

IITM Journal of Information Technology

Volume 5 (January - December 2019)

ISSN No. 2395 – 5457

SPECIAL ISSUE



IITM
Nurturing Excellence



INSTITUTE OF INNOVATION IN TECHNOLOGY & MANAGEMENT

Affiliated to GGSIP University, 'A' Grade by GNCTD & Assessed as 'A+' by SFRC

NAAC Accredited & ISO 9001-2008 Certified Institute



NCIRCSIT'2019

**NATIONAL CONFERENCE ON INNOVATIVE RESEARCH IN
COMPUTER SCIENCE AND INFORMATION TECHNOLOGY:
“ADVANCED COMPUTING TRENDS & CYBER SECURITY”**

MARCH 30, 2019

Editors: Dr. Geetali Banerji, Dr. Ankit Verma

IITM Journal of Information Technology

IITM Journal of Information Technology is a National Annual Journal of Information Technology intended for professionals and researchers in all fields of Information Technology. Its objective is to disseminate experiences, ideas, case studies of professionals in Management and Information Technology to propagate better understandings. Its focus is on empirical and applied research and reflections that are relevant to professionals of Management and Information Technology with academic standards and rigor within purview. The views expressed in the Journal are those of authors. No part of this publication may be reproduced in any form without the written consent of the publisher. The Journal intends to bring together leading researchers and Information Technology practitioners from around the country.

Editorial Board

Patron

Sh. J.C. Sharma

Sh. Ravi Sharma

Editor in Chief

Prof. (Dr.) C.P. Chawla

Editors

Dr. Geetali Banerji

Dr. Ankit Verma

Editorial Team

Dr. Prabhat Kumar Vishwakarma

Dr. Madhu Chauhan

Ms. Harsh Arora

Advisory Board

Dr. G Athithan, DS and DG (MED, CoS, & CS), DRDO

Dr. Arvinder Kaur, Professor & Dean, USIT, GGSIPU

Dr. Navin Rajpal, Professor, USIT, GGSIPU

Dr. B.V. Ramana Reddy, Dean, USET, GGSIPU

Dr. A. K. Saini, Professor, USMS, GGSIPU

Dr. Ajay Kumar, IAS, Additional Secretary, MeitY, GOI

Shri. Arvind Kumar, Scientist G, MeitY, GOI

Dr. A S Kamble, Senior Director, MeitY, GOI

Dr. Rakesh Maheshwari, Scientist G, MeitY, GOI

Sh. Pawan Duggal, Advocate, Supreme Court of India

Dr. Sushila Madan, Professor, University of Delhi

Dr. C. K. Jha, Professor & Head CSE, Banasthali University

Dr. Varun Monga, Professor & Dean, MVN University

Dr. Sunil Khatri, Director, Amity University

Dr. Shucheta Bhasin, Professor, Kurukshetra University

Dr. Jay Shankar Prasad, Professor & Head, MVN University

Dr. Kanak Saxena, Professor, SATI, Vidisha

Dr. Prem Chand Vashist, Professor, GL Bajaj ITM

Institute of Innovation in Technology & Management

Vision

The Institute aims to be a Centre of Excellence promoting value based Quality Education in the emerging areas of professional studies in Information Technology & Management.

Mission

The Institute endeavors to contribute towards meeting the growing demands for competent and trained Information Technology professionals, Software Engineers and World Class Managers determined to achieve excellence.

BLANK

IITM Journal of Information Technology

Annual Journal of Institute of Innovation in Technology & Management

Volume 5

January-December, 2019

CONTENTS

Research Papers & Articles

	Page No.
● Emerging Security Breaches and their Counter Measures <i>- Prabhat Kumar Vishwakarma</i>	3-8
● Smart City : The Evolutionary Innovation of IOT <i>- Stuti Kapoor, Aparna Bhardwaj, Vishakha Sehdev Verma, Ankit Verma</i>	9-14
● Three Level Secure App based Voting System using Biometric Linked with Aadhar Card for Universities <i>- Sushma Malik, Geetali Banerji</i>	15-21
● Block Chain <i>- Pragya Agarwal, Himani Verma, Govind Murari Upadhyay</i>	22-27
● Recommender Systems: A Detailed Analysis of Current Existing Protocols <i>- Kanika Bhalla, Tanvi Sharma, Shivangi Kharbanda</i>	28-33
● Shaping the Future with Crowdsourcing <i>- Himangini Khanna, R. Manisha, Vishakha Sehdev Verma, Ankit Verma</i>	34-38
● Web Usage Mining: Concepts & Applications <i>- Amanpreet Kaur, Charul Nigam, Priti Sharma</i>	39-46
● Software Development Life Cycle Models: A Brief Comparison Then and Now <i>- Himanshi Bhambri, Niharika Galgali, Himanshi Yadav</i>	47-52
● Autonomous Systems: The Era of Healing <i>- Saksham Bhambri, Muskan Ahuja, Lokesh Jain</i>	53-57
● Empirical Study of Privacy Preserving Data Mining Techniques <i>- Ankita, Neha Gulati</i>	58-63
● A Comparative Study of Automation Testing Tools <i>- Yamini Sharma, Akanksha Prasad</i>	64-68

	Page No.
● A Novel Approach to Assess the Quality of Student using Fuzzy Logic and Fuzzy C-Means Clustering - Lokesh Jain	69
● Delhi Youth Usage of Prominent Messengers: A Case Study Analysis - Sushma Malik, Mehak Tawakley, Megha, Mayank Gupta	69
● Analysis of Emerging Research Areas of Social Networks - Deepti Gupta, Archita Verma, Garvit Singh	70
● Blockchain Technology: Overview, Working and Application - Preetam Singh Choudhary, Dona Biswas, Vishakha Sehdev Verma, Ankit Verma	71
● Prediction of Movies Performance using Python - Deepanshu Saxena, Kailash Parihar, Lokesh Jain	71
● Virtual Reality in Actual Reality - Suruchi Sinha, Ranajoy Dutta, Himanshi, Harshit Dahiya	72
● A Big Data Approach towards Optimization of Railway Transportation using Machine Learning - Harsh Arora, Shivangi Kharbanda	72
● Mobile OS: A Case Study of iOS - Sushma Malik, Pulkit Bahl, Prakhar Sharma ^[2]	73
● Sixth Sense Technology: The Future of AI - Yashika Kohli, Geetika Lunj, Vishakha Sehdev Verma, Ankit Verma	73
● Fuzzy Linguistic Analysis of a Reheating-Furnace System Subject to Preventive Maintenance - Indu Uprety, Kalika Patrai	74
● Decoding the Concepts of Competitive Programming - Sushma Malik, Anant Kaushik, Mehak Tawakley, Mayank Gupta	74
● Learning Analytics and Educational Data Mining: Tools for Improving Quality Education in Higher Education Institutions - Monika, Chhavi Rana	75
● An Empirical Study on Agile Methods & Tools - Anand Sehgal	75
● Downloading Multiple Content From Websites using Parallel Downloading Technique - Shallu Juneja, Garvit Verma	76
● Performance Comparison of Proactive Routing Protocols in Wireless Sensor Networks using Qualnet Simulator - Puneet Garg	77
● Categories of Threats and Attacks to Information System in An Organization - Rashmi Jha, Divyansha Saini, Shurabhi Shanker, Nishant Nirmal	78

Emerging Security Breaches and their Counter Measures

Prabhat Kumar Vishwakarma

prof.dr.prabhat@gmail.com

Abstract

Right decisions at right time based on reliable data are the key factor in E-Business activities. In the Hi-tech arena of Digital Marketing, Big-Data and IoT, ensuring unauthorized access, protection and safe guarding of valuable online data is still a challenging task.

The purpose of this paper is to discuss emerging security breaches observed in E-Business and to propose their counter measures.

Key Words: IoT, E-Business, Big Data, Security Threats

I. Introduction

Internet is the backbone of global communication which has been the reason for the growth and acceleration of E-Business. The boundary less marketplace has given the challenges to the corporate world for thinking and promoting the business in a different way in terms of digital marketing, E-Logistics. Online rating, reviews about the products and services offered by dot com companies are now days are very common phenomena in business. Card payments, E-Receipts, paperless approach of business functionalities are strengthening the global economy and up to certain extent contributing toward the Ecological balance.

Continuous growth of Face Book, Whtasapps and other social networking services are leading towards Nonstop Internet user's expansion. In technical term expansion of Internet means more and more coupling of hardware and software resources. Therefore cyber world is turning toward Hi-Speed processing in the form of cloud computing, Big Data Processing. Undoubtly for maintaining E-Business momentum Intelligent business models are continuously needed. So in the next section of the paper focus will be on understanding emerging practices of Cloud Computing, Big Data and subsequently analyzing security issues and challenges underlying inside it.

Social Networking Sites(SNS) is a platform which facilitates information sharing among group of participants having electronic gadgets such as Android based smartphones, which is build up on Hi-Speed Internet Connections framework offered in the terms of 2G/ 3G/4G enabled services.

Ease of entry in E-Business along with optimizing the cost is greatly facilitated by Cloud Computing. Cloud Computing Technology provides the remote accessibility and utilization of Hardware and Software infrastructure without thinking so much about its other aspects. Cloud Computing services are majorly categorized as-

- IaaS- Infrastructure as a Service.
- PaaS-Platform as a Service
- SaaS-Software as a Service

The relative popularity of SNS among users till January 2019 with reference to www.statistica.com are depicted in graphical form in Figure 2 is as follows –

Due to heterogeneity in distributed computing environment, and need of fast computing in terms of Tera FLOPs and Peta FLOPs over a very large volume of unstructured and semi structured data usually in Tera Byte and Peta Byte, a term BIG DATA has been coined.

	Type	Properties
1.	Private cloud	<ul style="list-style-type: none"> • Outsource or own • Lease or buy • Separate or virtual data center
2.	Community cloud	<ul style="list-style-type: none"> • Private cloud for a set of users with specific demands • Several stakeholders
3.	Public cloud	<ul style="list-style-type: none"> • Mega scaleable infrastructure • Available for all
4.	Hybrid cloud	<ul style="list-style-type: none"> • Combination of two clouds • Usually private for sensitive data and strategic applications

Figure 1: Cloud Computing Deployment Model

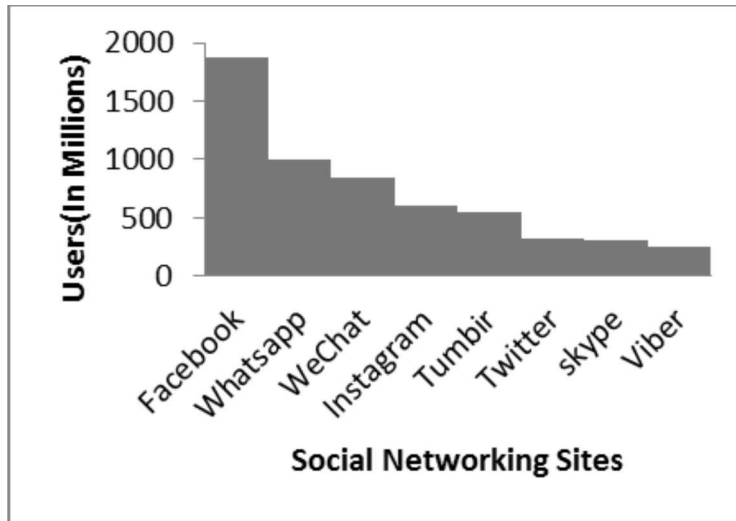


Figure 2: Users Vs SNS

The trend of processing a huge volume of data year by year is shown in figure 3.

Map-Reduce Frame Work presented by Google is gaining popularity for processing large amount of data on commodity hardware. Big Data may be characterized using the terms as tabulated in table 1.

II. Literature Review

During E-Business, privacy, confidentiality, non-repudiation, integrity and availability of the data needs to be ensured. In short proper data protection for avoiding data loss and for safeguarding the data from misuse, proper security mechanism is required to be implemented.

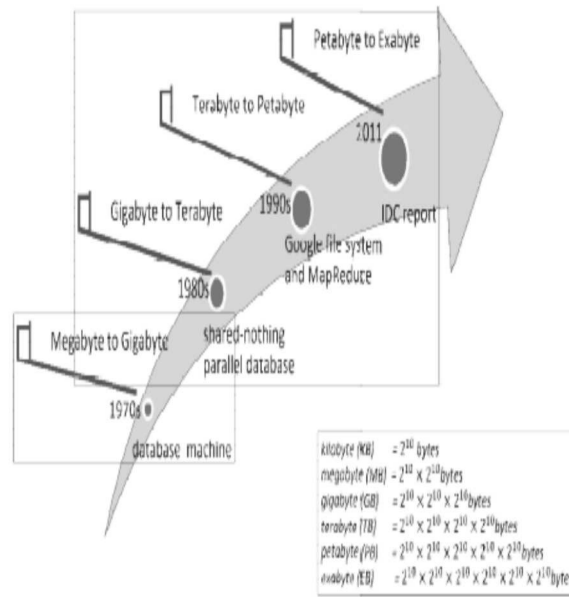


Figure 3: Increasing Data Trend

Table 1. Characteristics of Big Data

VOLUME	High volume of data such as streaming data, data collected through sensors.
VARIETY	Data in various formats such as Audio, Video, E-mail
VELOCITY	How fast data produced/processed to meet the required demand.

Data breach is the unauthorized practice by outside cyber attackers to penetrate the online defense of organization and exposing the data which were not expected to see or share. Data leakages may have the adverse impact of online business. It may damage the reputation of organization which may become the reason of losing the trust of company among customers. It can also form the basis to file a legal case against the breached firm by the customers. The consumers are the ultimate victims of data breaches, such cybercrimes may impose serious risks to the consumers for example the personal information stolen may be used to gain access to the account of consumer under attacker’s control which may be the reason of big money loss by misusing debit/credit cards. The insiders prevailing the privileged access to customer personal and confidential information may also indulge in unethical practice for undue monetary advantage. Personal digital privacy and protection from spyware

and malware are the major threats and without proper protection from these risks proper controlling and promoting online business will be a big challenge.

In fact it is the reality that due to transnational barrier protection policies and related laws are not imposed forcibly so controlling of unethical practices such as unauthorized access of others private data, misusing others private data are not very effective.

To commit illegal, unauthorized, malicious activity using the information and communication technologies is referred as E-crime.

III. Emerging Security Challenges

Spear-phishing attack exploits social engineering approach for improving the chance of success in the personal information stealing. In this approach the emails are tailored by the attacker to guess the interest of the targeted user and using those clues tries to

execute the financial scams. Suppose a user works in financial sector then possible approach may be to send the spear-phishing mail that looks as the new financial rules and regulations for stealing the personal information from the targeted users. Prior to such attack, it is thoroughly being analyzed the pattern, trends, profile and workforce of the targeted organization by the spear-phishing attacker.

Banking Trojans are the key areas for attackers to do malicious activities such as modifying banking online sessions and injection of extra fields for hacking the session or for stealing the sensitive details. The most common form of attack continues to be financial Trojans which perform a Man-In-The-Browser (MITB) attack on the client's computer during an online banking session.

While working in cloud computing zone it is quite common to get leverage of the technologies related with web services, web browsers as well as virtualization. Therefore the technical issues related with these technologies will get inherited into cloud computing infrastructure.

In cloud computing environment, initial interaction is through API and User Interfaces therefore they are the driving forces of cloud computing. Cloud API's IP address is vulnerable and unsafe between the cloud and user.

Shared technology issue in cloud computing is also one of the big issue as cloud provider provides services to several tenants so there is the need of multifactor authentication.

As availability of computing resources is needed to multitenant therefore it may be on the risk of DoS.

As Big Data Analytics is through open source software. So there may be quite possibility of fake data generation. There is lacking of security audits. It may also have the chances of mining sensitive information

Popularity of Social Networking, on one side created business in mobile technologies and communications on the other side it has created other security breach. It is a general tendency of a user that after purchasing electronic gadgets such as smart phones, they start downloading Apps for enhancing the functionality of smart phone. All the users of smart phones are not security conscious regarding their electronic gadgets. They do not so much of knowledge that which one app is fake. Fake app once installed, they start stealing personal information such as banking transaction data, debit/credit card details which may become the reason of severe financial loss Cyber. Attackers make apps similar to that of popular apps therefore for a user it is difficult to identify fake apps. Google has recently identified and deleted 29 fake beauty camera fake apps from its play store sharing pornographic contents. As per the report of Sofos Labs Fake app, through lucrative offers such as reward points, cash back, free data, interest free loans steal personal information for misusing it.

Noticeable facts about apps are enlisted in following table 2

Table 2. Apps Downloaded

1	197 Arab apps are downloaded in 2017
2.	21 lakhs apps are available on Apple store
3.	31 lakhs apps are available on Google play store.
4	Average 9 apps per day and 30 apps per month are used

IV. Counter Measuring Security Breaches

ARP (Address Resolution Protocol) and RTT (Round Trip Time) based sniffing detection tools can be easily used to counter measure the network sniffing attack. On the basis of information exchange, cooperative

and orchestrated association among various Intrusion Detection Systems, the DoS attack can be prevented.

First and foremost requirement for counter measuring security is the need of following guidelines of CSA

which better explains about the issue related with the leakage of customer data while using cloud computing infrastructure based on virtual network.

Up to certain extent the architecture ontology approach explained in several research papers is better option to counterfeit the security issues related with cloud based computing. It provides a controlled mechanism for secured storage, API and access management.

Need of imposing high ethical values and belongingness among employees so that data breach by internal attack may be minimized.

One should hover the links prior to click so that full length of URL may be checked whether it is known or unknown to the user. More Post, more vulnerable to external attacks, so only relevant posting on the net.

Good antivirus may be used for protection from fake apps, if someone want to know what data may be accessed then app brain detector app can be used. Android users and Apple users can download apps their play stores because they themselves controls the security aspects of the apps of their play store. Review of the app is also an alternate way of identifying fake app.

V. Conclusion

The dependency of the economy over IT cannot be denied. The efficiency and effectiveness of E-business is based on the fact that how are the resources utilized. Cloud based Computing is the best way of sharing the IT resources. In Education sector, cloud based computing are like a blessing for the enthusiastic learners who so ever have the issue of high investment on learning. The online commercial activities have the need of maintaing proper trade off between security

level and initial investment over it. Ethical and legal protection and in case any security issues are observed timmed resolution may create the more positive frame of mind toward acceptance of IT-Enabled Business Practices.

Big Data is promoted and geared up as open source; there is also the need of proper controlling and monitoring.

In Global Scenario, Authetication, accessing and revealing the confidential data needs to be taken in a non-compromising fashion. To hire and use of computing resources is cost effective but there is no surity that this data cannot be misused by third party consulting Cloud Computing Service providers. At the very first level it seems that while using cloud based services and even sharing confidential information if consumer is further not intersted to continue it then due to lack of clear legal cyber act the probability of being exploitation,blackmailing of cosumer based on sensitive data by service providers or by third party consumers associated with service providers , may increase.In fact there is a need of strict policy formulation and implementation on legal aspects about violation of data security or misuse of personal data resources available with cloud service providers, no matter whether consumer is further interested or not intersted to continue the cloud based services. Especially in legal context in Indian scenario the information technology act, cyber laws needs to be more strengthen in terms of misusing personal and confidential online resources available with cloud based computing environment. Otherwise Technology based, cost effective and transparent governance based on cloud computing will have no such remarkable impact in business growth as it is being expected from it.

References

1. Hasan Omar Al-Sakran, "Accessing Secured Data in Cloud Computing Environment", International Journal of Network Security & Its Applications (IJNSA) Vol. 7, No. 1, January 2015.
2. Mohammed J. Novel , "Approaches to Big Data Management" International Journal of Computer Science and Information Technology Vol. 3, Issue 1 January 2015 , pages 96-105
3. Subashini S, Kavitha V "A survey on Security issues in service delivery models of Cloud Computing". J Network Computing and Applications 34(1) (2011) pages 1–11.

4. D. Gollmann, "Securing Web Applications", Information Security Technical Report, vol. 13, issue. 1, 2008
5. R. Gellman, "Privacy in the Clouds: Risks to Privacy and Confidentiality from Cloud Computing," 2009.
6. Cloud computing use case discussion group, "Cloud Computing Use Cases: White Paper" available from: <http://cloudusecases.org>, 2012
7. Wang, W., Li, Z., Owens, R. & B. Bhargava, "Secure and Efficient Access to Outsourced Data", in Proc. of ACM Cloud Computing Security Workshop, 2009 pp. 55-65.
8. Yang, Ching-Nung & Jia-Bin Lai, "Protecting Data Privacy and Security for Cloud Computing Based on Secret Sharing", International Symposium on Biometrics and Security Technologies (ISBAST),(2013 pp 259 – 266.
9. Yang, Kan; Jia, Xiaohua; Ren, Kui & Bo Zhang,) "DAC-MACS: Effective data access control for multi-authority cloud storage systems", INFOCOM, Proceedings IEEE , 2013 pp 2895-2903
10. <https://www.zetta.net>
11. http://www.opengroup.org/cloud/cloud_for_business/p7.htm
12. <https://www.scnsoft.com>
13. Dainik Jagran, 10 Feb, 2019

Smart City : The Evolutionary Innovation of IOT

Stuti Kapoor

kapoorstuti05@gmail.com

Aparna Bhardwaj

aparna.bhardwaj.1993@gmail.com

Vishakha Sehdev Verma

vs4821@gmail.com

Ankit Verma

ankit.verma.aquarius@gmail.com

Abstract

In this paper, we will discuss about internet of things in smart cities. Nowadays it is the most effective and major concept in technology. This is mainly used in homes, transportation, parking, defence, shopping malls, hospitals, etc.

Key Words: Internet of Things, Sensors, Smart City, Smart parking

I. Introduction

The term “Internet of Things” was coined by Kevin Ashton in 1999 during his work at Procter & Gamble. The Internet of Things definition: “Sensors and actuators are embedded in physical objects which are linked through wired and wireless networks” [1]. Cities can be made more effective and efficient by the concept of Smart city technology. In the future, IoT-based technology will become so advanced that it will offer advanced levels of services and will change the complete lifestyle of people.

II. IOT (Internet of Things)

Internet of Things or IoT is a global network of physical devices connected to the Internet. These devices are equipped with sensors and other information transfer mechanisms [2]. The devices are usually combined with means of connection to control and/or processing units. Internet of Things theory states that inanimate equipment will be able to speak. It is massively scalable and efficient. It is based on machine-to-machine (also known as M2M) communication technology. It provides everyone with high transparency in almost every field of activity.

III. Working Process of IoT

There are four fundamental components of the IoT system, which tells us how IoT works. Figure 1 shows the Internal working of IOT.

- a) **Sensors/Devices** - First, sensors or devices collect very minute data from the surrounding environment. Then this collected data has various degrees of complexities. These complexities may range from a simple temperature monitoring sensor or a complex full video feed. Now this device has multiple sensors that bundle together to do more than just sense things [3]. For example, the smartphone has multiple sensors such as GPS, camera, accelerometer.
- b) **Connectivity** - Next, that data is sent to the cloud. The sensors/devices are connected to the cloud through various mediums of communication and transports such as cellular, Bluetooth, satellite, WiFi, low-power wide-area networks (LPWAN), or connecting directly to the internet via ethernet [4]. Users should choose the best connectivity option in the IOT system because every option has some specifications and trade-offs between power consumption, range, and bandwidth.

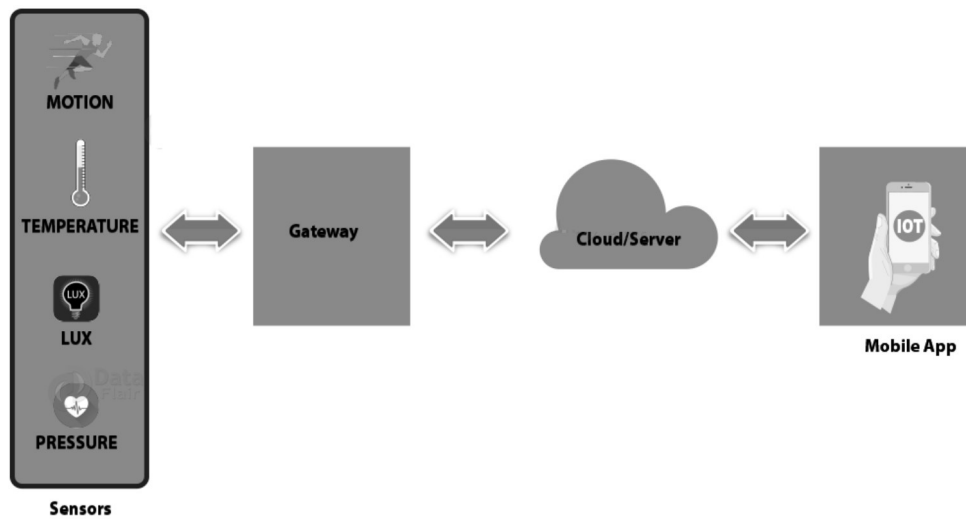


Fig. 1: Internal Working of IOT

c) Data Processing - Once the data gets to the cloud, the software performs processing on the acquired data. This includes checking the temperature reading is within an acceptable range [5]. Figure 2

shows the interaction between three components of IOT. It can sometimes also be very complex sometimes when it needs to as identify objects like intruders using computer vision on video.



