KEMENTERIAN PENDIDIKAN MALAYSIA

UPM

PROCEEDINGS OF THE

INTERNATIONAL SYMPOSIUM ON

ICT MANAGEMENT AND ADMINISTRATION (ISICTMA 2019)

in conjunction with

UNIVERSITY ICT ADMINISTRATION CONFERENCE 2019

powered by IDEC

Strengthening Smart Campus

31 July - 2 August 2019

Putrajaya Marriott Hotel, Malaysia

ppict2019.upm.edu.my

{Smart ; Secure ; Safe ; Shared ; Seamless }

PROCEEDINGS OF THE
INTERNATIONAL SYMPOSIUM ON ICT MANAGEMENT AND ADMINISTRATION (ISICTMA2019)

EDITORS:
Fatimah Sidi
Iskandar Ishak
Lilly Suriani Affendey
Marzanah A. Jabar
Noritah Omar

TECHNICAL EDITORS:
Saiful Ramadzan Hairani
Rizal Razak
Nur Mahirah Afiqah Abdul Rahman

ISBN 978-9676-960-456-6

Published by:

Database Technologies and Applications Research Group (DbTA),
Faculty of Computer Science and Information Technology,
Universiti Putra Malaysia,
43400 UPM Serdang,
Selangor Darul Ehsan,
Malaysia

All rights reserved. No part of this publication may be produced in any form or by any means without prior permission from copyright holders.

TABLE OF CONTENTS

FOREWORD BY VICE CHANCELLOR OF UNIVERSITI PUTRA MALAYSIA

Assalamualaikum Warahmatullahi Wabarakatuh and Salam Sejahtera,

First and foremost, it gives me great pleasure to extend my heartiest congratulations to Infocomm Development Centre (iDEC), Universiti Putra Malaysia, for advocating this year's University ICT Administration Conference 2019 (PPICT2019) with great success.

By giving much needed attention to the many emerging and pressing issues pertaining to education in Malaysia, UPM, in promoting technology to educators, has made the initial step by creating *Putra ® Virtual ID* (Putra®VID) and *Putra Future Classroom* as part of our initiative for a Smart Campus. The idea of a Smart Campus arises out of our desire to create a more technology-based learning and teaching ecosystem which will make the entire campus a more conducive and enjoyable learning environment for students.

Through the Smart Campus initiative, academics and students can explore and discover the best practices of the digital age in the face of the 4th Industrial Revolution (IR 4.0). This will ensure that our knowledge legacy will continue to be a guiding light, by creatively and innovatively addressing current challenges and anticipating future changes.

We hope that the technology implemented for the Smart Campus will continue to stimulate innovative learning and teaching, and invite more collaboration with research partners from industry, educational institutions and other communities. Such effort will enhance the university's reputation as a leading institution which provides quality learning and teaching in keeping with the values and philosophy of the national education.

I truly hope that the resilience, perseverance and brilliance shown by the organizers will remain constant and elevated to greater heights.

Thank you, and Wassalam.

WITH KNOWLEDGE WE SERVE

Prof. Datin Paduka Dr. Aini Ideris
Vice Chancellor,
Universiti Putra Malaysia
25th July 2019

FOREWORD BY PPICT2019 CHAIR

Assalamualaikum warahmatullahi wabarakatuh.

Dear Everyone,

First of all, we would like to welcome all presenters and participants to Universiti Putra Malaysia, and especially to the University ICT Administration Conference 2019 (PPICT2019). This is the first time that a conference that puts a focus on ICT administration for public universities has been organized.

The conference theme, "Memperkasa Kampus Pintar" or "Empowering Smart Campus" is an invitation to discuss case studies, efforts to promote smart campus initiatives, and new methods of transforming ICT as a powerful tool to support the transformation of Malaysian universities in becoming smart campuses in line with the fourth Industrial Revolution (IR4). The conference is mainly meant to serve as a platform for all stakeholders in the ICT industry (particularly those at public universities including practitioners and policy makers, academics, researchers, and students) to meet, discuss and deliberate on issues regarding technologies, innovations and research findings, and to showcase the latest inventions and innovations related to Smart Campus initiatives from all over Malaysia and abroad.

We at UPM, and in particular the InfoComm Development Center, UPM, are honoured to be given the trust to host this remarkable event. We endeavour to provide and present the best so that the event is successful in meeting its objectives, Insha'Allah.

We would also like express our thanks and gratitude to everybody who has directly and indirectly been involved in preparing and hosting PPICT 2019. Our special appreciation goes to: the Minister of Education, Yang Berhormat Dr. Maszlee Malik, Dato' Menteri Besar Selangor, YAB Tuan Haji Aminudin Shari, the Board of IPTA ICT Directors, MAPITA, Dr Iskandar Ishak, the Chair of ISICTMA2019 in conjunction with PPICT2019, Committee Members, Participants and all those involved are duly noted in the Conference Proceedings. We wish you a successful event and Happy conferencing.

Thank you, and Wassalam.

Associate Professor Dr. Fatimah Sidi
Director,
InfoComm Development Center (IDEC),
Universiti Putra Malaysia,
and Chair,
University ICT Administration Conference 2019 (PPICT2019).
25th July 2019

FOREWORD BY ISICTMA2019 CHAIR

Alhamdulillah, praise be to Allah S.W.T for granting us the opportunity to organize this symposium.

The use of Information and Communication Technology or ICT for management and administration is an important part in running an organization. Through ICT, jobs and tasks in organizations have been transformed to be more efficient in terms of time and cost due to the advancement of techniques and approaches that have evolved through many years of research. The introduction of smart technology based on high-speed network, artificial intelligence and Big Data for instance, has transformed ICT into a powerful tool to support administrating and managing organizations. This symposium is timely because it provides an excellent platform for researchers as well as practitioners of ICT management and administration to discuss and present original ideas, recent results and achievements on issues and challenges related to ICT management and administration.

I would like to take this opportunity to welcome all speakers, paper presenters and participants to this symposium and wish all of you a successful event.

I would also like to thank all parties who have contributed to the organization and success of this symposium.

Thank you, wassalam.

Dr. Iskandar Ishak
Chair,
International Symposium on ICT Administration and Management (ISICTMA2019)
and Leader,
Database Technologies and Application Research Group (DbTA),
Faculty of Computer Science and Information Technology,
Universiti Putra Malaysia.
25th July 2019

ORGANIZERS

Infocomm Development Centre (iDEC), Universiti Putra Malaysia

In Collaboration With

Council of Public Universities ICT Directors (MAPITA),
UPM Administrators Association (PPUPM),
Researcher Group of Database Technologies and Applications (DbTA),
Faculty of Computer Science and Information Technology UPM,
Bursar's Office and
Ministry of Education Malaysia

ORGANIZING COMMITTEE

SYMPOSIUM CHAIR

Dr. Iskandar Ishak
Learder of Database Technologies and Application Research Group (DbTA)
Faculty of Computer Science and Information Technology

PAPERS AND TECHNICAL SUPPORT

Assoc. Prof. Dr. Lilly Suriani Affendey
Assoc. Prof. Dr. Fatimah Sidi
Dr. Siti Nurulain Mohd Rum
Department of Computer Science
Faculty of Computer Science and Information Technology

Assoc. Prof. Dr. Marzanah A. Jabar
Department of Software Engineering and Information System
Faculty of Computer Science and Information Technology

PUBLICATION

Assoc. Prof. Dr. Noritah Omar
Department of English
Faculty of Modern Languages and Communication

Mr. Saiful Ramadzan Hairani
InfoComm Development Center (iDEC)

Ms. Nur Mahirah Afiqah Abdul Rahman
Faculty of Modern Languages and Communication

SMART SPATIUM: A DYNAMIC COLLABORATIVE WORKING SPACE FOR SMART CAMPUS

Salfarina Abdullah[1], Nur Farahsyazwani Nor Azmi[1] and Sazly Anuar[2]
[1]Universiti Putra Malaysia, [2]Universiti Kuala Lumpur Malaysia France Institute
salfarina@upm.edu.my

Abstract: Smart Spatium is an online dynamic collaborative working space developed for university students as a platform for sharing documents regardless of time and place. It eradicates the difficulty to establish consistent physical meetings by enabling the students to have active collaboration and discussion. Problems concerning work coordination, inability to make document editing, lack of back up platforms, and lack of mechanism that can provide evidence of contribution among team members will all be smartly tackled by this system. Committed to the betterment in active teamwork collaboration regardless of geographical distances, it allows students to commit and actively collaborate in every discussion and best of all; obtain a fair assessment among the team members. It also serves as a centralized platform for storing the folders containing assignments and projects so that the whole application looks more manageable hence can reduce any possible risks. Developed using Rapid Application Development (RAD) methodology, this application is highly suitable for newer generation in virtual learning and online community practice, thus greatly enhances the current education trend and methods.

***Keywords*:** Centralized Platform, Contribution Evidence, Online Education

INTRODUCTION

Katzenbach and Smith, (1993) developed a model consisting of five different types of team: *working groups* (in which members share information and activities but without clear team roles' definition), *pseudo teams*(they are labelled as a 'team' but, in reality, share very little responsibility and coordination of their teamwork), *potential teams* (in which members look as a collaborative team but there are still few factors lacking to become an effective one), *real teams* (where common goals accountability are shared among the team members), and *high performance teams* (in which all members really understand their roles, share common team goals and also inspire other member's personal expansion). Apparently, majority of university students often belong to the pseudo teams. Lack of motivation and coordination has become some sort of identity. So, this has inspired the idea for developing the Smart Spatium, which allows students' to actively collaborate in completing their team assignments and projects. This seems to be an effective way for the team assignments and projects to increase in quality of the task produced when everyone is able to see each other's progress toward completion. Hence, eliminates the one person in charge of validation, creating a whole and equal responsibility and workload to get fair assessments among team members.

Smart Spatium is also designed for students to have fun like they have in social media, where they can share and post comments on their timeline. Besides, they can also make post-comment and update information about their team assignments or projects. And at the same time, they will be notified for every posted comment. So, they will become more alert and focus on their assignments or projects until the end of their works.

We carefully studied the functionalities of the existing similar systems, and managed to identify some of their features and weaknesses, and took all of that into account when designing Smart Spatium. Ensuring our application fills in the existing gaps, is utmost importance for it to be recognized as an improved online work space application. Refer to Table 1. for the finding summary.

Table 1. Literature Review Finding Summary

	SOCIAL LEARNING			SOCIAL MEDIA		
Key functions	Edmodo	Google Docs	Dropbox Paper	Facebook	Google+	Skype
Edit the documents simultaneously	NO	YES	YES	NO	YES	YES
View the contribution of each team member	NO	YES	YES	NO	NO	NO
Post comment on timeline to share extra information	YES	NO	NO	YES	YES	NO
Comment on specific part of the documentation	NO	YES	YES	NO	NO	NO
Set alert notification for project submission date	YES	NO	YES	YES	YES	YES
Discuss with team members in the chat board	YES	YES	YES	YES	NO	YES
Checklist activities for assignment and project	YES	NO	YES	YES	NO	NO
Invite team member to collaborate	YES	YES	YES	YES	YES	YES
Provide a folder to have centralized platform to keep created document	NO	YES	YES	NO	NO	NO

METHOD

Smart Spatium was developed using Rapid Application Development (RAD) methodology. RAD is based on prototyping and iterative development with no specific planning involved. It employs minimal planning in favor of rapid prototyping and produces faster product delivery. It also emphasizes on building the prototype which provides an early look at the end product. The prototype is then refined in user interface design cycle until it is satisfied to be released as end product after validation and verification by clients.

Operating Environment

The development tools used in this project include Java Script (JS) for scripting and calling APIs involved in Smart Spatium and access the database that access the whole data of the system, Bootstrap for front-end development, and Firebase Database.

System Architecture

The system architecture we referred to when designing Smart Spatium is Model-View-Controller (MVC) architecture. Based on Figure 2, Model-View-controller (MVC) architectural pattern allows the application to be separated into three main logical components which are the model, the view, and the controller. The programming languages used are PHP and JavaScript that connected with PhpMyAdmin as my testing web server which also contain MySQL as my repository. MVC enables efficient code reuse and parallel development so that representations of information from ways information are presented to can be separated so it is accepted by the user.

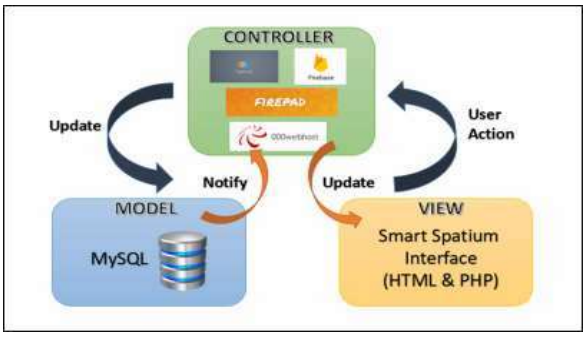


Figure 2: MVC Architectural Pattern

RAD methodology focuses more on testing even before this application started to be implemented. For Smart Spatium application, along with functional testing, we also employed system usability scale (SUS) survey for usability testing. For functional testing, we had tested every function, and all had passed successfully. As for the usability testing, the respondents of this survey were students from various universities and colleges. Normally SUS contains ten questions concerning the usability of the project. Based on the results of 20 respondents, we can conclude that Smart Spatium application had succeeded the usability testing.

FINDINGS AND DISCUSSION

Smart Spatium offers the following functionalities: 1) Student login system if they already have an account. If not, they will need to register as a first-time user of Smart Spatium. 2) Student to register into the system if they still do not have an account with Smart Spatium. 3) Student can post a comment to share information or updates about their assignment or project and an alert will be sent through notification. 4) Student can create a folder to keep their specific subject of documents. 5) Student can create new documents by inserting document title and date of creation. 6) Student can edit document simultaneously with their team members. This particular function consists of several other functionalities which are chat room for discussion, view contribution among their team members and invite their team members to collaborate. 7) The function of check listing activities for every assignments or projects.

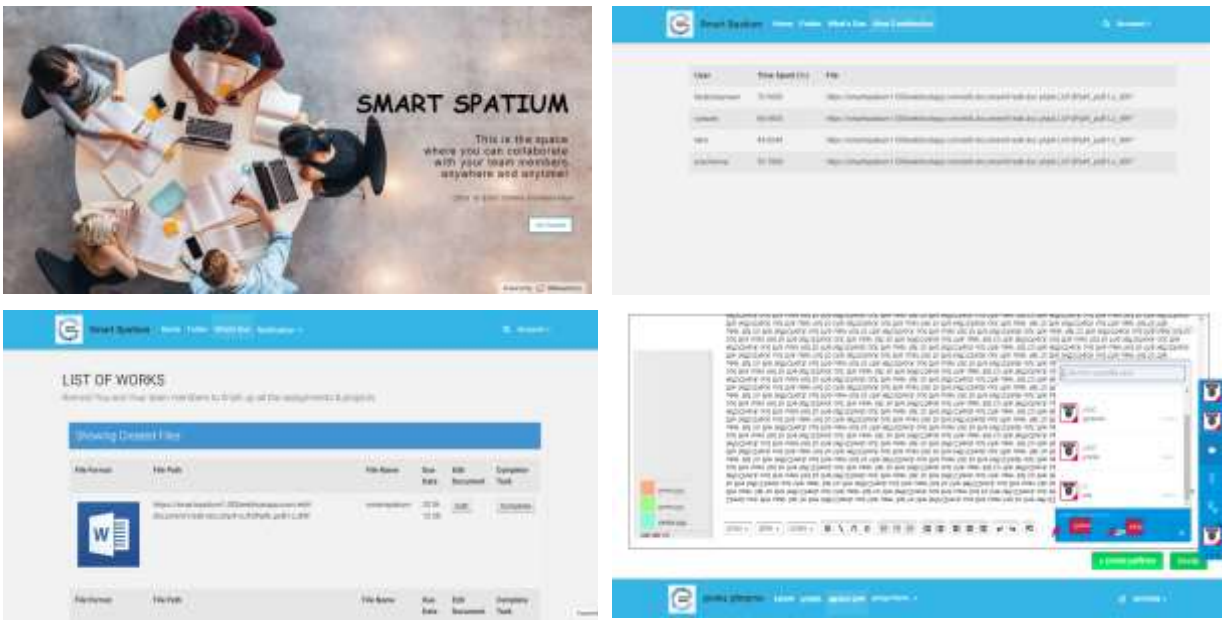


Figure 3: Some screenshots of Smart Spatium

There is a specific functiona that we would like to highlight, which is the function to view the contributions among their team members. We found this to be very important, useful and a must have feature especially in virtual learning application. As mentioned in the earlier section, one of the aims of this application is to promote more objective and transparent assessment based on the work contributions of the students. Note that it is never our intention to claim neither that this would be the replacement to the existing method for students' assessment nor this is the most effecient way that guarantees fair assessment. From our perspective, we are more keen to make it a function that assists and facilitates the process of assessing students' work contributions as a team member. We want it to be thought as a complimentary to the existing method being practiced currently in the university. Human intervention in students' assessment is still required and should never be an issue.

CONCLUSION

Based on grounded theory analysis, there are four common obstacles to collaboration: students' lack of collaborative skills, free-riding, competence status, and friendship (Le, H. et al. 2017). Smart Spatium application is built not only to solve problems faced by the students when collaboration is needed, but also to instill and cultivate the sense of community belonging that encourages work responsibility, and smart work coordination among their team members while doing assignments or projects. It has succeeded in solving three problems mentioned earlier, which are the difficulty to meet up regularly for discussion, no centralized platform for storing the folders' assignments or project, and finally the issue of providing evidence of contribution of each team member in completing a task. Overall, Smart Spatium has great potential to assist students working in groups in more efficient manner and thus promotes smart work collaboration and enhances the utilization of technology in higher learning education.

Nevertheless, we believe Smart Spatium is still on its way to perfection. There are some rooms for improvement including the security aspect on the system in terms of authorization in editing a document, the limitation that it has in providing real-time collaborative editing, which only allows for Word type of file, and also inability to provide a feature to comment on specific part of the document which we think might be easier for the students to focus on specific parts and reduces the time in finding their mistakes.

BENEFITS FOR HIGHER EDUCATION

Smart Spatium is a complementary to the existing system which involves conventional classroom, physical meetings and manual assessments. For higher learning institutions such as university and colleges, this would be an added value that not only facilitates and simplifies the students' learning process but also the lecturers' teaching and its related tasks including assessment. From time to time, the existing practice in higher education will gradually embracing more dynamic and efficient trend in teaching and learning especially with new technologies supporting collaborative teamwork and unbiased, objective assessment.

REFERENCES

Katzenbach, J., & Smith, D. 1993. The wisdom of teams: Creating the high-performance organization, Boston, MA: Harvard Business School Press.

H. Le, J. Janssen, and T. Wubbles. 2017. Collaborative learning practices: teacher and student perceived obstacles to effective student collaboration. Cambridge Journal of Education, 48(3).

