types of resonators such as ring and straight resonators (also called as an end coupled resonator or  $\lambda/2$  resonator) are manufactured. The earlier straight resonator was used for the detection of ethanol content in gasoline [2]. Microstrip ring resonator does not suffer from open-ended effects and can be used to give more accurate measurement [3], higher quality factor and lower insertion loss than the straight or  $\lambda/2$  resonator. The ring resonator has the smaller 3dB bandwidth and sharper resonance than the linear

resonator. Therefore the ring resonator is used for further investigation [4]. For ease of handling of a liquid, steel enclosure is fabricated along with planar ring resonator. The novel SMRR structure is developed for achieving better quality factor Q and less insertion loss  $S_{21}$ .

### A. Working Principle

The SMRR consists of the lower patch of the ring called as the fed patch which is excited by resonating frequency of 2.45GHz, and the upper inverted patch is known as the parasitic patch. Two resonances are associated with SMRR. First resonance is associated with fed patch and ground plane. Second resonance is associated with fed patch and parasitic patch [5]. When the size of the parasitic patch is nearly equal to the fed patch and the spacing between the fed patch and the parasitic patch is kept about 0.5 $\lambda$ , maximum energy coupling is obtained [6]. Due to dual resonance in SMRR, the maximum electromagnetic coupling is achieved which improves quality factor Q with minimum return loss [7].

### B. Design and Fabrication of SMRR

Two different substrates are used i.e., FR4 glass epoxy with thickness 1.6mm and RT Duroid 5880 with the thickness of 1.575mm. RT Duroid substrate is used for the fed patch and the FR4 substrate is used for the parasitic patch. Figure 1 shows a planar ring resonator as the fed patch of SMRR. 2, 3 & 4 rings patches are made on the parasitic patch. Figure 2 shows the 3D model of SMRR. The parasitic patch is shown with 4 rings and it is used as the inverted parasitic patch for the measurement. The optimized dimension of SMRR is as shown in the Table I.

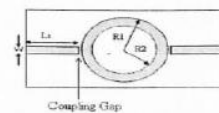


Fig. 1. Planar Ring Resonator as fed patch of SMRR

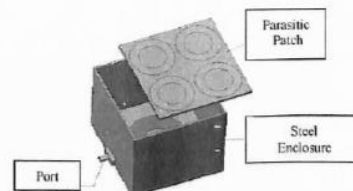


Fig. 2. 3D Model of SMRR

## Review on Computer Aided Detection systems of breast cancer

A.M.Solanke<sup>1\*</sup>, Dr.R.Manjunath<sup>2</sup>, Dr.D.V.Jadhav<sup>3</sup>

<sup>1,2</sup>Jain - a Deemed to be University Bangalore, India, [anjali.solanke1@gmail.com](mailto:anjali.solanke1@gmail.com), 7588951413

### Abstract

Breast cancer is life threatening disease for women. According to World Health Organisation breast cancer is second leading cause of death in the world. Many lives can be saved by early detection of breast cancer. Most widely used breast cancer screening technique is mammography. Mammography is used for detection and clinical evaluation of breast cancer. Computer aided detection techniques (CAD) are used to assist doctors and radiologists for analysing mammograms. CAD techniques play a very important role in early detection of breast cancer. In this paper total forty five papers are referred to present overview of signs of breast cancer, screening technique and survey of algorithms for detection of Micro-calcifications, masses and architectural distortion.

### Keywords

*Architectural distortion detection, CAD (Computer aided detection), Mammography, Mass, Microcalcifications*

### I. Introduction

Biomedical engineering plays an important role in healthcare technology. Healthcare technology brought a revolution in various domains of the medical field such as pathology, various screenings like X-ray, MRI, CT scan and surgical procedures [1]. Outcomes of these healthcare systems are quick treatment, early diagnosis and quality life. Biomedical engineering dramatically turned the style of diagnostic methods opted by physicians in the last half century. Wide varieties of tools are made available for improving diagnosis and disease treatments. These tools include medical imaging, computer-aided detection and medical instruments [2]. CAD systems are being used extensively by radiologists as it reduces human errors due to low contrast of medical images. Particularly, computer-aided detection is playing a major role in the detection and prevention of life-threatening diseases like breast cancer, lung cancer, skin cancer and many more. Here we will focus on breast cancer which is the second leading cause of death [3]. Deaths due to breast cancer can be decreased by early detection. High-risk patients are identified based on various factors like age, gender, past occurrences in family and density [4]. Mammography is widely used for breast screening and diagnostic procedures worldwide for the early detection of breast cancer. [5]. Masses, Microcalcifications, architectural distortion and bilateral asymmetry are signs of breast cancer in mammograms [6]. Usually benign masses are with definite shape like round or oval with regular boundaries, and low density. Masses are difficult to detect due to their density variation and shape. Microcalcifications are deposits of calcium which are very small in size and bright as compared to normal tissues. Their average diameter is of 0.3mm. Generally Clustered Microcalcifications are malignant [7]. Architectural distortion is nothing but distorted normal architecture of breast. Bilateral asymmetry is nothing but asymmetry between left and right breast parenchyma. Asymmetric small sized bright spots and contrast in both breasts is bilateral asymmetry. Radiologists may miss any abnormality due to human error. Therefore to reduce errors, researchers proposed computer-aided detection techniques for the detection of these abnormalities. As a result of this false positive cases are reduced and hence unnecessary biopsies can be avoided.