Fig. 2: IOT Components – Data Processing

- d) **User Interface** - Now the information is made useful to the end-user in some way [6]. This could be via an alert by email, text, notification, etc. For example, a text alert can help to indicate when the temperature is too high in the company's cold storage. Not only alerts but some actions can be performed automatically by establishing and implementing some predefined rules.

IV. Benefits of IOT in Technology

- a) **More effective, data-driven decision making** – Benefits “big data” and connected devices have permit cities access to information that's never been available before [7]. A well-designed data analytics blueprint gives city officials the capability to access and anatomize a huge amount of information – and easily glean meaningful, actionable insights.
- b) **Scalability** – The smart city system has an adaptable structure, so the solution can be adapted to any size of cities.
- c) **Increased digital equity** – Smart city technology has the ability to create a more equitable environment for citizens [8]. To ensure digital equity uses specially-advanced cameras with high resolutions up to hundreds of great bits of the

picture, making ready clear and sharp images under all lighting conditions.

- d) **Hardware-driven VCA** – Detection of traffic violations and a facilitate of city video surveillance is provided by automatic Video Content through Analysis applications.
- e) **Improved infrastructure** – Smart technology can provide cities with forecast analytics to identify areas that need to be fixed before there is any infrastructure failure.
- f) **Improved transportation** – Technologies such as intelligent traffic signals improve traffic flow, relieving traffic jam during peak travel times. Other smart transportation technologies, such as smart parking management, allow cities to capitalize on additional revenue streams.

V. Applications of IOT in Smart City

a) Road Traffic

Smart cities assure that their citizens get from point A to point B as secure and efficiently as possible. To attain this, municipalities turn to IoT development and implement smart traffic solutions [9]. Smart traffic solutions work on different types of sensors, as well as fetch GPS data from drivers'

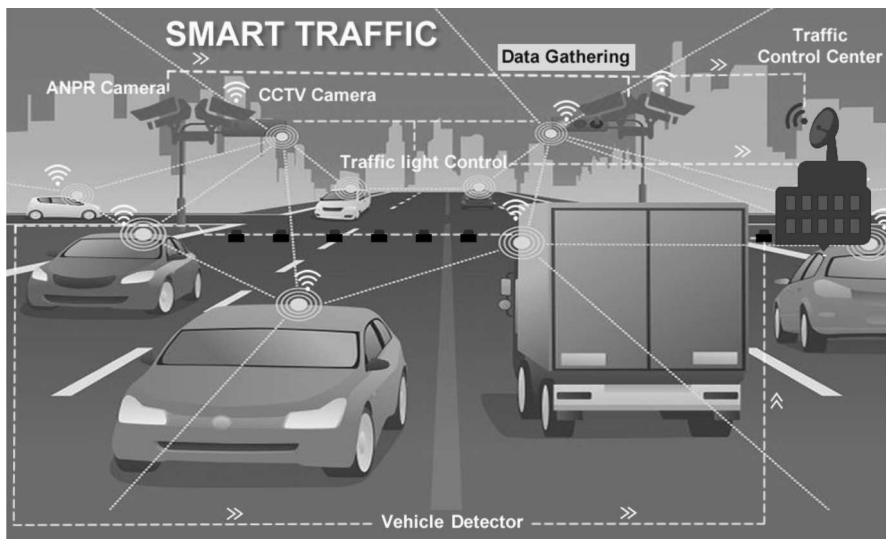


Fig. 3: Smart Traffic

smartphones to determine the number, location and the speed of vehicles. simultaneously, smart traffic lights connected to a cloud management platform permit monitoring green light timing and automatically change the lights based on the current traffic situation to obviate congestion. Figure 3 shows Smart Traffic. Additionally, using historical data, smart solutions for traffic management can predict where the traffic could go and take measures to prevent potential congestion.

b) **Smart Parking**

With the help of GPS data fetch from drivers' smartphones (or road-surface sensors placed in the ground on parking spots), whether the parking spots are occupied or available is determine by smart parking solutions and create a real-time parking map [10]. When the nearby parking spot becomes free, drivers receive a notification and go through the map on their phone to get a parking spot faster and easier instead of blindly driving around.

c) **Waste Management**

IoT-authorize smart city solutions help to advance waste collecting schedules by tracking waste levels and add route optimization and operational analytics. Firstly, each waste container notifies a sensor that gathers the data about the level of the waste in a container [11]. Once it is close to a specified threshold limit, the waste management solution receives a sensor record, further processes it, and sends a notification to a truck driver's mobile app. Thus, the truck driver deserted a full container, avoiding emptying half-full ones.

d) **Defence**

The IOT inclination has to expand defence utility: command and military intelligence and control system uses the multiple sensors that can be placed in the entire realm, allowing them to collect full situational acquaintance and control over diverse clash zones or battle areas [12]. The direction is towards development in urban scenarios where thousands of sensors could provide to military commanders with inflation situational awareness

and actionable intelligence to accomplish more powerful operations on the ground.

e) **Hospital**

Real-Time Remote Monitoring

IoT facilitates, connecting various monitoring devices and thus observe patients in real time [13]. Further, the signal scan sends out through these connected devices from home also, so it decreases the time needed for patient care in the hospitals.

- **Blood Pressure Monitoring** - A sensor-based smart system like Bluetooth allow coagulation system which can be used to monitor blood pressure levels of patients who go through hypertension. Such observing devices also help to decrease the possibility of cardiac arrests in critical cases.
- **Smart Pills** - Some pharmacy companies which have been making edible IoT like WuXi Pharma Tech, Proteus Digital Health, and TruTag, "smart" pills that help to observe health issues, medication controls, and adherence [14]. These kinds of Smart pills will support drug creation organizations to lower their risks.

f) **Home**

- **Security Systems** - While ability and conservation are exactly IoT advantages, its potential to have upgraded the control over home security is a fundamental focus. Smart locks, like Kwikset's Kevo, it is a Bluetooth enabled electronic deadbolt, and multiple connected home security systems, like an iSmart Alarm, iSmart Alarm offer multiple features including window and door sensors, video cameras and recording mechanisms and motion detectors [15]. All of these are connected to a mobile device and usable via the cloud, thus allowing you to access real-time information on the security status of your home.
- **Smart Appliances** - Smart refrigerators, such as LG's Smart Thin, enable you to scan grocery store receipts and make an inventory of your items and inform you if any item is about to expire [16]. More impressively, it recommends the recipes based on your refrigerator's items and informs when you

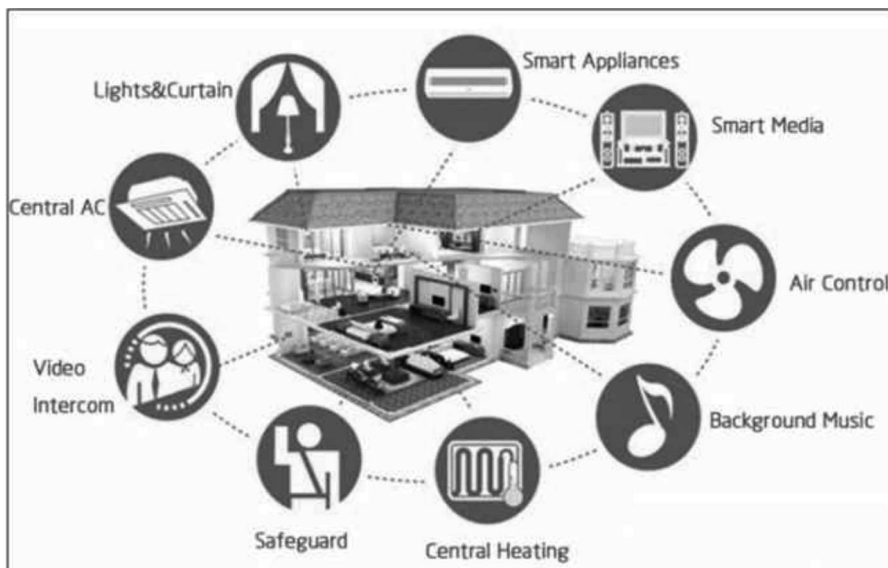


Fig. 4: Smart Homes

need to replace items. Smart ovens and smart refrigerators all are synch with your Smartphone and automatically set to the correct temperature according to the recipe selected from your database. While these appliance options seem a bit superficial and feature-based, there is also a conservation factor as well. Figure 4 shows different ways IOT can be useful in homes.

VI. Conclusion

IoT helps cities connect and manage multiple infrastructure and public services. From smart lighting

and road traffic to connected public transport and waste management – the range of use cases is highly diverse. What they have in common are the outcomes. Applying IoT solutions leads to reduced costs for energy, optimized use of natural resources, safer cities, and a healthier environment. However, to enjoy these benefits, municipalities should take a consistent approach to design functional and scalable smart city architecture. Well-designed, it will allow to reduce investments in IoT development and hasten the implementation of smart city solutions, still leaving space for expansion.

References

1. A. Verma and J. S. Prasad, "Performance Enhancement by Efficient Ant Colony Routing Algorithm Based on Swarm Intelligence in Wireless Sensor Networks," *Int. J. Wirel. & Mobile Comput.*, vol. 12, no. 3, 2017, pp. 232-238.
2. A. Verma and J. S. Prasad, "Optimum Path Routing Algorithm using Ant Colony Optimization to solve Travelling Salesman Problem in Wireless Networks," *Int. J. Wirel. & Mobile Comput.*, vol. 13, no. 2, 2017, pp. 131-138.
3. A. Verma and P. C. Vashist, "Enhanced Clustering Ant Colony Routing Algorithm Based on Swarm Intelligence in Wireless Sensor Network," *Proc. IEEE Int. Conf. Advances in Comput. Eng. & Applicat.*, 2015, pp. 150-154.
4. A. Verma and P. C. Vashist, "Evolution of Ant Colony Optimization and Swarm Intelligence in Wireless Sensor Networks," *Proc. DRDO Int. Conf. Inform. Tech. Yesterday, Today & Tomorrow*, 2015, pp. 286-289.

5. A. Zanella et al., "Internet of Things for Smart Cities", *IEEE Internet of Things J.*, vol. 1, no. 1, pp. 22-32, 2014.
6. J. Gubbi et al., "Internet of Things (IoT): A Vision Architectural Elements and Future Directions", *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645-60, 2013.
7. A. Gluhak et al., "A Survey on Facilities for Experimental Internet of Things Research", *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 58-67, Nov. 2011.
8. J. Jin et al., "An Information Framework for Creating a Smart City Through Internet of Things", *IEEE Internet of Things J.*, vol. 1, no. 2, pp. 112-21, 2014.
9. N. Mitton et al., "Combining Cloud and Sensors in a Smart City Environment", *EURASIP J. Wireless Commun. and Net.*, vol. 2012, no. 1, pp. 1-10, 2012.
10. I. Ganchev, Z. Ji, M. O'Droma, "A Generic IoT Architecture for Smart Cities", *25th IET Irish Signals & Systems Conf. 2014 and 2014 China-Ireland Int'l. Conf. Info. and Commun. Technologies*, pp. 196-99, 2013.
11. R. Ratasuk et al., "Narrowband LTE-M System for M2M Communication", *IEEE VTC-Fall*, pp. 1-5, 2014.
12. P. Papadimitratos et al., "Vehicular Communication Systems: Enabling Technologies Applications and Future Outlook on Intelligent Transportation", *IEEE Commun. Mag.*, vol. 47, no. 11, pp. 84-95, Nov. 2009.
13. J.-S. Lee, Y.-W. Su, C.-C. Shen, "A Comparative Study of Wireless Protocols: Bluetooth Uwb Zigbee and Wi-fi", *33rd Annual Conf. IEEE Industrial Electronics Society*, pp. 46-51, 2007.
14. X. Lin, A. Adhikary, Y.-P.E. Wang, *Random Access Preamble Design and Detection for 3GPP Narrowband IoT Systems*, 2016.
15. A. Botta et al., "Integration of Cloud Computing and Internet of Things: A Survey", *Future Generation Computer Systems*, vol. 56, pp. 684-700, 2016.
16. S. Sicaria et al., "Security Privacy and Trust in the Internet of Things: The Road Ahead", *Computer Networks*, vol. 76, pp. 146-64, 2015.

Three Level Secure App based Voting System using Biometric Linked with Aadhar Card for Universities

Sushma Malik

sushmamalikiitm@gmail.com

Geetali Banerji

geetalibanerji@gmail.com

Abstract

Election plays a vital role in Democratic Country. It is a process which allows the people to choose a perfect candidate who will lead the nation. A number of universities in India select their student representatives through election process. Universities spend lots of money on voting system to get a good candidate as the representative of the students. For better democratic system Universities need honest, transparent and fully secure voting system. In current scenario voting system is less transparent, due to which the chances of cheating are more during voting process. The main challenges of current voting system are authentication, security and protection of voters and voting data. That's why it becomes necessary to form a secure voting system. This paper proposes three level secure voting systems for University which is foolproof and designed to solve the problems present in existing voting system. This system will provide a single interface through which the voting process will be carried out smoothly. Through this system, a student can cast vote from anywhere in India. This system will also facilitate the live streaming of vote counts and it will save time. Students need not have to stand in a queue for voting and they can cast their vote without fear and hesitation.

Key Words: Aadhar Card, Biometric, Authentication, OTP.

I. Introduction

After getting the freedom from British government, Indian Government provided a right to Indian people to elect their leader. India, being one of the largest democratic countries in the world, where candidates are elected "for the people, by the people and of the people". A separate commission was introduced which named as Election Commission of India (ECI) for conducting and controlling the election process in country. The commission does not favor or support any political party. People, who are of 18 years and above age are eligible for the voting, ECI provides the voter ID cards to the people [11].

The voting system was introduced in the 18th century in India. Earlier it was Paper based and now E -Voting methods are used for voting. An E-Voting system is far better than the traditional voting technique in terms

of accuracy, convenience, flexibility, privacy, verifiability and mobility. In spite of improvement over traditional voting system Electronic voting systems has some shortcomings such as damage of machines due to lack of attention, mass update etc [5].

The following paragraphs discusses about these methods with their pro and cons:

A. Paper Based Voting

In this method, the ballot is used in the electoral system. The person whose age is 18 years and above, cast their vote in a piece of paper and elect the candidate [1].

Drawbacks

- Large number of man power is required.
- Time consuming process.

- Very difficult to collect and transport the ballot boxes to main centers.
- Chance of invalid voting.
- Errors can occur during counting process.
- More security is needed to complete this process in a secure manner.

B. E-Voting

E-voting means electronically voting without the use of paper and ballot boxes. In this method, data are stored in the electronic machine which allows the people to press a specific button to elect the desired candidate. The information stored here is in digital format [1].

Drawbacks

- One can modify the program of EVM and can change the result after polling is completed.
- Chance of illegal voting by using the EVM illegally.

This paper is organized as follows: Section I discuss about the voting system, Section II is about Literature Survey, Section III discusses about the problems faced by the current system, Section III introduces the proposed system and its salient features. Section IV discusses about future scope and the last section concludes.

II. Literature Survey

Sathya et.al. [1] Proposed the biometric voting system based on Aadhar card. The main objective of this system is to eliminate bogus voting and vote repetition, less election expenditure, more transparency and fast results.

Ankita et.al. [2] said that their project helps to increase voting percentage. In their voting system authentication can be done using fingerprint recognition to cast voter's votes. For the successful implementation of this system is very difficult; because it involves political issues, financial issues, and regional issues.

Sayali Shinde[3] proposed secured Online voting system uses AADHAR card and Voter Id for authentication. Their main proposal is to enable the

user to cast his/her vote using OVS without going to booth. User can cast his/her vote from his/her home or any way and to reduce the proxy vote and in booth capturing situation this system help us.

According to Jitendra Waghmare[4] the main purpose of system to make EVM machine authenticate with fingerprint identification and safe and secure AADHAR CARD based voting system to avoid misunderstandings which are going to takes place in elections. In this system recording and counting of votes will be faster, more accurate and less man power.

Nishigandha C [5] designs a voting system that helps everyone to cast their votes without any hesitation. This system also increases the voting percentage without manual counting. This system is more efficient than the existing one. They provide the Online Voting system application on android device. The security is in terms of providing the one time password (OTP).

S. Karthikeyan[6] said that their Electronic voting systems have many advantages over the traditional way of voting. The advantages are lesser cost, faster tabulation of results, improved accessibility, greater accuracy, and lower risk of human and mechanical errors.

D. M. Ujalambkar[7]design a system that can overcomes all the drawbacks of the ordinary voting machine and also provides additional security. Firstly this system checks the voter's age is 18 or not by using Aadhar database to identify invalid voters. One voter cannot vote more than once since system marks voters Aadhar id as Voted after he cast his vote. This system uses Aadhar id, therefore, there is no need for voter id. System use fingerprint of the user for the verification because fingerprint of every person is unique and hence system completely reduces the chances of rigging.

Tabish Ansari[8] said that their proposed system will automatically increase the transparency. It's more reliable and provides better scalability for large elections online voting is also an excellent mechanism that does not require voter location proximity. It leads to the easier verification of voters and candidates with Aadhar.

N.N.Nagamma [9] developed touch screen based EVM. It gives confidence in voting system; only the

legitimate voter is allowed to gain access to voting. The system is user friendly, in the sense that the user can easily understand the system even though the user does not know previously.

B. Madan Mohan Reddy [10] proposed the project that provides safety and security in voting process with RFID based Biometric Voting Method. And provides safety from alcoholic persons and Maoists who comes to polling booth to blast polling booth. System is linked to AADHAAR and biometric authentication to avoid vote's duplication. There is no scope to take place rigging in elections.

P.P. Shendage [11] developed a system which absolutely add to the accuracy of casting of votes and would nullify the chances of malpractices as it involves a total full proof electronic system. The system is also reliable in the sense that it can be used and implemented with ease and is flexible in nature. The main advantage of the system that the vote can be casted from any constituency, that is irrespective of the present location of the citizen, along with the authentication of voter with distributed server approach. The system can be used by the government of India to replace the currently used electronic voting system.

Saravanan.N [12] proposed the Iris based E-voting system using Aadhar database and this method is safe and secure. Iris is one of the unique identities of a human being this is being used in the aadhar system. By using arduino software iris of every individual is being captured. And compare it with Aadhar database. By using this method the voter will ensure if his/her vote has gone to correct candidate/party. The votes are going to be done automatically.

M.Abinaya [13] planned an internet electoral system that is healthier and quicker than previous systems. The new system prevents access to ill-gotten voters, provides simple use, transparency and maintains integrity of the option method. The system conjointly prevents multiple votes by a similar person and checks eligibility of the citizen. It conjointly permits someone to vote from anyplace given that the citizen is inside electoral limits.

R. Murali Prasad [14] proposed a online voting system and this system is managed in a easier way as all the

users should login by Aadhar card number and password and click on his/her favorable candidates to cast the vote. This features a larger security in the sense that voter high security password is confirmed before the vote is accepted in the main database of ECI. The extra feature of the model is that the voter will ensure if his/her vote has gone to correct candidate/party. The votes are going to be done automatically, therefore saving an enormous time and facultative ECI to announce the result at intervals a very short period.

III. Existing Problems of Voting System

- People do not vote as they are not in their home constituency on the Election Day: Most of the people of age between 18 to 25 years are students and some time their college is far away. At that time they don't want to spend 2-3 days to travel and large amount of money to cast a single vote. Similarly the people of age between 25 to 40 years are away from their native state because of their jobs and working area like Bangalore, Mumbai, Chennai and Delhi etc. from where they are not able to come home on festivals. Than how we can expect they come to cast a vote at that day. Some time it is the main reason for the low percentage of voting [2][3].
- Illegal Voting: It is a very common problem of every election procedure. One candidate casts the votes of all the members or few amounts of members in the list illegally. This can be done externally at the time of voting [2][6].
- Verifiability: it is not possible to verify that all votes have been counted correctly [6][4].
- No transparency: A voter could not check what happened to his/her vote i.e., whether it has been properly recorded in the system database or not [7].
- Small chips attached to the EVMs that can be controlled by fraudulent through radio waves or infrared can alter or manipulate the functioning of the machine leading to alter the vote results [7].

Indian Universities are still following the traditional method i.e. paper based voting. As we have already

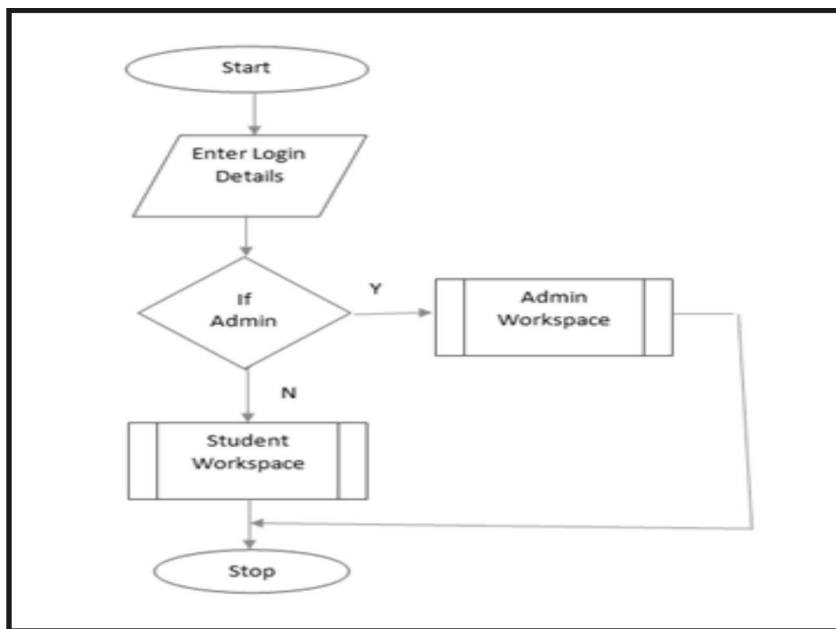


Fig. 1: Main Interface

discussed the problems faced by the paper based methods and EVM. This paper proposes App based system for the voting process in Universities which can be proved very beneficial for the students to cast their vote hassle free. The proposed system is a three level secure App based Voting System linked with biometrics and AADHAR card which is mandatory for all Indians.

IV. Three level secure App based Voting System for Universities

The system follows a three level security to make the voting process robust. It has Interface for two types of users: administrator and student. Initially the system asks to specify the type of user (Figure 1): administrator or student. In case of administrator after verification of the credentials the administrator interface will be made available. Administrator interface has following features (Figure 2):

- Add Candidate
- Add Voter
- Activate Voting period
- Count number of votes casted

- Declare Result

In case of student, the student gets the Student Interface (Figure 3) which asks the student for left thumb impression and Aadhar number. It is matched with the Aadhar database containing the scanned thumb impression and Aadhar number. If both the input matches, then the system verifies the age. The voter is eligible if and only if the age is 18 or above. The status of the voter recorded in the system will determine whether the candidate has already voted or not. If the voter tries to recast, the system will refuse. Once it makes sure the voter is a genuine voter, it asks for the telephone number where the OTP can be send. If the OTP provided by voter is correct then he/she can vote by clicking on the candidate's image and the control will exit from the system immediately. The voting will be active for the declared period by the government, beyond that duration the system will not allow to accept the input.