IMPLEMENTATION OF ZCHAIN4U FOR CASHLESS PAYMENT AND DIGITAL CERTIFICATION VERIFICATION USING BLOCKCHAIN TECHNOLOGY

Zuriati Ahmad Zukarnain and Nurul Nadia Abdol Rahman
Department of Communication Technology and Network, UPM, Serdang 43400 Malaysia
zuriati@upm.edu.my, nurulnadiaabdolrahman@gmail.com
corresponding author: zuriati@upm.edu.my

Abstract: Blockchain technology is being embraced globally as a secure and confidential means of transaction, inspired by the Merkle tree, it stores information in the blocks in chains. The aim of the Zwallet is to create a more friendly interface for the end-user while at the same time maintaining the cryptographic security functions of a blockchain ledger. The end-user will be able to to view the current balance, purchase tokens, view transaction history and send token to other address. Transaction charges, downtimes and third-party involvement has pushed towards alternative means of financial transaction. Built on uses Vue.js, Zchain4U as a digital wallet is essentially a cryptocurrency similar to a bank account. It allows users to receive cryptocurrency, save it, and then send it to others. A user wallet was successfully created and token was purchased then saved in the account which was used to send and receive tokens by simulating a blockchain transaction. Wireless transfer and interface with interswitch are areas of future research. Whereas Due to its security, Blockchain technology is also used as a technique in determining student authentication of digital scroll. Zscroll is designed to verify the legitimacy of a document in an easy but fraud-proof method. Technological advancement has made forgery of important documents such as certificates, identity cards, and passports easy. Traditional methods are costly, time-consuming and could lead to damage or loss of original documents. Furthermore, it's more difficult to detect a well-forged document. Digital certificates that are kept safe on a blockchain thus an employer can easily authenticate the certificate using an easily available open-source software. The system issues a digitally signed academic certificate, with time-stamping and the hash value is generated and saved to a public domain where anyone can use a given public key to asertain the veracity of the said certificate or document. This guarantees that records can be verified instantly without the need to contact the issuing institution while integrity of documents is assured.

***Keywords*:** blockchain, cashless, digital-scroll, digital-signature, e-wallet

INTRODUCTION

Blockchain is these days known for money transfers which use cryptographic technologies, thus creating cryptocurrencies for online business. Businesses such as IT companies, hotels and also some restaurants accept Bitcoin for payment. It employs SHA-256 hash function, public and private keys, plus it consists of the blocks as the blocks consist of data in Blockchain. Those blocks hold the hash function of the verification method for data transactions. The benefits include less time to transfer money to the receiver and prevents double-spending. Since the sale and purchase used this kind of technology, normal users can also purchase cryptocurrency. We believe in the future, the position of cryptocurrencies depends on their controllers such as Blockchain, previously it was called Genesis Block.

Blockchain Technology used in the transaction system without cash payment system is to ensure that transactions run quickly and safely. Though Blockchain technology is also used as a technique in determining student authentication digital scroll. Our scope focuses on the security of flow of transactions involving cryptographic methodologies for data transmission from sender to receiver. The need of a third-party in conventional e-banking processes which functions to verify or regulate transactions between clients shows there is no privacy between the sender and receiver on the value of data transmission. Also, the requirement of a charge known as transaction fee which is still employed by modern internet banking procedures is unfair to customers who transact yet lose part of their money for the transactions. Another challenge is downtime. The standard e-banking systems usually have times in which the clients cannot access the system, due to maintenance or security

reasons, improvised downtime, system failure, etc., which renders the verification mechanism inoperable thereby denying clients successful transactions.

Certificates issued institutions are among the most important documents for graduates. Many government institutions and organizations require original documents for verification purposes. But the parties involved may experience reluctance in handling these important documents as they could get damaged, lost in the mail, etc. Technological advancement has made forgery of important documents such as certificates, identity cards, and passports easy. Verification of certificates using traditional methods is costly time-consuming. Also, security is the contemporary concern of people while using technology because it exposes us to fraud, data theft, and stealing (Bezhovski, 2016).

The aim of the Zwallet is to create a more friendly interface for the end-user while at the same time maintaining the cryptographic security functions of a blockchain ledger. The end-user will be able to to view the current balance, purchase tokens, view transaction history and send token to other address. The project objectives propose a possible solution by creating a new Blockchain infrastructure (Zchain), with Zwallet and Zscroll to enable secure authentication without third party or any regulations. As for Zwallet, the Blockchain does not need a financial intermediary like a bank where transactions will need transaction fees. Also, a cryptography hash functions technique will be used to verify the Proof-of-Work (POW) of the all peers in the chain of a peer-to-peer network, which also gives users ability to use the service to transact at any time without doubt.

For Zscroll, the aim is to create a solution for academic Certificate issuing and verification using Blockchain technology which only stores the proof that a digital asset has been certified (or signed) by an institution on the Blockchain. If anybody would want to verify the legitimacy of a digital asset they can simply verify the digital asset by vetting it using the proof provided. Hence, the Blockchain's role in Zscroll solution is to provide an immutable storage container for legitimate proofs.

Cryptocurrency was the first blockchain application while Bitcoin gave the individual full possession over user money in the digital realm via ownership of a cryptographic key necessary for the transfer of funds. More complex transaction types can be built into cryptocurrency that replicates the full suite of financial instruments, such as loans, bonds, and stocks, now offered by most institutions (Peck, 2017). Application of cryptocurrency or e-wallet is being adopted due to its anonymity, in other word, security. Bonneau et al. (2014) proposed Mixcoin, a protocol to facilitate anonymous payments in Bitcoin and similar cryptocurrencies. They build the accountability mechanism to expose theft, ensure that aggregational transactions will not be ambiguous. This system has been upgraded by Valenta & Rowan (2015) guaranteeing that the input/output address mapping for any user is kept hidden from the mixing server. These authors suggested a system called Blindcoin which changed the Mixcoin engineering protocol using blind and open logs. These logs allow third parties to verify the integrity of the statement when a blind signature is used. The mixing service then upgrades to decentralized, Ziegeldorf et al. (2015) based on a combination of decryption mix-nets with threshold signature called as CoinParty. The authors proved CoinParty is secure against malicious adversaries and achieves anonymity by orders of magnitude higher than related work as they quantified by analyzing transactions. Decentralized currency network has been updated to completely decentralized Bitcoin mixing protocol that allows users to utilize Bitcoin in a truly anonymous manner by Ruffing, Moreno-Sanchez, & Kate (2014). The authors introduced CoinShuffle which only a small communication overhead for its users, while completely avoiding additional anonymization fees and minimizing communication overhead. To increase privacy, ZeroCoin is introduced to hide transaction value and address balances by Androulaki & Karame (2014). Then the anonymity has been improved by (Saxena, Misra, & Dhar, 2014) they suggested the use of composite signatures to prevent linking between sending and receiving addresses. Wilson & Ateniese (2015) proposed adopted Bitcoin technology to enhance the Pretty Good Privacy (PGP) mechanism so that Blockchain key server is used to improve the user's trustworthiness.

A novel approach to detect emotions like happy, sad or angry in textual conversations using an LSTM based Deep Learning model. The input user utterance is fed into two LSTM layers using two different word embedding matrices Glove and SSWE. One layer uses a semantic word embedding, whereas

the other layer uses a sentiment word embedding. These two layers learn semantic and sentiment feature representation and encode sequential patterns in the user utterance, both representations are then concatenated and passed to a fully connected network with one hidden layer which models interactions between these features and outputs probabilities per emotion class. This model is not contexted aware hence the emotion classification is not stable enough proposed by Srikanth, Chatterjee, & Agrawal (2017). The researchers (Yadav, Sain, Sheth, Ekbal, Saha, & Bhattacharyya, 2018) proposed an effective architecture that uses a Convolutional Neural Network (CNN) as a data-driven feature extractor and a Support Vector Machine (SVM) as a classifier. Here, they showed that the use of medical sentiment features along with extracted features from CNN improves the model performance. This model lacks implicit sentiments analysis. The researchers (Yadav, Ekbal, Saha, & Bhattacharyya, 2018) presented a benchmark setup for analyzing the sentiment with respect to users' medical condition considering the information, available on social media in particular. Thereafter, they have presented a deep convolutional neural network-based classification framework to predict the possible medical sentiment category for both 'medication' and 'medical condition' classification schemas. This model lacks medical sentiment specific lexicon and to capture implicit, metaphoric & sarcastic phrases. Other researchers (Duc-Anh & Matsumoto, 2018) proposed a new multilabel corpus named Emotional Movie Transcript Corpus (EMTC). In their approaches, no machine learning techniques were used and the accuracy was not that acceptable. Suhasini & Badugu (2018) have proposed a method that detects the emotion or mood of the tweet and the twitter message under the appropriate emotional category. The first approach is the Rule-Based Approach (RBA), their minor contributions in this approach are pre-processing, tagging, feature selection, and Knowledgebase creation. Feature selection is based on tags. Their second approach is Machine Learning Approach (MLA), in this, the classifier is based on a supervised machine learning algorithm called Naïve Bayes which requires labeled data. Naïve Bayes is used to detecting and classifying the emotion of a tweet They have compared the accuracy of both the approaches, observed that with the rule-based approach we are able to classify the tweets with the rule-based the accuracy around 85% and with the machine learning approach the accuracy is around 88%. Machine learning approach performance is better than rule-based approach, the performance has been improved as they have removed the error data while training the model. The approaches are involved with the concepts of Natural Language Processing, Artificial Intelligence, and Machine Learning for the development of the system.

METHODS

For cashless process, a node initiates a connection to a blockchain P2P network for authentication providing inverse hash function calculation. A UI e-wallet with a decentralized and distributed database which was created will use a token algorithm to validate the node's e-wallet details. If correct, the node proceeds to perform the transaction else, the node is returned to the starting point. Also, ECC (Elliptic Curve Cryptography) was used to create a secure identity of nodes and also transaction between entities. The first node (user) sends a block to both nodes. Both nodes will verify the first key of the public node and sign the personal key and the first signature of the digital node. The second node sends a signed and verified block to all three nodes that also confirm and sign the primary and second digital signals and then pass through and so forth. A Zchain4U as a digital wallet is essentially a cryptocurrency similar to a bank account. It allows users to receive cryptocurrency, save it, and then send it to others. To implement our Zwallet, we must create a new final shape that users can ask for block specification if the hash is known. This project uses Vue.js to implement the part and account of the explorer's block which is a website used to visualize the situation of the blockchain (easily check the rejection or confirmation if the transaction is entered into the block).

Digital certificates that are kept safe on a blockchain have significant advantages over regular digital certificates. Anyone with access to blockchain can easily authenticate the certificate using an easily available open-source software. Hence, intermediary parties are no longer needed. Thus, the certificate can be authenticated even after the organization is dissolved or it no longer has access to the issued record. Issued records and received certificates on a blockchain can be removed only if all copies on all nodes that host the software are destroyed.

RESULTS AND DISCUSSION

A Zwallet was successfully created, a user done installed into smart phone to access it. User then embedded token into the Zwallet and then saved in the account. This token used to transact by successfully sending and receiving part of the token which led to change in balance of wallet account. Ways to adding the token could easy by scan a QR code from the reload system, other was set the QR code from Zwallet account, sent to other user as token request. The QR code can be generated and used to ease transaction by make a payment at counter, and demanding or indicating a willingness to sending tokens for transaction to take place. Here is how the interface for user to request payment look like.

Figure 1. Payment request set QR code from receiver to sender of token

In brief, the system issues a digitally signed academic certificate, with time-stamping – adds another layer of security, this model uses a peer-to-peer network-based system to issue a digital stamp affixed to the document and indicates the precise day, month and year when the digital certificate was issued. The hash is then hosted on the blockchain and the private university key is used to create a record that indicates the certificate was issued to the said student on the said date. Public keys of educational institutes are easily available on the internet or in public directories or they can be provided by the graduates. The employer uses verification software to scan the digitally signed document and public key. The verification software will use the digital signature and the public key to generate a hash and compare with the hash of the original certified copy. The verification software will also check if the digital signature is linked to the public key of the educational institution. Both steps are done without requiring students to disclose their private keys. If both conditions are true, the certificate is passed valid. If any or both conditions are false, the verification software indicates the certificate is fake.

Figure 2. Valid result

Figure 3. Invalid result

CONCLUSIONS

Blockchain technology has helped in increasing the efficiency of administrative and inventory-related tasks as well as networking and information storage. Hence, Zwallet gives the user flexibility of creating and maintaining an e-wallet with blockchain technology which can be used securely for transactionsThe following are the recommendations for further studies that will enable drive towards a more effective electronic wallet system: There should be an option of wireless transfer from third-party bank accounts to the electronic wallet account. There should be interfacing with the interswitch payment portal for additional security and reliability. There should be an avenue for communication between the administrator and users (customers and merchants). This will enable users to adequately lodge complaints or request assistance at any point in time.

Significance guarantees that cryptographically-sealed data that cannot be falsified are issued. All data are stored securely and can be referenced. Spending unplanned time for the transmission of official records to individuals is not needed. Records can be verified instantly without depending on issuing authorities, the records also are not lost because transactions are recorded on the blockchain, and can be checked by third parties anytime. Official records can be owned and shared easily. Calling separately for the deletion of certificates that are expired or damaged is not needed. As recipients and consumer's benefits, cryptographically sealed data are owned and shared. The latest record can be verified because of the easily integrated blockchain lookup services. Blockchain-based model for graduation certificate verification will reduce the incidence of certificate forgeries and ensure that the security, validity, and confidentiality of graduation certificates would be improved. The advantage of Zscroll is that all the information that is required to validate and authenticate the certificate is hosted on the blockchain itself. In order to validate the certificate, the prospective employer needs not contact the university at all.

REFERENCES

Androulaki, E., & Karame, G. O. (2014). Hiding transaction amounts and balances in Bitcoin. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-08593-7_11

Bezhovski, Z. (2016). The Future of the Mobile Payment as Electronic Payment System. *European Journal of Business and Management*.

Bonneau, J., Narayanan, A., Miller, A., Clark, J., Kroll, J. A., & Felten, E. W. (2014). Mixcoin: Anonymity for Bitcoin with Accountable Mixes BT - Financial Cryptography and Data Security. *International Conference on Financial Cryptography and Data Security*.

Peck, M. (2017). Reinforcing the links of the blockchain. *Ieee*. https://doi.org/10.1016/j.jacc.2017.06.060

Ruffing, T., Moreno-Sanchez, P., & Kate, A. (2014). CoinShuffle: Practical decentralized coin mixing for bitcoin. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-11212-1_20

Saxena, A., Misra, J., & Dhar, A. (2014). Increasing anonymity in bitcoin. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-662-44774-1_9

Suhasini, M., & Badugu, S. (2018). Two Step Approach for Emotion Detection on Twitter Data. *International Journal of Computer Applications*. https://doi.org/10.5120/ijca2018917350

Valenta, L., & Rowan, B. (2015). Blindcoin: Blinded, accountable mixes for bitcoin. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-662-48051-9_9

Wilson, D., & Ateniese, G. (2015). From pretty good to great: Enhancing PGP using bitcoin and the blockchain. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-25645-0_25

Yadav, S., Ekbal, A., Saha, S., & Bhattacharyya, P. (2018). Medical Sentiment Analysis using Social Media: Towards building a Patient Assisted System. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*.

Ziegeldorf, J. H., Grossmann, F., Henze, M., Inden, N., & Wehrle, K. (2015). CoinParty: Secure multi-party mixing of bitcoins. In *CODASPY 2015 - Proceedings of the 5th ACM Conference on Data and Application Security and Privacy*. https://doi.org/10.1145/2699026.2699100

Srikanth, U. G. R., Chatterjee, A., & Agrawal, P. (2017). A Sentiment-and-Semantics-Based Approach for Emotion Detection in Textual Conversations. *Indian Institute of Technology Kharagpur.*

Yadav, S., Sain, J., Sheth, A., Ekbal, A., Saha, S., & Bhattacharyya, P. (2018). Leveraging Medical Sentiment to Understand Patients Health on Social Media. *Indian Institute of Technology Patna.*

Duc-Anh, P. & Matsumoto, Y. (2018). Multilabel Corpus in Movie Domain for Emotion Analysis in Conversational Text. *NARA Institute of Science and Technology.*

EVALUATION OF MISSING VALUES IMPUTATION METHODS TOWARDS THE EFFECTIVENESS OF ASSET VALUATION PREDICTION MODEL

Mohd Izham Mohd Jaya[1], Fatimah Sidi[2], Lilly Suriani Affendey[2], Iskandar Ishak[2] and Marzanah A. Jabar[3]

[1]Bahagian Dasar Transformasi, Aras 5, MAMPU, Blok B2, Kompleks JPM, Putrajaya
[2]Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia.
[3]Department of Information System and Software Engineering, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia
izham@mampu.gov.my, fatimah@upm.edu.my, lilly@upm.edu.my, iskandar_i@upm.edu.my, marzanah@upm.edu.my
corresponding authors: izham@mampu.gov.my, fatimah@upm.edu.my

Abstract: Missing values is a common problem found in dataset from any field of research. A data value in a dataset can be missing due to numerous reasons such as non-response items in the interview and survey, equipment malfunction, human error and faulty data transmission. The occurrence of missing values in a dataset need to be managed using appropriate imputation methods to estimate the approximate values to replace the missing values. The problem of missing values also led to a data quality problem which then resulted inaccurate decisions. In this work, we compared and evaluated various imputation methods including deletion of records with missing value (DEL), mean values imputation (MEAN), k-Nearest Neighbor (KNN), Predictive Mean Matching (PMM), MissForest and Ontology-based Framework for Financial Decision Making (OFFDM) towards the effectiveness of asset valuation prediction model. In portfolio management, asset valuation prediction model is used to aid the decision making process. Additionally, we adopted MissForest method in the OFFDM which aim to improve the OFFDM. We conducted several experiments using different dataset derived from different imputation methods to measure the accuracy, Root Mean Squared Error (RMSE) and F-measure of the prediction model which being built in Artificial Neural Network (ANN). We found that dataset derived from DEL resulted the lowest accuracy and the highest RMSE. Whereas, the adoption of MissForest method in OFFDM resulted the highest accuracy and second lowest RMSE value. The selection of imputation methods is depended on the severity of the task in hands as each method is different in its complexity and efficiency. Imputation method such as MissForest is efficient but required more computational resources. On the other hand, simpler methods such as DEL is still popular due to its simplicity but less efficient.

***Keywords*:** missing value, imputation, data quality.

INTRODUCTION

Achieving high quality data enable the organizations to lower their operational cost, improve customer satisfaction and minimize risk of false decision in decision making (Strong et al., 1997). Furthermore, in decision making, high quality data is not only important to ensure correct decision, but it also provide the foundation to get a positive outcome (Ghasemaghaei & Hassanein, 2015). A long progress in data quality research defined high quality data as a state of data that is fit for use and able to meet the purposed given by data user (Lee & Strong, 2003). In order to correctly measure and improve data quality, several data quality dimensions has been proposed in previous research such as completeness, accuracy, consistency and timeliness (Wand & Wang, 1996; Ballou & Pazer, 1985). In this paper, we focus on the dimension of data completeness as it is commonly discussed in various domain such as finance (Du & Zhou, 2012), health (Stockdale & Royal, 2016; Curé, 2012) and education (Cox et al., 2014). Data completeness dimensions is define as the occurrence of missing values in which data records has the undesirable null values (Liu et al., 2016; Wechsler & Even, 2012). Due to this definition, we can conclude that data is complete when all necessary value pertaining to the data exist (Jayawardene et al., 2013).

The emergence of researches in handling missing value problem proposed various imputation methods. Each of these methods aim to introduce a new value which are closest to the missing value with a minimal disturbance to the dataset pattern and resulted analysis. In this work, our intention is to compare existing missing value imputation methods such as deletion of records with missing value (DEL), mean values imputation (MEAN), k-Nearest Neighbor (KNN), Predictive Mean Matching (PMM), MissForest and Ontology-based Framework for Financial Decision-making (OFFDM). Additionally, we also adopted MissForest method in OFFDM (OFFDM MF) instead of KNN as proposed in Du & Zhou (2012). We provide the evaluation of each imputation methods based on the effectiveness of the prediction model used in our experiment. In order to do so, we impute the dataset using each imputation methods and run a separate experiment using the prediction model. In this work, we evaluate the effectiveness of the prediction model using the percentage of the prediction accuracy, Root Mean Squared Error (RMSE) and F-measure.