### II. Mammography:

X-ray film screen mammography is one of the most recommended and widely used imaging methods for the diagnosis of breast diseases. Mammography machine is as shown in Figure 1(a). It consists of X-ray tubes, detector, anti-scatter grids and compression device. There are two types of mammography. Screening



As per Revised Syllabus of

## SAVITRIBAI PHULE PUNE UNIVERSITY

Credit System Scheme-2015

S.E. (IT) Semester - I

# DISCRETE STRUCTURES

**Dr. H.R. Bhapkar**

M.Sc. SET, Ph.D. (Mathematics)  
MIT Art, Design and Technology University's  
MIT School of Engineering, Lonikalbhor, Pune  
Email : drhrbhapkar@gmail.com  
Mobile : 9011227141

**Dr. Parikshit N. Mahalle**

M.E. Ph.D. (Computer Engineering)  
Smt. Kashibai Navale  
College of Engineering  
Vadgaon, Pune.  
Ph. No. - 9822416316

**Dr. Anant M. Bagade**

M.Tech., Ph.D. (Computer Science & Engineering)  
Information Technology Department  
Pune Institute of Computer Technology  
Pune.

**Dr. Vijaykumar Bidve**

M.Tech., Ph.D. (Computer Engineering)  
Marathwada Mitra Mandal's College of Engineering  
Pune.  
Ph. No. - 9552500740

 **TECHNICAL<sup>®</sup>  
PUBLICATIONS<sup>™</sup>**  
An Up-Thrust for Knowledge

Price : ₹ 235/-

ISBN 978-93-332-1234-2



9 78 93 33 21 23 42

As per Revised Syllabus of  
**SAVITRIBAI PHULE PUNE UNIVERSITY**

---

# **Software Testing and Quality Assurance**

B.E. (IT) Semester - I (Elective - II)

---

**Dr. Vijaykumar S. Bidave**


Ph.D (C.S.&E.), M.Tech. (IT), B.E. (C.S.& E.),  
Associate Professor & Head,  
Information Technology Department,  
MMCoE, Karvenagar, Pune.

**Sheetal S. Koul**

M.E. (Computer Network), B.E. (IT)  
Assistant Professor, Information Technology Department  
SKNSITS, Kusgaon, Lonavala, Pune.

 **TECHNICAL  
PUBLICATIONS**  
An Up-Thrust for Knowledge

Website : [www.technicalpublications.org](http://www.technicalpublications.org)

 <https://www.facebook.com/technicalpublications>

(i)



1<sup>st</sup> Edition

w.e.f. from 2018-2019



Strictly as per the new Credit System Syllabus (2015 Course)  
**Savitribai Phule Pune University**  
w.e.f. academic year 2018-2019



# Software Testing and Quality Assurance

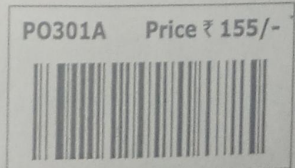
(Elective - II)

(Code : 414457C)

Semester VII - Information Technology

Dr. K. S. Wagh

Swapnil S. Shinde



1<sup>st</sup> Edition  
w.e.f. from 2018-2019

Strictly as per the New Credit System Syllabus (2015 Course)  
**Savitribai Phule Pune University**  
w.e.f. academic year 2018-2019

SPPU

# Software Testing and Quality Assurance

(Elective II)

(Code : 410245(B))

Semester VII - Computer Engineering

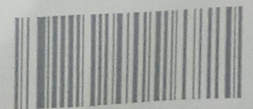
K. S. Wagh

Swapnil S. Shinde

 **Tech-Max** Publications, Pune  
Innovation Throughout  
Engineering Division

26

P0310A Price ₹ 125/-



9



# Implementing a Hybrid of Efficient Algorithms For Mining Top-K High Utility Itemsets

Ingle Mayur Rajendra

Department of Information Technology  
Marathwada Mitra Mandal's College of Engineering  
Pune Maharashtra, India  
ingle.mayur07@gmail.com

Shri Chaitanya Vyas

Department of Information Technology  
Marathwada Mitra Mandal's College of Engineering  
Pune Maharashtra, India  
chaitanyavyas.it@mmcoe.edu.in

Sanika Sameer Moghe

Department of Information Technology  
Marathwada Mitra Mandal's College of Engineering  
Pune Maharashtra, India  
sanikamoghe5@gmail.com

Deepali Deshmukh

Department of Information Technology  
Marathwada Mitra Mandal's College of Engineering  
Pune Maharashtra, India  
deepalideskmukh.it@mmcoe.edu.in