Advantages:

-

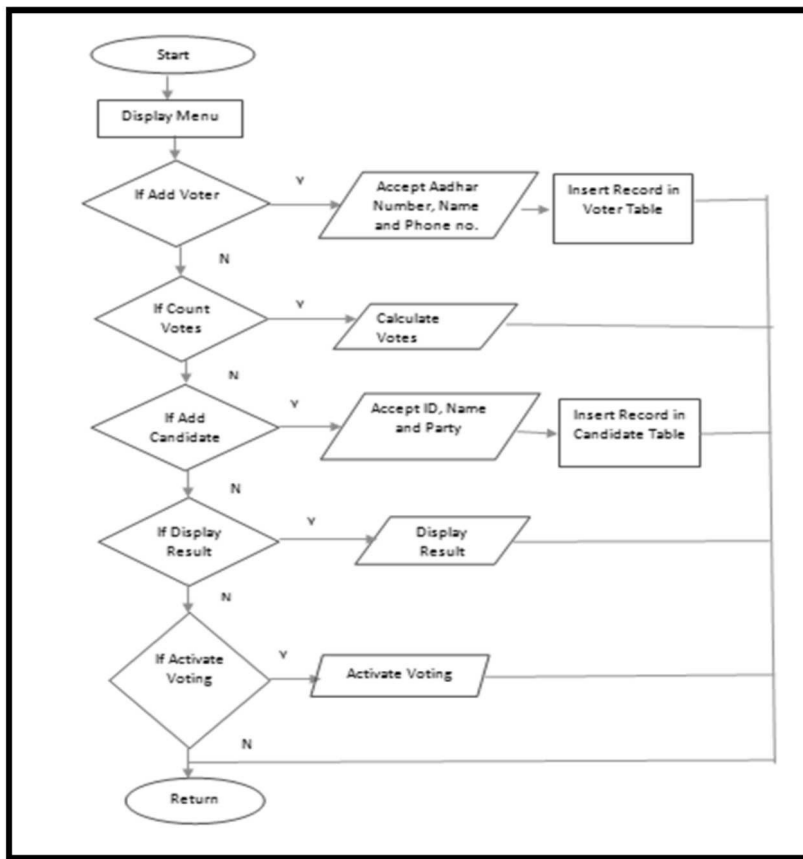


Fig. 2: Administrator Interface

- Authentication of a student is done on the basis of fingerprint and Aadhar number so there is no scope of illegal voting.
- Once a person has casted the vote, he/she will not be allowed to vote.
- No Manpower is required.
- Declaration of Results can take place immediately after the voting period is over.
- Nobody can change the software/data.
- Transport related problems at the time of voting day will be reduced.

Disadvantages:

- Each student must have smart phone with high speed internet connection.

Future Scope

- Android based app is to be developed.
- This system can be implemented for election at government level.
- The system should further have interface for various Indian local languages. It will become beneficial for the voters who belong to rural area and will be able to cast their vote in their mother tongue.

V. Conclusion

This paper discusses about the voting process of India in past and present. It elaborates the problems present in earlier and existing voting system. The paper introduces a three level secure App based voting system linked with Aadhar Card and biometric (left thumb

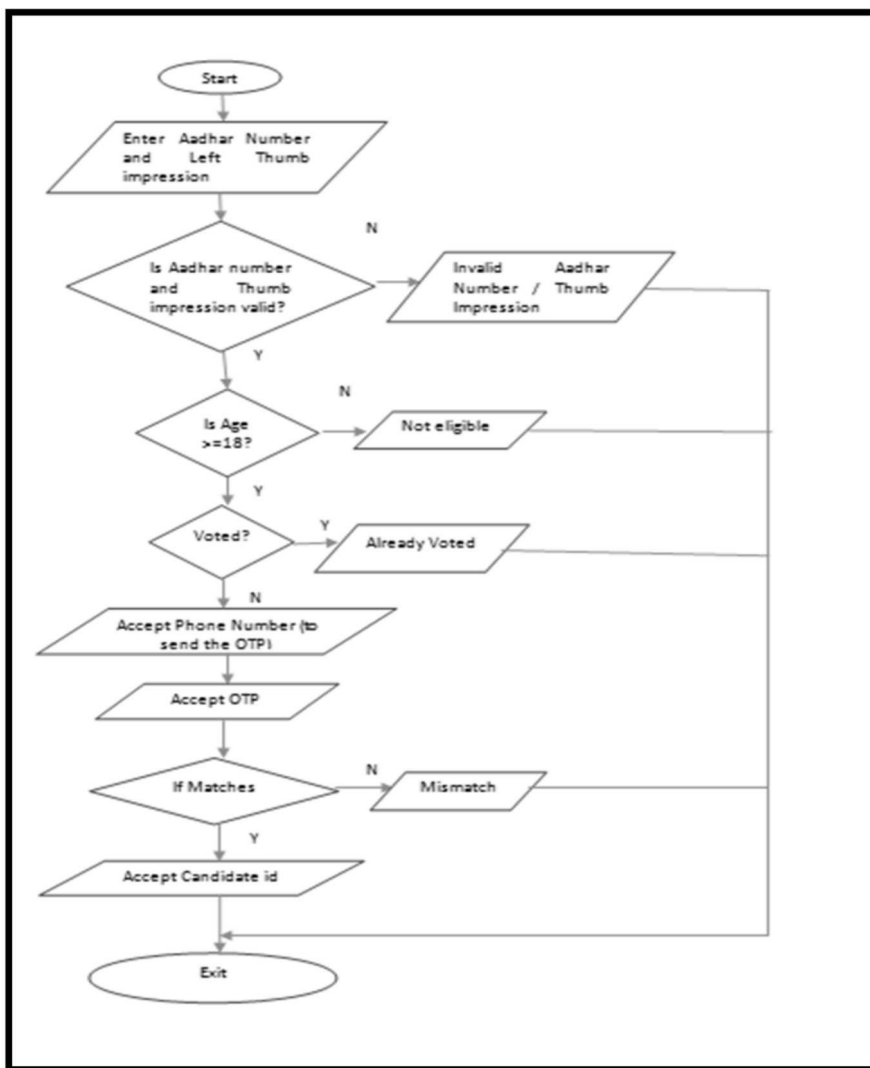


Figure 3: Student Interface

impression) and provides additional security with the onetime password (OTP) on the mobile number. Through this voting process, everyone can cast their votes without any problem and it will also increase the percentage of voting. Counting process is time consuming in case of Manual/EVM but with this App,

it will be made available instantly as soon as the voting period is over. It is expected that the proposed online voting App will increase the transparency in the voting process without human intervention. It can be very well envisage that the future of election at government level will be based on mobile app only.

References

1. Sathya.G, Jeevanantham.C, Sangeetha.K, Venmathi.V, Ramya.P, "Bio Metric Authentication System Based On Aadhar Card", International Journal of Pure and Applied Mathematics Volume 117 No. 9 2017, 7-11 ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version).

2. Ankita R. Kasliwal, Jaya S. Gadekar, Manjiri A. Lavadkar, Pallavi K. Thorat, Dr. Prapti Deshmukh, “ Aadhar Based Election Voting System”, IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, PP 18-21.
3. Sayali Shinde, Nikita Sherla, Sampada Zadgaonkar, Sapna Hulsurkar, S. P. Mone, “Online Voting System Linked with AADHAR Card”, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 6, Issue 3, March 2018, ISSN(Online): 2320-9801 ISSN (Print) : 2320-9798.
4. Jitendra Waghamare, Dhananjay Manusmare, Bhagyashri Dongre, Manisha Bannagare, Chahande, Kunal Sontakke, “Implementation of Aadhar Card Based EVM Machine with GSM Module”, IJES, 2017.
5. Nishigandha C, Nikhil P, Suman P, Vinayak G, Prof. Vishal D, “Online Voting System Using Aadhar Card and Biometric”, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 6, Issue 5, May 2017.
6. Dr. S. Karthikeyan, J. Nithya, “Secured Electronic Voting Machine Using Biometric”, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Vol. 2, Issue 2, 2017, ISSN: 2456-3307.
7. D. M. Ujalambkar, Kashinath Raut, Kiran Chavan, Ashish Bhingardive, Yogesh Kolape, “Secured Aadhar Based Electronic Voting System”, International Journal for Research in Applied Science & Engineering Technology, Volume 5 Issue XII December 2017, ISSN: 2321-965.
8. Tabish Ansari, Brijesh Chaurasia, Niraj Kumar, Nilesh Yadav, Sonalii Suryawanshi, “ Online Voting System linked with AADHAAR Card”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 6, Issue 9, September 2017, ISSN (Online) 2278-1021 ISSN (Print) 2319 5940.
9. N.N.Nagamma, Dr.M.V.Lakshmaiah and T.Narmada, “Aadhar based Finger print EVM System”, International Journal of Electronics Engineering Research, Volume 9, Number 6 (2017) pp. 923-930, ISSN: 0975-6450.
10. B. Madan Mohan Reddy, D. Srihari, “ RFID Based Biometric Voting Machine Linked To Aadhaar For Safe And Secure Voting”, International Journal of Science, Engineering and Technology Research (IJSETR) Volume 4, Issue 4, April 2015.
11. P.P. Shendage, P.C.Bhaskar “ Distributed Server Approach for Novel E-Voting System with Biometric Authentication Using Aadhaar Card”, International Journal for Scientific Research & Development| Vol. 5, Issue 03, 2017 | ISSN (online): 2321-0613.
12. Saravanan.N, Pavithra.K, Nandhini.C, “Iris Based E-Voting System Using Aadhar Database”, International Journal of Scientific & Engineering Research, Volume 8, Issue 4, April-2017, ISSN 2229-5518.
13. M. Abinaya, K.Gowthami, M.Abinaya, “Online Voting System Powered By Aadhar Authentication”, IJARIII, Vol-3 Issue-2 2017, ISSN (O)-2395-4396.
14. R. Murali Prasad, Polaiah Bojja, Madhu Nakirekanti, “ AADHAR based Electronic Voting Machine using Arduino”, International Journal of Computer Applications, Volume 145 – No.12, July 2016, ISSN: 0975-8887.

Block Chain

Pragya Agarwal

pragyaagr1999@gmail.com

Himani Verma

vermahimani828@gmail.com

Govind Murari Upadhyay

govindmurari.upadhyay@gmail.com

Abstract

Block chain is the technology which is being used to develop the Cryptocurrency, like bit coin. As a section of the 4th industrial revolution and the information technology, block chain technology can be applied in different fields such as economic, judicial and education sector. This paper focuses on its capability in education system and Cryptocurrency and also explores how block chain technology is being implemented to resolve some education and payment problems. This article introduces what is block chain, how its work, features, benefits of block chain technology and few of its current applications in Cryptocurrency and education system.

Key Words: bit coin, block chain, Cryptocurrency, hash function ledger, nodes, wallet.

I. Introduction

Traditional supply of the chain system has become the heritage system with the development of the technology.

Use of the Excel sheets to upgrade data of the recent stage in the supply chain has to be replaced more by the automatic system requiring less human intrusion. Automation its own self calls for a system that holds and manages data with the least possible use of a human. The current systems involved in the supply of chain management system are indeed, websites that organized upon the company servers that hold the key data relating to the up to date and present undergoing processes of each phase in the supply chain. People working on the different stages have to use the same website threshold to update or fill the data that would complete the metadata of the process. This can be explained by a sample of a manufacturer's intranet application that reserves a spot for people to update or upgrade data in different sectors like raw material stock to the export to retailer. All the data mentioned in the example flows to and from one server to other and data archived that exists in the company.

The centralized server and the intranet application depend on them are localized ownership of that company only. The company is just one of the large numbers of nodes, the supply chain tree has and the centralized intranet application concern to a particular company and what it does is the point of interruption of supply chain.

II. What is Block CHAIN?

You must have followed banking, investment, or Cryptocurrency over the recent years, you may have knowledge of the block chain, and it is the data storage technique after the bit coin. To know more about the block chain, you may have encountered with a definition "block chain is a distributed, decentralized, public ledger.

Block chain technology is called as distributed ledger technology. It allows its participants to protect the execution of transactions, initiate the transaction and transfer of the resources at a low-amount. In this technology transactions are gathered into blocks and stored in a cryptographic chain blocks and in a Sequential manner and permit the deriving ledger to be used by various servers.

III. What is Ledger?

Definition of Ledger is— it is a book of record that keeps all the financial transactions of the organization like a register.

Distributed ledgers make the use of individual computer to store, share and synchronize transactions in their corresponding electronic ledgers. Block chain organize data into blocks, which are bind together in an adjoin mode.

There are 2 types of ledger system mention below:

A. Centralize Ledger

A centralized ledger also called as general ledger. A general ledger is the main center of any accounting system which holds all the financial and non-financial data of an organization. All the systems are linked to the main center point which holds all the data.

B. Distributed Ledger

Distributed ledgers use individual computer to record, share and synchronize transactions in their corresponding electronic ledgers. Block chain organize data into blocks, which are linked together in an adjoin mode.

IV. History of Block Chain

Before its use in Cryptocurrency, it had humble beginning as a thesis of the computer science — specifically, in the area of cryptography and the data structures.

The first form of the block chain was the hash tree, is called as a Merkle tree. This data structure was licensed by Ralph Merkle in 1979, and functioned by verifying and managing the data between computer systems. In a peer-to-peer system of computer, accurate data was important to make sure nothing was modified during transaction. It also helps to ensure that wrong data was not delivered. Primarily, it is used to maintain and prove the nobility of the data being shared.

In 1991, the Hash tree was used to develop a – protected chain of blocks – a sequence of data records, each connected to the previous one. The latest record in this chain should contain the history of the thorough chain and the block chain was created.

In 2008, Satoshi Nakamoto gave the idea of the distributed block chain. It would hold a secure history of data interchange, use the peer-to-peer network to timestamp and verify every exchange and that could be run autonomously without a third party. This has become the foundation for the Bit coin

V. Characteristic

Its basic features are summarized as follows:

- **Decentralization:** Each distributed node in the block chain system is descents distributed and trust each other through mathematical formula rather than compulsory central management organization, so as to make sure that each node has equal rights and responsibility.
- **Security and Transparency:** Two asymmetric codes, Public and Private Key are used in the encryption process. Integrated with the hybrid encryption mechanism and Data Encapsulation Mechanism, the security requirements and its ownership verification needs are satisfied.
- **Smart Contracts:** The block chain has a transforming programming system which supports each node to build an efficient and intelligent contract. Intelligent contract automatically judge the situations and function of each node to execute the agreement, and automatically runs the contract items that meet with the conditions.
- **Verifiability:** Block chain system uses timestamp technology to increase time dimension for data information to ensure that the data information storage and exchange can be verified.

VI. How Block chain WORKS?

For understanding how block chain, firstly we need to perceive the meaning of **Hash Function**.

A. Hash Function

Let just imagine that there are 10 persons in a room, has decided to create a different currency. They have to follow the flow of the resources and one person – let's say him Rahul has decided to keep a record of all activities in a register.

One person –say Jay has decided to lift money and to hide this, he has to change the entries in the register.

Rahul noticed that someone had hamper with his register. He decided to terminate this activity. Then he founded a program called a Hash function that converts the text into a set of numbers and letters.

Like: 18c177926650e5550973303c300e136f22673b74.

A hash is a line of numbers and letters, generated by hash functions. A hash function is a mathematical function that gets input of a varying number of characters and transforms it into a string with a constant number of characters. Even a slightly small change in a string will generate an entirely new hash.

After every record, he recorded a hash. The new register was follow.

Jay decided to alter the entries again. So he checked the register and modified the record and created a new hash.

Rahul observed that someone had again passed the entry through the register. So, he decided to make the record of each and every transaction complex. After each record, he had added a hash generator from the record+last hash. So each entry depends of the last entry.

If Jay tried to alter the record, he has to modify the hash in all the other previous entries. But Jay really wanted some more money, so he had spent the full night computing all the hashes.

B. Block chain Working

Each block in a block chain system contains some information along with the hash of its last block or record. If the data inside the block is changed then the hash of the block will also be changed or modified. The relation of the blocks to its unique hash key makes the block chain secure. While transactions took place in a block chain, there are nodes in the network that authenticate these transactions. In block chain, these nodes are called as miners and they used the idea of proof-of-work in order to work and validate proceedings in the network. In a series, for a transaction

to be authentic, each and every block must refer to the hash of its last block. The transaction will execute only, if the hash is absolutely correct. If a hacker attempts to hack the network and modify information of any particular block, the hash linked to the block will also get changed.

The violation will be recognized as the modified hash will not match with the original hash. This ensures that the block chain is transformable but if any alternation is being made to the chain of blocks will be reflected throughout the complete network and will get notice easily. How block chain allows transactions to take place:

- A block chain system makes the use of public and private keys in order to create a digital signature to ensure security and permission.
- Once the identity is assured through the keys, the need for authorization arises
- Block chain allows participants in the network to complete the mathematical authentication and reach an accord to agree on and a particular value.
- While making a transaction, the sender uses their private key and announced their transaction information over the network and the block is created consisting of information such as digital signature, timestamp and the receiver's public key.
- This block of information is podcasted through the system and the verification process starts.
- Miners in the network initiated solving of the mathematical code related to the transaction to retrieve it. For solving this code, the miners are required to invest their computing potential.
- On solving the code, the miners receive benefit in the form of bit coins. These types of problems are termed as proof-of-work mathematical problems.
- Once the majority of systems in the network come to accord and agree to a common solution, the block is time stamped and inserted to the last block chain. This block can consists of anything from money to data to messages.
- After the new block is being inserted to the chain, the last copies of block chain are altered for all the systems in the network.

C. Proof of Work

The lay down of the transaction in a block is known as the proof of work challenge, and is carried out by special nodes called miners.

Proof of Work is a system which required some of the work from the service requestor. Producing a proof of work is a non-systematic process with less probability, so normally many trial and errors is required for a verified proof of work to be generated.

D. Miners

Asserts new transactions and record them in the distributed record. Miners participate to solve the tough mathematical problem on the basis of a cryptographic hash algorithm. The solution got is called the Proof-Of-Work. When a block is 'solved', the transactions are considered to be committed.

Miners receive the reward when they resolve the highly complex mathematical problem. There are two types of rewards: new bit coin or transaction fees. The amount of bit coins created decreases every 5 years. Today, a latest created block can create 12.5 bit coins. This number will keep on decreasing until no bit coin will be supplied. This will occur around on 2140, when approximately 21 million bit coins would have been created.

VII. Block Chain in Cryptocurrency.

Cryptocurrency is the system which allows the secure payment of the online settlements that are break down in terms of a virtual "token," that represents the record entries to the system itself. "Crypto" refers to as the fact that different cipher algorithms and cryptographic techniques, like public-private key pairs and hashing functions are utilized.

The first ever block chain based Cryptocurrency was Bit coin, which still persists as the most popular and the most worthwhile. Currently, there are many alternate crypto currencies with different functions and specifications such as Lite coin, Namecoin and Peercoin, etc. Some of these are duplicates of Bit coin while others are separated.

A. Benefits of Cryptocurrency

Cryptocurrency hold the promise of making it simple

to carry the funds directly in between the two parties in a transaction, without the involvement of a trusted third party such as bank or Credit Card Company. These transfers are implemented through the use of the public and private keys for security reasons. In this modern cryptocurrency system, a user uses "wallet," or some account address and has a public and a private key which is being used to sign transaction. Fund transfer is done with least processing fees, allowing users to avoid the high fees charged by many banks and the financial institutions for wired transfer.

B. Drawbacks Of Cryptocurrency

The semi-unidentified behavior of the cryptocurrency transaction makes them capable for a host of criminal activities, such as money concealment and tax evasion. Some cryptocurrencies are more exclusive than others. Bit coin, is a comparatively bad choice for controlling illicit business online, and forensic study of bit coin transactions will the help the high authorities to arrest criminals. The other aspects of a cryptocurrency environment are not safe to the danger of hacking. In Bit coin's history, many online transactions have been the subject of hacking and stealing of blocks.

VIII. Block Chain In Education Sector

In all of the world's developed countries, the field of education has been a primary focus. That is because all the upcoming areas like agriculture, science and any every other area is dependent on education only.

The Block Chain is spreading its influence not only in the field of digital currency but also in other areas as well. So let us see the various applications of Block Chain in education sector.

A prominent educational use is to hold the records of the achievement and the credits, such as gradation and certificates. The certificate information would be added to the Block Chain by granting institution which, the student can retrieve, share with others, or link from an online resume. It provides a perpetual public data, safeguard against modifications to the institution or loss of its personal data.

The Block Chain provides public proof that a student can receive an award from an institution, but does not

of itself; authenticate the reliability of either party. A university can grant a bogus certificate or a student can still cheat in exam. The Block Chain resolves this problem of swiftly and reliably checking the happening of an event, such as awarding of a degree, but not its authenticity. It minimizes the chances of making fake diplomas or certificates and thus plays a major role in validating the identity.

A. Issuing of Certificate:

When block chain is used in the issue of certificates, then the credentials can be checked without the involvement of the III party or intermediary, but we can add value to the already existent digital certification ecosystem. By issuing certificates on block chain, a student can receive the certificate privately without the consultation of involvement of the third party.

B. Certification of Credentials:

It focuses on mutual relations and it is possible to obtain them via “peer to peer” processes, better known as P2P. It is based on the reality that the students with whom the assignment is done can give credentials to their peers, certifying some skill enhanced during the group learning. We come to know about the “open badge”, which can support the knowledge and capability acquired during the course.

C. Uses and Prospect of Block Chain in Education:

Many educational bodies have taken a step forward on improvising the sphere, enabled by current technologies.

- The Massachusetts Institute of Technology (MIT) has issued hundreds of digital certificates and diplomas for its students through the usage of Block chain.
- University of Nicosia, the largest private higher education institute in Cyprus implemented a full-functioning Block Chain library for storing all the information about students’ grades, their diplomas and certification.
- Federation of State Medical Board has introduced a program to entitle their students to get their certificates digitally over their course of training to reduce falsification of diplomas.
- Sony Global Education has developed educational schemes in collaboration with IBM that uses Block Chain technology to secure and share students’ records.
- Publishing: Block Chain have many applications in publishing industries, from breaking into the industry to rights management to stealing other person’s work. New opportunities are uplifting the playing field of writers and motivate the association among writers, editors, translators and publishers.
- Another reason for developing this technology in education is the growing vogue of online courses. Institute that aim internet users are generally forward-looking in terms of new technologies that bring them closer for the practice of Block Chain.
- Block Chain removes intermediaries: Use of Block Chain also eradicates the intermediaries. This further reduces any possibilities of biasing as the results will not be shared with any third party who may try to alter the data.
- Block Chain makes dematerialization of documents possible: Since there is no paperwork involved and no central party will be required, this will help in reducing the unnecessary tariff and also saves time and human labor.
- Block chain maintains the users’ privacy by presenting all the information recorded within it by ID numbers. User can access his personal information through it by uniquely defined private key. The other people cannot access this information until or unless he has this private key. This is how Block chain ensures privacy and security of the users’ information.
- Thus, this technology is capable to transform the education sector and with new applications being developed every day to capitalize on this opportunity. The benefits of Block Chain are wide-reaching ranging from managing the monetary through to Block Chain enabled education platforms. It’s an exciting area of emerging technology.

IX. Advantages

- Decentralized data distribution
- Increased transparency
- More flexible system
- Increased data availability
- Improved watch over the data flow
- Enables easier trace back to find flaws in products or processes

X. Conclusion

Block chain is majorly a distributed ledger technique, which uses the cryptographic mechanism and distributed algorithms to develop the characteristics of decentralization, attributable, permanent, and financial assets. Block chain technology allows the Bit coin cryptocurrency. It is a decentralized ecosystem

for transactions, in which every transaction is stored to a common ledger which is being visible to everybody. The aim of the Block chain is to give security, privacy, and clarity to all the users. However, these feature set up many technical challenges and limitations that should be concern. Its cryptocurrency properties have the future to activate many innovative ideas in education. It can hold a whole, trustworthy set of records of all educational activities including the activity and results in authorized and unofficial learning environment too. It can also hold information about teachers and their behavior till the performance thus providing a reference for teaching evaluation. In other words, for both students and mentors, block chain has major prospective applications in comprehensive design, behavior recording, and analysis. Simultaneously, it brings out many scopes for researchers, developers, and educators.