METHODS

Portfolio management is an example of financial decision-making process which aim to decide on the selection of securities to hold and funds allocation (Du & Zhou, 2012). Fourteen financial ratios as described in Table 1 is computed based on the financial variables in financial statement to support asset valuation in portfolio management. The computed financial ratios have various degree of missing values as it is derived from the financial dataset that contained missing values.

Table 1. 14 financial ratios used to support asset valuation

Financial ratio	
Earnings before interests and taxes / total assets	Current assets / current liabilities
Net income / net worth	Inventories / current liabilities
Gross profit / total assets	Interest expenses / sales
Net income / gross profit	Selling, general and administrative expenses / sales
Current liabilities / total assets	Account receivable / sales
Total liabilities / total assets	Account payable / inventories
Long term debt / total equity	Account payable / sales

Experiment is conducted to investigate and compare the effect of seven missing values imputation methods namely: DEL, MEAN, KNN, PMM, MissForest, OFFDM and OFFDM MF towards the effectiveness of asset valuation prediction model in portfolio management as depicted in Figure 1. Missing values pattern in the dataset is analyzed and financial ratio is ranked according to the percentage of missing values occurrence. Missing values of the highest ranked ratio is then imputed using various missing values imputation methods which resulted seven different datasets. Consequently, the imputed dataset is then used as the training dataset for the asset valuation prediction model. The generality of the prediction model is evaluated using ten-fold cross validation. In order to evaluate the effectiveness of each imputation methods, three metrics namely: prediction accuracy, RMSE and F-measure are adopted. The asset valuation prediction model in portfolio management is build based on Artificial Neural Network (ANN) in WEKA environment. Consistent parameter and configuration is used to build the prediction model in order to ensure valid comparisons between missing values imputation methods. The learning rate of the prediction model is set to 0.1 and the number of nodes in the hidden layer is set to half of the total inputs and outputs. Financial ratios as mentioned earlier are used as the input. Class labels in the prediction model is determined based on the value of Relative Change in Stock Earning (RCSE) calculated as the following (Du & Zhou, 2012):

$$RCSEj\ (t+1) = [Pi(t+1) + Di(t+1))-(Pi(t) + Di(t))]/Pi(t) + Di(t))$$

Where;

Pi(t) = Price of *i*th stock during time period t

Di(t) = Dividend of *i*th stock during time period t.

The calculated value is classified into three class labels which are DOWN (RCSE less or equals to .011), NOCHG (RCSE less or equals to .104) and UP (RCSE more than .105).

FINDINGS AND DISCUSSION

All imputation methods used in this study performed better than DEL in terms of prediction accuracy and RMSE. This suggested that, deleting instances with missing values in the dataset is not the best option to handle missing values. Instead, imputation methods should be used in order to obtain better prediction result.

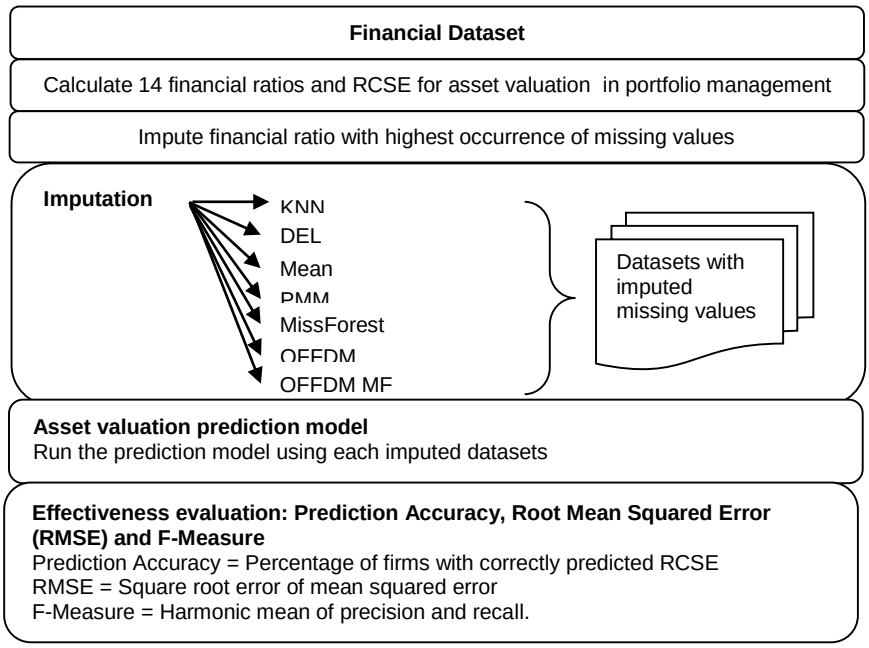


Figure 1. Experiment setup

As depicted in Figure 2, prediction accuracy attained from dataset imputed with OFFDM MF showed the largest prediction improvement from DEL, in which instances with missing value were deleted.

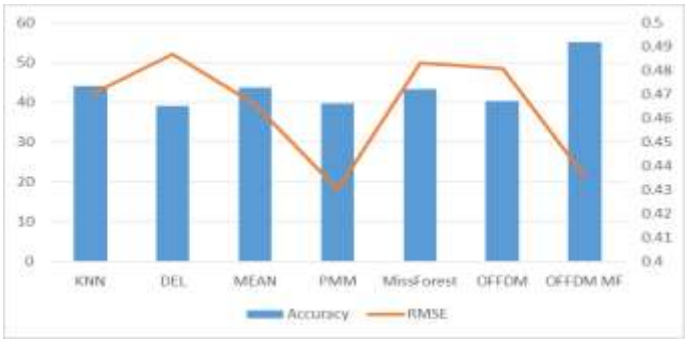


Figure 2. Accuracy percentage and Root Mean Squared Error (RMSE) of the prediction model on the imputed dataset.

The prediction accuracy is 16% improved when OFFDM MF is used as compared to DEL. Whereas, the prediction accuracy for OFFDM is 14.9% less than OFFDM MF. Large difference in prediction accuracy between OFFDM MF and OFFDM is due to the adoption of MissForest method in OFFDM MF instead of KNN method. The prediction model loaded with data derived from OFFDM MF also resulted higher F-measure value in UP, NOCHG and DOWN as shown in Figure 3. When DEL is used, disparity between the F-measure of UP, NOCHG and DOWN is huge compared to other imputation methods. Overall, the resulted values of F-measure for UP is higher compared to other class labels regardless of the imputation methods used. Consistent performance of F-measure values is observed in DOWN and NOCHG for all imputation methods used. High prediction accuracy is

observed when MEAN method is used to impute the data. However, as all the missing values carried the same value, the variance in the dataset cannot be retained.

The implementation of OFFDM and OFFDM MF required the ontology construction. Thus, the performance of ontology itself is important. It is quite a challenging task to find the online financial data sources to be included into OFFDM and OFFDM MF. Moreover, most of the online financial data sources such as Yahoo!Finance and Google Finance has a limited archived period which is less than 5 years. This limit the ability of OFFDM and OFFDM MF to handle missing values in the dataset of more than five years old. Furthermore, the lack of APIs from the online financial data sources became another hurdle to implement the framework. The absence of APIs caused difficulties to automate data fetching from the online data sources. On the other hand, adoption of KNN, MEAN and MissForest do not require specific skills and knowledge about the dataset. Though, the adoption of KNN method required user to determine the number of neighbour, k, used in the imputation. Analysis of the imputed data is needed to tune the number of k as the value of k has an essential influence on the imputation performance.

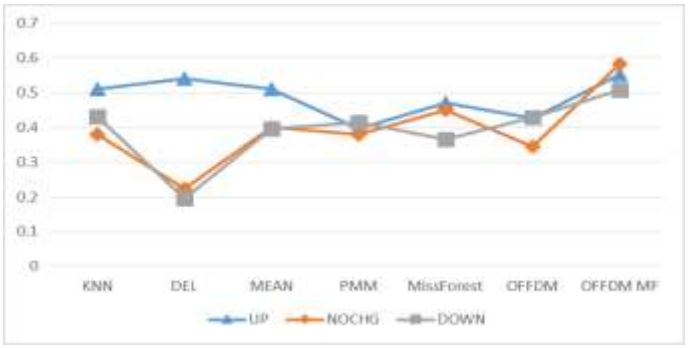


Figure 3. F-measure value for each class label of the prediction model on the imputed dataset.

CONCLUSIONS

In summary, data imputation process regardless of the imputation methods being used is important to increase the effectiveness of prediction model. We conclude that OFFDM MF is the best method to impute missing values in financial data as it led to the highest prediction accuracy. The selection of imputation methods is also depended on the severity of the task in hands as each method is different in its complexity and efficiency. Imputation method such as MissForest is efficient but required more computational resources. On the other hand, simpler methods such as DEL is simple but less efficient.

BENEFIT FOR HIGHER EDUCATION

This study suggests the effect of imputation methods towards prediction model performance which could give guidance for student and implementer in selecting imputation method.

ACKNOWLEDGEMENT

Authors would like to thank the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia for its support. This research is funded by Ministry of Higher Education of Malaysia under the Fundamental Research Grant Scheme (FRGS) - Mygrants Reference Code is FRGS/2/2014/ICT07/UPM/02/1. This support is gratefully acknowledge

REFERENCES

Ballou, D. P., & Pazer, H. L. (1985). Modeling Data and Process Quality in Multi-Input, Multi-Output Information Systems. *Management Science*, *31*, 150–163.

Cox, B. E., McIntosh, K., Reason, R. D., & Terenzini, P. T. (2014). Working with Missing Data in Higher Education Research: A Primer and Real-World Example. *The Review of Higher Education*, *37*(3), 377–402.

Curé, O. (2012). Improving the Data Quality of Drug Databases using Conditional Dependencies and Ontologies. *Journal of Data and Information Quality*, *4*(1), 1–21.

Du, J., & Zhou, L. (2012). Improving financial data quality using ontologies. *Decision Support Systems*, *54*(1), 76–86.

Ghasemaghaei, M., & Hassanein, K. (2015). Online information quality and consumer satisfaction: The moderating roles of contextual factors – A meta-analysis. *Information & Management*, *52*(8), 965–981.

Jayawardene, V., Sadiq, S., & Indulska, M. (2013). An Analysis of Data Quality Dimensions. *ITEE Technical Report No. 2013-01*, *01*, 1–32.

Lee, Y. W., & Strong, D. M. (2003). Knowing-Why About Data Processes and Data Quality. *Journal of Management Information Systems*, *20*(3), 13–39.

Liu, Y., Li, J., & Zou, Z. (2016). Determining the Real Data Completeness of a Relational Dataset. *Journal of Computer Science and Technology*, *31*(4), 720–740.

Stockdale, M., & Royal, K. (2016). Missing data as a validity threat for medical and healthcare education research: problems and solutions. *International Journal of Healthcare*, *2*(2), 67–72.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data Quality in Context. *Communications of the ACM*, *40*(5), 103–110.

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, *39*(11), 86–95.

Wechsler, A., & Even, A. (2012). Assessing Accuracy Degradation Over Time With a Markov-Chain Model. In *Proceedings of the 17th International Conference on Information Quality (ICIQ-12)* (pp. 99–110).

VERSIONING APPROACH FOR DATA WAREHOUSE SCHEMA

Sharmila Mat Yusof[1,2], Fatimah Sidi[2], Lilly Suriani Affendy[2], Hamidah Ibrahim[2], Iskandar Ishak[2] and Marzanah A. Jabar[2]
[1]School of Computing, College of Arts and Sciences, Universiti Utara Malaysia, Sintok, Kedah, Malaysia.
[2]Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang, Selangor, Malaysia.
ysharmila@uum.edu.my, fatimah@upm.edu.my, lilly@upm.edu.my, hamidah.ibrahim@upm.edu.my, iskandar_i@upm.edu.my, marzanah@upm.edu.my
corresponding authors: ysharmila@uum.edu.my, fatimah@upm.edu.my

Abstract: A Data warehouse (DW) is a repository designed for querying and analyzing data. The main aims of a DW is to provide the most accurate and historically correct information to users to support the business analysis and decision-making. The input to DW is dependent on its business requirements and data sources. With the nature of dynamic business requirements and autonomous data sources, the requirement of the DW is inevitably changes over times. As the DW is very critical for organization's business analysis and strategic decision-making, the dissemination of the changes is crucial to the DW system in order to supply a precise and reliable information. Thus, the objective of this paper is to present an approach to cater for DW change management problem. We apply the approach on TPC-H benchmark data to illustrate its practicality. The application shows that the proposed approach could assist designer to manage the DW schema changes.

***Keywords*:** Data Warehouse, Schema change, Versioning.

INTRODUCTION

A Data warehouse (DW) is a database designed for querying and analyzing data. The main aims of a DW is to provide the most accurate and historically correct information to users to support the business analysis and decision-making (Inmon, 2002). With the nature of dynamic business requirements and autonomous data sources, the requirement of the DW is inevitably changes over times. As the DW is very critical for organization, the dissemination of the changes is crucial to the DW system. A common solution to the problem is to handle the schema changes at the physical level (Roddick, 1995; Thakur & Gosain, 2011; Ahmed et al., 2014). The disadvantage of this approach is that it incurs high maintenance cost and complex OLAP server administration (Subotic et al., 2014). Thus, the objective of this paper is to propose an approach to handle DW schema when the data sources or requirements change. The approach follows a hybrid methodology using the Multidimensional (MD) model with an ontology-based method as its conceptual design. The approach could assist DW designer to handle the changes at the early phase of design. Alternatively, the impact analysis can be done prior propagating the changes to physical level (Thenmozhi & Vivekanandan, 2014). With the proposed approach, the expectation is that the quality of the logical schema could be improved.

From literature, the current proposed approaches are the evolution and versioning methods (Saroha & Gosain, 2014; Solodovnikova et al., 2015). The evolution method resulted in the history loss of data due to the deletion of the previous structures (Body et al., 2003; Wrembel & Morzy, 2005; Saroha & Gosain, 2014). The drawback of the evolution method give arise to the versioning method. Versioning approach creates a new schema version assigned with a timestamp or user-defined identifier (Jovanovic et al., 2012). However, the current versioning approach are using the metadata to store all the changes occurred. This approach leads to the need of huge storage space and is not effective.

Nowadays, the concept of ontology has emerged as a promising key component in information systems (Elamin & Feki, 2014). Ontology is a structured and formal depiction of knowledge domain. Using of ontology for designing a DW helps organizations to solve semantic heterogeneity problems between data sources and user requirements. The main reason of using the ontology-based

approach is the concepts can be represented using Web Ontology Language (OWL) format. OWL supports semantic reasoning that make the process easier to be automated (Pardillo & Mazon, 2011).

An approach for handling changes of DW is proposed by using versioning method with traceability mechanism to maintain history loss. We believe that the *traceability* mechanism should be included at conceptual level in order to ease the schema versioning and to produce a quality logical schema. We test our framework on Transaction Processing Council H Benchmark (TPC-H) (Council, 2008) and Star Schema Benchmark (SSB) (O`Neil et al., 2009) data to illustrate its practicality. TPC-H is a decision support benchmark that represents our data source. The Star Schema Benchmark is a variation of the TPC-H benchmark, which models the DW for the TPC-H schema.

METHODS

In the proposed approach, requirements, data sources and DW schema are presented by ontologies. In case of the various format of data sources involved (i.e. relational, xml and text data), the relevant tools are used (i.e. RDBOnto, JXML2OWL and OntoLT) to create the respective ontologies (Thenmozhi & Vivekanandan, 2012). The DW designer is the main person to conduct changes of the DW schema. If there is any changes to the requirements, designer will update the Requirement Ontology (RO). For requirement change, the changed entities are conciliated with the Data Source Ontology (DSO) to ensure its existence. During the conciliation process, the entities of the requirement are checked and tagged. Then, the tagged entities in data sources are checked with the DW Ontology (DWO) to obtain the related MD elements that would be affected by the changes. Finally, depending on the type of change (add, edit, delete), the changes are propagated to the DWO. For any changes in data sources, the designer will directly update the DSO. The changed entities and type of change for data sources are identified and the related DW elements are checked. Then, depending on the type of change, the related operations are carried out and changes are propagated to DWO. The affected elements by the change operations are fact, dimension, attributes and relationships.

The main element of the model is the Change Manager and tagging are used to ensure the adaptation process are properly handled. The proposed approach is presented in Figure 1. The RCT and DCT will store the changed entity, the operation type and changed date. The CT will store the changed concept that exist id DSO, the mapped concept in DWO, the operation type and changed date. The DWT capture the version number, start valid date and end valid date. The Change Manager will use the tagging information from the Requirement Change Tag (RCT), Data source Change Tag (DCT) and Conciliation Tagging (CT) to detect (Change detection), conciliate and verify the changes among the Requirement Ontology (RO), Data Source Ontology (DSO) and DW Ontology (DWO). Then, the appropriate operations are conducted and the DW ontology are adapted (Change adaptation). The updated DWO will contain the version information through the use of DW tagging (DWT). Consequently, the logical schema of the DW could be generated according to the respective version from the same conceptual schema. The distinctive feature of the proposed approach is that the DW schema can be created ad-hoc when it is needed. This feature leads to the unnecessary of maintaining huge storage of the versions.

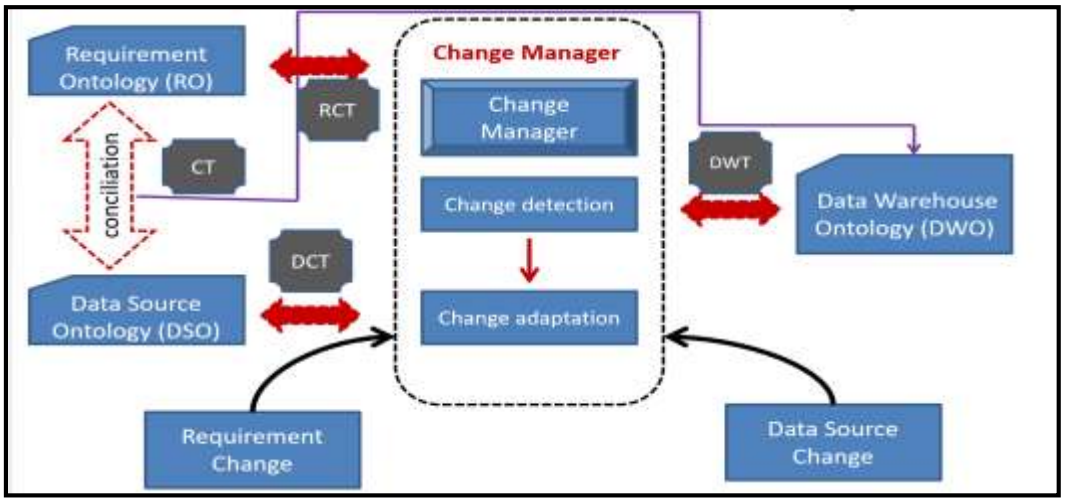


Figure 1. Overview of the proposed approach

FINDINGS AND DISCUSSION

The following sections describe our approach in detail.