Sachin Sakhare

Department of Information Technology  
Marathwada Mitra Mandal's College of Engineering  
Pune Maharashtra, India  
sachinsakhare.it@mmcoe.edu.in

Prof Sudhanshu Gonge

Department of Information Technology  
Marathwada Mitra Mandal's College of Engineering  
Pune Maharashtra, India  
sudhanshugonge@mmcoe.edu.in

## Abstract—

*Data mining is a methodical process of discovering data patterns and models in large data sets that involve methods at the intersection of the database system. This paper issues the popular problem of the extraction of high utility element sets (HUI) in the context of data mining. The problem of these HUIs (set of elements of high usage and value) is mainly the annoying mixture of frequent elements. Another addressable issue is the one of pattern mining which is a widespread problem in data mining, which involves searching for frequent patterns in transaction databases. Solve the problem of the set of high utility elements (HUI) requires some particular data and the state of the art of the algorithms. To store the HUI (set of high utility elements) many popular algorithms have been proposed for this problem, such as "Apriori", FP growth, etc., but now the most popular TKO algorithms (extraction of utility element sets) K in one phase) and*

*TKU (extraction of elements sets Top-K Utility) here TKO is Top K in one phase and TKU is Top K in utility. In this paper, all the aforementioned issues have been addressed by proposing a new framework to mine k upper HUI where k is the desired number of HUI to extract. Extraction of high utility element sets is not a very common practice. Although, it is indefinitely being used in our daily lives, e.g. Online Shopping, etc. It is part of the business analysis. The main area of interest of this paper is implementing a hybrid efficient Algorithm for Top K high utility itemsets. This paper implements the hybrid of TKU and TKO with improved performance parameters overcoming the drawbacks of each algorithm*

**Keywords—:** utility mining, high utility item-set, top k- pattern mining, top- k high utility item-set mining.

---

---

## I. Introduction

Today we found that search Engine plays a vital role in providing faster result which helps users in getting the appropriate content on the web. A hit is recorded when user searches for data. Search engines are medium through which group of items are matched according to the interest of user. Search Engines are a medium through which group of items are matched according to the interest of user. In search engines when user fires a query it is termed as Query Search and in data mining frequent items concept is being used for the web. Frequent Itemset means a set of items that occur frequently. When the user searches for Query result. The primary goal of frequent item set mining is to discover hidden patterns unexpected trends in the data. With the utility mining each item is associated with a utility and occurrence count in each transactions. In order to address the limitation of Frequent item mining High utility itemset mining is used. In high Utility mining [HIM] we find the itemset that generate a high profit in a database when they are sold together, in which the user has to provide a value for a threshold called 'minutil'. The HIM algorithm outputs all the High utility itemset that is the itemset that generate atleast minutil profit.

Nowadays data is being generated in huge amount. Data has many forms mainly data is classified as structured data and unstructured data. Many research studies have been proposed on Various TOP K pattern mining such as Top k frequent itemset, top k frequent closed itemset, top k association rule. Every algorithm in top k mining is different it performs the task of mining but with some performance parameter that are different based on the aim of algorithm.

## II. Related work

In the paper[1] Vincent S Tseng, proposed the two Efficient algorithms for mining Top K High utility Itemset without the need of setting the min utility value. The paper explains the usage of both

algorithms and its performance. The TKO Algorithm when given an input dataset requires less time to execute but the result is not accurate and in case of TKU when given same input which was given to TKO it result requires more time for execution but gives accurate result

Paper[2]Chowdhury Farhan Ahmed represented three variations of tree structure for utility pattern mining for handling incremental databases. Further the author used pattern growth approach to avoid level wise candidate generation.

In the paper[3]Yuqing Lan, worked on article Mining high Utility itemsets over uncertain databases, in which the author proposed an Efficient mining Algorithm named as UHUI apriori which has solved the problem of mining high utility itemset(MHUI) over uncertain databases, in which each item has a utility.

Paper[4] proposed a system on Top K High Utility itemset mining based on utility list structures In this paper pattern growth method was used. In addition to it Author used the TKUL Miner Algorithm with using utility list to avoid Additional scanning which is necessary step for the existing Top K HUIM algorithms.

Adinarayanareddy B., O. Srinivasa Rao, MHM Krishna Prasad,[5]suggested improvedUP-Growth high utility itemset mining. The compact tree structure, Utility Pattern Tree i.e. UP-Tree, maintains the information of transactions and itemsets and avoid scanning original database repeatedly. UP-Tree scans database only twice to obtain candidate items and manage them in an efficient data structured way. Applying this UP-Tree to the UP-Growth algorithm takes more execution time for Phase II. Hence they presents modified algorithm aiming to reduce the execution time by effectively identifying high utility itemsets

Ramaraju C.,Savarimuthu N[6] proposed a conditional tree based novel algorithm for high utility itemset mining. A novel conditional high utility tree (CHUT) compress the transactional databases in two stages to reduce search space and a new algorithm called HU-Mine is proposed to mine complete set of high utility item sets