References

1. Chapter 10The Block chain in IoT Carlos Davila and Jacob Tarnow
2. 2018 3rd International Conference for Convergence in Technology (I2CT) the Gateway Hotel, Pune, India. Apr 06-08, 2018
3. XU Hao1+, SHI Xiao-Hong2, YI Dian 3 123(Information engineering College, Shanghai Maritime University, Shanghai 201306, China)
4. Cryptocurrency REVIEWED BY JAKE FRANKENFIELD Updated Dec 10, 2018
5. Exploring block chain technology and its potential applications for education Guang Chen1,2, Bing Xu1, Manli Lu1 and Nian-Shing Chen3* Chen et al. Smart Learning Environments (2018) Proof of authentication:IoT-friendly blockchains Deepak Puthal and Saraju P. Mohanty
6. How Does the Block chain Work? BLOCK CHAIN TECHNOLOGY EXPLAINED IN SIMPLE WORDS MICHELE D'ALIESSI FOLLOW How does block chain work in 7 steps — An easy explanation. JIMI S.FOLLOW MAY 6, 2018
7. Block chain: What is Mining? DAMIEN COSSET JAN 4 '18

Recommender Systems: A Detailed Analysis of Current Existing Protocols

Kanika Bhalla

kanikajethwani@gmail.com

Tanvi Sharma

tanvisharma1074@gmail.com

Shivangi Kharbanda

shivangikharbanda4@gmail.com

Abstract

A recommender system is associated data Retrieval technology that improves access and proactively recommends relevant things to users by considering the users' expressly mentioned preferences and objective behaviors. A recommender system is one in every of the foremost techniques that handle data overload drawback of data Retrieval by suggesting users with acceptable and relevant things. Today, many recommender systems are developed for various domains but, these aren't precise enough to fulfil the knowledge wants of users. Therefore, it's necessary to make prime quality recommender systems. In planning such recommenders, designers face many problems and challenges that require correct attention. This paper investigates and reports this trends, issues, challenges, and analysis opportunities in developing high-quality recommender systems. If properly followed, these problems and challenges can introduce new analysis avenues and therefore the goal towards fine-tuned and high-quality recommender systems may be achieved.

Key Words: Recommender systems, protocols

I. Introduction

On the net, wherever the quantity of selections is overwhelming, there's ought to filter, place and with efficiency deliver relevant info so as to alleviate the matter of data overload, which has created a possible downside to several web users. Recommender systems solve this downside by exploring through giant volume of dynamically generated info to supply users with personalized content and services. This paper explores the various characteristics and potentials of various prediction techniques in recommendation systems so as to function a compass for analysis and apply within the field of recommendation systems. Recommender systems square measure one in all the foremost roaring and widespread application of machine learning technologies in business. Machine learning algorithms in recommender systems are

typically classified into two categories – content based and collaborative filtering methods although modern recommenders combine both approaches. Content primarily based strategies square measure supported similarity of item attributes and cooperative strategies calculate similarity from interactions. Below we tend to discuss principally cooperative strategies sanctionative users to find new content dissimilar to things viewed within the past. The explosive growth within the quantity of obtainable digital info and also the range of tourists to the net have created a possible challenge of data overload that hinders timely access to items of interest on the Internet. Information retrieval systems, such as Google, Devil Finder and Altavista have partially solved this problem but prioritization and personalization (where a system maps available content to user's interests and

preferences) of information were absent. This has inflated the demand for recommender systems over ever before. Recommender systems are information filtering systems that deal with the problem of information overload by filtering vital information fragment out of large amount of dynamically generated information according to user's preferences, interest, or observed behavior about item. Recommender system has the flexibility to predict whether a specific user would like associate item or not supported the user's profile. Recommender systems square measure helpful to each service suppliers and users. They scale back dealing prices of finding and choosing things in an internet searching surroundings. Recommendation systems have additionally proven to boost higher cognitive process and quality. In e-commerce setting, recommender systems enhance revenues, for the fact that they are effective means of selling more products. In scientific libraries, recommender systems support users by allowing them to move beyond catalog searches. Therefore, the necessity to use economical and correct recommendation techniques at intervals a system that may offer relevant and dependable recommendations for users cannot be over-emphasized. Recommender system is outlined as a call creating strategy for users underneath complicated info environments. Also, recommender system was outlined from the angle of E-commerce as a tool that helps users search through records of information that is expounded to users' interest and preference . Recommender system was outlined as a way of helping and augmenting the human process of exploitation recommendations of others to form decisions once there's no ample personal data or experience of the alternatives.

Recommender systems handle the matter of knowledge overload that users unremarkably encounter by providing them with customized, exclusive content and repair recommendations. Recently, various approaches for building recommendation systems have been developed, which can utilize either collaborative filtering, content-based filtering or hybrid filtering. Collaborative filtering technique is that the most mature and therefore the most typically enforced. Collaborative filtering

recommends items by identifying other users with similar taste; it uses their opinion to recommend items to the active user.

Collaborative recommender systems are enforced in several application areas. A recommender system provides a group of things (e.g., content, solutions, or other users) that's most relevant to a selected user of the system. Typically, recommender systems deliver the goods this by predicting connection scores for all things that the user has not seen yet. Items that receive the highest score get recommended (typically the top-N things or all things on top of a threshold t). The prediction is made by considering each the traits of the item and user. Typically, systems look at similarities between items, similarities between users, or relations between particular styles of things and explicit styles of users. The performance of a recommender system is decided by gazing the advice accuracy, that is, the error between given and expected results.

II. Types of Recommender Systems

A. Collaborative Filtering

One of the first collaborative filtering recommender systems is Tapestry, by Goldberg et al. This system was designed to retrieve email messages from Usenet mailing lists, relevant to a user's particular interests. He observed that conventional mailing lists are too static and rarely form a perfect match to a user's demands. Tapestry relies on what the authors termed collaborative filtering techniques, which are still widely used today. In collaborative filtering, each user rates content items. These ratings confirm similarity between either users (similar users like similar things) or things (users like items just like extremely rated items). Different metrics exist to compute similarity. Recommended for this user square measure those things that square measure rated highest by his most similar peers or contain those things that square measure rated most just like his favorite things.

B. Content-based

Content-based recommender systems use item similarity to determine recommendations. Unlike the collaborative filtering method, item similarity is

computed by item metadata. Examples of data area unit room for restaurants, genre for movies, and artist for music. Recommended area unit those things that area unit most just like the user's favorite things. An example of a content-based recommender system is Newsweeder, by Lang.

C. Demographic

When detailed information about the user's preferences is not available, demographic information can lead to somewhat personalized recommendations. Grundy, by wealthy, is associate degree example of this. Demographic information may include age, gender, country of residence, education level, etc. The demographic data is matched to a stereotype, and the items attached to this stereotype are recommended. Personalization for the user is proscribed because of the generalization to a stereotype.

D. Knowledge-based

When requiring a recommendation, the user enters his preferences in the recommender system. The system then outputs a (number of) potential recommendations supported (expert) data contained within the system. Possibly, the user can give feedback, and the recommendation is refined. After a few iterations, the recommendation is tailored to the user. Entree is an example of such a system, built to help diners find a suitable restaurant. In learning knowledge-based recommender systems, feedback from the user is fed back into the system to add to the knowledge.

E. Context-aware

In many application domains, contextual information is available, which can be used to improve recommendations. Common examples of contextual information are location, group dynamic, time, date, and purpose. While user preferences and domain information are comparatively static, context is highly dynamic in nature. Every recommendation, even for the same user, may have a completely new context.

Adomavicius and Tuzhilin provided a discussion on contextual information in recommender systems. They showed three ways in which such contextual information can be added to existing recommender systems:

- (1) Use a prefilter to remove content items (and information associated with them) that do not fit the context from the system.
- (2) Use a postfilter to remove recommendations that do not fit the context.
- (3) Add the context to the model of the recommender system and use the discourse info throughout the advice method.

F. Ensemble

Ensembles of recommender systems mix many of constant kind of recommender system to improve performance. The idea behind ensembles is to get multiple opinions before making a decision. Schlar et al. detailed the use of ensembles on collaborative filtering.

G. Hybrid

Hybrid recommender systems, like ensembles, combine multiple recommender systems. However, during a hybrid system, multiple differing kinds of recommender systems are combined. A comprehensive overview of different hybridization techniques is given by Burke. As concluded by Burke, given a certain hybridization, not all basic recommender systems can be (straightforwardly) combined.

H. Social

The rise of on-line social networks multiplied the supply of a user's social information (e.g., the friendship network). Because friends typically share interests, the data they provide to the recommender system is additional possible to fit with the user. Or or else, the social data may be accustomed infer communities of similar users. As an example, Konstas et al. utilized the social information in LastFM to improve collaborative filtering.

III. Information in Recommender Systems

The different types of information typically used in recommender systems:

A. Behavioral Information

It is the implicit information that the recommender system can gather whereas the user interacts with the

broader system. For example, product views in an exceedingly web shop or not absolutely observance a picture show on a video on demand web site.

B. Contextual Information

It describes to the context during which a recommendation question is made. Common samples of discourse info area unit location, grouping, time, date, and purpose

C. Domain Knowledge

It specifies the relationship between a user stereotype and content items. Domain data is sometimes static however will amend over time.

D. Item Metadata

It is the descriptive information about content items. Examples of information area are unit room for restaurants, genre for movies, and creative person for music.

E. Purchase or Consumption History

It is the list of content that has antecedently been purchased or consumed by the user.

IV. Phases of Recommendation Process

A. Information Collection Phase

This collects relevant information of users to generate a user profile or model for the prediction tasks

including user's attribute, behaviors or content of the resources the user accesses. A recommendation agent cannot operate accurately till the user profile/model has been well made. The system must apprehend the maximum amount as potential from the user so as to produce cheap recommendation right from the onset. Recommender systems trust differing types of input like the foremost convenient top-quality express feedback, which includes explicit input by users regarding their interest in item or implicit feedback by inferring user preferences indirectly through perceptive user behavior. Hybrid feedback may also be obtained through the mixture of each express and implicit feedback. In E-learning platform, a user profile is a collection of personal information associated with a specific user. This data includes psychological feature skills, intellectual abilities, learning styles, interest, preferences and interaction with the system. The user profile is often accustomed retrieve the required data to make up a model of the user.

Thus, a user profile describes a straightforward user model.

The success of any recommendation system depends for the most part on its ability to represent user's current interests. Accurate models are indispensable for obtaining relevant and accurate recommendations from any prediction techniques.

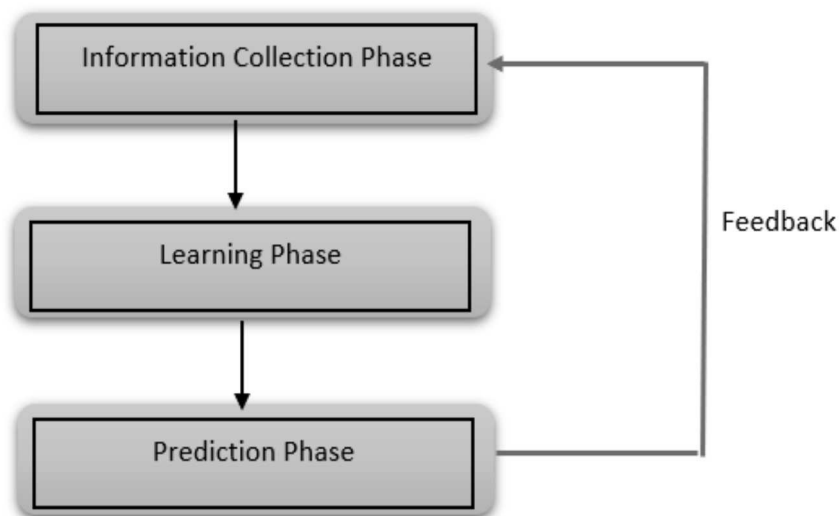


Fig. 1. Recommendation Phases

B. Learning phase

It applies a learning rule to filter and exploit the user's options from the feedback gathered in info assortment part.

C. Prediction phase

It recommends or predicts what reasonably things the user could like. This can be created either directly supported the dataset collected in data assortment

section that may well be memory based mostly or model based or through the system's determined activities of the user.

V. Evaluation Protocol

The evaluation protocol is then designed thanks to the mapping between the 4 core functions and the associated performance measures as summarized below:

Table 1: Adapted measures for each core function

Functions	Quality Criterion	Measure
Decide	Accuracy of the rating prediction Penalization of extreme errors to minimize the risk of wrong decisions	RMSE
Compare	Good predicted ranking for every couple of items of the catalog	COMP % of compatible rank indexes
Discover	Selection for a user the most preferred items in a list of items Identification of good/bad recommendations Precise, useful, trusted recommendation	(Precision, not recommended!) Average Measure of Impact (AMI)
Explore	Precise recommendations Identification of good/bad recommendations	Similarity matrix leading to good performances, in accuracy, relevancy, usefulness and trust

The following notations are adopted: a log (u, i, r) corresponds to a user u who rated an item i with the rating r . U is the set of all the users, I is the set of all the items. Given a dataset D of logs and an algorithm A , the evaluation protocol we propose is as follow:

- **Initialization:** Randomly split the dataset into 2 datasets train and test Use the train dataset to generate a model with the algorithm A .
- **Evaluation**
 1. For each log (u, i, r) of the test set:
 - 1.1 compute the predicted rating of the model.
 - 1.2 compute the predicted rating error
 2. Use the RMSE which gives an indicator of the performance of the Help to Decide function.

3. For each user u of U :
 - 3.1 sort all u 's logs of the test set by ratings
 - 3.2 sort all u 's logs of the test set by rating prediction.
 - 3.3 compute COMP comparing the indexes of u 's logs and the indexes of the predicted ratings of the logs.
4. Use the averaged COMP as an indicator of the Help to Compare function.
5. For each item i of I , compute $\text{count}(i)$ which is the number of logs in the train set referencing i .
6. For each user u of U :
 - 6.1 compute the predicted rating of each item i of I .

- 6.2 select the top-N highest predicted rating items noted $iu,1$ to iu,N which are the Top-N recommended items.
- 6.3 compute the rating average of u , noted.
- 6.4 for each recommended item iu,j of u :
 - 6.4.1 check if a corresponding log (u, iu,j,r) exists, If so the recommendation of iu,j is evaluable else skip the step 6.4.2.
 - 6.4.2 If $r \geq$ then the recommendation is considered relevant (and irrelevant in the other case).
- 6.5 compute the Precision and the AMI for the evaluable recommendations
7. Use the Precision and the AMI, averaged by users, as the indicators for the Help to Discover Function
8. Specify a way to compute efficiently, using the model of the algorithm A, the similarity between every couple of items (i,j) .
9. Compute the similarity matrix of all the couple (i, j) for $I \times I$.
10. Use this similarity matrix as the kernel of an item-item K-Nearest Neighbor model, then run the protocol for the steps 1 to 7 for RMSE, COMP, AMI and Precision to obtain a 4-dimensional indicator of the quality of the Help to Explore function.[6]

VI. Conclusion

In recent years information square measure generated at a dramatic pace. So we need well developed and technically developed recommender systems. To this finish during this paper, we have a tendency to survey the types of recommender systems, phases of recommender systems, evaluation protocols. We believe that in future researchers will pay more attention to these applications to resolve issues of huge information effectively and efficiently.

References

1. Recommendation Systems: Principles, methods and evaluation by Isinkaye, Y.O.Folajimi, B.A. Ojokoh
2. Evaluating Recommendation Systems by Guy Shani and Asela Gunawardana
3. Privacy in Recommender Systems by Arjan J.P. Jeckmans, Michael Beye, Zekeriya Erkin, Pieter Hartel, Reginald L. Lagendijk, and Qiang Tang
4. Paper Recommender Systems: A Literature Survey by Joeran Beel , Bela Gipp , Stefan Langer , and Corinna Breitingger
5. Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols by Pedro G.Campos, Fernando Diez, Ivan Cantador.
6. "Toward a New Protocol to Evaluate Recommender Systems" by Frank Meyer, Françoise Fessant, Fabrice Clérot and Eric Gaussier.

Shaping the Future with Crowdsourcing

Himangini Khanna

khannahimangini@gmail.com

R. Manisha

immanishar1999@gmail.com

Vishakha Sehdev Verma

vs4821@gmail.com

Ankit Verma

ankit.verma.aquarius@gmail.com

Abstract

Crowdsourcing is evolving as a distributed problem-solving and business production model in recent years. This paper gives a survey on crowdsourcing which is categorized according to their advantages, disadvantages, implementations and success stories. Also, on how Crowdsourcing leads to engaging a 'crowd' or group for a common goal of often innovation, problem-solving, or efficiency powered by new technologies, social media. Crowdsourcing can take place on many different levels and across various industries. As our connectivity is growing, it is now easier than ever for individuals to collectively contribute, whether with ideas, time, expertise, or funds to a project or cause. This collective mobilization is crowdsourcing.

Key Words: crowdfunding, crowdsourcing, peer-production

I. Introduction

The term crowdsourcing was coined by Jeff Howe in 2006 while he was working at the Wired magazine to describe how businesses were using the Internet to "outsource work to the crowd", which quickly led to the broadcasting of "crowdsourcing". Howe, first published a definition for the term crowdsourcing in a companion blog post to his June 2006 Wired article, "The Rise of Crowdsourcing", which came out in print just days later:

"Simply defined, crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call [1]. This can take the form of peer-production (when the job is performed collaboratively) but is also often undertaken by sole individuals. The crucial prerequisite is the use of the open call format and the large network of potential laborers [2].

But crowdsourcing goes all the way back to the 1700s even before the age of the internet and crowdsourcing people used to work in groups and collaborate. In 1714, the British government offered a reward to anyone who came up with a method to determine a ship's longitude. Also, in 1884, a group of scholars created the first Oxford dictionary. It seemed like an impossible task until various experts volunteered to write about different topics according to their expertise.

Crowdsourcing is the process of connecting with a larger group of people via the internet to address their knowledge, expertise, and resources. Crowdsourcing is the act of taking a job traditionally performed by a designated employee and outsourcing it to an undefined, generally large group of people in the form of an open call. The classic and earliest example of crowdsourcing is Wikipedia. It is the world's largest and most dynamic encyclopedia by some distance. Since the success of Wikipedia, crowdsourcing has

become increasingly commonplace in business, government, and society.

In the crowdsourcing model, tasks are segregated to networked people to complete such that a company's production cost can be greatly reduced [3]. In 2003, Luis von Ahn and his colleagues introduced the concept of "human computation", which uses the human abilities to perform computation tasks that are difficult for computers to process [4]. Later, the term "crowdsourcing" was coined by Jeff Howe in 2006. Since then, a lot of work in crowdsourcing has focused on different aspects of crowdsourcing such as discovering the upsides and downsides of its, and the usage either implicitly or explicitly.

II. The Crowdsourcing

Crowdsourcing is a business model or function that relies on a large group of users as third parties for outsourcing certain tasks. Crowdsourcing can add significant value to a product or service, and can also generate valuable connections between users and the company. It is a process of engaging a "crowd" or group for a common goal often innovation, problem-solving or efficiency.

It can take place on various different levels and across

various industries. It is a collective mobilization; this phenomenon can provide organizations with access to new idea and solutions, deeper consumer engagement, opportunities for co-creation, optimization of tasks and reduced costs [5]. The internet and social media have brought organizations closer to their stakeholders, laying the groundwork for new ways of collaborating and creating value together like never before. Crowdsourcing touches across all social and business interactions. It is changing the way we work, hire, research, make and market. Governments are applying crowdsourcing to empower citizens and give greater voice to the people. In science and health care, crowdsourcing can democratize problem-solving and accelerate innovation. With education, it has the potential to revolutionize the system, just as crowdfunding is currently challenging traditional banking and investing processes.

III. Importance of Crowdsourcing

Crowdsourcing adapts the principle of "more heads are better than one". We can see Crowdsourcing almost anywhere as it taps into the global world of ideas, helping Companies work through dynamic environment [6]. Due to broad participation in crowdsourcing, it charges relatively no cost at all.

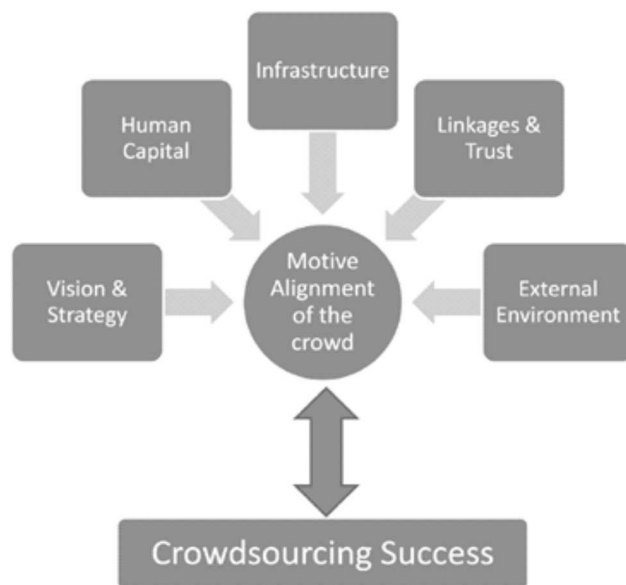


Figure 1: Crowdsourcing Success

Solutions are generated from volunteers or free-lance professionals who get paid only are we use their ideas. It makes good business sense for all of us to think in terms of how we can tap into those global worlds of ideas or a global crowd of creativity.

IV. Present Scenario of Crowdsourcing

The web-based crowdsourcing efforts today include voting, funding, creative sourcing, mobile crowdsourcing, data collection, talent searching, and micro work. Figure 1 shows the success of crowdsourcing.

The most prevalent forms of crowdsourcing are crowdfunding where people can raise capital from the people who believe in their frontier's ideas and beliefs such as Milaap, Ketto etc. Mobile crowdsourcing is an activity that takes places on a smartphone or other portable devices where data is collected from the crowd for a better user experience apps such as Google Maps, Uber and lift uses the state of collection method [7]. Creative outsourcing is when customers can select the best design for a website, logo or graphic from a pool of submitted entries. The selected designer would then give payment for e.g., Fiverr and 99designs. Crowdsourcing can be used by companies, the government, non-profit organizations (NGOs), groups or individual for whatever need they may have.

V. Upsides and Downsides of Crowdsourcing

a) Upsides

- Crowdsourcing can be used by anyone.
- It allows us to access larger online labor forces where we can select the workers or we just post the work and let the worker find us.
- It allows us to ask the crowd to find a solution.
- It allows us to find and organize available knowledge and finally we can get feedback and ideas from the crowd.
- There is limited financial risk.

b) Downsides

- There is no guarantee that the results are credible.
- The manager could waste time in managing the large scale of outsourced workers.

- It is not always easy for crowd members to collaborate on a project.

VI. Users of Crowdsourcing

When it comes to practice crowdsourcing, there are lots of options available to organizations- which is to put up tasks or questions out to a large group (either profit or non-profit) for their feedback and inputs. The group or can say "crowd" could be comprised of commoners, consumers, entrepreneurs, specialists, developers or interns. These types of users will be provided with objectives to discuss, review and a plea for their inputs. Objectives can be anything depending upon the need from a food campaign review to brand awareness.

VII. Ways of Implementing Crowdsourcing

- a) PEPSICO - When it comes to PepsiCo, it stand out to be expanding its use of crowdsourcing which is a known food snack and beverage corporation and generating new ideas for healthy ingredients or flavors, This snack food market has been booming with US sales up 12% in the last four years for a total of \$145 billion today. Now what they are actually practicing is just a simple trick, gathering a relevant and interested crowd via social media and introducing campaigns like "Do Us a Flavour" are great ways to innovate existing product lines [8]. The first thing crowdsources do is to gather out the flavor suggested from millions of participants. Next, they filter out the most voted flavor, basically capturing the upper tail end of "good" ideas. Hence, with zero investment, tracking down the trending recipe leading to speeding up the iteration phase, ultimately, the company estimates that they have been able to reduce their time to product launch to 10 months from the original 15-20 months [9]. They also saw a 12% increase in sales from the first competition, suggesting this new process also produces better results.
- b) TRUECALLER - There is a mobile application, which purely uses the concept of crowd-sourcing implicitly, for everyone's customized use i.e. True Caller, a worldwide number

lookup service. It seems magical that how an app just in seconds let the user know about the unknown or unsaved mobile numbers identification, spam calls unwanted call or privacy/security concern. Now, how does it happen? Basically, whenever we install the app in our phones, leading to the syncing of own contacts with the True Caller servers; they actually use our data only to provide us with the required data we want. A unique app embedded with crowdsourcing creating a huge market, as of now, Truecaller has 150 million registered users in India and 100 million daily active users, according to its own interior account.