1. Requirement Ontology (RO)

   For the proposed framework, the requirements are assumed to have been done earlier based on the i* framework. Thus, the information requirements should provide the information of Strategic goals, Decision goals, Information goals and Information requirements. OWL ontology is used to describe the semantics of different entities involved in this approach. The reason for using OWL as the utility, instead of XML, UML or others, is that OWL supports the semantics reasoning. Thus, we define RO as:

   RO = {SG, IG, DG, IR, BP, MS, CS}:
   ➔ SG ≈ OWL classes for strategic goals
   ➔ IG ≈ OWL classes for information goals
   ➔ DG ≈ OWL classes for decision goals
   ➔ IR ≈ OWL classes for information requirements
   ➔ BP ≈ object properties for business process
   ➔ MS ≈ data type properties for measures
   ➔ CS ≈ OWL classes for contexts

2. Data Source Ontology (DSO)

   The ontology for data sources are created using the general semantic mapping as the following:

   - Entity/table is equivalent to an ontology class.
   - If table/entity has a relationship of 1: M between them, the parent-child class relationship is created for an ontology.
   - If table/entity is connected to two other tables (either 1:M or M:1), the class superclass/subclass is created.
   - The attributes/column elements are equivalent to the class properties.
   - The primary key is equivalent to data type property of the class.
   - The foreign key is equivalent to object property of the ontology.
   - The attributes is equivalent to respective class properties

Thus, we define DSO as:

DSO = {CL, DTP, OTP}:
➔ CL ≈ OWL classes for database tables
➔ DTP ≈ data type properties for table attributes
➔ OTP ≈ object type properties for relationship between tables

3. Change Manager

When there are changes for data sources or requirements, the designer will update the respective ontology using ontology editor (i.e. protege). The changes of the ontology will be captured by protege and saved in Change and Annotation Ontology (ChAO). In case of changes in requirements, the DSO need to be checked to ensure the affected data or concept exist. If the data does not exist, the requirements need to be refined or discarded. In case of the addition change, the MD concept in DWO need to be checked for its existence. If the data exist then the MD element of the data would be checked againts the list of the MD elements in the DWO. To determine whether the class found is either fact, dimension or level, the rules are folowed as specified in Table 1.

Table 1. Multidimensional element rules

Class	**Features**
Fact	• Range class exist in DSO and belongs to dimension in DWO • Its data properties are checked whether it contains certain threshole of numerical values to qualify as fact • Class has M:1relationship with range class
Dimension	• Domain class belongs to fact in DWO • Domain class has 1:M relationship with fact
Level	• Domain class belongs to dimension or level in DWO

Source: Thenmozhi & Vivekanandan, 2014

4. Data Warehouse Ontology (DWO)

The Star Schema Benchmark (SSB) is a variation of the TPC-H benchmark that models the DW for the TPC-H schema. Thus, we define DWO as:

DWO = {FC, FP, MS, DM, DMP, RP}:
➔ FC ≈ OWL classes for fact
➔ FP ≈ OWL classes for fact properties
➔ MS ≈ data type properties for fact measures
➔ DM ≈ OWL classes for dimensions
➔ DMP ≈ data type properties for dimension properties
➔ RP ≈ object properties for facts and dimensions relationship

For the illustration on TPC-H benchmark data, one example is when there is a change in the requirement that needs to add the new decision goal of increasing the revenue by promotion activities. The decision goal is added to the requirement ontology. The new contexts for the goal are Customer and promotion while the measures are ExtendedPrice and Discount. The changes will be logged in ChAO. The Change Manager will detect the changes and tag the affected concept. Then, it will conciliate with the DSO and if found, the match concepts will be tagged. The match concept then will be conciliated with DWO and tagged. Finally, using the tagged concepts, the Change Manager will adapt the DWO accordingly. Figure 2 shows the DWO after the adaptation take place.

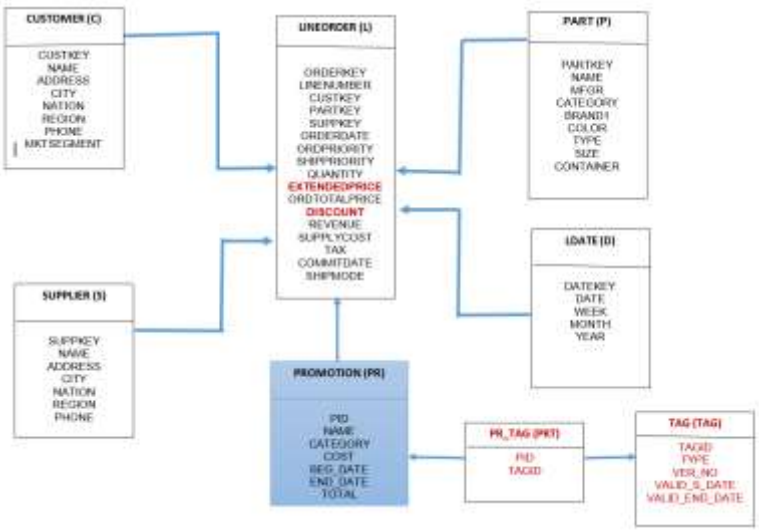


Figure 2. Updated Data Warehouse Ontology

CONCLUSION

The DW is considered as the core component of the modern decision support systems. As the data sources and business requirements from which the DW is derived frequently change, it may have its impact on the DW schema. The existing works on DW evolution such as schema evolution and schema versioning mostly focus on changing the schema structure at the physical level. The proposed approach handles evolution of the DW schema at the ontological level. The ontological representation of the data source, requirements and DW schema helps us to provide automation for the evolution process. The impact of the evolution to DW schema could be examined and the decision to carry out the changes to DW schema at the physical level is up to designer. The proposed approach has been applied to the TCP-H and SSB data and it shows the practicality of the proposed approach.

BENEFIT FOR HIGHER EDUCATION

The potential benefits of using a DW for Higher Education institution is the timely access to data for analysis and important decision-making, which is the main goal of DW system. The Institutions is able to identify certain worrying trends in their student enrollment and analyse why they are occurring. They can also determine their strengths and weaknesses among themselves and their competitors. All the information on trends and competitors would enable them to identify exact targets for marketing activities. As the DW system is continually prone to changes, the approach to solve the above mentioned problem would certainly further benefit the Higher Education institution for sustainability.

ACKNOWLEDGEMENT

Authors would like to thank the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia for its support. This research is funded by Ministry of Higher Education of Malaysia under the Fundamental Research Grant Scheme (FRGS) - Mygrants Reference Code is FRGS/2/2014/ICT07/UPM/02/1. This support is gratefully acknowledge

REFERENCES

Ahmed W., Zimányi E., & Wrembel R. (2014) A Logical Model for Multiversion Data Warehouses. In: Bellatreche L., Mohania M.K. (eds), *Data Warehousing and Knowledge Discovery. DaWaK 2014*. Lecture Notes in Computer Science, pp. 23-24, vol 8646. Springer, Cham.

Body, M., Miquel, M., Bedard, Y., & Tchounikine, A (2003). Handling Evolutions in Multidimensional Structures. In Proceedings 19th International Conference on Data Engineering (ICDE)(pp. 581). Bangalore, India: IEEE.

Council, T. P. P. (2008), TPC-H benchmark specification. Retrieved March 9, 2013, from http://www.tpc.org/tpch/

Elamin, E. & Feki, J. (2014). Toward an Ontology Based Approach for Data Warehousing. The International Arab Conference on Information Technology (ACIT2014), Oman, December 2014.

Inmon, W.H. (2002). *Building the Data Warehouse*. Retrieved from https://leseprobe.buch.de/images-adb/e2/e7/e2e7f926-314c-4e83-86a1-e481fafaf580.pdf

Jovanovic, P., Romero, O., Simitsis, A., and Abello, A. (2012). ORE: an iterative approach to the design and evolution of multi-dimensional schemas. In: Song, I.Y., Golfarelli, M., (eds.) DOLAP, pp. 1–8. ACM.

O'Neil, P., O'Neil, B, & Chen, X. (2009). The Star Schema Benchmark, Retrieved August 13, 2015, from https://www.cs.umb.edu/~poneil/StarSchemaB.PDF

Pardillo, J. & Mazon, J. N. (2011). Using Ontologies for the design of data warehouses. International Journal of Database Management Systems (IJDMS), 3(2), 73-87.

Roddick, J. F. (1995). A survey of schema versioning issues for database systems. *Information and Software Technology*, 37(7), pp. 383-393, doi: 10.1016/0950-5849(95)91494-K

Saroha, K., & Gosain, A. (2014, September). *Multi-version Data Warehouse: A Survey*. Paper presented at the 5th International Conference – Confluence The Next Generation Information Technology Summit. doi: 10.1109/CONFLUENCE.2014.6949236

Solodovnikova, D., Laila, N., & Natalija, K. (2015, September 8-11). *Handling Evolving Data Warehouse Requirements.* Paper presented at ADBIS 2015: 19th East-European Conference on Advances in Databases and Information Systems, France. doi:10.1007/978-3-319-23201-0_35

Subotic, D., Poscic, P. & Jovanovic, V. (2014, September). *DataWarehouse Schema Evolution: State of the Art.* Paper presented at the Central Europian Conference on Information and Intelligent Systems, Varazdin, Croatia. Varazdin*, Croatia:* University of Zagreb, Faculty of Organization and Informatics Varaždin.

Thakur G. & Gosain, A. (2011). DWEVOLVE: A requirement based framework for Data Warehouse evolution. *ACM SIGSOFT Software Engineering Notes*, 36(6), pp. 1-8, doi:10.1145/2047414.2047433.

Thenmozhi, M., & Vivekanandhan, K. (2014). An Ontological Approach to Handle Multidimensional Schema Evolution for Data Warehouse. International Journal of Database Management Systems (IJDMS), 6(3), 33-52.

Thenmozhi, M., & Vivekanandhan, K. (2012). An Ontology based Hybrid Approach to Derive Multidimenional Schema for Data Warehouse. International Journal of Computer Applications, 54(8), pp. 36-42.

Wrembel, R., & Morzy, T. (2005). Multiversion Data Warehouses: Challenges and Solutions. IEEE 3rd International Conference on Computational Cybernetics – ICCC 2005, Mauritus, 13-16 April, 2005. Mauritius:Mauritius: IEEE.

APPROACHES IN SENSOR-BASED SYSTEM DATA MODELLING: A SURVEY

Muhd Amirul Hakim Zailan[1], Iskandar Ishak[1], Fatimah Sidi[1] and Thinagaran Perumal[1]
[1]Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Malaysia.
hakim_s09@hotmail.com, iskandar_i@upm.edu.my, fatimah@upm.edu.my, thinagaran@upm.edu.my
corresponding author: iskandar_i@upm.edu.my

Abstract: Internet of Things (IoT) is a new technology live that helps to improve our everyday life by offering new services. Through IoT, recent developed technologies exist such as Big Data, Cloud Computing and Monitoring. Several approaches have been proposed in literature to reduce energy consumption and reduce communication overhead in sensor-based system such as Wireless Sensor Network (WSN). To optimize WSN, data aggregation is needed to secured data transmission at both sensor nodes and base station. In WSN, data aggregation process should be optimized in an energy efficient manner and providing quality data for processing is a challenge as sensor based system has smaller specification of hardware and software compared to bigger machines. In this paper, we reviewed all of the methods that have been proposed to model data on sensor-based systems. We also discussed the strengths and issues of each approaches.

***Keywords*:** Sensor-based system, data modelling, data management, Internet of Things.

INTRODUCTION

Sensor-based system is consists of a base station and a number of small wireless electronic devices called sensor nodes. IoT is one of the sensor-based system technologies that includes Wireless Sensor Network or WSN which rising rapidly because it involved in everyday life. Those tiny sensor nodes can monitor, read and collect physical or environmental condition such as pressure, temperature, humidity, motion, light, ultraviolet and etc. The collected data will be transmitted over network to the base station for next process. In recent research, there are several challenges need to overcome in implementing WSN such as localization, deployment, coverage, data integrity, reliability and consistency. The most important constraint is energy resources for WSN (Tayeh, Makhoul, Laiymani, & Demerjian, 2018). Each and everything around us become smarter since the device can communicate each other and carry out task independently. The term “intelligent” or “smart” is being used to describe the usage of IoT devices around us to help and assist whenever in need. Big Data is one of the terms used to describe the large amount of data produced by all interconnected IoT devices as said as IoT and Big Data are interdependent technologies and should be develop together (Plageras, Psannis, Stergiou, Wang, & Gupta, 2018).

The purpose of this paper is to analyse and compare current method in maximizing the usage of IoT by using data model, data prediction, and statistical analysis to store sensor data without having a problem in storage and energy dissipation.

APPROACHES IN DATA MODELLING FOR SENSOR-BASED SYSTEM

Several literatures have been reviewed based on the methods proposed by the researchers with regards to data modelling in sensor based system. This paper looked into six types of approaches in modelling data in sensor based systems that includes Dual Prediction Mechanism, Trust weighted Secure Data Aggregation, Gaussian Markovian & Joint Gaussian distribution, Regression Data Model, Dual Kalman Filter and polynomial regression modelling.

Dual Prediction Mechanism (DPM)

In (Tayeh et al., 2018), they proposed a combination of adaptive sampling technique with Dual Prediction Mechanism (DPM) which only requires two measurements to be built and one measurement to be updated. To reduce the number of data sampled by sensors, an adaptive sampling algorithm is being used which is the Kruskal-Wallis statistic model. The Kruskal-Wallis test

takes as input a group of data sets to identify whether there is a difference between at least two of these sets.

DPM is used to conserve energy and maximize the lifetime of the network. A large number of radio operations in the network cause the energy dissipation in wireless sensors network. The lifetime of the network can be extended significantly if the radio communication operations are optimized. Hence, the DPM analyses the history of previously collected information, extract to moving trend of data in order to approximately estimate future readings. Both sensor nodes and base station are deployed with same prediction model and using the same historical data, the sensor nodes and base station regularly make the same prediction of any future observation. As long as the prediction matches the readings, this technique allows the sensor nodes to avoid transmitting its sensed data to the base station. In this DPM method, it requires only two measurements to build the prediction model, and a single measurement to correct it rather than different types of prediction model that need tens to thousands number of data readings to train the model depending on its learning capacity.

Trust Weighted Secure Data Aggregation (TESDA).

(Padmaja & Marutheswar, 2018) proposed a new method for data aggregation to reduce redundancy and energy consumption. There are several data aggregation techniques in WSN such as Lossless Data Aggregation refers to concatenating individual data items into larger packets. It is only effective when the system load is not excessive. Next, Lossy Data Aggregation which is the amount of communicated data must be forcibly reduced if the total communication load exceeds system capacity. Structured Free Data Aggregation provides efficient data aggregation without explicit maintenance of a structure. Lastly, Centralize Approach is an address centric approach where each node sends data to a central node via the shortest possible route using a multi-hop wireless protocol. Employing energy inefficient nodes in data aggregation affects lifetime of sensor network. Thus, aggregation process in WSN should be optimized in energy efficient manner.

Simple data aggregation process is suitable only in attacker free environment. It is necessary to introduce a data aggregation mechanism that filters out attacker's contribution during data aggregation. Nodes behaviour need to be observed in every round of data aggregation and it should reflect in subsequent rounds to filter out the impact of attacker contribution at the final result. If the aggregator compromised, it affects entire aggregation accuracy. An optimized and secure data aggregation protocol that is resilient to false data injection attack launched by compromised sensor and aggregator is proposed. The proposed protocol support energy efficient clustering and performs secure data aggregation process along with trustworthiness estimation using Trust weighted Secure Data Aggregation (TESDA) algorithm. The data aggregation process is optimized by performing aggregation in energy efficient manner through clustering.

Gaussian Markovian & Joint Gaussian distribution

A statistical data model is used to predict the value of sensor without performing actual sensing. Apparently, it benefits the energy consumption on sensor and reducing network traffic. The prediction is useful when the actual reading of sensor cannot be obtained. There are two dominant WSN data models which are Markov Gaussian models where the collection sensors are modelled as Gaussian and evolving through a Markov chain. Second is Regressive model where are a low order Auto-Regressive (AR) model is used for prediction. Existing work has established a statistical model can be very accurate and effective in providing valuable energy saving. In previous work, the main drawback is an implicit assumption that the sensor data used to build the model consists of 100% good readings. Any anomalies or incorrect readings need to be filtered out manually. (Wang, Damevski, & Chen, 2015) proposed a dynamic sensor data model based on clustering of sensor attributes. The model will provide validation and prediction of sensor readings in one framework. Their method uses independent Gaussian Markovian (GM) model and a joint Gaussian distribution for prediction accuracy between two models.

Regression Data Model

One of the earliest data mining techniques that have been used in wireless sensor networks is prediction. Prediction model is a model to predict the future data states of sensor nodes based on current data. There are two types of prediction can be defined, classification and regression. Classification is used to predict nominal or discrete values while regression is used to predict continues or order values (Samarah, 2015). The main goal of regression based prediction algorithm is to build a regression model to predict the future reading of sensor nodes and exempt the nodes from sending their readings to the base station. To maintain regression models, there are two paradigms which are dual prediction scheme and single prediction scheme. The same prediction model is built at the base station and sensor node for dual prediction scheme. During operation time, the sensor node will update the base station with the new value if the sensed value has a significant difference from the predicted value. For single prediction scheme, a single model is built at the base station or the sensor node, based on the reading of a subset of sensor nodes.

Dual Kalman Filter (DKF)

Dual Kalman Filter (DKF) is a regression-based data gathering framework introduced by R. E. Kalman. (Jain, Chang, & Wang, 2004) By using DKF, it can reduced communication overhead and online data smoothing helps to provide query answers for noisy data. The mechanism is based on input characteristics and output is sensitive to input values. The prediction gives better results, reducing load adaptively rather than dropping chunks of data indiscriminately. The traditional Kalman Filter (KF) is a linear algorithm that estimates the internal state of a system based on two mechanisms; Prediction/Estimation and Correction. Kalman Filter comprises a set of mathematical equations that provides a recursive solution to the least-squares method. DKF framework is by having multiple KF at the main server without having effect to the performance significantly. Those KF operate on the same parameters and precision width. This technique saves communication overhead by collecting the sensor data with the same reading unless the reading changes or if there are anomalies presented.

Polynomial Regression Modelling

Polynomial Regression Model in sensor-based system is applied using the characterization of the relationship between strains and drilling depth. The parameters of the model are estimated using least square method and after fitting, the model is evaluated using common indicators used to evaluate the accuracy of the regression model. Regression analysis involves identifying the relationship between a dependent variable and one or more independent variables. A model of the relationship is hypothesized and estimates of the parameter values are used to develop and estimated regression equation (Ostertagová, 2012). Least squared method is an extension of the procedure used in simple linear regression. Mean Squared Error (MSE) is a measure of how well the regression fits the data. The Root Mean Squared Error (RMSE) and MSE are measures of the size of the errors in regression and do not give indication about the explained component of the regression fit. Mean Absolute Percentage Error (MAPE) is the most useful measure to compare the accuracy of forecasts between different items or products since it measures relative performance. It is commonly used in quantitative methods of forecasting. If calculated value of MAPE is less than 10%, it is interpreted as excellent accurate forecasting whereas over 50% is inaccurate forecasting.

FINDINGS AND DISCUSSION

An adaptive sampling technique with DPM is proposed by (Tayeh et al., 2018) is a technique that requires two measurements and update one measurement. The adaptive sampling technique is using Kruskal-Wallis statistic model to form the backbone of the adaptive sampling algorithm to reduce the number of data sampled by each sensor nodes. It takes input a group of data sets to identify the differences between two data sets. Based on Kruskal-Wallis test, when a sensor node notices high variance differences, it increases its sampling rate in order to prevent missing important measurements and decreases it sampling rate when the variance is less than the threshold. DPM is a

model that analyses the history of previously collected information and extract the moving trend of data in order to approximately estimate future readings.