- c) HAIKUJAM - Think about a person who is feeling down over a recent heartbreak and a failed startup using crowdsourcing to get things off from his mind. simply sound impractical but it is true. when an Economic student Dhru, opened his notebook wrote a random line passes it to his friend to complete it and that friend to the other, leading to a syncnon-sync poetry. this went for hours and turned out to be fun and mentally relaxing Dhru thought it would be cool to write like this with people around the world. so he asked his true friends to create a collaborative creative app naming HaikuJam.

VIII. Conclusion

After discussing and reviewing the recent marketing sensation of crowdsourcing, we believe that its elevation in our society today can be largely connected to its appeal to the consumer's central values, in addition to growing sensible trends. Marketers channelize and hold the mentality of the customers who often consider and acknowledge the group influence as well as the need for an outlet of creativity.

With the rise of crowd product development and marketing, the contests which are conducted on the social platforms and otherwise and marketing schemes definitely seemed to be the advantage of increasing brand loyalty and awareness.

From our analysis of concerned consumer behaviors towards crowdsourcing, we do believe that this marketing sensation can be long-lived, due to a number of reasons; firstly, the gradual increase in the usage of digital and social media platforms will always be profitable to crowdsourcing.

In the long term, crowdsourcing seems to be expanding and being practiced by more established companies and institutions. Also, people who are having their small networks and are entrepreneurs and start-up planners value their inputs, now started to practice and get associated with it, for their future benefits.

References

1. A. Verma and J. S. Prasad, "Performance Enhancement by Efficient Ant Colony Routing Algorithm Based on Swarm Intelligence in Wireless Sensor Networks," *Int. J. Wirel. & Mobile Comput.*, vol. 12, no. 3, 2017, pp. 232-238.
2. A. Verma and J. S. Prasad, "Optimum Path Routing Algorithm using Ant Colony Optimization to solve Travelling Salesman Problem in Wireless Networks," *Int. J. Wirel. & Mobile Comput.*, vol. 13, no. 2, 2017, pp. 131-138.
3. A. Verma and P. C. Vashist, "Enhanced Clustering Ant Colony Routing Algorithm Based on Swarm Intelligence in Wireless Sensor Network," *Proc. IEEE Int. Conf. Advances in Comput. Eng. & Applicat.*, 2015, pp. 150-154.
4. A. Verma and P. C. Vashist, "Evolution of Ant Colony Optimization and Swarm Intelligence in Wireless Sensor Networks," *Proc. DRDO Int. Conf. Inform. Tech. Yesterday, Today & Tomorrow*, 2015, pp. 286-289.
5. Estellés-Arolas, Enrique; González-Ladrón-de-Guevara, Fernando (2012), "Towards an Integrated Crowdsourcing Definition," *Journal of Information Science* 38 (2): 189-200.

6. Groves, R. M., Singer, E., Corning, A. (2000). Leverage-salience theory of survey participation. *Public Opinion Quarterly* 64 (3): 299-308.
7. Hars, A., and Ou, S. (2002) Working for free? Motivations for participating in open-source projects. (2002)*International Journal of Electronic Commerce*, (6) 3, 25-39.
8. Hertel, G., Niedner S. & Herrmann, S. (2003), Motivation of software developers in Open Source projects: an internet-based survey of contributors to the Linux kernel. *Research Policy*, 32 pp. 1159-1177.
9. Heider, F. (1958) *The Psychology of Interpersonal Relations*. Mahwah, NJ: Lawrence Erlbaum.

Web Usage Mining: Concepts & Applications

Amanpreet Kaur

amanpreetkaur250899@gmail.com

Charul Nigam

charul.nigam@gmail.com

Priti Sharma

priti.shr18@gmail.com

Abstract

With the growth of technology at a faster pace, World Wide Web has also grown exponentially. With the increase in popularity of web, its complexity also increases because of the presence of large amount of data. It became difficult for the users to retrieve the information of their interest from such a huge nugget. It was Oren Etzioni who first coined the term Web mining in his paper in 1996. The Web Mining technique acts as a device to bring out of this trouble. It helps in automatic discovery and retrieval of information from the internet. The paper explains in detail the process of web usage mining and the different techniques used for pattern discovery, the process and the architecture of web usage mining.

Key Words: Web Usage Mining, World Wide Web, Pattern Discovery, Data Mining, Pattern Analysis

I. Introduction to Web Mining

Web mining is an exciting discipline in the domain of data mining as well as in classification/clustering. This area of research is so huge today partly due to the interests of various research communities, the tremendous growth of information sources available on the Web and the recent interest in E-commerce. Web Mining is defined as the application area of data mining which deals with the extraction of useful and interesting facts from World Wide Web. It is used to provide the solution of various problems such as finding relevant information, creating information from the data available on web, learning about consumers or individual users, personalization of information and so on. Web mining makes use of various data mining techniques to automatically discover web and retrieve information from the web document. Web Mining uses several data mining techniques to retrieve the useful facts from internet. However, along with Data mining techniques various other techniques such as artificial intelligence, information retrieval, natural language processing,

information extraction, machine learning can also be applied.

Identifying the usage patterns of users is very important in utilizing the vast information available in the World Wide Web.

Web Mining is divided into three categories:

- Web Content mining
- Web Structure Mining
- Web Usage Mining

We are describing the most successful applications in this section:

- Personalized customer experience in B2C E-commerce – Amazon.com
- Web search – Google
- Web wide tracking - DoubleClick
- Understanding auction behavior – eBay
- Personalized portal for the web – MyYahoo
- CiteSeer –digital library and autonomous citation indexing

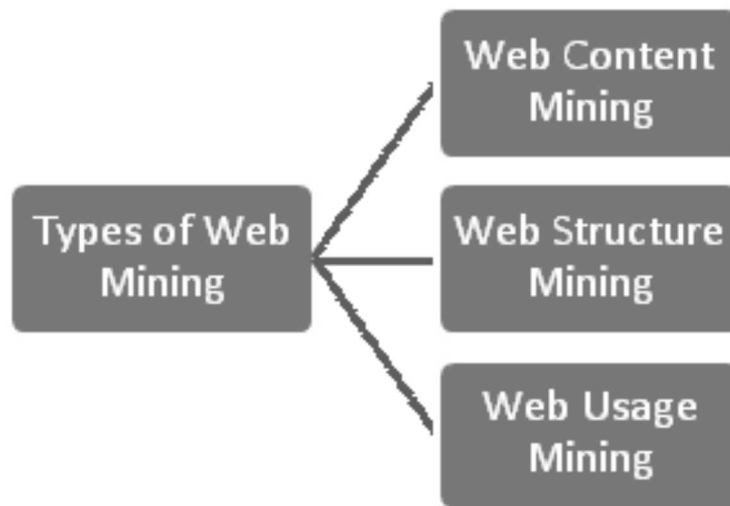


Fig: 1 Types of Web Mining

Web mining helps the web designers in discovering the knowledge from the information available in the web. Also it helps the users in getting the fast retrieval of the information they are looking for. Three major areas of web mining are

Web content mining - Trying to get useful information from the text, images, audio and video in web pages

Web structure mining - Trying to understand the link structures of the Web which will help in categorization of Web pages.

Web usage mining- Trying to get useful information from the server logs to understand what the users are looking for. It also helps in personalization of web pages.

Though all the three categories of web mining are interlinked, we are going to discuss about the web usage mining. Web usage mining helps the web masters to understand what the users were looking for so that they can develop the strategies to help the user to get the required information quickly. Web mining is generally implemented by using the navigational traces of users which give the knowledge about user preferences and behavior. Then the navigational patterns were analyzed and the users were grouped into clusters. The classification of navigational patterns into groups helps to improve the quality of

personalized web recommendations. These web page recommendations were used to predict the web pages that are more likely to be accessed by the user in near future. This kind of personalization also helps in reducing the network traffic load and to find the search pattern of a particular group of users.

Data mining techniques like, clustering, sequential pattern mining and association rule mining were used in web mining. All these techniques were used to extract interesting and frequent patterns from the information recorded in web server logs. These patterns were used to understand the user needs and help the web designers to improve the web services and personalization of web sites.

II. Web Usage Mining

The ease and speed with which business transactions can be carried out over the Web has been a key driving force in the rapid growth of electronic commerce. Specially, ecommerce activity that involves the end user is undergoing a significant revolution. The ability to track users' browsing behavior down to individual mouse clicks has brought the seller and end customer closer than ever before. It is now possible for a seller to personalize his product message for individual customers at a massive scale, a phenomenon that is being referred to as mass customization. The scenario described is one of many possible applications of Web

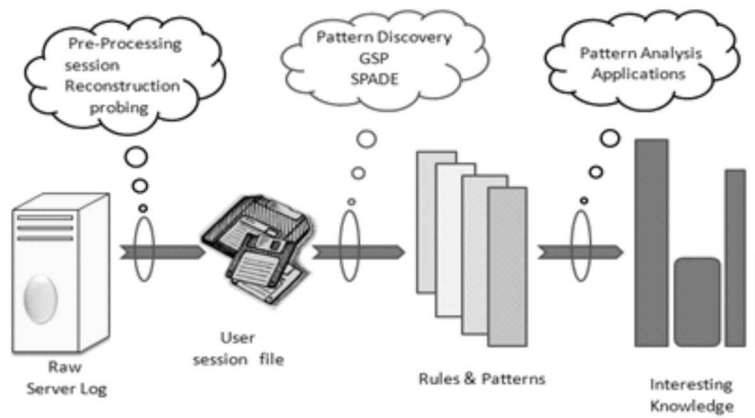


Fig: 2 Phases of Web Usage Mining

Usage mining, which is the process of applying data mining techniques to the discovery of usage patterns from Web data, targeted towards various applications. This paper provides an up-to-date survey of Web Usage mining. As shown in Figure 1, there are three main tasks for performing Web Usage Mining or Web Usage Analysis. This section presents an overview of the tasks for each step of the process.

Pre-Processing

Preprocessing consists of converting the usage, content, and structure information contained in the various available data sources into the data abstractions necessary for pattern discovery.

Usage preprocessing: It is arguably the most difficult task in the Web Usage Mining process due to the incompleteness of the available data. Unless a client side tracking mechanism is used, only the IP address, agent, and server side clickstream are available to identify users and server sessions. Some of the typically encountered problems are:

- Single IP address/Multiple Server Sessions - Internet service providers (ISPs) typically have a pool of proxy servers that users access the Web through. A single proxy server may have several users accessing a Web site, potentially over the same time period.
- Multiple IP address/Single Server Session - Some ISPs or privacy tools randomly assign each request from a user to one of several IP addresses. In this

case, a single server session can have multiple IP addresses.

- Multiple IP address/Single User - A user that accesses the Web from different machines will have different IP address from session to session. This makes tracking repeat visits from the same user difficult.
- Multiple Agent/Singe User - Again, a user that uses more than one browser, even on the same machine, will appear as multiple users.

Content Preprocessing: Content preprocessing consists of converting the text, image, scripts, and other files such as multimedia into forms that are useful for the Web Usage Mining process. Often, this consists of performing content mining such as classification or clustering. While applying data mining to the content of Web sites is an interesting area of research in its own right, in the context of Web Usage Mining the content of a site can be used to filter the input to, or output from the pattern discovery algorithms. For example, results of a classification algorithm could be used to limit the discovered patterns to those containing page views about a certain subject or class of products.

Structure Preprocessing: The structure of a site is created by the hypertext links between page views. The structure can be obtained and preprocessed in the same manner as the content of a site. Again, dynamic content (and therefore links) pose more problems than static

page views. A different site structure may have to be constructed for each server session. This paper describes the kinds of mining activities that have been applied to the Web domain. Methods developed from other fields must be taken into consideration the different kinds of data abstractions and prior knowledge available for Web Mining. For example, in association rule discovery, the notion of a transaction for market-basket analysis does not take into consideration the order in which items are selected. However, in Web Usage Mining, a server session is an ordered sequence of pages requested by a user.

Pattern Discovery

Pattern discovery focuses on to uncover patterns from the abstractions produced as a result of preprocessing phase. It focuses on applying various methods and techniques developed from several fields such as data mining, machine learning, statistics and pattern recognition.

Steps in pattern discovery:

- **Statistical Analysis:** Statistical techniques are the most common method to extract knowledge about visitors to a Web site. By analyzing the session file, one can perform different kinds of descriptive statistical analyses (frequency, mean, median, etc.) on variables such as page views, viewing time and length of a navigational path.
- **Association Rules:** Association rule generation can be used to relate pages that are most often referenced together in a single server session. In the context of Web Usage Mining, association rules refer to sets of pages that are accessed together with a support value exceeding some specified threshold. These pages may not be directly connected to one another via hyperlinks.
- **Clustering:** Clustering is a technique to group together a set of items having similar characteristics. In the Web Usage domain, there are two kinds of interesting clusters to be discovered: usage clusters and page clusters. Clustering of users tends to establish groups of users exhibiting similar browsing patterns.

- **Classification:** Classification is the task of mapping a data item into one of several predefined classes. In the Web domain, one is interested in developing a profile of users belonging to a particular class or category. This requires extraction and selection of features that best describe the properties of a given class or category.
- **Sequential Patterns:** The technique of sequential pattern discovery attempts to find inter-session patterns such that the presence of a set of items is followed by another item in a time-ordered set of sessions or episodes. Web marketers can predict future visit patterns which will be helpful in placing advertisements aimed at certain user groups.
- **Dependency Modeling:** Dependency modeling is another useful pattern discovery task in Web Mining. The goal here is to develop a model capable of representing significant dependencies among the various variables in the Web domain. As an example one may be interested to build a model representing the different stages a visitor undergoes while shopping in an online store based on the actions chosen (I e. from a casual visitor to a serious potential buyer).

Pattern Analysis

The last step of web usage mining process is pattern analysis. This phase separates the interesting and uninteresting patterns from the overall patterns discovered during pattern discovery phase. Result of pattern analysis phase is used in various applications such as system performance improvement, site modification, personalization, e-commerce etc. Patterns can be analyzed by using following techniques described below:

- **Knowledge Query Mechanism:** Structured Query Language (SQL) is most commonly used language for knowledge query mechanism. This language is applied in order to extract the useful patterns from discovered patterns.
- **OLAP/Visualization tools:** Patterns are also analyzed by using OLAP tools in which discovered facts are placed on to data cubes for performing

various OLAP operations such as roll up and drill down and interesting facts are retrieved. OLAP provides an integrated framework for analysis which allows changes in aggregate levels. Output of OLAP queries acts as an input to data mining or data visualization tools. Graphing patterns or assigning colors to different values are used as visualization technique for same purpose.

- **Intelligent Agents:** Various agents are also devised that helps in examining the patterns in web usage mining. These agents perform the work of analyzing the discovered patterns.

III. Architecture of Web Usage Mining

The system WEBMINER implements different parts of this general architecture, in the architecture the Web Usage Mining is divided into two parts, the first one is transformation of data on WWW into appropriate transaction form based on domain, which includes preprocessing the data, identifying the transactions and

integrating the data components. The second one is completely domain independent and involves application of general data mining techniques and pattern matching techniques.

Cleaning of data is the primary step involved in the Web usage mining process. Basic level data integration also can be performed at this level. Available Web usage mining products are often closed-architecture proprietary systems. Both commercial and non-commercial systems don't offer many configuration options. The methods they use, the kind of data sources they analyze, the types of reports they generate are often fixed. For example, he has to prepare raw log server data, performing all cleaning and filtering tasks. Then, he is obliged to identify users, sessions and transactions, preparing data to be analyzed, and so on. The problem is then stated in this way: he consumes a lot of his efforts working in accessory tasks, instead of devoting all attention to his main concern, the mining method or algorithm to be tested.

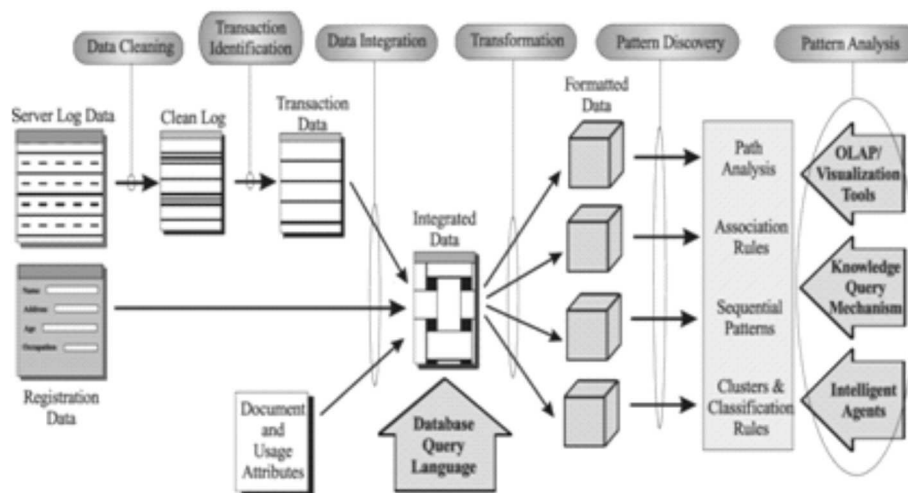


Fig: 3 Architecture of Web Usage Mining

IV. Literature Review

P. Lopes & B. Roy had their main focus on providing good quality product recommendation to all the users specially unregistered ones of e commerce. The proposed system (i.e. users interact with web portal and users click stream data is maintained in raw log files. Multiple preprocessing and data cleaning tasks

are performed to extract valuable information from raw log files and transform it in structured form, the cleansed data is further used to discover patterns, hidden rules and provide top product recommendation to all users of E-commerce site.) dynamically provides recommendations per changing user's behavior by constructing patterns from the historical data. Results

prove 80 to 85 percent accuracy for registered users 65 to 70 percent for unregistered users as per their search history. According to them the proposed system has potential to attract new customers and retain the old ones. It can be tested for many others applications like restaurant recommendations, hotel recommendations.

A. Bhargav & M. Bhargav had their main focus on the proposed model which has been programmed to return the results in the form of accuracy and class density along with the classification obtained from the input dataset. The proposed model has been designed to detect and classify the input dataset to the given classifier by using the results obtained from the online student data portal. The experimental results have been obtained from the simulation model in the form of the enlisted performance parameters. The proposed model output has been designed in the way to perform all of the operations in the sequential order as per the system design. The simulation model detects the abnormalities in the given API data. The proposed model has been designed to detect and track the slow learners in the given database which is observed on the basis of the input dataset. The proposed model results show the effectiveness in the automatic classification of the student data. The proposed model has performed better for the classification of the student data.

R. Nayak and A. Bose paid focus on Data mining and web mining which according to them is the study of exploring patterns in huge volumes of raw data, refer to techniques that help us to find content of web and retrieve the user's interest and needs respectively. This paper focuses on the comparison of mining algorithms. As huge data is adding in repository every second so there is need of quality information to satisfy upcoming needs of user. In future Web Service provisioning and web service discovery should be made by analyzing user's interest.

S.K. Pani had main focus on research issues and drawbacks of the existing techniques. More research work need to be done on the web mining domain as it will rule the web in the near future. Web mining along with semantic web known as semantic web mining is

to be concentrated that is evolving which helps us to overcome the cons of web mining. Through various algorithms and techniques have been proposed still work has to be done in discovering new tools to mine the web.

Liu kewen had main focus on Web sites are of much use for users. Web sites are built, deployed and maintained to serve with various function to user. At what extent this functions, features which were thought of is implemented can be identified, verified by careful inspection at the log data captured while user interacts with web site. Content recommender systems make use of this data to recommend a most appropriate content through the application of various subjective and/or objective, procedural, algorithmic or heuristic processes, methods or techniques.

V. Applications of Web Usage Mining

The main objective of web usage mining is to collect data about the user's navigation patterns. This information can be used to improve the Web sites in the user view. Three main applications of this mining are studied in this section.

- The privatization of web content: Web usage mining techniques can be used for personalization of web users. For example, user behavior can be immediately predicted by comparing her current survey patterns with survey patterns extracted from the log files. Recommendation systems which have a real application in this area are, suggest links that direct the user to his favorite pages. Some sites also organize their product catalogues based on predicted interests of specific user and represent them.
- Pre-recovery: The results of web usage mining can be used to improve the performance of Web servers and Web-based applications. Web usage mining can be used for retrieving and caching strategies and thus reduce the response time of Web servers.
- Web site designing: Usability is one of the most important issues in designing and implementing web sites. The results of web usage mining can help to appropriate design of web sites. Adaptive

web sites are an application of this type of mining. Website content and structure are dynamically reorganized based on data derived from user behavior in these sites.

- **Personalization:** Website Personalization is gathering the personal user data, some privacy issues is a major anxiety. Personalization Consortium endorses and guide the growth of one-to-one marketing carried out. Behind of personalization technologies include collaborative filtering, in which filtering is applied with different sites for selecting the relevant information that may apply to the specific group of customers. User profiling information is collected from different web sites can result of personalized web page .By using data analysis can be predict the users future prediction and find the session with multiple websites with interesting links.
- **System improvement:** The major system improvement life cycle is planning, analysis, development and implementation. It should support the user demand to build a system. Developing the system with the security can avoid the intrusion and to restrict the user's access to certain online contents. Understand the customer demand and retaining the customized products. Improve some satisfaction with the help of browsing behavior.
- **Link Prediction:** Link prediction is used for analyzing the nodes in a network, from the large network suggest that information can be extracted

from the network topology.

- **Business Intelligence:** Web usage mining provides data to improve the customer, sales and marketing field. It is the technology to access the data from various data sources, for business advantage, the data is gathered, stored and analyzed in organization can improve the customer needs and demands. Some decision about the business can be made to success. Thus the disciplines of Business intelligence includes decision support, data mining, online analytical processing (OLAP), querying and reporting, statistical analysis and forecasting. Some of the business intelligence tools are Bizz Score Suite, IBM, Cognous, Series10, Web FOCUS, QlikView, Tableau Software, Style Intelligence, Board Management Intelligence Toolkit, AS Enterprise BI Server can retrieve, analyse and generate report.

Conclusion

In this paper we briefly reviewed about Web mining, types of Web mining, applications of web mining too. We also briefed the various Web usage mining concepts, its techniques and applications, Architecture and process of WUM. Web Mining is not a new term, but in the recent years its growth day by day is touching great horizons. It has spread in almost all areas now-a-days. As huge data is adding in repository every second so there is need of quality information to satisfy upcoming needs of user. In future Web Service provisioning and web service discovery should be made by analyzing user interest.

References

1. M. Heydari et al., "A graph based web usage mining method considering client side data", International Conference on Electrical Engineering and Informatics, IEEE, 2009.
2. Liu Kewen, "Analysis of Preprocessing methods for web usage mining", International Conference on measurement, Information and Control, IEEE, 2012
3. S. K. Pani et al., "Web Usage Mining: A survey on pattern extraction from web logs", International Journal of Instrumentation, Control &Automation, Vol.1, Issue 1, 2011.
4. P. Lopes and B. Roy, "Dynamic Recommendation System Using Web Usage Mining for E-commerce Users," Procedia Comput. Sci., vol. 45, pp. 60–69, 2015.

5. T. T. S. Nguyen, H. Y. Lu, and J. Lu, "Web-page Recommendation based on Web Usage and Domain Knowledge," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 10, pp. 2574–2587, 2014.
6. R. Geng and J. Tian, "Improving web navigation usability by comparing actual and anticipated usage," *IEEE Trans. HumanMachine Syst.*, vol. 45, no. 1, pp. 84–94, 2015.
7. R. Nayak and A. Bose, "A Data Mining Based Method for Discovery of Web Services and their Compositions," *Real World Data Min. Appl.*, vol. 17, pp. 325342, 2015.
8. A. Bhargav and M. Bhargav, "Pattern discovery and users classification through web usage mining," in *IEEE International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICT)*, 2014, pp. 632–636.

Software Development Life Cycle Models: A Brief Comparison Then and Now

Himanshi Bhambri

himanshi.clouds@gmail.com

Niharika Galgali

niharikagalgali@gmail.com

Himanshi Yadav

himanshiyadav70@gmail.com

Abstract

Foundation for every development lies in the model which well describes the product. So, this research paper provides a tour of the main software development life cycle (SDLC) models. Software Development Life Cycle is a methodology that describes a sequence of steps taken to develop a software. The life cycle commences with planning and terminates with the finished product. The paper focuses on the various software development models used in the software life cycle. Here we perform a comparative study of traditional model with the current agile models. These models are represented diagrammatically so as to enhance its effect.

Key Words: Software Engineering, SDLC, Traditional Software Models, Agile Model, Waterfall Model, Spiral Model.

I. Introduction

The way earth revolves around sun, in a similar fashion our lives also revolves around software. So software is best defined as a set of instructions and its associated documentations that tells a computer what to do or how to perform a task.

Software Development Life Cycle provides sequence of operations for software developers to develop the software in a manner such that it is completed within deadlines and quality of the product of software is maintained as per the standards laid down by the developers [1] In the primitive years, the software industry experienced many failures which led to the popularity of hardware due to its reduced cost. Hence, need of the hour is to develop a effective and efficient software. Software Development Life Cycle (SDLC) is concept used in software engineering to describe a procedure for planning, creating, coding, testing and implementation of user requirement specification [2]

Software development consists of a plethora of activities that are sub divided into phases. Planning, designing, coding, testing and maintenance are the integral phases of SDLC.

A. Requirement Gathering & Analysis: “Better planning produces better results”. So in software development, requirement gathering is the most crucial phase of SDLC. In this phase, the developers conduct a survey of information regarding development of software. Studying of existing software, conducting interviews, collecting answers from questionnaires are some of the methods through which developers accumulate information.

B. Designing: Designing is the successor of the requirement gathering phase. In this phase a blueprint is laid down by the developers and architects to accomplish the desired result. It lists problem definition, constraints, references and other aids to proceed further with the development.

C. Coding: Here, the actual implementation of the layout made commences. The programmers contribute their time, efforts to code the software. The layout formed is decomposed into modules and every individual performs the task assigned. Database administrator works on database, programmers apply their logics and brain on coding, front end developer makes significance by making interactive GUI.

D. Testing: The Solution program is tested against various test cases. Unit testing, integration testing and system testing are the 3 types of testing that plays a vital role in making a software bug free and with negligible/ minimal loopholes.

E. Maintenance and Support: This phase symbolizes the termination of software development. The software is delivered according to the user's satisfaction. The development team is informed about the hindrances/ obstacles encountered in the system, if any, and corrective measures are taken.

Traditional and agile are the two categories of Software Development Life Cycle.

II. Traditional Models

The traditional model uses a linear approach where the stages of the software development process must

be completed in a sequential order. This means that a stage must be completed before the next one begins. The various traditional models are discussed below.

A. Waterfall Model

The immemorial waterfall model was introduced by Winston Royce in 1970. It is successive model that segregate software development into various postures .During SDLC phase, each phase is sketched out to implement the activities. In view of this each phase is built before proceeding to next phase. After gripping to next phase, there is no flipside for going back. In waterfall model the succeeding phase is dependent on the result of the proceeding frame. The supremacy of waterfall model is that it acquiesce segmentation and control. Development proceeds from concept, through design, implementation, testing, installation, troubleshooting and it comes to a halt at the operation and maintenance. This is the agenda which is set up for each stage of development. Despite in some situation it induces some delay in project finalization because precautionary succeeding onto the consequent stage we need to replete the preceding stage. The obstacle can't be fixed until you reach the maintenance stage since there is little room for alteration once stage is integrated.

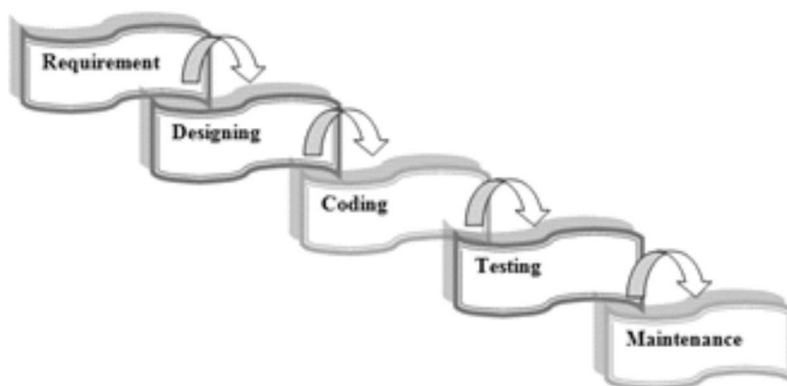


Figure-1: Waterfall Model

B. Prototype model

Prototype model is not a standalone, complete development methodology, but rather an approach to handling selected parts of a larger, more traditional development methodology [4] Software prototyping

is most suited when the customers is ambiguous about the project requirements. In this model, customer's feedback plays a pivotal role as the prototype is built according to their requirements, tested fully content with the model. Clients are interviewed which leads

to development of a sketchy high level paper model. This document lays down a foundation by supporting the basic functionality. Refinement of the prototype is carried out to eliminate the obstacles. Approval of the customer marks the success of the model. The developer should tread lightly while using the prototyping approach as one mistake can result in catastrophic damages. There are two types of prototyping:

1. Rapid Throwaway Prototyping : This technique

enhances the exploring ideas and customer feedback. Developer prototype may or may not be shortlisted to the accepted prototype. The final prototype is of optimum quality as customer feedback helps in removing faults and bugs.

2. Evolutionary Prototyping : This model follows an incremental approach as the prototype gets refined incrementally until customer's satisfaction. Evolutionary prototyping has an advantage over rapid throwaway as it saved time and effort.

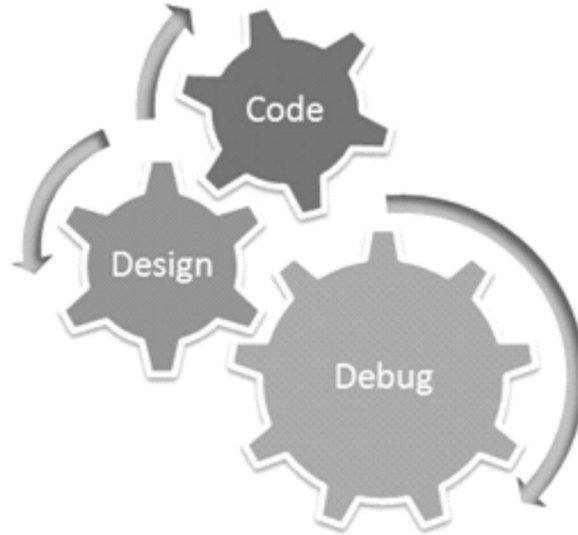


Figure-2: Prototype Model

C. Iterative Model

Waterfall model and prototype model complement each other in iterative model. Development of the software is initiated by specifying and implementing a part of the software and not full-fledged specification of requirements. The requirements gathered from the client are iterated sequentially until the software is implemented. The principle of this approach is to divide the phases of life cycle into various builds. After every iteration, the development module goes through the requirements, design, implementation and testing phases. Rigorous validation, verification of each version of software is the key to a successful iterative model.

Effortless error detection at an early stage enables to take corrective measures which add onto the usefulness

of iterative model. The pitfall of this approach is that resources can be consumed quickly by repeating the process again and again. Incomplete requirement gathering during the lifecycle hampers the smooth flow of developing effective software. This model is applied on voluminous software projects as decomposing small scale projects in a tedious job/task.

D. Spiral Model

In 1986 paper "A Spiral Model of Software Development and Enhancement" by Barry Boehm was the first one who interprets the Spiral Model. The spiral model is effective for large, expensive and complicated projects as it provides support for Risk Handling. Since it subsumes all the other SDLC models so the another name given to this model is Meta Model. The spiral

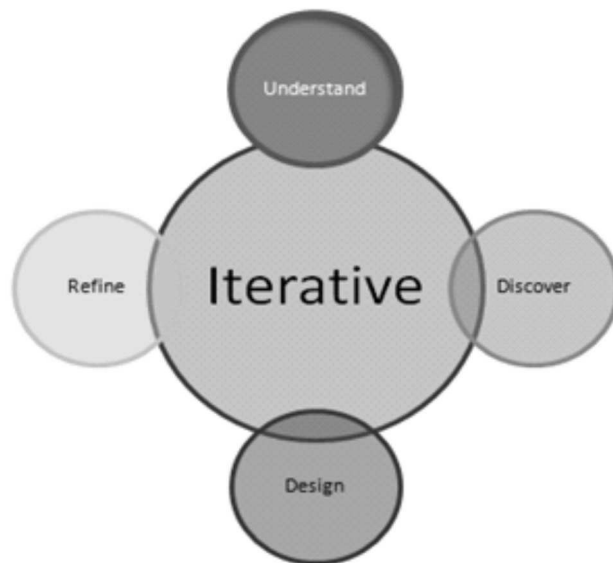


Figure-3: Iterative Model

model proceeds with an fundamental pass through a definitive waterfall model, applying a subset of the overall requirements to flourish a robust prototype. The cycle is proposed again, inserting new functionality and releasing the succeeding prototype after each assessment. This procedure sustain with the prototype welcoming larger and larger with each updation. Hence it is termed as “Spiral Model”. The phases in the waterfall model are used in spiral model also. In spiral model, a spiral is used which has many cycles. There are 4 phases:

- a. Planning -: in this phase, objectives are determined and alternatives are identified.
- b. Risk analysis-: in this, alternative is evaluated, identify and resolve the risks.
- c. Development -: develop, verify next level product.
- d. Customer evaluation-: in this, customer evaluate the prototype.

Spiral model concedes different components of

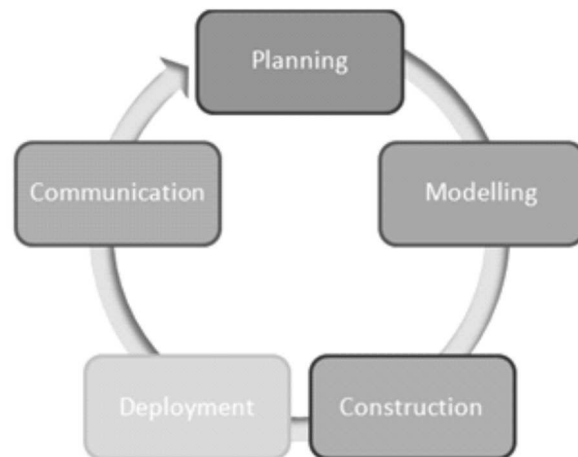


Figure-4: Spiral Model

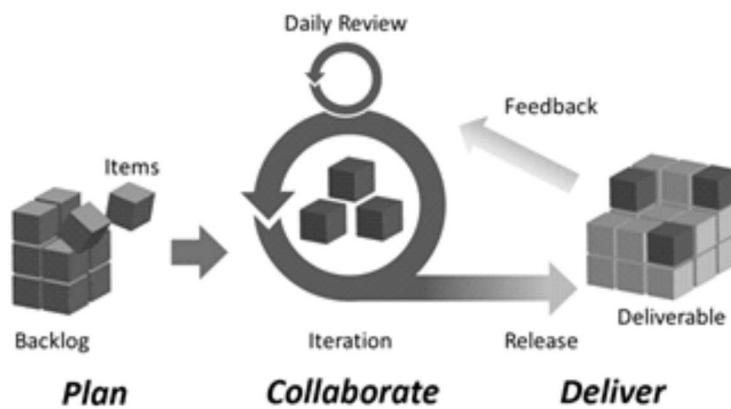


Figure-5: Agile Model

product can be elevated when they turn out to be feasible. This confirms that there is no conflict with preceding requirements and design. With accesses that have multiple software builds and releases which let on accomplishing an orderly transition to a maintenance activity, the spiral model is the most consistent model. It takes a lot of time and more cost for the development. There is a possibility of proceeding with indefinite loop of the spirals and to outright such software there is a very strict management. Time estimation is arduous because the number of phases is unfamiliar at the initial stage of the project.

E. Agile Model

In 2001 the Agile Manifesto and the agile team imported the agile software development model. Through iterative and incremental development, it focuses on early and continuous dispatch of software, so as to achieve customer satisfaction.

The agile model is hybrid model it is uses advantages of the both iterative and incremental model [3] First, the iterative approach plays its part by building the modules and delivering it after each iteration. This is followed by the incremental approach which incorporates the features required by the customer. The main attributes of the agile model are:

Table-1: Comparison between Traditional & Agile Model

S.No.	Basis	Traditional Model	Agile Model
1.	Flexibility	Provides little or no scope for making changes to the product.	Offers a lot of flexibility in terms of making changes to the product.
2.	Ownership	Project Manager	Team members
3.	Problem Solving	Manager resolves the unexpected issues.	Team members have the authority to make decisions on their own.
4.	Monitoring Progress	There are no frequent check-ins.	Encourages team members to have checkpoints at regular intervals.
5.	User Requirements	Clearly defined before coding or implementation.	Interactive input
6.	Organizational Structure	Linear	Lterative
7.	Project Scale	Large Scale	Small or medium scale

- 1) Incremental: Small software releases, accompanied by rapid development cycles
- 2) Co-operative: Closer customer-developer interaction
- 3) Adaptive: flexible enough to adapt to last moment changes.

One of the major advantages of agile over waterfall model is that it requires minimum planning as compared to waterfall. Agile model makes an assumption that user's needs are dynamic in nature i.e they change rapidly. Feedback plays a vital role in agile model as features can be added or removed as per user's discretion.

Customer ambiguous requirement can derail the software development process unless and until it is lead by/ guided by experienced/senior programmers instead of a newbie/amateur/novice.

III. Traditional Vs Agile

The following table compares the traditional model

with the Agile Model

IV. Conclusion

To sum up we have come across, various software models used in software engineering. Each model have their own pros and cons. Traditional approach consists of four models i.e, waterfall model, spiral model, iterative model and prototype model whereas the current approach in software engineering is agile approach. After analyzing the advantages and disadvantages of every model, we have reached to a conclusion that the selection of model depends upon the user requirements and type of industry.

There should be a tradeoff between the development time and the quality of the product. Customers don't expect a bug free product but they expect a user-friendly product that results in Customer Ecstasy! [5]

References

1. Shubhmeet Kaur, "A Review of Software Development Life Cycle Models", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 11, November 2015 ISSN: 2277 128X.
2. Mohit Kumar Sharma, "A study of SDLC to develop well engineered software", International Journal of Advanced Research in Computer Science, Volume 8, Issue 3 March – April 2017 ISSN No. 0976-5697.
3. Sahil Barjtya¹, Ankur Sharma², Usha Rani³, "A detailed study of Software Development Life Cycle (SDLC) Models", International Journal Of Engineering And Computer Science, Volume 6 Issue 7 July 2017, ISSN:2319-7242.
4. S. Maheshwari and D. C. Jain, "Comperative analysis of different type of models in software development life cycle," International Journal of Advanced Research in Computer Science and Software Engineering, Vol.2, Issue.5, pp. 285-290, 2012.
5. Rajendra Ganpatrao Sabale, Dr. A.R. Dani "Comparative Study of Prototype Model For Software Engineering With System Development Life Cycle" IOSR Journal of Engineering (IOSRJEN) ISSN: 2250-3021 Volume 2, Issue 7(July 2012)

Autonomous Systems: The Era of Healing

Saksham Bhambri

saksham1999274@gmail.com

Muskan Ahuja

ahujamuskan22@gmail.com

Lokesh Jain

nsitlokesh.jain@gmail.com

Abstract

In this paper, we are going to learn about the new approach of the hardware systems which have the concept of self-healing and self-protection systems. Earlier system was very small in size and can only be maintained by the IT experts, but now the time has been changed and the requirements of these systems are increasing day by day. The term and characteristic of self-healing and self-protecting, applied to systems, is often seen from different domains of computer science, such as fault tolerance or network and service management. Self-healing approaches are meant to complement existing approaches to security. The aim of the self-healing software feature is to fast recover the application and keep it running and available for 24/7 as optimal as possible. This survey provides an overview of self-healing software and system that is especially useful in all of those situations in which the involvement of humans is costly and hard to recover and needs to be automated with self-healing.

Key Words: Self healing system, security, automation, self protection.

I. Introduction

The biological systems of a human body have the ability to self-heal and self-protect itself from any outsource danger, it has the tendency to adopt necessary characteristics for his survival in different environmental condition. Similarly, Paul Horn from IBM has given us a new approach in 2001 to develop automated computing systems. These systems should have the ability to self-heal, self-protect and self-manage themselves from any fault or error and recover itself with minimum or no human interaction [1].

Self-healing and self-protecting capabilities are emerging as an exciting and potentially valuable technique in the software and hardware world. We believe that it is only a matter of time before we encounter a fully functional self-healing and self-protecting system with our own eyes. Therefore we believe it is important for us to understand what are self-healing and self-protecting systems, why we are

trying to enhance them in the first place, and how they can perform a valuable role in the future [6].

What, exactly, is a self-healing system? For the purposes of our discussion, a self-healing and self-protecting software system is a software architecture that enables the continuous and automatic monitoring, diagnosis, and remediation of software faults. Generally, such architecture is composed of two high-level elements: the software service whose integrity and availability we are interested in improving, and the elements of the system that perform the monitoring, diagnosis, and healing. The self-healing components can be viewed as a form of middleware — although, in some systems, such a separation is difficult to delineate [2], [3].

II. Characteristics of Autonomous Systems

Characteristics provide self-management potential to autonomic system and there are also known as the self-properties of the systems [4]. These characteristics are

either major or minor according to their contribution in the system. Both major and minor characteristics are explained in this section as shown in Fig. 1.

A. Major Characteristics

Autonomic systems have four major characteristics these are known as self-CHOP (configuration, Healing, Optimizing, protecting). This section explains all four of them.

- Self-Configuring

In this the system is able to remove faulty components, add new components, and configure them all by themselves without any help of humans. These components will be able to adapt the necessary environment provided by user for desired performance and quality.

- Self-Healing

By this property Autonomic system can cure itself and its components. System is able to detect all the faults and errors and try to remove it using corrective mechanism. ROC (Recovery Oriented Computing) is the process for performing self-healing in these systems. It provides various methods for recovery of system from faults and errors.

- Self-Optimizing

By this autonomic system can optimize themselves and their components for better efficiency and satisfy requirements of the users. Major aspects of this characteristic are Resource utilization and workload management.

- Self-Protecting

As biological systems have the capability of protecting themselves from danger and extreme calamities, autonomic systems have similar self-protecting mechanism because of this they are capable of defending their system and its components from various hacks and malware defects.

B. Minor Characteristics

Autonomous systems have four minor characteristics which are *self-aware*, *open*, *context aware* or *environment aware* and *anticipatory*. These are explained below.

- Self-Aware

The system is well aware of their internal states, components and their behaviors.

- Open

These systems are open which means these systems

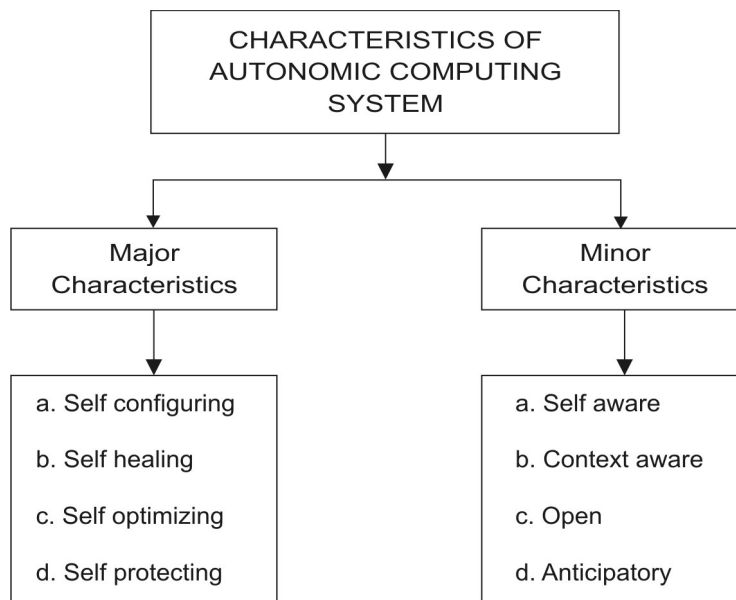


Fig. 1: Characteristics of Autonomic Computing System

can perform their tasks in hermetic environment as well as heterogeneous environment.

- Context aware

These systems are well aware of the environment in which they are being executed and they are capable enough to react to the changes in their executable environment.

- Anticipatory

These systems can use their optimized resources while keeping their complexity hidden from the user.

III. Life Cycle of a Self-Healing System

It is clearly visible that the process of self-healing software is a difficult task [5]. The self-healing life cycle is a crucial part of the self-healing system as shown in Fig. 2.

- Monitoring:** in this the system monitors all the necessary information gathered by it and categorizes it for processing.
- Analyzing:** analyzing the information helps you to determine the action to be taken by comparing the data with the systems requirements.
- Diagnosis:** It is a process of detecting all the errors and faults in the system by comparing on different techniques.

- Healing:** The process of healing consists of Planning and execution.

- A plan is carried out to successfully remove the errors and faults.
- The plan is then executed to achieve the desired results from the system.

- Knowledge:** this phase is most important as it contains the combined knowledge of all the above process.

IV. Advantages of Self Healing Systems

Self-healing systems can benefit all kind of organizations [7]. Following are some of the top benefits of automating self-healing systems.

- Cost saving:** these systems can save a lot of time which was preoccupied because of the service maintenance of the software and hardware. And we all know time is equally precious like money.
- Real time fixes without delay:** Self-Healing helps resolve all error's in just a matter of time. Whereas it might take a quite a while for any human to do the same thing.
- No need to involve people:** When the system can solve all the errors by itself, then you did not have to waste your time and energy on these things.



Fig. 2: Life Cycle of Self-Healing System

- d) **Less manual work for IT:** You can use the brain of your staff on more pressing matters than to solve bugs of the software.
- e) **Customer satisfaction:** these systems can give the best experience to the user, it will help you earn trust and respect from your clients and because of the satisfaction they will repeat business over time.

V. Challenges of Self Healing Systems

After 2001 when Paul horn from IBM gave this approach, it catches the eye of many companies and researchers all around the world. Many projects on this topic have been developed but none of them is purely safe and working. There are so many problems encountered by the researchers while developing these systems [8], [9]. These are stated below:

- a) Self-Healing is a new field for many of the researchers and companies. While making these systems it is a big challenge to make estimation about the possible problems and error's.
- b) Self-Optimization: an autonomous system should optimize all its components and use its resources, this is a big challenge for a system because its performance should not degrade when it adapts to different platforms.
- c) Self-management: all the elements of the self-healing system should have a strong relationship with each other; they should manage work together as a whole system. It is a challenge to face the unpredictable IT conditions and maintain a strong relationship with each other.
- d) Robustness: Huebscher et al. [12] stated that robustness is one of the major challenges for these systems. The components of the system should be capable enough to perform various functionalities without changing its core structure.
- e) For making an autonomous system some special tools are required. Selecting the best and right tool

is also a challenge.

- f) Uncertainty is also a big challenge for self-healing systems. It can change the policies and goals of the system in future.
- g) There are so many more challenges faced by the researchers while making an autonomous system.

VI. Conclusion and Future Scope

Autonomic computing is the new era of the IT environments all over the world. The main problem IT is facing is the "Complexity" of these systems which we have discussed in the challenges of this paper. After 2001 many researchers have started to show their interest in this new approach, and it is only a matter of time when we are going to encounter the first autonomous system which has all the self-managing properties in it.

The field of autonomous systems is an extremely vast in research. Rather than describing all the specific research topics in details, we outline two general research areas: Fault detection and fault recovery, in which self healing system will appropriate.

In Fault detection, one of the most effective factors of a self-healing system is their ability to detect faults and errors, and ways to implement this in less memory and less cost is necessary. In future we believe that self-healing systems will have a more detailed analysis of the software behaviour than today. This will be necessary to detect the faults which have occurred in the past such as software crashing, software hacking, etc. Future systems can learn from the existing profiles of the systems of today [10].

Fault recovery includes the ability to recover from a fault or an error one of the most important but complicated factor of the self healing approach. The success of self-healing systems will depend on this factor. In this the system has to select the correct method from recovering from the fault and but it also has to adapt these recovery methods to recover from different other faults and errors [11].