The same prediction model is deployed at both sensor nodes and base station to make the same prediction regularly of any future observation. It allows the sensor nodes to avoid transmitting its sensed data to the base station as long as the prediction matched the readings. To build the prediction model, only two measurements are sufficient and single measurement to correct it. The adaptive sampling algorithm reduces the sampling rate of a sensor when the difference between collected measurements is not significant. Thus, it enable sensor node to avoid collecting redundant and superfluous information. The transmission reduction algorithm reduces the number of data transmitted to the base station by using prediction model that can forecast future measurements within a narrow error range. The efficiency of the prediction model is at peak when data is smoothly changing with low variance between measurements. Both algorithm thus not overlap each other and does not affect each other results. Thus these two techniques are compatible. By combining the adaptive sampling technique and transmission reduction algorithm, the overall complexity remain unchanged and able to achieve lower energy consumption.

TESDA is a protocol that supports energy efficient clustering and performs secure data aggregation process along with trustworthiness estimation. Data aggregation process is optimized by performing aggregation in energy efficient manner through clustering. Sensor network is divided into clusters and each energy efficient cluster head aggregates data collected from its cluster and transmit to base station. Each sensor attaches its ID and residual energy in its message. The node that receives the message adds the sender in its neighbor list. Each sensor node compares the residual energy of all of its neighbors. The neighbor that has high residual energy is selected as clusterhead. The cluster member then attached its clusterhead ID in the message. Each node checks the clusterhead ID in the message and its own ID is same on receiving. If same, it adds the corresponding sender into its member list. In order to balance the energy consumption, clusterhead roles are rotated in every round. Sensed result of every sensor is submitted to clusterhead. It derives the aggregated result by taking weighted average of collected information. Attacker will compromises the sensor and alters the result to very low or high to distort the aggregation result. Thus from compromised result, it reduce aggregation result as clusterhead computes aggregation result from the reported value. When the clusterhead transmit the result to base station, it leads to false decision making. The final aggregated result at base station is the trust weighted summation of data reported by the clusterhead in the round.

Gaussian distributions are widely used to describe non-deterministic events and sensor modelling. A multivariate Gaussian is the extension of the normal PDF in R space (Wang et al., 2015). Gaussian distributions are very computing friendly which are preserved under various manipulations. The marginalization of a joint-Gaussian stills a Gaussian distribution with a reduced dimension. It allows calculating the posterior probability with limited computing power at the sensor side. Most of the prediction and model update can be performed by dealing with the mean and covariance. As in BBQ model, new observations at the current time slot will trigger the updating of the density function for the next time slot. The procedure is based on the join-probability of a cluster between two consecutive time slots. The validation procedure uses the current model to determine the legitimacy of an observation. Firstly, it will be checked against the overall static model of the given attribute. Then, a correlation test will be done with voltage reading of the same sensor and finally another correlation test with the most correlated cluster model. If the observation is within 90% range of the static model, it is deemed as legitimate reading without further testing.

(Samarah, 2015) proposed model is based on the idea that all the values of the vectors will be distributed around a line in dimensional space. The model building process starts within the sensor by computing the vector by averaging the values of all the time period's vectors. After that, the Euclidean distance between time period's vector and average vector is computed. The model usage starts by the sensor nodes sending the minimum and maximum vectors to cloud system through the base station. The cloud constructs the prediction model based on the equation given. The prediction model then is evaluated with comparison to linear regression model. The error rate computed is based on Mean Squared Error formula.

DKF (Jain et al., 2004) is used to model data for data streaming application. In most object moving tracking system, the future location of the object is modelled as linear function of time. The trajectory of the moving object is a line segment on the space time domain. With the KF formulation, it can be easily simplified or generalize to suit many different system models and external observations. When there are no measurements, the filter predictions would be based on the latest trend that the object has been following. Next, the electrical power load of a particular zone updated every hour. At the datasets, it indicates that the measurements follow a sinusoidal trend, where the load reaches its peak value during working hours and drops during night and early morning hours. In network monitoring, the data is extremely noisy revealing no visually-identifiable trend. It is difficult for any prediction algorithm to aid in conserving network communication overhead. By using KF at the remote source, this filter smooths the data online before operating on it. The filter smooth the data based on parameter and feeds data to other KF at the remote source.

By applying polynomial regression, the curve fitting will be fit with linear, quadratic and cubic polynomial to see if it can reduce the model by few terms. A more common use of a regression analysis is prediction, providing estimates of values of the dependent variables by using the prediction equation. By using best polynomial regression model, the prediction interval for strains will show the predicted values at 95% prediction bounds (Ostertagová, 2012).

Table 1: Techniques for Data Model in Sensor-based system

Technique	Features
Dual Prediction Mechanism (DPM)	Conserve energy
Trust Weighted Secure Data Aggregation (TESDA).	Reduce redundancy Conserve energy secured
Gaussian Markovian & Joint Gaussian Distribution	Conserve energy
Regression Data Model	Save communication overheard
Dual Kalman Filter (DKF)	Save communication overheard
Polynomial Regression Modelling	Accurate forecasting

CONCLUSION

As a conclusion, several methods have been explained from previous literatures which have different method to model the data, to minimize communication overhead and to reduce the energy consumption. There are pros and cons from each methodologies and those methodologies can be used either single use or combining the method to produce new method in order to be used in sensor-based system. In sensor-based system, those problems such as high-energy consumption and data transmission overhead will affect the end result. It may alter the end result and provide fragile information. With such fragile information, it may produce inaccurate results and need to re-run the experiments to obtain sturdy information. In this era of digital information, all of the information needs to be accurate and precise for ability to predict the future values and provide valuable information from the expected result. The literatures mostly focused on the accuracy of the modelling approach and also the energy consumption but lack in terms of other perspective such as security and decision making aspects. Researchers need to aggressively explore and produce better approaches, techniques and algorithms to overcome the situations and provide solutions for the problems stated.

REFERENCES

Čolaković, A., & Hadžialić, M. (2018). Internet of Things (IoT): A review of enabling technologies, challenges, and open research issues. Computer Networks, 144, 17–39. https://doi.org/10.1016/j.comnet.2018.07.017

Jain, A., Chang, E. Y., & Wang, Y.-F. (2004). Adaptive stream resource management using Kalman Filters. Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data - SIGMOD '04, 11. https://doi.org/10.1145/1007568.1007573

Ostertagová, E. (2012). Modelling using polynomial regression. Procedia Engineering, 48, 500–506. https://doi.org/10.1016/j.proeng.2012.09.545

Padmaja, P., & Marutheswar, G. V. (2018). Energy efficient data aggregation in wireless sensor networks. Materials Today: Proceedings, 5(1), 388–396. https://doi.org/10.1016/j.matpr.2017.11.096

Plageras, A. P., Psannis, K. E., Stergiou, C., Wang, H., & Gupta, B. B. (2018). Efficient IoT-based sensor BIG Data collection–processing and analysis in smart buildings. Future Generation Computer Systems, 82, 349–357. https://doi.org/10.1016/j.future.2017.09.082

Samarah, S. (2015). A data predication model for integrating Wireless Sensor Networks and cloud computing. Procedia Computer Science, 52(1), 1141–1146. https://doi.org/10.1016/j.procs.2015.05.148

Tayeh, G. B., Makhoul, A., Laiymani, D., & Demerjian, J. (2018). A distributed real-time data prediction and adaptive sensing approach for wireless sensor networks. Pervasive and Mobile Computing, 49, 62–75. https://doi.org/10.1016/j.pmcj.2018.06.007

Wang, J., Damevski, K., & Chen, H. (2015). Sensor data modeling and validating for wireless soil sensor network. Computers and Electronics in Agriculture, 112, 75–82. https://doi.org/10.1016/j.compag.2014.12.0

VISUAL ANALYTICS FOR GOVERNANCE AND INFORMED DECISION MAKING: AN OVERVIEW FOR LAND USE LAND COVER CHANGE MANAGEMENT

Lilly Suriani Affendey[1], Norwati Mustapha[1], Iskandar Ishak[1], Fatimah Sidi[1] and Nurul Hayat Hamdan[1]
[1]Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 Serdang Selangor
lilly@upm.edu.my, norwati@upm.edu.my, iskandar_i@upm.edu.my, fatimah@upm.edu.my
corresponding author: lilly@upm.edu.my

Abstract: Governance describes the structures and decision-making processes that allow an organization or group of people to conduct affairs. A municipal government generally is one that provides basic city-type services to a local community, and one of the services is land use planning. Keeping track of the Land use and land cover (LULC) change information is important for planning and management activities as well as for monitoring. However, time-base geospatial data results in difficulties for analysis. Traditional static imagery can assist in analysis however the resulting visualizations are often highly specific to a particular data question and must be rebuilt to answer new questions. A solution to this problem is the creation of dynamic interfaces that use time as a mapping for the temporal component of the data. These kinds of visualizations are particularly important when the data is geospatial as well as time-dependent, since effective static visualizations of such multidimensional data are difficult to create. Visual analytics can facilitates the analysis of geospatial data on LULC changes over a period of time. Such analysis may help highlight the complexity of LULC interactions, provide better ways to communicate complex insights so that decision makers can quickly absorb the meaning of the data and take action. Various sources need to be exploited in order to assemble the LULC data which include publicly available national and international statistics and databases, land cover classification information, historical satellite imagery, and aerial photographs. Interactive visualization strategies are recommended since interaction strategies support further scalability and complexity of visual information. Using advanced visual interfaces, users may directly interact with the data analysis capabilities of the visual analytics application, allowing them to make well-informed decisions in complex situations.

Keywords: visual analytics, decision making, geospatial, land use

INTRODUCTION

The ability to obtain, process, and share basic data and insights gained from processed data is key to making informed decisions to improve productivity and economic returns as well as broader social and environmental outcomes. However, this increases the need for enhanced data management and analytical capabilities. While the capacity to collect and store new data grows rapidly, the ability to analyze these data volumes increases at much lower pace. This gap leads to new challenges in the analysis process, since analysts, decision makers, or engineers; depend on information "concealed" in the data. Opportunity and innovative analytics can be applied to identifying opportunities and improving the future [1, 2, 3, 4]. The emerging field of visual analytics focuses on handling massive, heterogeneous, and dynamic volumes of information through integration of human judgments by means of visual representations and interaction techniques in the analysis process.

The rest of the paper is organized as follows: introduction to visual analytics, overview of Land Use Land Cover, results and discussion, and finally the conclusion.

VISUAL ANALYTICS

Visual analytics employs interactive visualizations to integrate users' knowledge and inference capability into numerical/algorithmic data analysis processes. Fig. 1 by Keim et al., illustrates the entire visual analysis process. The process starts by transforming the data for further exploration. After that, a visual or an automatic analysis method is adopted separately. When automatic analysis methods are applied, approaches such as data mining methods are used to estimate models for

characterizing the data. When visual data exploration is used, users directly interact with the visual interface to analyse and explore the data. The combination and interaction between visual and automatic analysis methods are the key feature of visual analytics which helps distinguish the visual analytics process from other data analysis process. Thus visual data exploration together with automatic model-based analysis can often lead to better analysis results [10, 11]. The detail process of visualization includes:

i. Mapping – mapping means how to visualize information or how to encode information into visual form.
ii. Selection – selection means to select data among those data, which is available according to the given task or job. Selection of data is directly dependent on the aim to get through visual graphics or pictorial representation.
iii. Presentation – presentation means how to manage, organize information in the available space on the screen effectively.
iv. Interactivity – interactivity means what are the provided facilities to organize, explore, and rearrange the visualization. User-friendly interactivity enables a user to best explore, understand, and interpret the data or information, which improve their exploration capabilities.
v. Human factors – involves usability and accessibility factors.
vi. Evaluation – to find out whether the visualization method has effectiveness or not, the goal is achieved or not.

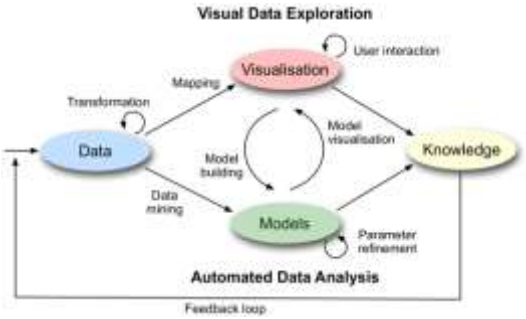


Fig. 1: Visual Analysis Process (Keim, et.al, 2013)

OVERVIEW ON LAND USE LAND COVER CHANGE MANAGEMENT

Land Use refers to the human activities that are directly related to the land. Meanwhile, Land Cover describes the vegetation, water, natural surface, and man-made features of the land. Land use and land cover areas are classified into nine major categories namely as urban or built-up land, agricultural, rangeland, forest, water areas, wetland, barren land, tundra, and perennial snow or ice. The impacts of land Use Land cover (LULC) change on carbon dynamics, climate change, hydrology, and biodiversity have been recognized. However, the LULC models require both historical and current land-cover maps coupled with data representing the driving forces of change. Availability of data, especially spatially and temporally consistent data representing those forces, is a primary challenge for modelling. Nevertheless, remote sensing data play as increasingly important role in LULC modelling. The foundational elements included in the LULC modelling are geographic context, regional land-use history, representation of drivers of change, and representation of local land-use patterns. While the classes of data needed are current and historical land-use data, environmental data, and scenario data [5, 6, 7]. Thus the LULC data are both temporally based and spatial. Geovisual analytics approach was proposed by [8] to deal with challenges in uncertainty arising from errors and inaccuracies from multiple datasets in the remote sensing and GIS domain. To design a dynamic visualization of geospatial and time dependent data is challenging. The demand for specialized visualizations to answer specific questions for even specific questions has long been recognized as a problem in static visualization design. There are several stages in the evolution in complexity of time-based data visualizations, from the simple and ubiquitous time series graph, to static visualizations that support complex relationships between data and time including conditional, variable, or even unknown time values for data points. Ultimately, the more elaborate the visualization, the more specialized it is. Visualization can be considered dynamic in 2 ways: (a)

animation provides a playback mechanism whereby the data is displayed over time to the analyst, and (b) interaction provides mechanisms for the analyst to change the data representation and/or control the playback. Dynamic visualization tools that incorporate both animation and interaction have more flexibility in the kinds of questions they can assist in answering than those without. In addition, dynamic visualization also resolve the lack of visual variables to represent the temporal component of the data, since the sequential order of frames in the dynamic visualization represents the relative temporal relationships between data elements [9, 10, 11, 12, 13]. Since dynamic visualization requires complex computation, research on lightweight clients were conducted to support information retrieval[14]. Lucas visual browser web tool was implemented on a server-client architecture, which integrates several web technologies. The system manage the more complex processing tasks on the server offering lightweight clients for different devices types. Nevertheless, the interactions were limited to selecting specific region of interest and querying points inside the region.

RESULTS AND DISCUSSION

As an initial step for the research project, we explored the satellite images and maps data, as well as tools such as ENVI and ArcGIS. Fig. 2, shows the web-based Geographic Information System (GIS) prototype that can visualize maps and calculate the percentage of LULC changes over a period of years for a specific region in Kota Tinggi, Johor. The land use and land cover data were collected from aerial photography acquired by the satellite (LandSat).

Fig. 2: The web-based Geographic Information System (GIS) prototype

Fig. 3 shows the interface for a user to perform the Map Classification according to the Class Scheme.Fig. 4 shows the base map with the class scheme, namely Badan Air, Tanah Kosong, Hutan, Pertanian, Infrastruktur & Utiliti & Kemudahan Awam. Meanwhile Fig. 5 shows the interface for the calculation of LULC changes for the chosen years.

Fig. 3: Map Classification Interface

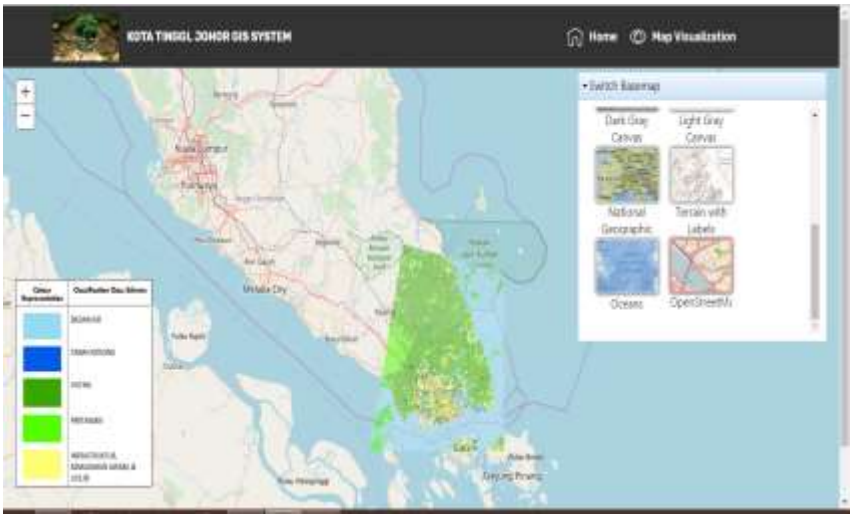


Fig. 4: Kota Tinggi base map with the class scheme

Fig. 5: Details of LULC Calculation

Future work shall include advance visual analytics for predicting possible LULC changes.

CONCLUSIONS

Visual analytics for temporally based geospatial data provide informative insight for decision-making. Thus, the objective of this research is to design and develop a visual analysis method which facilitates the analysis of geospatial data on LULC changes over a period of time. Such analysis may help highlight the complexity of LULC interactions, provide better ways to communicate complex insights so that decision makers can quickly absorb the meaning of the data and take action. Interactive visualization strategies are useful since interaction strategies support further scalability and complexity of visual information. Using advanced visual interfaces, users may directly interact with the data analysis capabilities of the visual analytics application, allowing them to make well-informed decisions in complex situations.

REFERENCES

Almeida, F. & Calistru, C. “The main challenges and issues of big data management”, International Journal of Research Studies in Computing, Vol. 2, No.11, pp. 11-20, April 2013.

IBM, “Big Data Analytics:Deriving Meaning From the Data Explosion”, Deep Dive Series, Infoworld Special Report, 2013.

Informatica, “Preparing For the Big Data Journey: A Strategic Roadmap to Maximizing your Return from Big Data”, White Paper, 2013.

Alexandros Labrinidis & H. V. Jagadish, “Challenges and Opportunities with Big Data”, In Proceedings of the VLDB Endowment, Vol. 5, No. 12, pp.2032-2033, Turkey, August 2012.

Terry Sohl & Benjamin Sleeter. “Role of Remote Sensing for Land-Use and Land-Cover Modeling”, In book: Remote Sensing and Land Cover: Principles and Applications, Chapter: Role of Remote Sensing for land-use and land-cover change modeling, Publisher: Taylor and Francis CRC Press, Editors: Chandra Giri, pp.225-239, 2013.

Lin Cassidy, Michael Binford, Jane Southworth & Grenville Barnes, “Social and Ecological Factors and Land-use Land-cover Diversity in Two Provinces in Southeast Asia”, Journal of Land Use Science , 5(4), 277-306, 2010.

Amna Butt, Rabia Shabbir, Sheikh Saeed Ahmad, Neelam Aziz, “Land use change mapping and analysis using remote sensing and GIS: A case study of Simly watershed, Islamabad, Pakistan”, The Egyptian Journal of Remote Sensing and Space Sciences, 18, pp.251-259, 2015.

Christoph Kinkeldey, “A Concept fro Uncertainty-Aware Analysis of Land Cover Change Using Geovisual Analytics”, ISPRS International Journal of Geo-Information, 3, pp.1122-1138, 2014.

Peter Fox & James Hendler, “Changing the Equation on Scientific Data Visualization” Science 331, 705-708, 2011.

Theresa-Marie Rhyne, “Top Scientific Visualization Research Problems”, Visualization Viewpoints, IEEE Computer Graphics and Applications, 2004.

Daniel Keim, Huamin Qu, & Kwan-Liu Ma, “Big Data Visualization”, IEEE Computer Society, July/August 2013.

Guo-Dao-Sun, Ying-Cai Wu, Rong-Hua Liang & Shi-Xia Liu, “A Survey of Visual Analytics Techniques and Applications: State-of-the-Art Research and Future Challenges”, Journal of Computer Science and Technology, 28(5), 852-867, 2013.

Muzammil Khan & Sarwar Shah Khan, “Data and Information Visualization Methods, and Interactive Mechanisms: A Survey”, International Journal of Computer Applications (0975-8887), 34(1), 2011.

Kevin Alonso, Daniela espinoza-Molina, Mihai Datcu, “ Lucas Visual Browser: A Tool for Land Cover Visual Analytics”, In Proceedings of 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2015.

DEVELOPMENT PROCESS OF HOSPITAL INFORMATION SYSTEM (HIS)

Siti Nor Haslinda Zainudin[1], Wan Nurfaezah Shafiqah Wan Mohd Hawari[2],Farah Nabiha Ahmad Radzi[3], Norasyikin Sarju[4] and Nur Athirah Khairil Mokhtar[5]
[1] Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 Serdang Selangor
[2,3] Department of Software Engineering and Information System, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 Serdang Selangor
[4] Department of Multimedia, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 Serdang Selangor
[5] BIT Software (M) Sdn. Bhd, 4809-3A-29, CBD Perdana 2, Jln Perdana, Cyber 12, 63000 Cyberjaya, Selangor
haslindazainudin96@gmail.com, faezahshafiqah00@gmail.com, fnabiharadzi@gmail.com, norasyikinsarju@gmail.com, athirah.mokhtar@bit.com.my
corresponding authorsnorasyikinsarju@gmail.com, athirah.mokhtar@bit.com.my

Abstract: Hospital Information System (HIS) is essentially a computer system that can manage all the information to allow health care providers to do their jobs effectively[1]. This system can manage a wide range of hospital administration and management processes for the multispeciality hospital. In order to develop the system, it is necessary to analyze the requirements thoroughly to make sure the system meets the user requirement. Currently, the Hospital Information System contains Registration, Outpatient, Inpatient, Appointment and Medical Report modules. These modules act as different roles and components that totally run and deploy on a separate process. In developing HIS, it's important to ensure the system can work effectively and efficiently by conducting testing. This system applied to three types of users which are doctors, nurse and system administration.

***Keywords*:** hospital information system, develop, manage

INTRODUCTION

Healthcare is a very important part of our society and it is imperative for healthcare providers to do their jobs in an efficient and effective manner. Each day hundreds of thousands of patients enter healthcare facilities challenging the administration to run the show smoothly. The employees have to manage and integrate clinical, financial and operational information that grows with the practice. Previously, this data was organized manually, which was time-consuming and failed to deliver the desired level of efficiency. Most professionally run hospitals and clinics now rely on hospital information systems (HIS) that help them manage all their medical and administrative information. Hospital Information System (HIS) is essentially a computer system that can manage all the information to allow health care providers to do their jobs effectively. HIS includes many applications addressing the needs of various departments in a hospital. The hospitals that have switched to HIS have access to quick and reliable information including patients' records illustrating details about their demographics, gender, age, etc. (2018, January 11) Retrieved from: http://www.emrconsultant.com/emr-education-center/emr-selection-and-implementation/hospital-information-systems-his/. The importance of these systems emerges from the importance of their role in keeping all types of patient data and information including key data about the patient and other comprehensive medical data; recording all medical services that have been provided to the patient such as investigations, diagnoses, treatments, follow up reports and important medical decisions (Ismail et al., (2010).

It is accessible either by an administrator, nurse, and doctor. Only they can add data and retrieving data from the database. Besides, the data are well protected for personal use and makes the data processing very fast. This system is easy to use and is designed and developed to deliver real conceivable benefits to hospitals. By a simple click of the mouse, they receive important data pertaining to hospital finance systems, the diet of patients, and even the distribution of medications. With this information, they can monitor drug usage in the facility and improve its effectiveness (2018,

January 11) Retrieved from: http://www.emrconsultant.com/emr-education-center/emr-selection-and-implementation/hospital-information-systems-his/.

HIS is a software product suite designed to improve the quality and management of hospital management in the areas of clinical process analysis and activity-based costing. This HIS enables to develop an organization and improve its effectiveness and quality of work. Managing the key processes efficiently is critical to the success of the hospital helps to manage the processes. The rest of the paper is organized as follows: Introduction to Hospital Information System, related modules, the development process of the system, results and discussion, and finally the conclusion.

METHODS

The development of HIS is using a Scrum Methodology. Scrum methodology is a simple framework that facilitates team collaboration on complex projects and it emphasizes teamwork in project management (2018, January 11) Retrieved from: https://www.projectmanager.com/blog/scrum-methodology. Scrum methodology advocates for a planning meeting at the earlier of the sprint, where team members figure out how many items they can commit to, and then create a sprint backlog, a list of the tasks to perform during the sprint (2018, January 11) Retrieved from: https://www.mountaingoatsoftware.com/agile/scrum. Sprint methodology is implemented in HIS by planning a meeting on a date that has been set by the Scrum Master or in HIS case, the solution architect. On each day of the sprint, all team members will attend the meeting and the team members will share what they have worked on the given deadline and identify any impediments related to the items or the task assigned.

RELATED MODULES

For now, HIS consists of several modules that make the system function as each module have their own functionality. Explanation of each module is as below.

Registration Module

Registration modules are intended to save patient details by registering the patient to the system before treatment is done. Among the functionalities included in this module are Manage Patient Information and Add to Queue.

Outpatient Module

The Outpatient module is a module that will do all the clinical task involving Nurse and Doctor task. It can be, after registration or an appointment, there is a case where one or more of patient got conflict with appointment details, so in the outpatient module will solve the issue by managing the record well. Submodule included in Decision Module are View Clinical Queue, View Patient, Check-in Patient, Create Encounter, and Discharge.

Inpatient Module

The Inpatient module is a module that will do all the tasks that involve the admitted patient. It can be, after an appointment or outpatient, there is a case where one or more of patients have a conflict when nurse want to assign ward and bed to the patient, so in inpatient module will solve the issue by managing the record well. Submodule included in Inpatient Module are Create Inpatient Progress, Check-In Patient, Manage Ward, Ward Transfer, Create Inpatient Progress and Discharge.

Appointment Module

The Appointment modules will handle the doctor scheduling by save all the unavailable time for doctor and doctor leave. Submodule involve in this module are Create Appointment, Manage Room Appointment, and Scheduling (management).

Medical Report Module

The Medical Report module is a module that handles the application for Medical Report. All application of Medical Report will handle by front desk. Once the application is complete doctor will create Medical Report for the applicant.

DEVELOPMENT PROCESS OF THE SYSTEM

For development, there are five stages that involved to make sure the system being developed successfully. Explanation of the development process is as below.

Analyzing Requirements

In analyzing the requirements stage, the requirement is analyzed from the documented requirements received from the client which is called as Statement of Compliance (SOC).In SOC there are several items such as Registration, Outpatient Management, Inpatient Management and more. In this process, each point in the items going through and analyze. In this process also, other types of documents such as hardcopy form are being gone through and then being transforms into the interactive web form.

Designing Prototype

After the requirements have been analyzed, the next phase is to design the screen of the system. The prototype is designed based on the requirement that has been validated by the project consultant. Besides, the prototype for each module is designed using Justinmind Prototyper application.

Developing User Interfaces

Development phase starts after the mock-up screens are verified by the project manager. During the development phase, the mock-up screens are transformed into an actual website page. Hence, for the development for the front-end, the following tools or technologies are used; Hypertext Markup Language (HTML), Cascading Style Sheets (CSS) and Bootstrap 4.0 templates. In the earlier part of the development phase, the overall design of the system is developed by the web designer, then the static HTML page is developed without any functionalities based on the mock-up screen.

Binding User Interface with Application Program Interface (API)

In this process, the interfaces being binds with API (ASP.NET) using Angular JS (open-source front-end web framework). The process binding using Angular JS is to synchronize the model and the view. It is to able interfaces to interact with the database. This process of binding use the concept Model-View-Controller (MVC) where model represents the data, the view is user interfaces and the controller is the processes that handle input.

Testing

After done with the binding process, the system undergoes testing. Process of testing being done by the developer themselves and also by the client. Testing by the developer is to make sure the system function as what it should and to fix any bugs that occur before release the system to being testing by the client. Testing by the client is done mostly to make sure the function, user interfaces, and the system flow meet with what the client wants. Once there is any function that did not meet want client want, the process of developing interfaces and binding being done again to fix anything that client want to fix.

RESULTS AND DISCUSSION

According to the research that has been done by Jamali (2016), the system is designed for multispecialty hospitals, to cover a wide range of hospital administration and management processes. The system is an integrated end-to-end Hospital Management System which provides relevant information across the hospital to support hospital administration, effective decision making for patient care and critical financial accounting, in a seamless flow. As for now, there are five modules includes in HIS which the modules are registration module, outpatient module, inpatient module, appointment module, and medical report module.

Figure 1 shows the screenshot of the main page of HIS. The modules can be accessed from the main page. Figure 2 shows the screenshot of the registration module for search patient. From the search patient page, the user able to register patients, view patient information, edit patient information, add the patient to the clinical queue. Figure 3 shows the screenshot for Outpatient module: Nurse Assessment Vital Sign. The nurse assessment vital sign page can be accessed by both nurse and

doctor, it is meant to record the patient's vital signs for monitoring purposes. Figure 4 shows the screenshot of Inpatient module for ward management (register ward). Ward management (register ward) is for managing the process for patient's check-in into ward. Figure 5 shows the screenshot for appointment module for creating an appointment. Figure 6 shows the screenshot for medical report module for creating a medical report application.

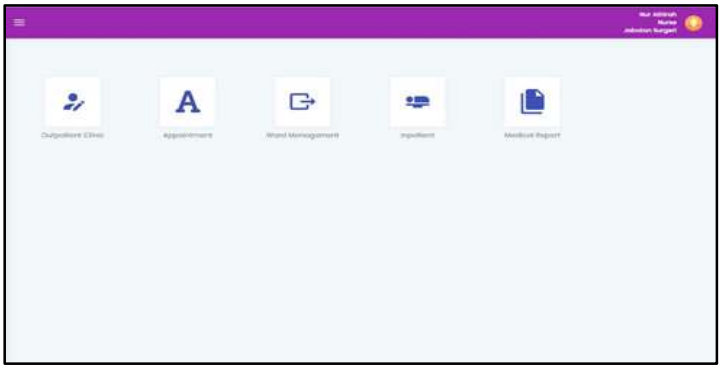


Figure 1: Main Page

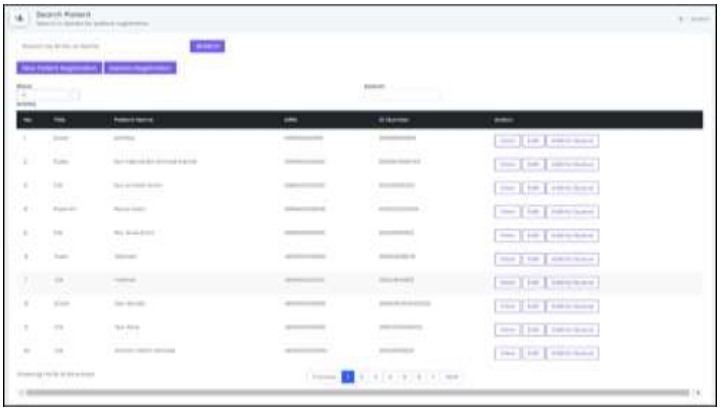


Figure 2: Registration Module – Search Patient

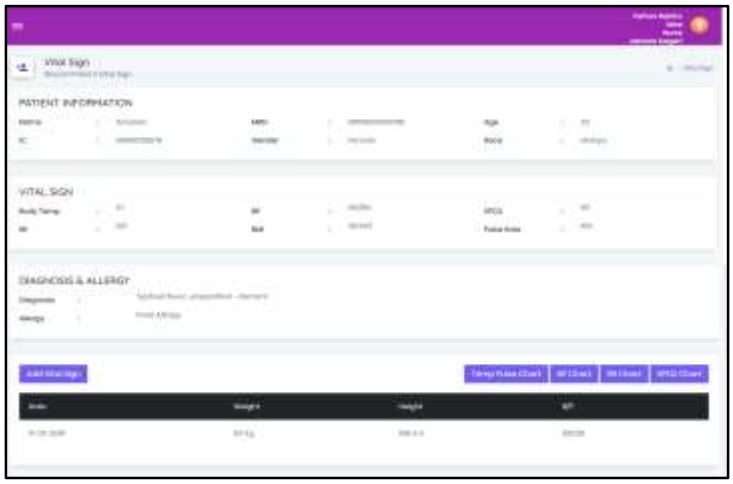


Figure 3: Outpatient Module: Nurse Assessment Vital Sign (Nurse & Doctor)

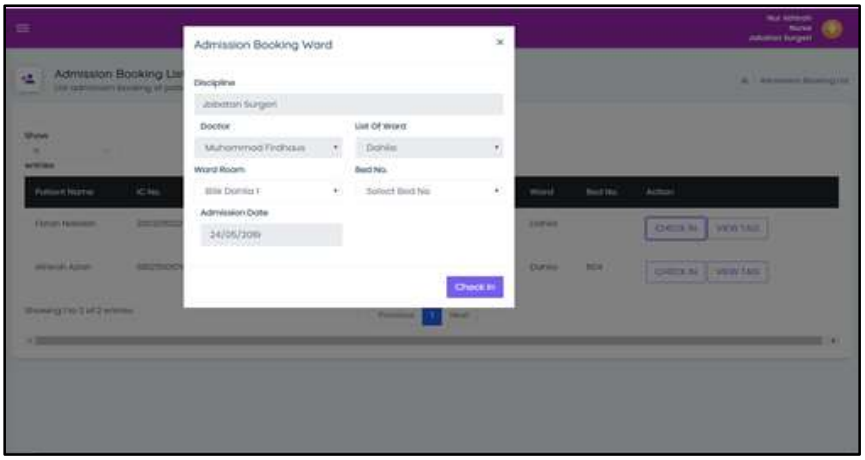


Figure 4: Inpatient Module: Ward Management (Register Ward)

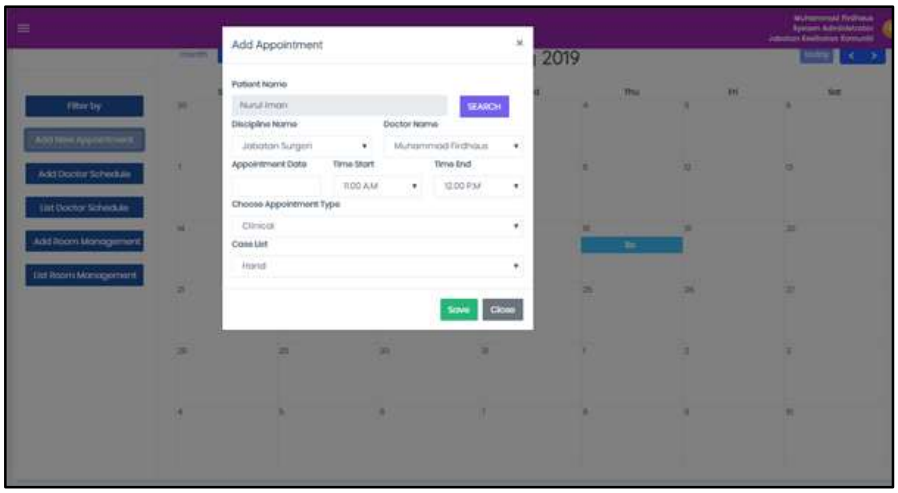


Figure 5: Appointment Module - Create Appointment

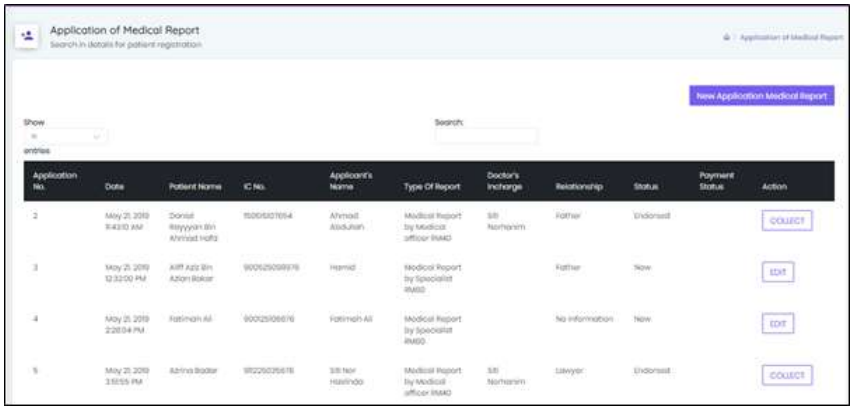


Figure 6: Medical Report Module – Medical Report Application

BENEFITS FOR HIGHER EDUCATION

Hospital Information System were developed with its own purposes and can give benefits to others. This system can attract more university students to work in the hospital and they can use HIS for their educational purposes especially when doing their practical at the hospital. The system has well designed information retrieval and viewing, with ease of data entry and speeds up the clinical decision-making process (Malliarou & Zyga, 2009) that make the system easy to use for a first time user. Next, this system has improved on the clinical management (2018, January 11) Retrieved from: https://www.ncbi.nlm.nih.gov/m/pubmed/28679914/#fft which involved doctors, nurses and patients as it can help improve patient satisfaction on the clinical process and reduce patients waiting time. Plus it also make doctors and nurses works easier. Then, this system has improved on accessing patient information (2018, January 11) Retrieved from: https://www.ncbi.nlm.nih.gov/m/pubmed/28679914/#fft By using patient's identity card, the information of the patient can be accessed and retrieved easily.

CONCLUSIONS

Hospital information systems (HIS) have become very advanced and new innovations are continuously being introduced. Hence, the developed HIS is meant to improve the quality and management of hospital management in the areas of clinical process analysis and activity-based costing. Moreover, this system must be user-friendly and should include training by the vendors. A good HIS offers numerous benefits to the hospital including the delivery of quality patient care and better financial management. The technology changes quickly and if the system is not flexible it will not be able to accommodate hospital growth. An effective HIS also delivers benefits such as enhances information integrity, reduces transcription errors, reduces duplication of information entries and optimizes report turnaround times.