References

1. Taranjeet Singh and Avadhesh Kumar “Survey on Characteristics of autonomous System”, International Journal of Computer Science & Information Technology (IJCSIT) Vol 8, No 2, April 2016.
2. Rodosek, Gabi Dreo, et al. “Self-Healing Systems: Foundations and Challenges.” Self-Healing and Self-Adaptive Systems 09201 (2009).
3. Amjad A. Hudaib, Hussam N. Fakhouri, Fatima Eid Al Adwan, Sandi N. Fakhouri “A Survey about Self-Healing Systems (Desktop and Web Application)” Communications and Network, 2017, 9, 71-88.
4. Keromytis, Angelos D. “Characterizing self-healing software systems.” Proceedings of the 4th international conference on mathematical methods, models and architectures for computer networks security (MMM-ACNS). 2007. Brown, Richard E. “Impact of smart grid on distribution system design.” Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE. IEEE, 2008.
5. Tianfield, Huaglory, and Rainer Unland. “Towards autonomic computing systems.” Engineering Applications of Artificial Intelligence 17.7 (2004): 689-699.
6. Ansari, Zubair. “System and method for active diagnosis and self healing of software systems.” U.S. Patent No. 7, 293, 201. 6 Nov. 2007.
7. Nami, Mohammad Reza, and Koen Bertels. “A survey of autonomic computing systems.” Autonomic and Autonomous Systems, 2007. ICAS07. Third International Conference on Autonomic and Autonomous Systems, IEEE, 2007.
8. Frey Jr, Alexander H. “Self-healing computer system storage.” U.S. Patent No. 6,530,036. 4 Mar. 2003.
9. Park, Jeongmin, Giljong Yoo, and Eunseok Lee. “Proactive self-healing system based on multi-agent technologies.” Software Engineering Research, Management and Applications, 2005. Third ACIS International Conference on. IEEE, 2005.
10. Mcfarland, Norman R. “Self-healing control network for building automation systems.” U.S. Patent No. 7,139,239. 21 Nov. 2006.
11. Amin, Massoud. “Toward self-healing infrastructure systems.” Computer 33.8 (2000): 44-53.
12. Huebscher, Markus C., and Julie A. McCann. “A survey of autonomic computing – degrees, models, and applications.” ACM Computing Surveys (CSUR) 40.3 (2008): 7.

Empirical Study of Privacy Preserving Data Mining Techniques

Ankita

anckita21@gmail.com

Neha Gulati

gulatineha92@gmail.com

Abstract

Data mining is a way to analyze already existing data and produce new facts from that data. This large amount of data can also be misused if proper protective techniques are not used. Nowadays, people are very much concerned about their data as well as privacy. If data falls in wrong hands then sensitive information of the user can be misused. In order to make our personal data secure, privacy preserving algorithms are used. In this paper, various techniques of privacy preserving are discussed. They are: De-identification, k-anonymity, Perturbation, Randomized response, Condensation and Cryptography.

Key Words: Condensation, Cryptography, De-identification-anonymity, Perturbation, Randomized response.

I. Data Mining

Data mining is a set of automated techniques used to extract buried or previously unknown pieces of information from large databases. Successful data mining makes it possible to unearth patterns and relationships, and then use this “new” information to make proactive knowledge-driven business decisions.[1]

The problem of privacy protective data processing has become crucial in recent years thanks to the increasing ability to store personal information concerning users and also the increasing sophistication of data mining rule to leverage this data. Variety of techniques like classification, k-anonymity, association rule mining, clustering are prompt in recent years so as to perform privacy protective data processing. What is more, the problem has been mentioned in multiple communities like the info community, the applied math disclosure management community and also the cryptography community. data processing techniques are developed successfully to extracts data so as to support a range of domains selling, foretelling, medical diagnosing, and national security. However it's still a challenge to

mine sure styles of information while not violating the data house owners' privacy. For example, a way to mine patients' private information is Associate in Nursing current downside in health care applications. As data processing become additional pervasive, privacy issues are increasing.

II. Techniques of Privacy Preserving

A. De-Identification

This is an approach in which the removal of identifier attributes like Full Name, License no, Voter Id, Aadhar Card no is removed from the collected data so that recognition of an individual could not be done. Let us consider an example of a hospital where data of patients is recorded as follows:

Table-1: Patient Record

S. No.	Name	Aadhar Card No.	Gender	Age	Area	Diagnosis
1.	Ayush Khurana	123456721543	M	24	Pitam Pura	Stone Problem
2.	Priyanka Tripathi	153489321897	F	45	Ashok Vihar	Diabetes

S. No.	Name	Aadhar Card No.	Gender	Age	Area	Diagnosis
3.	Lokesh Sharma	165897 432569	M	27	Bhajan Pura	Kidney Disease
4.	Nitish Goel	265417 654398	M	55	Adarsh Nagar	Lung Cancer
5.	Deepika Iyer	369412 579863	F	49	Paschim Vihar	Diabetes
6.	Anamika Sharma	421587 963125	F	59	Malviya Nagar	Lung Cancer
7.	Pooja Sharma	452189 635713	F	29	Karam Pura	Stone Problem
8.	Om Prakash	654321 908786	M	48	Ashok Vihar	Heart Disease
9.	Jai Kumar Sharma	126742 416894	M	71	Patel Nagar	Lung Cancer
10.	Charu Malia	236984 123367	F	47	Mayur Vihar	Diabetes
11.	Kamlesh Jain	128645 796308	M	36	Nanak Pura	Lung Cancer
12.	Lal Prakash	415896 574236	M	32	Vasant Vihar	Heart Disease
13.	Pragya Prakash	765490 874378	F	53	Bhajan Pura	Stone Problem
14.	Jai Yadav	487218 052776	M	58	Anand Vihar	Kidney Disease
15.	Kashish Sharma	760912 675543	M	63	Paschim Vihar	Diabetes
16.	Leela Rani	267456 103894	F	70	Arya Pura	Heart Disease
17.	Neena Gupta	598743 765009	F	69	Nathu Pura	Stone Problem
18.	Manish Goel	609217 864390	M	57	Nirman Vihar	Kidney Disease

Now to protect the patient's privacy, hospital removes identifier attributes like Name, Aadhar Card No. The anonymous records are shown in following table:

Table-2: Patient record after removal of privacy details

S.No.	Gender	Age	Area	Diagnosis
1.	M	24	Pitam Pura	Stone Problem
2.	F	45	Ashok Vihar	Diabetes
3.	M	27	Bhajan Pura	Kidney disease
4.	M	55	Adarsh Nagar	Lung Cancer
5.	F	49	Paschim Vihar	Diabetes

S.No.	Gender	Age	Area	Diagnosis
6.	F	59	Malviya Nagar	Lung Cancer
7.	F	29	Karam Pura	Stone Problem
8.	M	48	Ashok Vihar	Heart Disease
9.	M	71	Patel Nagar	Lung Cancer
10.	F	47	Mayur Vihar	Diabetes
11.	M	36	Nanak Pura	Lung Cancer
12.	M	32	Vasant Vihar	Heart Disease
13.	F	53	Bhajan Pura	Stone Problem
14.	M	58	Anand Vihar	Kidney Disease
15.	M	63	Paschim Vihar	Diabetes
16.	F	70	Arya Pura	Heart Disease
17.	F	69	Nathu Pura	Stone Problem
18.	M	57	Nirman Vihar	Kidney Disease

Here we only have quasi-identifiers (gender, area, age) and sensitive information (diagnosis). By looking at this we cannot identify which patient has been diagnosed with which disease. Now suppose a friend of Priyanka visit the hospital and provided with the Table2 record. Now as she knows that age of Priyanka is 45 and she lives in Ashok Vihar she can easily identify that Priyanka has Diabetes. As we can see, though records do not contain any identifier attributes but still sensitive information can be learned by the use of quasi-identifiers and the external data.

B. K-Anonymity

This method was introduced to address the risk of de-identification of anonymised data through external data. For this approach, there should be at least k records in data table that share set of attributes that may become identifying for each individual. K-anonymity can be attained by suppression and generalization.

In Suppression technique, the quasi-identifiers are replaced by symbols like * constant values. The following table shows the anonymised data using suppression. Here 'Gender' attribute is replaced by symbol '*'.

Table-3: Patient Record after suppression

S.No.	Gender	Age	Area	Diagnosis
1.	*	24	Pitam Pura	Stone Problem
2.	*	45	Ashok Vihar	Diabetes

S.No.	Gender	Age	Area	Diagnosis
3.	*	27	Bhajan Pura	Kidney disease
4.	*	55	Adarsh Nagar	Lung Cancer
5.	*	49	Paschim Vihar	Diabetes
6.	*	59	Malviya Nagar	Lung Cancer
7.	*	29	Karam Pura	Stone Problem
8.	*	48	Ashok Vihar	Heart Disease
9.	*	71	Patel Nagar	Lung Cancer
10.	*	47	Mayur Vihar	Diabetes
11.	*	36	Nanak Pura	Lung Cancer
12.	*	32	Vasant Vihar	Heart Disease
13.	*	53	Bhajan Pura	Stone Problem
14.	*	58	Anand Vihar	Kidney Disease
15.	*	63	Paschim Vihar	Diabetes
16.	*	70	Arya Pura	Heart Disease
17.	*	69	Nathu Pura	Stone Problem
18.	*	57	Nirman Vihar	Kidney Disease

In our next table, we have used Generalization technique to hide some data by using more general values for quasi-identifiers. Here we have generalized the attribute 'Age', as shown in below table:

Table-4: Patient Record using generalisation

S.No.	Gender	Age	Area	Diagnosis
1.	M	>20 < 30	Pitam Pura	Stone Problem
2.	F	>40 < 50	Ashok Vihar	Diabetes
3.	M	>20 < 30	Bhajan Pura	Kidney disease
4.	M	>50 < 60	Adarsh Nagar	Lung Cancer
5.	F	>40 < 50	Paschim Vihar	Diabetes
6.	F	>50 < 60	Malviya Nagar	Lung Cancer
7.	F	>20 < 30	Karam Pura	Stone Problem
8.	M	>40 < 50	Ashok Vihar	Heart Disease
9.	M	>70 < 80	Patel Nagar	Lung Cancer
10.	F	>40 < 50	Mayur Vihar	Diabetes
11.	M	>30 < 40	Nanak Pura	Lung Cancer
12.	M	>30 < 40	Vasant Vihar	Heart Disease
13.	F	>50 < 60	Bhajan Pura	Stone Problem
14.	M	>50 < 60	Anand Vihar	Kidney Disease
15.	M	>60 < 70	Paschim Vihar	Diabetes
16.	F	>60 < 70	Arya Pura	Heart Disease
17.	F	>60 < 70	Nathu Pura	Stone Problem
18.	M	>50 < 60	Nirman Vihar	Kidney Disease

The best efficient way to anonymized data is using both techniques together i.e., suppression and

generalization. The following table shows the anonymised data.

Table-5: Patient Record using suppression and generalization

S.No.	Gender	Age	Area	Diagnosis
1.	*	2*	* Pura	Stone Problem
2.	*	2*	* Pura	Stone Problem
3.	*	2*	* Pura	Kidney disease
4.	*	3*	* Pura	Lung Cancer
5.	*	5*	* Pura	Stone Problem
6.	*	7*	* Pura	Heart Disease
7.	*	6*	* Pura	Stone Problem
8.	*	4*	* Vihar	Diabetes
9.	*	4*	* Vihar	Heart Disease
10.	*	4*	* Vihar	Diabetes
11.	*	4*	* Vihar	Diabetes
12.	*	3*	* Vihar	Heart Disease
13.	*	5*	* Vihar	Kidney Disease
14.	*	6*	* Vihar	Diabetes
15.	*	5*	* Vihar	Kidney Disease
16.	*	>=55	* Nagar	Lung Cancer
17.	*	>=55	* Nagar	Lung Cancer
18.	*	>=55	* Nagar	Lung Cancer

It blocks the external data linking attack. It is although not completely protecting user's privacy as sensitive data is still exposed. Examining the above table, the equivalence class which has quasi-identifiers as (*Nagar, >=55) the private information has only single values i.e, Lung Cancer. If an intruder knows who lives in *Nagar and has age >=55 regardless of the gender can easily find that he/she has Lung Cancer. This can also be known as Homogeneity Attack.

K-anonymised data is also exposed to another type of attack. Let us suppose some intruder knows a person who is admitted in this particular hospital & the patient lives in * Pura and is of 24 years of age. The intruder also knows that the patient doesn't drink much of water & is addicted towards salty foods. Based on this knowledge, the intruder learns that the patient's information is in 1, 2 and 3 record number. Without extra information, he cannot find out that whether the patient has stone problem or kidney disease. However, it is well known that people who are more

addicted to salt & drink less water have a high tendency to have stone problem. Therefore, the intruder can conclude with some assurance that patient has stone problem. This type of attack is also known as Background Knowledge attack.

In the above table, the quasi identifiers are Gender, Age and Area and $k=3$. Assuming $QI = \text{set of quasi identifiers}$, we can say $QI1 = QI2 = QI3$, $QI4 = QI5 = QI6 = QI7$ and $QI8 = QI9 = QI10$.

C. Perturbation Approach

Data perturbation is a famous strategy for security protecting information mining. The real test of information annoyance is adjusting security insurance and information quality, which typically considered as a couple of contradictive components. Geometric data perturbation strategy is a blend of Rotation, Translation and Noise expansion perturbation method. It is particularly helpful for information proprietors to distribute information while safeguarding protection – delicate data. Commonplace models incorporate distributing miniaturized scale information for research reason or redistributing the information to the outsider that gives information mining administrations.

Data perturbation approach is classified into two: the probability distribution approach and the value distortion approach. The probability distribution approach substitute the data with a new example from the similar division or by the division itself, and the value distortion approach perturbs data elements or attributes directly by either additive noise, multiplicative noise, or some other randomization procedures. There are three types of data perturbation approaches: Rotation Perturbation, Projection Perturbation and Geometric Data Perturbation.[2]

● Different Methods of Data Perturbation

- a) Noise Additive Perturbation
- b) Condensation-based Perturbation:
- c) Random Projection Perturbation:
- d) Geometric data perturbation:

D. Randomized Response Techniques

Randomized Response technique was first introduced

by Warner [3] in 1965 as a technique to solve the following survey problem: to estimate the percentage of people in a population that has attribute A, queries are sent to a group of people. Since the attribute A is related to some confidential aspects of human life, respondents may decide not to reply at all or to reply with incorrect answers.

E. Condensation Approach

Condensation approach develops compelled groups in dataset and after that creates pseudo information from the measurements of these bunches. It is called as buildup due to its approach of utilizing dense insights of the bunches to produce pseudo information [4]. It develops gatherings of no homogeneous estimate from the information, to such an extent that it is ensured that each record lies in a gathering whose estimate is in any event equivalent to its namelessness level. In this way, pseudo information is produced from each gathering in order to make a engineered informational index with indistinguishable total conveyance from the first information. The utilization of pseudo-information gives an extra layer of insurance, as it winds up hard performing antagonistic assaults on manufactured information. This approach helps in better protection conservation as thought about to different methods as it utilizes pseudo information instead of altered information. Also, it works even without updating information mining calculations since the pseudo information has indistinguishable configuration from that of the first information. At the same time, information mining results get influenced as expansive measure of data is lost as a result of the buildup of a bigger number of records into a solitary factual bunch substance.

F. Cryptography

Cryptography is related with the way toward changing over customary plain content into indiscernible content and the other way around. It is a technique for putting away and transmitting information in a specific structure with the goal that those for whom it is expected can peruse and process it. Cryptography shields information from burglary or adjustment, yet can likewise be utilized for client confirmation.

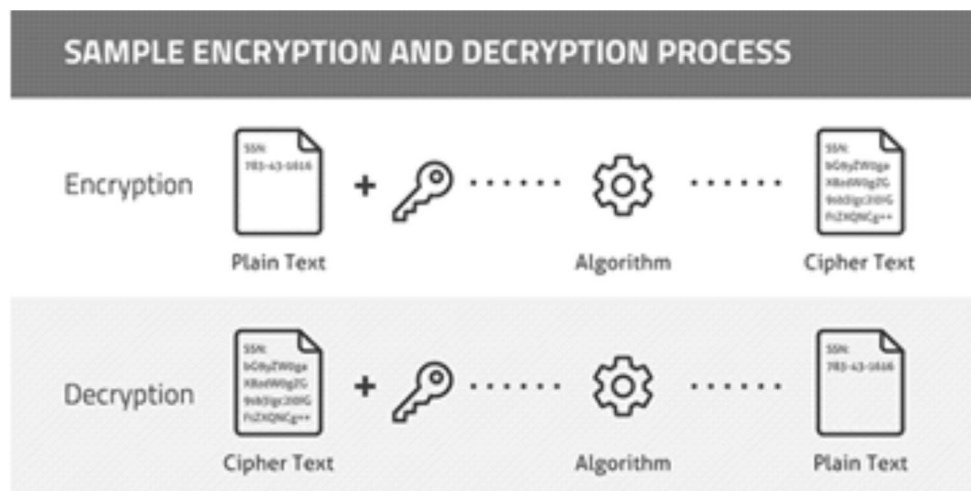


Fig. 1: Encryption and Decryption Process

Three types of cryptographic techniques used in general.

1. Symmetric-key cryptography
2. Hash functions.
3. Public-key cryptography

Symmetric-key Cryptography: Both the sender and beneficiary offer a solitary key. The sender utilizes this key to encode plaintext and send the figure content to the beneficiary. On the opposite side the beneficiary applies a similar key to unscramble the message and recoup the plain content.

Public-Key Cryptography: This is the most progressive idea in the last 300-400 years. In Public-Key Cryptography two related keys (open and private key) are utilized. Open key might be uninhibitedly disseminated, while its combined private key, remains a mystery. General society key is utilized for encryption and for unscrambling private key is utilized.

Hash Functions: No key is utilized in this algorithm. A settled length hash esteem is figured according to the plain content that makes it incomprehensible for the substance of the plain content to be recuperated. Hash capacities are likewise utilized by many working frameworks to scramble passwords.

Advantages

Cryptography is an essential information security tool.

It provides the four most basic services of information security –

- Confidentiality – Encryption strategy can watch the data and correspondence from unapproved disclosure and access of data.
- Authentication – The cryptographic procedures, for example, MAC and computerized marks can secure data against ridiculing and imitations.
- Data Integrity – The cryptographic hash capacities are assuming imperative job in guaranteeing the clients about the information uprightness.
- Non-repudiation – The advanced mark gives the non-revocation administration to prepare for the debate that may emerge because of forswearing of passing message by the sender.

All these key administrations offered by cryptography has empowered the lead of business over the systems utilizing the PC frameworks in amazingly proficient and successful way.

Drawbacks

Aside from the four crucial components of data security, there are different issues that influence the viable utilization of data –

- A emphatically scrambled, valid, and carefully marked data can be hard to get to notwithstanding

for a real client at a urgent time of basic leadership. The system or the PC framework can be assaulted and rendered non-useful by a gatecrasher.

- High accessibility, one of the key parts of data security, can't be guaranteed using cryptography. Different techniques are expected to prepare for the dangers, for example, disavowal of administration or complete breakdown of data framework.
- Another crucial need of data security of specific access control additionally can't be acknowledged using cryptography. Regulatory controls and methodology are required to be practiced for the equivalent.
- Cryptography does not make preparations for the vulnerabilities and dangers that rise up out of the poor structure of frameworks, conventions, and techniques. These should be settled through legitimate structure and setting up of a protective foundation.
- Cryptography comes at expense. The expense is as far as time and cash –

Addition of cryptographic methods in the data handling prompts delay.

The utilization of open key cryptography requires setting up and support of open key framework requiring the great looking money related spending plan.

- The security of cryptographic procedure depends on the computational trouble of scientific issues. Any achievement in tackling such scientific issues or expanding the registering force can render a cryptographic strategy powerless.

III. Conclusion

The essential goal of PPDM is elevating calculation to hide delicate information or over protection. This touchy information doesn't get uncovered to unapproved gatherings or intruder. In information mining there exists an exchange of among utility and protection of information. When we achieve one it definitely prompts the hindering effect on the other. Numerous PPDM procedures in presence are audited in the paper. At last, it is closed with the way that there is no single PPDM method in presence that eclipses each other system with connection to every conceivable criteria, for example, utilization of information, execution, trouble, similarity with techniques for information mining, etc. A specific calculation may work better when contrasted with another, on a particular standard. Different calculations might be found to work superior to each other on given measure. Specialists are doing broad research in guaranteeing that the delicate information of an individual isn't uncovered just as not bargaining the utility of information with the goal that the information can be valuable for some reasons.

References

1. Ann Cavoukian, Information and Privacy Commissioner, Ontario, "Data Mining Staking a Claim on Your Privacy", 1997.
2. Nimpal Patel and Shreya Patel, "A Study on Data Perturbation Techniques in Privacy Preserving Data Mining," International Research Journal of Engineering and Technology (IRJET), Volume: 02 Issue: 09, Dec-2015.
3. S. L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. The American Statistical Association, 60(309): 63–69, March 1965.
4. Aggarwal C, Philip S Yu, "A condensation approach to privacy preserving data mining", EDBT, 183-199, 2004.

A Comparative Study of Automation Testing Tools

Yamini Sharma

yamini.sharma.1910@gmail.com

Akanksha Prasad

akankshara93@gmail.com

Abstract

Software Testing is a sequence of events involved in executing any program or application with the intent of finding software bugs. In this paper, the three commonly used tools for automation testing are described and compared, they are: Selenium, TestComplete and QTP.

Key Words: Manual Testing, Automation Testing, Selenium, TestComplete, QTP.

I. Introduction

Software Testing is vital stage in Software Development Life Cycle. It is the way towards validating and verifying the technical and business requirements. It focuses on finding defects, providing quality product, and meeting customer requirements. If testing is to be done at an early stage, it reduces cost and time, and produces quality software to the client. Testing is a never ending process and some aspects are taken into consideration for stopping testing: Deadlines, Test Case Completion, Management decision etc. In Testing, one can never assume that there are no defects or the application is 100 percent bug free even if thorough testing is done.

So, to summarize we can say that:

- i) Software testing is required to check the dependability of the product.
- ii) Software testing guarantees that the framework is free from any bug that can bring about any sort of disappointment.
- iii) Software testing guarantees that the item is in accordance with the prerequisite of the customer
- iv) It is not possible to create software with zero defects without incorporating software testing in the development cycle.
- v) Regardless of how well the product configuration looks on paper, when the improvement begins and you begin testing the item you will discover bunches of deformities in the plan.

II. Types of Testing

This segment depicts the diverse kinds of testing that might be utilized to test a product amid SDLC:

- a) Manual Testing
- b) Automation Testing

Manual testing incorporates testing a product physically, i.e., without utilizing any computerized apparatus or any content. In this sort, the analyzer assumes control over the job of an end-client and tests the product to recognize any surprising conduct or bug. There are distinctive stages for manual testing, for example, unit testing, integration testing, system testing, and user acceptance testing.

Automation testing, which is also known as Test Automation, is when the tester writes scripts and uses software to test the product. This process involves automation of a manual process. Automation Testing is used to re-run the test scenarios that were performed manually, quickly, and repeatedly.

III. Fundamentals of Automation Testing

It is not possible to robotize everything in a software. The zones at which a client can make transactions, for example, the login structure or feedback forms, any territory where huge number of clients can get to the product all the while ought to be automated.

Automation Testing is applicable when: there are large projects, testing is to be done on same areas at

regular interval, and requirements do not change frequently.

Automation is done with the help of any scripting language and an automation software or tool.

A. Process of Automation

- a) Identifying zones for automation within a software
- b) Identifying appropriate test automation tool
- c) Developing test scripts and test suites
- d) Executing scripts
- e) Making result reports

B. Software Testing Tools

The following tools can be used for automation

testing:

- HP Quick Test Professional
- Selenium
- IBM Rational Functional Tester
- SilkTest
- TestComplete
- Testing Anywhere
- WinRunner
- LoadRunner
- Visual Studio Test Professional
- WATIR

C. Flow of Automation Testing

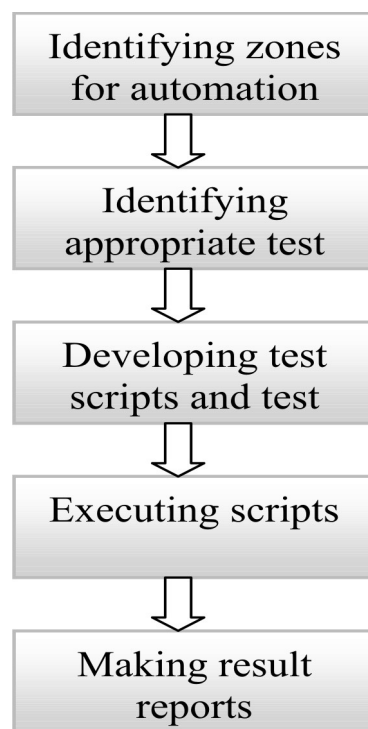


Figure 1: Flow of Automation Testing

D. Criteria for Selection of Tool

Following specifications are considered for selection a tool:

- Information driven capability: It helps you test your application against multiple data sets and data values.
- Troubleshooting and Logging Capability: It helps with *debugging* the exact step and the page where the failure occurred.
- Platform Independence Capability: It supports all patterns or standards of an industry.

Table 1: Comparison of Automation Tools

S.No.	Basis	Selenium	Testcomplete	QTP
1	Test Development Platform	Cross-Platform	Windows	Windows
2	Application Under Test	Web Apps	Windows Desktop, Web, Mobile Apps	Windows Desktop, Web, Mobile Apps
3	Scripting Languages	Java, C#, Perl, Python, Javascript, Ruby, PHP	Javascript, Python, Vbscript, Jscript, Delphi, C++, And C#	Vbscript
4	Programming Skills	Advanced Skills Needed To Integrate Various Tools	Not Required. Recommended For Advanced Test Scripts	Not Required. Recommended For Advanced Test Scripts
5	Ease Of Installation And Use	Require Installing And Integrating Various Tools	Easy To Setup And Run	Easy To Setup And Run
6	Object Storage And Maintenance	Xpath, UI Maps	Built-In Object Repository, Detecting Common Objects	Built-In Object Repository, Smart Object Detection And Correction
7	Devops/ALM Integrations	No Additional Libraries	Many	Many
8	Cost	Free	License & maintenance fees	License & maintenance fees
9	License type	Open source (Apache 2.0)	Proprietary	Proprietary
10	Test Analytics	No	No	No

- Email Notification Capability: It is used for easy customer communication and feedback.

IV. Different Automation Testing Tools

Automation Testing is the automation process of Manual Testing. Following automation tools are discussed and compared below:

A. Selenium Testing Tool

Selenium is a free (open source) computerized testing suite for web applications crosswise over various platforms. Selenium isn't only a solitary instrument yet a suite of software's, each taking into account distinctive testing needs of an association.

It has four parts:

1. Selenium Integrated Development Environment (IDE)
2. Selenium Remote Control (RC)
3. WebDriver
4. Selenium Grid

B. TestComplete Testing Tool

TestComplete is a testing tool that gives you a chance to make, manage and run tests for any Windows, Web or Rich Client programming.

It is available with two licenses and a free trial version valid for 30 days. TestComplete offers a wide range of test automation capabilities:

1. Keyword Testing
2. Test Record and Playback
3. Data Driven Testing
4. Scripted Testing

C. QTP Testing Tool

HP QTP is an automation testing tool which helps testers to execute automated scripts to identify any errors, defects or gaps in contrary to the expected results of the application under test.

Its features are:

1. Support for OS and browser
2. UFT extension in Chrome store
3. Support for windows runtime objects
4. Newly supported technologies
5. Extended support for various technologies

V. Comparison of Automation Tools

A characterizing factor for effectively applying test automation in software based projects is picking and utilizing the right automation tools.

The table below provides a comparison of the tools based on the key features of software automation.

References

1. Richa Rattan, "Comparative Study Of Automation Testing Tools: Quick Test Professional & Selenium", *Vsrd International Journal Of Computer Science & Information Technology*, Vol. 3 No. 6 June 2013.
2. Manjit Kaur, Raj Kumar, "Comparative Study Of Automated Testing Tools: Test Complete And Quick Testpro", *International Journal Of Computer Applications (0975-8887) Volume 24-No. 1*, June 2011.
3. Zhi Quan Zhou, Bernhard Scholz, Giovanni Denaro, "Automated Software Testing And Analysis: Techniques, Practices And Tools", *Proceedings Of The 40th Annual Hawaii International Conference On System Sciences (Hicss'07) 0-7695-2755-8/07*.
4. Chayanika Sharma, Sangeeta Sabharwal, Ritu Sibal by "A Survey on Software Testing Techniques using Genetic Algorithm", *International Journal of Computer Science Issues*, Vol. 10, Issue 1, No 1, January 2013.
5. Dudekula Mohammad Rafi, Katam Reddy Kiran Moses, Kai Petersen Mika V. Mantyla by "Benefits and Limitations of Automated Software Testing: Systematic Literature Review and Practitioner Survey".

VI. Conclusion

There is no one perfect tool for every scenario, as the selection of tool depends on the project requirements. When cost is an important factor, Selenium is a choice because of open source availability, no licensing and maintenance fees. Also, it is open for integration with other tools and frameworks to enhance its capabilities.

When User Support is the main factor, Unified Functional Testing (UFT) is probably the most popular commercial tool for test automation because it serves a dedicated user support through its large user community. Also, QTP provides support for many platforms including SAP, .NET, Java, Web Services, Windows and Web. QTP integrates well with its Test Management tool and has inbuilt provision for "Object Repository".

UFT is used to automate legacy based app and selenium for browser based workflows.

TestComplete offers a choice among multiple scripting languages and if the budget allows, TestComplete can become a great tool for a team of less skilled developers. By buying modules on demand, you can customize the pricing and leverage excellent technical support.

6. Pradeep Kumar Singh, Om Prakash Sangwan by “A Study and Review on the Development of Mutation Testing Tools for Java and Aspect-J Programs”, I.J. Modern Education and Computer Science, 2014.
7. A. Sinha, S. Sutton, and A. Paradkar by “Text2Test: Automated inspection of natural language use cases” (2010).
8. Manjit Kaur and Raj Kumari : “Comparative Study Automated Testing Tools: Test Complete and QuickTest Pro”, International Journal of Computer Application, Volume 24 No.1 June 2011.
9. Vishal Sangave, Vaishali Nandedkar by “A review on Automating Test Automation”, International Journal of Advance Research in Computer Sciece and Management Studies, Volume 1, Issue 12, December 2014.
10. Singh M, Gupta P.K., Mishra S., Automated Test Data Generation for Mutation Using Aspect J Programs, In Proceedings of ICIP-2011, IEEE, 2011.

A Novel Approach to Assess the Quality of Student Using Fuzzy Logic and Fuzzy C-Means Clustering

Lokesh Jain

nsitlokesh.jain@gmail.com

Abstract

In the current educational system, it is very tough to judge the quality of the students based on their score in the exam. The difficulty level of exam is also one of the factors to ensure the quality of the student. In this paper, we proposed a new method based on fuzzy logic to evaluate the performance of the student as well as also apply the fuzzy C-means Clustering to categorized them into four categories (Excellent, good, average, weak). Fuzzy logic has been used to model the qualitative aspects of human data, and decision making as done by human beings by applying the rule base. The proposed approach is also better in terms of precision, recall, F1-score, and accuracy with respect to other methods.

Key Words: Fuzzy Logic, Student's performance, Fuzzy C-means clustering.

Delhi Youth Usage of Prominent Messengers: A Case Study Analysis

Sushma Malik

sushmamalikiitm@gmail.com

Mehak Tawakley

mehaktawakley@gmail.com

Megha

meghasingh.1607@gmail.com

Mayank Gupta

mittalmayank095@gmail.com

Abstract

Messengers are widely used for chatting and are the most popular services over mobile phone networks. According to a recent study, it was found that youths now a day's tend to connect with their peers via messages rather than meeting in person and in order to achieve that, they use instant messenger that provides them a platform to connect in a convenient way through a user friendly interface and also loads them with various other features such as images, video and audio sharing, money transfer, live video chat, group chats etc and all for no cost. Along with above, instant messengers also help them to fulfill their academics and work needs. This paper is intended to explore impact of using Instant Messaging Applications on youths.

Key Words: Instant Messenger, Instant Messaging, IM, Social Media, Internet Relay Chat, Computer-mediated communication, youth

Analysis of Emerging Research Areas of Social Networks

Deepti Gupta

deeptigupta72@gmail.com

Archita Verma

architaverma2016@gmail.com

Garvit Singh

garvitsingh292@gmail.com

Abstract

Social network analysis is gaining a striking attention from the preceding decade. The techniques for Social Network Analysis have turned up as the most successful applications of internet. It is essentially considered branch of learning analytics. People are showing huge interest in social network for information, news and getting opinion of other users on different subject matters. The heavy dependence on social network sites causes them to generate massive amount of data that have mainly three computational issues i.e. size, noise and dynamism. Due to these issues, examining the data manually is very complex. In today's era, it has turned into a widespread routine to not to frame a proper sentence with appropriate grammar and spelling. This may lead to distinct forms of uncertainties similar to lexical, syntactic and semantic and just because of this form of fuzzy data, it is hard to find out the actual data order. Hence, we use the techniques of data mining for analyzing them. These techniques also provide information retrieval and machine learning. In this paper, we have discussed the opportunities and challenges that arise from this transformation of social networks, and provide the framework to better recognize the hastily mounting fields of these social networks.

Key Words: Social network Analysis, data mining, information

Blockchain Technology: Overview, Working and Application

Preetam Singh Choudhary

spreetam27@gmail.com

Dona Biswas

donabiswas45@gmail.com

Vishakha Sehdev Verma

vs4821@gmail.com

Ankit Verma

ankit.verma.aquarius@gmail.com

Abstract

The invention of blockchain technology attracts lots of people. Soon realized that the technology can be used in a different thing like storing medical records, creating a digital document etc. In this paper, we explore how it works on a basic level and what problems it solves.

Key Words: Blockchain, Cryptocurrency, Hash

Prediction of Movies Performance Using Python

Deepanshu Saxena

deepanshusaxena21@gmail.com

Kailash Parihar

kailashparihar1211@gmail.com

Lokesh Jain

nsitlokesh.jain@gmail.com

Abstract

This paper details our analysis of the Internet Movie Database (IMDb), a free, user-maintained, online resource of production details for over 390,000 movies, television series and video games, which contains information such as title, genre, box-office taking, cast credits and user's ratings. We gather a series of interesting facts and relationships using a variety of data analyzing techniques. In particular, we concentrate on attributes relevant to the user ratings of movies, such as best movies, best genres, movies released per year, popular language among people and which country is producing maximum no. of movies. This paper also reports on the techniques used, giving their implementation and usefulness. We have found that the IMDb is difficult to perform data analysis upon; due to the format of the source data so we collected it, arranged it and converted into CSV format. Afterwards we implemented the data analyzing techniques and generated the results.

Key Words: IMDB, prediction, python, data analysis, movies, Hollywood.

Virtual Reality in Actual Reality

Suruchi Sinha

suruchi.sinha90@gmail.com

Ranajoy Dutta

ranajoydutta7@gmail.com

Himanshi

himanshikohli110016@gmail.com

Harshit Dahiya

harshitdahiya816@gmail.com

Abstract

Virtual Reality (VR) is an interactive and computer-generated experience taking place within a simulated environment. Virtual reality (VR) is a rapidly growing computer interface that strives to immerse the user completely into an experimental simulation, thereby greatly enhancing the overall impact and providing a much more intuitive link between the computer and the human participants. In this paper we discuss about how virtual reality is setting its roots in all domains. We start with history of VR and proceeds with discussion about ongoing technologies. Then comes the implementation of virtual reality in various fields such as education, healthcare, gaming, entertainment etc.

Key Words: Pygmalion's Spectacles, Stereoscopic photos, Virtual Reality.

A Big Data Approach towards Optimization of Railway Transportation using Machine Learning

Harsh Arora

harsh.k.a.579@gmail.com

Shivangi Kharbanda

shivangikharbanda4@gmail.com

Abstract

Current train delay prediction systems do not take advantage of the optimizing tools and techniques for handling and extracting useful and actionable information from data collected by the railway information systems. Compared with high speed trains and metro subways, heavy haul train operations are more challenging due to their complexities. The most important issue is safety and efficiency of heavy haul train transportation. Over the time, the issues encountered by the heavy haul trains have attracted outmost attention of numerous researchers due to their importance in improving transportation system in railways. Nowadays, Big data analytics (BDA) has risingly attracted a wise attention of analysts, researchers and practitioners in railway transportation and engineering. The public transportation industry has been at the forefront in using and implementing Analytics and Big Data because they are generally closed systems that involve large volumes of data. This paper gives an overview of Big Data and its technologies in railways using machine learning. This paper also gives an insight of the applications of Big data in Railways.

Key Words: Big data, Machine learning, Railways

Mobile OS: A Case Study of iOS

Sushma Malik

sushmamalikiitm@gmail.com

Pulkit Bahl

pulkitbahl11@gmail.com

Prakhar Sharma

sharmaprakha1912@gmail.com

Abstract

It takes a toll to build the building blocks of an entity, especially when that entity is the best-selling range of the most valuable company in the world. Apple's reputation is attributable to heaps of revolutionary products and services such as flagship I-phones, the quirky I-pods, and the sleek I-pads, that bent the technological frontier of the world in their distinctive way, but one thing that unites all of them is Apple's home-grown operating system (OS), iOS. Drawing more hearts than your favorite lullabies or a James Bond classic, the iOS nurtures the idea of "Simplicity and consistency with effectiveness", and hence, deserves to be the subject matter of this research paper. In our most humble attempt to characterize the operating system and its impact on the world, we have highlighted factors such as history of Apple's classic homemade operating system – iOS, scope and limitations of iOS app development, world renowned security features, comparison with its rival – Google's Android followed and some astonishing facts and figures about Apple and its iOS.

Key Words: Android, Apple, iOS, OS.

Sixth Sense Technology: The Future of AI

Yashika Kohli

yashika14.kohli@gmail.com

Geetika Lunj

geetikalunj@gmail.com

Vishakha Sehdev Verma

vs4821@gmail.com

Ankit Verma

ankit.verma.aquarius@gmail.com

Abstract

Sixth Sense is a technology which is stealing attention from everyone worldwide by the present and future generations. Sixth Sense is a portable, easy to manage gestural interface that elevates the physical world with the computerized world surrounding us. It incorporates hardware elements connected wirelessly to other computing devices. It uses enabling surfaces, walls and physical objects as an interface. Sixth Sense delivers the information from the surrounded surface. The main objective is to put together components of the physical world to the digital computerized world. In this paper we have demonstrated and simplified about Sixth Sense Device-Functioning, Parts Assembled, Idea of Name, Present State, Future Vision, Applications/Usage, Advantages.

Key Words: Sixth Sense, Artificial Intelligence, Networks

Fuzzy Linguistic Analysis of a Reheating-Furnace System Subject to Preventive Maintenance

Indu Uprety

induuprety@yahoo.co.in

Kalika Patrai

patrai.kalika@gmail.com

Abstract

This paper presents the stochastic model of a 'Reheating-furnace system' subject to preventive maintenance (PM) by listing out different failure and repair modes to evaluate the availability of the system under fuzziness. The system consists of three operating Reheating-furnaces, roll table and pulpit in standby configuration. A Markov model has been developed to represent the stochastic process of a dynamic system whose states change as time moves on. As the Markov model requires some parameters which may not be exactly known during the calculation, this allows uncertainty-based parameters that are represented as fuzzy numbers. This paper presents the combination of fuzzy theory and Markov model and uses fuzzy linguistic variables to handle the uncertainty in system performance. Each linguistic variable is represented by a fuzzy triangular number. As a consequence several curves which have associated degrees of confidence are obtained as a measure of fuzzy availability of a multi-state system subject to PM using the α -cut (interval of confidence) approach.

Key Words: Preventive maintenance, steady-state availability, fuzzy Markov, fuzzy linguistic approach, α -cut

Decoding the Concepts of Competitive Programming

Sushma Malik

sushmamalikiitm@gmail.com

Anant Kaushik

anant.kaushik2@gmail.com

Mehak Tawakley

mehaktawakley@gmail.com

Mayank Gupta

mittalmayank095@gmail.com

Abstract

Ever since programming has come into action, programming contest too have gained some attention. Everyone who loves programming, loves to take part in programming contest, either as a participant or as a coach. As a participant we wish to gain more knowledge, if we start it from earlier stages and as a coach, we aim to get better results for future from our participants and like to prepare them too. It takes time to learn all basic algorithms and logics. Also, implementing them in real life cases needs proper knowledge and practice. Competitive programming aims for same. It helps us to prepare for future and explore us to various real-life problems for which we can use algorithms to solve. This paper aims to decode the concepts of competitive programming.

Key Words: Competitive programming, code, coder, mind sport, practice, programming.

Learning Analytics and Educational Data Mining: Tools for Improving Quality Education in Higher Education Institutions

Monika

monikahooda555@gmail.com

Chhavi Rana

chhavi1jan@yahoo.com

Abstract

In this digital era, the higher education world is becoming complex, scaled and more challenging. The key challenge that Higher Education Institutions (HEIs) and schools have to face is not to be inundated with the quantity of data. Then HEIs must need a strategy to handle such an enormous data about students and learners to gain insights into this data to make decisions about learners, teachers, administrators and policy-makers. Learning analytics turn out to be a very useful tool for HEIs and schools that help them in providing quality education by uncovering big data's potential. As the Indian education sector is growing faster, LA is the next big thing needed to amplify quality education and employability skills in students. This paper focuses on the understanding the significance of Learning Analytics (LA) and Educational Data Mining (EDM) in higher education. The paper main objective is to outline major reasons why HEIs should implement LA and EDM methods and various ethical issues of LA and EDM in the context of higher education. This review paper finally concludes a roadmap for potential direction for the advancement and quality education in HEIs and schools by implementing LA and EDM methods.

Key Words: Learning Analytics; Educational Data Mining; MOOC; higher education; Big Data Analytics.

An Empirical Study on Agile Methods & Tools

Anand Sehgal

anandsehgal05@yahoo.co.in

Abstract

Software Development in the recent years is focusing on agile methods. This is a generic term. It is covering a number of specific methods with names Kanban, Adaptive Software development, Crystal, Extreme Programming, Feature Driven Development, and Scrum. All these are related to RAD (Rapid Application Development). Each of them has a different emphasis.

Key Words: PRAD, SCRUM, FDD, DSDM, Agile Metrics, Agile Tools

Downloading Multiple Content From Websites Using Parallel Downloading Technique

Shallu Juneja

shallujuneja@mait.ac.in

Garvit Verma

garvitcse@gmail.com

Abstract

A Multiple File Downloader is a computer program dedicated to the task of downloading possibly unrelated stand-alone files from the Internet for storage and then securing them in an encrypted format so that it is only readable by the user when decrypted through his/her password. Thus, any other person who does not know the password to the wallet is creating a multiple file downloader for storing downloaded files of any format and keeping them inaccessible. Different users can create their own wallets and store them password protected. This is unlike a World Wide Web browser, which is mainly intended to browse web pages, composed of a multitude of smaller files, where error-free moving of files for permanent storage is of secondary importance. (A failed or incomplete web page file rarely ruins the page.) The typical download manager at a minimum provides means to recover from errors without losing the work already completed, and can optionally split the file to be downloaded into 2 or more segments, which are then moved in parallel, potentially making the process faster within the limits of the available bandwidth. (A few servers known to block moving files in parallel segments, on the principle that all users should share server capacity equally.) Multi-source is the name given to files that are downloaded in parallel. Download managers are useful for very active Internet users. For dial-up users, they can automatically dial the Internet Service Provider [HYPERLINK "http://en.wikipedia.org/wiki/Internet_Service_Provider"](http://en.wikipedia.org/wiki/Internet_Service_Provider) at night, when rates or tariffs are usually much lower, download the specified files, and hang-up. They can record which links the user clicks on during the day, and queue these files for later download. For broadband users, download managers can help download very large files by resuming broken downloads, by limiting the bandwidth used, so that other internet activities are not affected (slowed) and the server is not overloaded, or by automatically navigating a site and downloading pre-specified content (photo galleries, MP3 collections, etc.) this can also include automatically downloading whole sites and regularly updating them (see Mirroring). Many download managers' support Metalink, an XML file listing mirrors, checksums, and other information useful for downloading.

Key Words: Multi-source; Checksums; Metalink

Performance Comparison of Proactive Routing Protocols in Wireless Sensor Networks Using Qualnet Simulator

Puneet Garg

puneetgarg.er@gmail.com

Abstract

Wireless Sensor Networks are becoming the new area of interest for the emerging researchers now-a-days because of its vast applications in the real-time situations. Wireless Sensor Networks offer a great variety of routing protocols. These protocols can be categorized into three major categories viz. Proactive, Reactive and Hybrid. Among these categories, the most important one is proactive routing protocols. Proactive routing protocols are also sometimes known as Table driven protocol as every node maintains a table for its routing before any communication over network. Proactive routing protocol covers DSDV, WRP and OLSR. The sole objective of this paper is aimed to compare all available proactive routing protocols in wireless sensor networks. It is believed that this paper shall help new researchers in selection of the suitable routing protocol in all aspects for their aimed research. For comparing the available routing protocols, Qualnet simulator is being used since it is recognized worldwide for its accurate and most precise results. Jitter, Average throughput, packet delay, transmission error are taken as performance metrics for comparing the mentioned protocols.

Key Words: Hybrid, Proactive, Qualnet, Reactive

Categories of Threats and Attacks to Information System in An Organization

Rashmi Jha

dr.rashmijha22@gmail.com

Divyansha Saini

dishusaini01@gmail.com

Shurabhi Shanker

dr.surabhashanker@gmail.com

Nishant Nirmal

nirmalnishant645@gmail.com

Abstract

The Security of Information systems is being highly challenged by the Propagation of web -based applications including e-commerce and a variety of information services. Information systems are frequently exposed to variety of threats which leads to damages that might result into significant financial losses. It ranges from small losses to entire information system destruction. It is essential that security of an information system should protect the confidentiality, integrity, and availability of the system. Now a day, organizations are struggling to understand what the threats to their information assets are and how to obtain the necessary means to counter them which continues to pose a challenge. Further improve our understanding of security threats; this paper addresses a security threat classification model which allows us to study the threats class impact over time. This paper also proposes different criteria of information system security risks classification and gives a review of most threat classification models. We define a model for information system security threat classification in order to propose a classification architecture that supports all threat classification principles and helps organizations implement their information security system.

Key Words: Information system security; Threat classification; Threat; criteria; security risk

National Conference on Innovative Research in Computer Science and Information Technology: “Advanced Computing Trends & Cyber Security”

The purpose of the Conference is to bring together the academicians, researchers, scientists, engineers and practitioners from various institutions, universities, colleges and industry to discuss the relevance of various futuristic technologies emerging in Artificial Intelligence, IOT, Machine Learning, Networking, Big Data, Data Mining, Cyber Security and also areas related to innovations in Information Technology throughout the world. This Conference is aimed to expand its coverage in the areas specified where expert talks, young researcher's presentations will be placed in every session will be inspired and keep up your enthusiasm.

Conference Objectives

- To present the latest research and results of experimental work related to Computer Science and Information Technology
- To endow opportunities for delegates to exchange new ideas and application experiences.
- To address the real-world opportunities and challenges faced by today's IT professionals.
- To promote high level of interaction among stakeholders in theoretical, experimental and applied concepts.

Conference Themes

Topics covered by the Conference include, but are not restricted to, the following:

Advanced Computing Trends

- Computer Architecture and Engineering
- Web Security, Client and Server Security
- Applications of Cryptographic Techniques
- Self-protecting and Self-Healing System
- Artificial Intelligence and Robotics
- Soft Computing
- Software Engineering
- Human Computer Interaction
- Internet of Things (IoT)
- Social Networking
- Crowd Sourcing
- Image Processing
- Theoretical Computer Science
- Tracking and Locating Mobile Users
- Energy Saving Protocols
- Analysis of security protocols
- Communication and Networks
- Authentication and authorization of Systems
- Bots and Botnets
- Big Data Analytics

Cyber Security

- Current Trends in Ethical Hacking
- Web-Resource Protection
- Cyber Forensics and Cyber Laws
- Crypto-Systems
- Security aspects in Open Source Software
- Authentication and Authorization strategies in Networked Environment
- Management of Confidentiality and Privacy
- Securing Computing Infrastructures
- Security Breaches in Big Data Analysis
- Machine Learning applications in security
- Biometric Technologies
- Steganography
- Trojans, Vulnerability and Zombie Drones
- Emerging IT Security Models and Practices
- Intrusion detection and prevention
- Cryptography
- Diagnostics and Forensics for Security
- Information Assurance and Security
- Feature extraction and selection techniques



Institute of Innovation in Technology and Management

**Affiliated to GGSIP University, 'A' Grade by GNCTD & Assessed as 'A+' by SFRC
NAAC Accredited & ISO 9001-2008 Certified Institute**

D-27/28, Institutional Area, Janakpuri, New Delhi-110058
Tel: 011-28520890, 28520894, Tele Fax: 011-8520239
E-mail: director@iitmjp.ac.in Website: <http://www.iitmjp.ac.in>