REFERENCES

"Hospital Information System (HIS)", 21 August 2013. Accessed on 11 January 2018. [Online] Retrieved from: http://www.emrconsultant.com/emr-education-center/emr-selection-and-implementation/hospital-information-systems-his/

Ismail, A., Jamil, A. T., A Rahman, A. F., Abu Bakar, J. M., Mohd Saad, N., & Saadi, H. (2010). The Implementation of Hospital Information System (HIS) in Tertiary Hospitals in Malaysia: A Qualitative Study. *Malaysian Journal of Public Health Medicine,10*(2), 16-24. Retrieved July 15, 2019, from https://www.mjphm.org.my/mjphm/journals/Volume10.2/3.%20Implementartion%20hospital%20(Last)(16-24).pdf

Shortliffe, E. H., & Barnett, G. O. (2014). Biomedical data: Their acquisition, storage, and use. In Biomedical informatics (pp. 39-66). Springer London

Khalifa, M., & Alswailem, O. (2015). Hospital Information Systems (HIS) Acceptance and Satisfaction: A Case Study of a Tertiary Care Hospital. *Procedia Computer Science,63*, 198-204. DOI:10.1016/j.procs.2015.08.334

Jamali, M. C. (2016). An Analysis on Transforming Developments in Electronic Hospital Management & Hospital Information Systems. *Journal of Clinical Trials & Patenting,1*(1). DOI:10.13188/2573-3834.1000005

"Understanding Scrum Methodology". Accessed on 11 January 2018. [Online] Retrieved from: https://www.projectmanager.com/blog/scrum-methodology

"Scrum Methodology and Project Management". Accessed on 11 January 2018. [Online] Retrieved from: https://www.mountaingoatsoftware.com/agile/scrum

Malliarou, M., & Zyga, S. (2009). Advantages of Information Systems in Health Services. *Sport Management International Journal, 5(2).* DOI: http:dx.doi.org/10.4127/ch.2009.0040

Khalifa M. Stud Health Technol Inform. 2017. "Perceived Benefits of Implementing and Using Hospital Information Systems and Electronic Medical Records". Accessed on 11 January 2018. [Online] Retrieved from: https://www.ncbi.nlm.nih.gov/m/pubmed/28679914/#fft

A STUDY ON PHYSICAL ACCESS CONTROL USING QR CODE AS VIRTUAL IDENTIFICATION SYSTEM FOR DOOR ACCESS CONTROL

Ashvini Priya Rajangeran[1], Puvanah Sandra Sekaran[1], Fatimah Sidi[1], Saiful Ramadzan Hairani[2], Mohd Faizal Daud[2] and Rizal Razak[2]
[1]*Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 Serdang Selangor*
[2]*InfoComm Development Centre (iDEC), Universiti Putra Malaysia, 43400 Serdang Selangor*
ashvini.priya@yahoo.com, buvanah.segar5@gmail.com, fatimah@upm.edu.my, ramadzan@upm.edu.my, faizal@upm.edu.my, rizalr@upm.edu.my
corresponding author: fatimah@upm.edu.my

Abstract**:** The increase in criminal activities in today's world is the cause of tremendous development in safety measures across the nation. One of the widely used safety measures is physical access control. Authentication methods of physical access control such as keys, biometrics (fingerprint, retinal scan), smart cards and One Time Password (OTP) require additional equipment and a high cost to purchase, install and to maintain it, whereby many organizations depending on their specific industry and its requirement along with budget allocation, may choose to opt for these features or not. Unfortunately, tools like keys and smart cards are not unique to specific users and could be possibly misplaced due to human error. OTPs are normally received as text messages on mobile devices and it could be accessed by other parties than the actual individual under minimal supervision by the owner of the device. In this innovative era, smartphones play a noteworthy part in our daily life. This device is easily detectable when it has been misplaced since it is being habitually utilized everyday. Hence, smartphone is utilized in our research due to its deliberate availability. An element that will be corporated on the device, is the Quick Response (QR) code. A QR code can now, not just solely save massive records, but additionally has the error-correction capacity. The QR code has been provided to each UPM students in the PutraVID (Putra Virtual Identification) application as a physical access control tool. The PutraVID has implemented the usage of QR code in the campus for everyday use. Although the QR code is utilized in a huge scale in the campus, there is no particular system to differentiate UPM students and the public, as they not more wearing their matric card or staff ID card as they used to prior to the introduction and implementation PutraVID. The campus environment became vulnerable to attacks as it became an opportunity for unauthorized people to trespass the facilities and caused damages. In order to cope with this issue, QR code is generated in PutraVID which is unique to individuals, as it contains their information such as matric number and name. This information is encrypted into a two-dimensional matrix code, used to scan and access facilities especially restricted areas. The QR code will not be the same for two people, as each student has their individual information stored in it, in order to avoid outsiders from having access to the campus environment. Hence, the entire ecosystem for this system is researched and developed in this paper inclusive of the hardware and software part which proficiently handles the problem. The system develops a centralized door access control management system which is used to store the time stamp of students and staff, and introduces an IoT-based door access control system using the QR code of PutraVID.

Keywords**:** QR Code, Physical Access Control, Mobile, Authentication, Centralized, IoT-based, Smartphone, Matrix code, PutraVID, Door Access Control System.

INTRODUCTION

Quick Response (QR) codes are used to encode data in both vertical and horizontal directions. QR-Codes are basically two-dimensional bar codes where users can use smartphones to access the encoded data. The process begins with the built-in camera scanning and capturing the image of the QR-Code and then decodes it using a QR-Code reader application. Their easy to use feature has gained more interest as they are considered as an enabling technology for several applications such as storing information, authentication, identification, and many others.

Enabling authentication for access into facilities is one of the interesting features of QR-Codes. For instance, QR-Codes are used in campuses of universities in order to limit the access of people entering certain property, building or restricted areas such as laboratories and faculties. Ever since PutraVID has been introduced in Universiti Putra Malaysia (UPM), it has been used at a larger scale at the campus, thus becoming a necessity in order to limit the access to certain areas that are prone to thefts and vandalism on the campus property because it is easily accessible to the public.
Through this project, we propose an IoT-based door access control system using PutraVID with a centralized database which will be used to store the time stamp of students and staff, moving in and out of the facilities. The proposed system will benefit the university since it does not have an existing access control system to monitor the activity of people at the campus. This project aims to highlight the vulnerabilities, architecture and development of IoT-based Door Access Control System using QR-Code of PutraVID, to be implemented in our campus facilities.

BACKGROUND AND RELATED WORKS

i. QR Code PutraVID
QR Code is matrix code or two-dimensional barcode, outlined for encrypting or decrypting the content in a faster manner. The UPM community (students and staff) can utilize their smartphone for QR code of PutraVID and pick up get to without having to contribute in extra tokens or contraptions. They are having different QR Code which is stored in their database. The example of PutraVID QR Code and barcode are shown in Figure 1.

Figure 1: QR Code of PutraVID

ii. One Time Password (OTP)
QR codes have been utilized to supply physical access control by combining them with confirmation strategies. Kao et al. (2011) conducted a project using One Time Password (OTP) method in physical access controls, using the QR-Code. OTP is considerably secured compared to other methods because the password is only valid once. The architecture method used for their project is matrix code and RSA algorithm using One Time password (OTP). In order to authenticate, the client generates a QR code encoding a scrambled password, which is then scanned by the client's gadget to confirm his rights.

METHODS

For this project, the "waterfall model" is used which is very simple to understand and easy to use as shown in Figure 2. In this model, each stage must be completed before the testing stage. Only then, the quality stage could be started because there will be no overlapping in the stages.

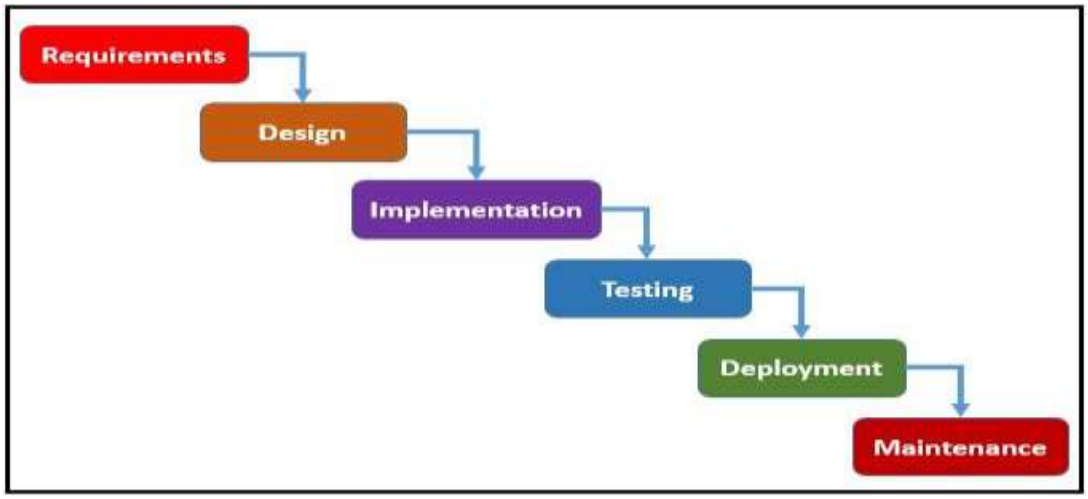


Figure 2: Waterfall Model

These stages are requirements, design, implementation, integration, and maintenance with the following explanation:

Stage 1: Requirement

In this process, before understanding the needs of users, we carried out literature research first. Collect all the requirements which are related to QR code of PutraVID and do brainstorming to understand the requirements. Requirements are collected using several methods such as interviews, observations and analyses on the existing documents. One-to-one interviews are to be conducted with the stakeholders, in order to understand their expectations and how they would want the system to be, as the end product. Observations are done by monitoring the staff while they are accessing the existing access control, setting an appointment with the management of Faculty of Computer Science and Information Technology and IDEC of UPM in order to observe their way of organizing and managing the students and staff's data. Next, by analyzing existing documents, the knowledge on how others integrated the each of components to one another will be gained.

In this study, the UPM community needs to access the faculty's lab and other restricted areas in UPM by simply scanning the QR-code from an authorized mobile phone through PutraVID mobile application. By using the QR code of PutraVID, the process which should be done involves:

i. User scans their QR-code using a QRCode scanner which will be checked in the main database of UPM for authentication permission using Internet of Things (IoT) through PutraVID API.
ii. Simultaneously, time stamp of the user will be stored in the centralized database.

Stage 2: Software Design

After the requirement specifications from first stage are finalized, a framework is drawn to facilitate the procedure of implementations. This framework is done based on the requirements collected from the first stage. This design stage is divided into two phases which are logical and physical phase. Logical phase is also known as the data modeling mapping phase.

2.1 Flowchart of proposed system

A flowchart could be a sort of chart that represents a calculation. The proposed framework could be an IoT-based Door Access Control System using QR-Code of PutraVID. The flowchart of proposed system depicted in Figure 3.

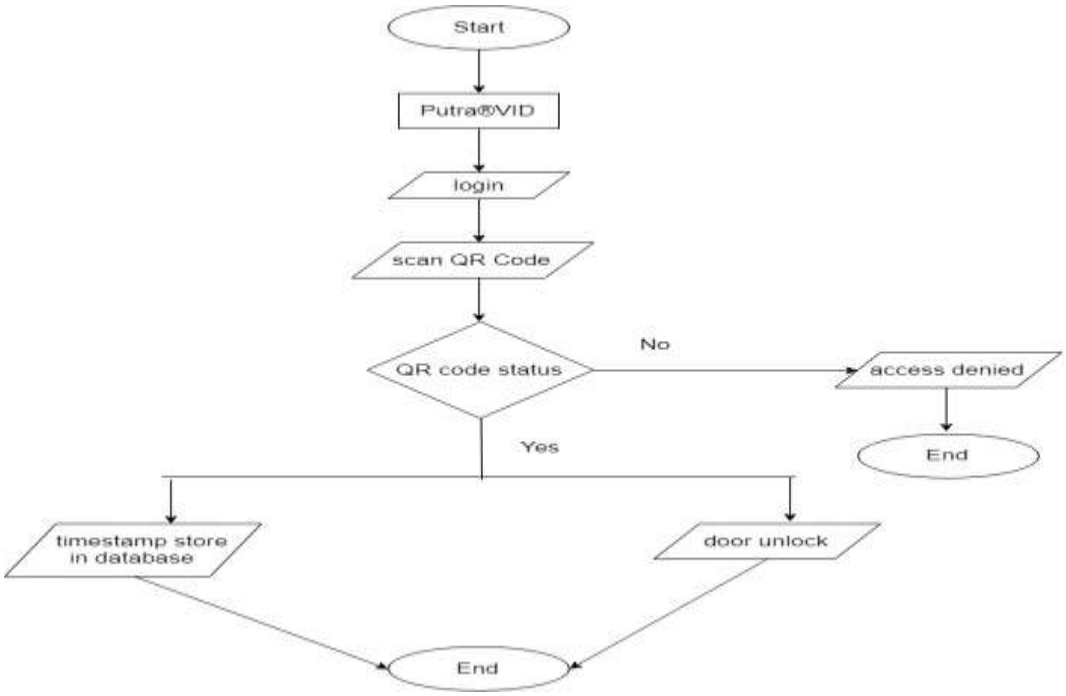


Figure 3: Flowchart of the proposed system

Stage 3: Implementation

The diagram of the system is verified and the implementation of the system will start according to the design of the framework. Each of components according to our project such as data validation, data verification and system managements are called units. Individual task separation based on each component is below:

3.1 Data Validation

- Processing the embedded part of the system.
- Designing the hardware such as QR-Code reader.
- Testing the functionality of hardware.
- Creating the hardware an IP address (IoT).
- Setting up its function using Raspberry pi.

3.2 Data Verification

- Design a website for the usage of users which are students, staff and administrators.
- Integrate the centralized database to the website.
- Assign roles and permission to the users.
- Test and manage the functionality of the website.

3.3 Data Verification

- Create an API to compare the QR-Code that has been scanned with the information exist in database.
- Request permission for access, deny access if needed or else grant the access permission.
- Store the timestamp of the permission requested into the centralized database.

Stage 4: Testing

The implemented system is now verified and tested by the stakeholders such as project managers or the end users themselves. The team will perform testing of the functional and non-functional activities to ensure that the system meets the requirements. The entire framework is tested for any failures or issues and also testing activities must be recorded in order to make further amendment and improvement. We can proceed to the next stage if all the requirements are met and the stakeholders are satisfied.

Stage 5: Deployment

Once the functional and non-functional testing is done, the system is deployed in the respective environment. The system will be up for the implementation in the UPM environment, whereby the usage of the Virtual Identification for door access is used. The hardware will be installed in all the restricted areas in each faculty.

Stage 6: Maintenance

This stage will do regular updating, troubleshooting if there any issues and verification of the system. After all this process is done, the patches will be discharged. Finally, this phase will deliver the final product or system in customer environment.

DISCUSSION

This system is created and tested in a controlled campus environment and the participants are UPM students. The door access control system using PutraVID will be created in a way where all the vulnerabilities and problems has been addressed. The system should also be developed and documented in a way that when the system is passed to the next person in charge, it is easy for them to understand the system to conduct continuous maintenance and upgrades. There are three modules to be developed in the system which are Data Validation Module, Data Verification Module and System Management Module as depicted in Figure 4.

The overall process of the system starts with a QR code image, generated from the PutraVID and the user scan their QR-code using a QR-Code scanner which will be checked in the main database of UPM for authentication permission using Internet of Things (IoT) through PutraVID API. The costly installment of QR code generator is avoided as the existing PutraVID application generates its own QR code for each of the user. If the QR-Code is invalid, then the access will be denied, or the access will succeed and the door will be unlocked as the permission will be granted. At the meantime, the time of access will be recorded in a centralized database where it will be monitored and administrated by authorized people through web server.

In this paper we have presented the approach of centralized database that enables communication between faculties. IoT-based embedded system is used in the making of the QR code scanner helps to connect the scanner to the WiFi and share data in one centralized database. The centralized database used is accessible by each of the faculty administrator. Hence students from one faculty that is required to attend classes in another faculty can still be able to be present to the class. The centralized database also helps to detect pattern and trends in access as there might be peak hours in weekdays where the traffic of movement in and out is higher compared to other time in a day. It is during the peak hour where criminal activity tends to happen. Therefore, the records will be helpful in order to monitor suspicious activities across the premises.

There will be several users of the web server such as students, staff, faculties and database administrator. A different user has different permission and roles assigned to them. Students can view only their access time into the property, while staff can view the number of students who have accessed the door of certain property such as the laboratory. Faculty act as the administrator that can view all the access made throughout the day at the faculty including students and staff to monitor but

they can't delete any queries. Database administrators are the ones that will assign the roles for each one of them and the only one that can update and delete the queries.

The data stored in the centralized database is the timestamp of the user checking in and out of the property. The system is able to produce various reports printed and shared, to be utilized for faculty management and reference purposes as well. Maintenance staff and general workers who are outsiders, will be required to get assistance or verification from the faculty in order to access the premises. They are not provided with QR Code to prevent any security violence in future, also the existing PutraVID system is specific for students and staff only. This entire ecosystem prevents unauthorized access into the premises including forged verification. The potential of this system can be expanded for future improvisation.

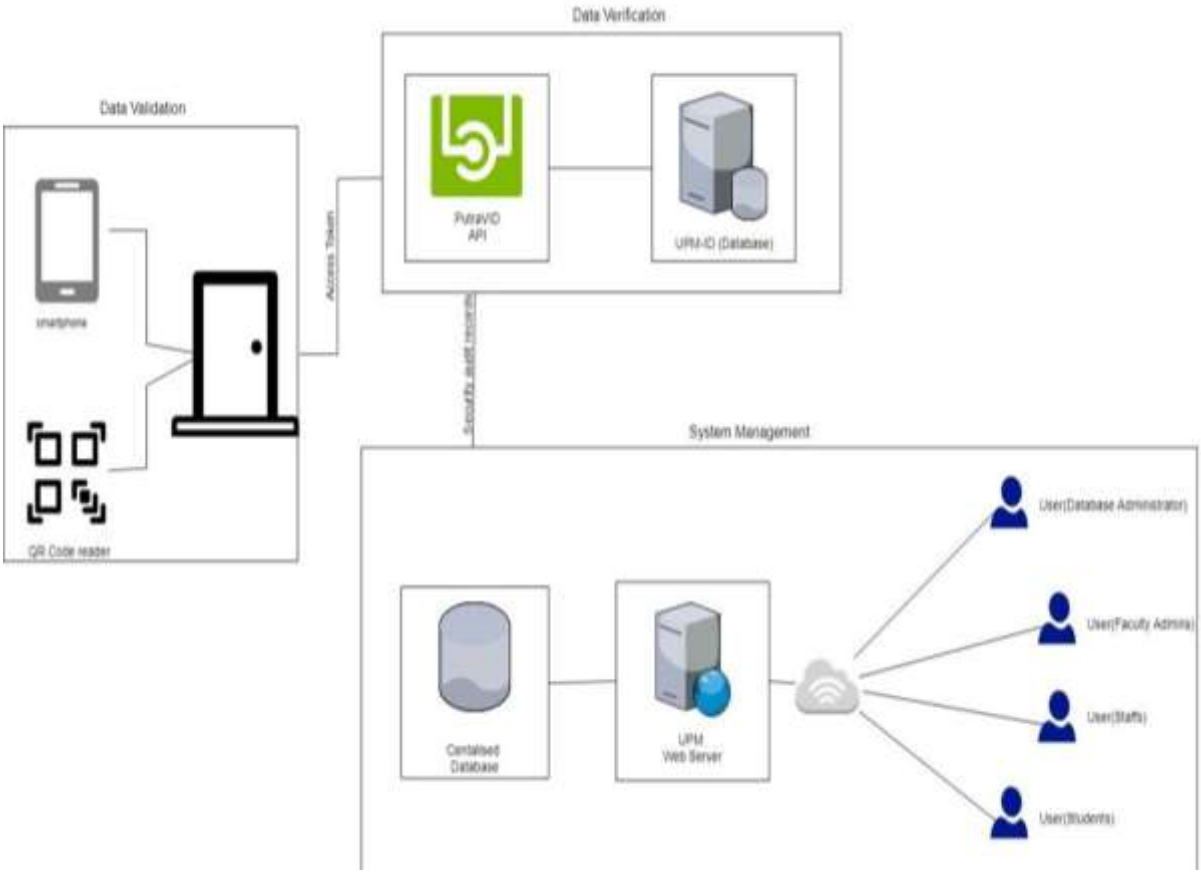


Figure 4: Proposed System Architecture

FUTURE WORKS

This paper is mainly used to fill the gap found in the previous researches such as lacking of researches in combined discipline of embedded system, database management and computer system. Yet, there are few possible improvements for our proposal in future such as:-

i. Using the timestamps as their proof of attendance and to record the working hours of employees in large organization
   This system can be improvised and used in large organizations in order to keep track of their employee's movement in and out and also to make sure to be a proof of the working duration. The process has much more integrity as the smartphones cannot be passed around to fraud the timing.
ii. Integrate technology of blockchain with the access control system
   Since blockchain is decentralized, the distributed ledger which will be built on the system will provide security and trust. Transactions can be recorded chronologically using cryptography and each one time-stamped and linked to the previous one makes it hacker-proof. The chain form makes is hard for manipulation because the entire chain link need to be identified and manipulate. It has the potential to improve data integrity and digital identities for safer IoT devices which eliminates data being tampered. It enables the end user to be in control of the data.

CONCLUSION

Implementing Door Access Control System using PutraVID can bring numerous benefits to Universiti Putra Malaysia including providing a secure space and a step forward to encourage smart campus. This paper has discussed the vulnerabilities of the system and how to overcome those vulnerabilities by improvising the security features of the system. This system will utilize the data gathered by providing a space to manage and maintain the data. It has made the life on campus easier as the students and staff does not have to bring cards everywhere there go as they can use their smartphones to gain access to the faculties and buildings. This system is also one way of utilizing the existing PutraVID application as the application has already provided the users with their very own QR codes and scanner. Hence, the cost can also be reduced as there will be no need for card purchasing in future as they can use the smartphones and in case they lose their phone, they can still use other phones to log in to PutraVID and gain access to the facilities.

REFERENCES

Sung, S. et al. (2015). Security Analysis of Mobile Authentication Using QR-Codes. *Computer Science & Information Technology (CS & IT).*

Kao, Y.W. et al. (2011). Physical Access Control Based on QR Code. *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*.

Matulevičius, R. and Lakk, H. (2015). A Model-Driven Role-Based Access Control for SQL Databases. *Complex Systems Informatics and Modeling Quarterly, no. 3*.

"Mobile Access Control". *Kisi*, Accessed on 13 July 2019. [Online] Retrieved from: https://www.getkisi.com/guides/mobile-access-control.

"OVERVIEW." *Matrix Security Solutions - Employee Attendance & Surveillance SecuritySystem,* Accessed on 13 July 2019. [Online] Retrieved from: http://www.matrixaccesscontrol.com/solutions-qr-code-based-access-control.html.

Voon, M.J., Yeo, S.M. and Voon, N.H. (2016). Campus Access Control and Management System. *ResearchGate.*

Kim, Y.G. and Jun, M.S. (2011). A Design of User Authentication System Using QR code Identifying Method. *International Conference on Computer Sciences and Convergence Information Technology (ICCIT).*

AN APPROACH TO SECURE EPUB EBOOK CONTENT

Yunus Parvej Faniband[1], Aslam Nandyal[2], Iskandar Ishak[3], Fatimah Sidi[3], Marzanah A. Jabar[3]
[1]King Fahd University of Petroleum and Minerals, Research Institute, Center for Communications and IT Research Dhahran-31261, Saudi Arabia
[2]St. Francis Institute of Technology SFIT Borivali-west, Mumbai-400103, India
[3]Faculty of Computer Science and Information Technology, Universiti Putra Malaysia
yunusrf@gmail.com, iskandar_i@upm.edu.my
corresponding author: iskandar_i@upm.edu.my

Abstract: EPUB is an industry standard e-book format. Digital content owners and producers use Digital Rights Management (DRM) to protect and regulate the mechanism of distribution and usage of those digital assets. The features of content protection may range from limiting simple copying or transferring the content, and enabling reading to an authorized reader hardware/software, to complex enterprise distribution policies involving user and device authentication for a specific duration. Commercial E-Publishing services offer to help authors to self-publish or distribute their eBooks with DRM on agreed royalty terms. These E-Publishing services mandate extensive guidelines and procedures to make the content readable and use a proprietary format with DRM. Since International Digital Publishing Forum (IDPF) has no standardized algorithm for copyright protection of EPUB ,there is a challenge to secure the EPUB content following the copyright recommendations. In this paper, we present an overview of the proposed EPUB Content protection and control system and explain a mechanism to protect the EPUB content in a cross-platform open source EPUB editor. We have implemented a prototype of the system using the 'Sigil' EPUB authoring tool.

***Keywords*:** ebook, secured, EPUB, digital publication, encryption.

INTRODUCTION

Usage of E-books is on the rise and they are becoming serious competitors to the classically printed books. E-books offer valuable features, such as embedded audio and videos, search functionality, and built in dictionary, which are absent in classic printed books. EBook file can be read using a dedicated eReader hardware or eReader software. With the availability of appropriate reading software, every desktop computer, laptop, handheld, or mobile phone can function as an eReader device (e.g., Kindle reader). The most widely used eBook formats are AZW (Kindle Format 8), MOBI (Mobipocket or PRC, based on the Open eBook standard), LRF (Sony broadband BBeB), iBook (Apple) and EPUB. Some of the formats support DRM (Digital Rights Management) restrictions that prohibit reading eBooks on other devices or reader software, or sharing eBooks with other people. EPUB [1] is a free and open eBook standard created by the International Digital Publishing Forum (IDPF) [2], based on a variety of other technologies and standards like Open eBook and XHTML. EPUB is a reflow-able format in which the content can adapt its presentation to the screen. The page size remains the same when the user increases or decreases the font size. Due to the availability of digital eBook stores on the Internet, digital books can be easily downloaded into devices. However the risk of illegal copying, piracy and transferring/copying digital assets without authorization has resulted in a need to developing a system to secure digital intellectual property. Hence DRM is a system to safeguard copyrighted digital assets and regulate the distribution and usage of those digital assets. Content owners may opt to apply DRM schemes to protect their content being copied, distributed or printed. Adobe Digital Editions is one example of a DRM protection scheme for EPUB books.

With respect to the content of DRM provided in the form of an e-book, such as EPUB-based content, IDPF [2] has no standard enforcing DRM of EPUB-based content, but has recommendations based on XML security standards of the World Wide Web consortium (W3C) [1], [3].

The copyright protection of e-book content, such as EPUB has become problematic because a content provider gives user's rights to be provided with content in stages according to the users, or intends to prevent a non-permitted user from accessing content. However, as mentioned above, IDPF

merely suggests recommendations for copyright protection and has no standardized algorithm for copyright protection. Therefore, there is a high demand for a copyright management algorithm that can be applied to EPUB. There are many methods to implement DRM in EPUB. An existing EPUB copyright protection system employs a method of encrypting a whole '.epub' file using one encryption key or separately encrypting components of an '.epub' file using a single encryption key. Therefore, it is impossible to provide content in stages according to access rights.

METHODS

EPUB specification [1] defines the standard of distribution and interchange format for digital content. EPUB specifies the rules to construct and encode the organized and semantically upgraded web assets (HTML5, CSS, images, media and other metadata) and distribute them in the form of a zip archive as an '.epub' file extension. EPUB3, the latest release based on the latest HTML5 standard (supporting audio, video and interactivity) , defines a set of four specifications as listed in Table-2. The details of encryption information of EPUB file are represented in optional 'encryption.xml' file.

This xml file has 'encryption' as the root element and exists within the 'META-INF' directory (Figure.3) at the root level of the container file system. The encryption element contains child elements of type 'EncryptedKey' and 'EncryptedData' as defined in the process for encrypting data and representing the result in XML [3]. Each 'EncryptedData' element describes how one or more files within the container are encrypted. Consequently, if any resource within the container is encrypted, 'encryption.xml' must be present to indicate that the resource is encrypted and to provide information on how it is encrypted.

The proposed system is a lightweight content protection system, which involves the content creation module, and the reading system. When the encryption of e-book content is performed while the EPUB standard of IDPF is satisfied, it is recommended to generate an 'encryption.xml' file [1] containing key information and algorithm information for encrypting contents of the container present in META-INF directory. EPUB Open Container Format (OCF) uses XML Encryption to provide a framework for encryption, allowing a variety of algorithms to be used according to the XML security standards of the W3C [3]. An optional 'signatures.xml' file holds digital signatures of the container and its contents for preventing forgery and falsification of content, and the 'rights.xml' file holds information of DRM. The OCF specification does not mandate the use of designated algorithm for content encryption, e-signature, and rights control, but recommends using a specific algorithm and rights expression language.

The proposed EPUB content protection system allows the content creator to create a protected content and ex-porting (save) it for a secure sharing. The import (reader) module allows the file to be unlocked and renders the content. The system secures individual text files, media files and the associated assets in zip archive with encryption. The proposed implementation based on Sigil follows the embedded control architecture with message- push [4], where the control set is embedded in the digital information and always comes with the digital information within its digital container. Instead of developing the EPUB editor from scratch, an open source version of Sigil EPUB editor (built on Qt, a cross-platform application framework) is enhanced to support the following central aspects of the system, as follows:

1) Content encryption during Export/Save.
2) Content decryption during Import/Read.

Sigil software design allows the system to implement content protection system in two ways:

1) Using Qt C++: This method uses the Qt C++ en-gine and Crypto library to implement the required encryption and decryption mechanism.
2) Developing Python plugin: The content protection system can also be achieved by implementing the Python plugins.

- Qt UI. Sigil is based on Qt, a cross-platform application framework and the Qt UI framework provides a variety of user interface controls such as Widgets for desktop applications.

- Qt WebKit. Sigil has a WYSIWYG editor feature and uses the QtWebKit module for the BookView and Preview windows.

- Python plugin support module. Sigil supports the addition of new features through standalone plugins written according to standard plugin framework [5] [6] supporting python language.

- Third party helper modules. Sigil depends on some of open source libraries for features like HTML5 parsing (Gumbo parser) [7], regular expression pattern matching (PCRE, Perl-compatible regular expression library) [8], data compression and unzip (zlib, minizip) [9], spell checking and morphological analysis (hunspell) [10].

- Crypto. Crypto++ [11] (also known as CryptoPP) is a free and open source C++ class library of crypto-graphic algorithms and schemes which provide complete cryptographic implementations. This module is integrated as a contribution of this work and used extensively using a wrapper API in the proposed EPUB Content Protection System and the evaluation modules of java (using Java Native Interface (JNI)).

- Individual file encryption. The export/save module of the EPUB editor is enhanced to implement an EPUB multiple key encryption method that encrypts individual files with content encryption key (CEK). Maintaining the CEK for individual files helps in supporting simple license issuing, delivery methods and decryption process in client readers. The epub Export module invokes the encryption calls from the Crypto++ library and encrypt each of the allowed individual files and formalizes and repackage the encrypted unit files according to the EPUB standard.

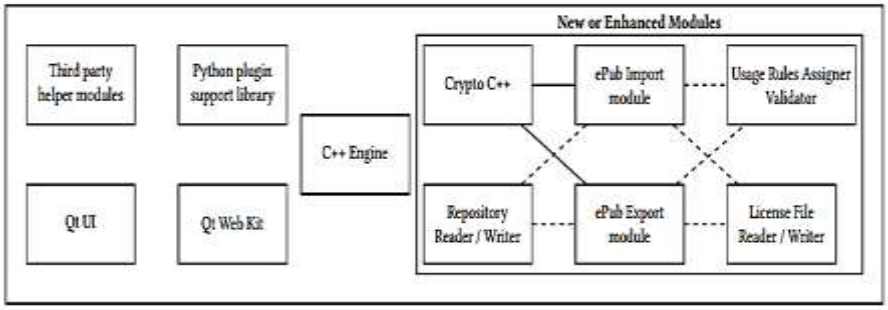


Fig. 1. Epub Architecture

The Encryption process of EPUB content is described below.

1) The export module pops up a dialog to create a pass-phrase for the whole content.

2) The contents of each file are encrypted with a random symmetric key K. According to the EPUB standard [1] the subsequent files are considered as reserved files and must not be encrypted, regardless of whether default or specific encryp-tion is requested (mimetype, container.xml, man-ifest.xml, metadata.xml, signatures.xml, encryp-tion.xml, rights.xml and content.opf)
3) Create a random key and a random initialization vector (IV) of 16 Bytes i.e.(128 bits). A key, in the context of symmetric cryptography, is something that is kept secret. The random IV ensures that each message is encrypted differently.
4) Encrypt the file with a block length of 128 bits and key lengths of 128.

5) Store the content encryption key in a repository with a unique mapping reference name: 'UniqueFilename-Keyfile.dat'.
6) Store the IV in a repository with a unique mapping file name: 'UniqueFilename-ivfile.dat'.
7) Write the identification information of components, key and IV information to the encryption reference XML file

FINDINGS AND DISCUSSION

To evaluate our implementation, the import module in Sigil EPUB Editor itself is enhanced to read the encrypted content. Test content of EPUB-3.0 [12] samples files are created and tested after encryption. Additionally, an Android Java test client is separately implemented. This module tests the files after decryption and writes the output zip EPUB file into a different folder without encryption, for manual evaluation. Any of the general EPUB readers can be used for testing by loading the resulting un-encrypted EPUB file. The java testing module has 480 lines of code. Different types of test data (audio, video, vertical writing, different language content like Arabic, Japanese), involving EPUB3 sample test content [12], are used to evaluate the system. The proposed system can successfully encrypt the content and package the contents into a zip archive. The import module of system can also open the content in a readable format after reopening the encrypted EPUB zip archive.

Fig. 2. Epub Test Suite Application using encryption

CONCLUSION

The main contribution of this work is a proposed system for protecting and controlling the EPUB content. In addition, a prototype is implemented using the open source Sigil desktop Editor, and enhanced to accommodate the proposed EPUB protection system. The implementation is tested with the decryption/reading module in the editor and a separate Android Java client. Our proposed simple system works well for protecting personal content and sharing. In the future, the current implementation can be integrated with an extra licensing mechanism and feature to limit the content reading based on certain rules. The proposed solution can be enhanced further to enable online protection by storing the encryption and other metadata on cloud servers, instead of packaging within the EPUB zip archive.

REFERENCES

[1] "Epub open container format (ocf) 3.0," 2016, [Online; accessed 5-October-2016]. [Online]. Available: http://www.idpf.org/epub/30/spec/epub30-ocf.html
[2] "The international digital publishing forum (idpf) - epub," 2016, [Online; accessed 5-October-2016]. [Online]. Available:http://idpf.org/epub/
[3] "Xml encryption syntax and processing," 2016, [Online; accessed 5-October-2016]. [Online]. Available: https://www.w3.org/TR/2002/REC-xmlenc-core-20021210/Overview.html
[4] J. Park, R. Sandhu, and J. Schifalacqua, "Security architectures for controlled digital information dissemination," in Computer Security Applications, 2000. ACSAC'00. 16th Annual Conference. IEEE, 2000, pp. 224–233.
[5] "Sigil plugin framework," 2016, [Online; accessed 8-October-2016]. [Online]. Available: https://goo.gl/0DkOR1
[6] "Sigil plugin index," 2016, [Online; accessed 8-October-2016]. [Online]. Available:http://www.mobileread.com/forums/showthread.php?t=247431
[7] "Gumbo parser - an html5 parsing library in pure c99," 2016, [Online; accessed 8-October-2016]. [Online]. Available: https://github.com/google/gumbo-parser
[8] "Pcre - perl compatible regular expressions," 2016, [Online; accessed 8-October-2016]. [Online]. Available: http://www.pcre.org/
[9] "zlib - a massively spiffy yet delicately unobtrusive compression library," 2016, [Online; accessed 8-October-2016]. [Online]. Available: http://www.zlib.net/
[10] "Hunspell - spell checker," 2016, [Online; accessed 8-October-2016]. [Online]. Available: http://hunspell.github.io/
[11] "Crypto++ library 5.6.5," 2016, [Online; accessed 8-October-2016]. [Online]. Available: https://www.cryptopp.com/
[12] "Epub 3 samples table," 2016, [Online; accessed 8-October-2016]. [Online]. Available: https://idpf.github.io/epub3-samples/samples.html

SVM-BASED APPROACH FOR DETECTING MISLEADING ONLINE NEWS ARTICLES

Normala Che Eembi[1], Iskandar Ishak[1], Fatimah Sidi[1] and Lilly Suriani Affendey[1]
[1]Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
Selangor, Malaysia
normala.jamil@yahoo.com, iskandar_i@upm.edu.my, lilly@upm.edu.my, fatimah@upm.edu.my
corresponding author: iskandar_i@upm.edu.my

Abstract: Since its existence in the 1990s, online news has been the primary source of news content for newsreaders. Unfortunately, based on several findings, readers tend to judge on specific event based on the news headlines rather than its contents. With the advancement of mobile and web technologies, it is easier to spread the news to others through this medium habits that can cause negative impacts towards individuals, organizations, or nations that are victimized by the news. Therefore, it is an important task to determine the truth about information being spread to the public, such as online news. To solve this problem, multiple methods have been developed to detect misleading online news. In this works, we aim to improve deception detection method on online news based by simplifying the pre-processing and improve features selection techniques to improve the SVM-based deception detection approach accuracy. The experimental results showed that the proposed approach managed to improve the efficiency above 90%.

Keywords**:** veracity, deception, truthfulness, accuracy, headline, content, online news.

INTRODUCTION

Based on the negative impact of the spreading of misleading news, it is crucial to have a proper deception detection technique for online news. The problems happen when readers cannot find differences between true or false of news stories in online news. The catchy headline of the news story also can make people confuse while the content is dissimilar. In this paper, we present an improved approach for deception detection based using Support Vector Machine. The goal of this research is to improve the accuracy of deception detection over online news.

To date, several studies have explored the identification of deceptive content in online news (Pérez-Rosas & Mihalcea, (2015), Chakraborty et al., (2016), Anand, Chakraborty, & Park, (2017)). However, the previous works focus only on detecting deception on online news based on the topic rather than focusing on the headline. Unlike previous studies, this research targets the identification of false online news by combining headlines and content.

We improved the pre-processing part where lemmatization is introduced. Lemmatization is an approach where the process of grouping the inflected forms of a word so they can be analysed as a single item, identified by the word's lemma, or dictionary form. Different than stemming method, lemmatization depends on the identification of an intended part of speech and meaning of a word in a sentence correctly, as well as within the larger context surrounding that sentence, such as neighbouring sentences or the entire document. In this paper, we also compared the available deception detection approach to improve deception detection accuracy for online news veracity. Through experiments, we compared the performance of classifiers using Support Vector Machine (SVM), Stochastic Gradient Descent (SGD), Logistic Regression (LR) and k-Nearest Neighbour (k-NN).

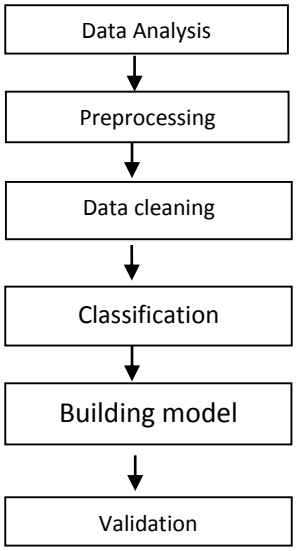


Fig. 1. Research methodology

METHODS

In general, four main research stages have been carried out throughout this study (Figure 1). The first stage of the research stage is data analysis, where it consists of two mains research activities involving literature review and selection of public datasets. In building the research problem, the gaps and problem statements are gathered from the previous works. The analysis of the real-world setting datasets is then studied and selected at this stage. Pre-processing at the second stage involves the activities of feature extraction of the public datasets for fake news datasets. Data cleaning and other feature selection are required in these stages. Classification is at the third stage involves the classification activities using deception learning technique with different base classifiers. At the fourth stage, the build model is predicted based on the deceptive classification technique to cater to the research problems. At the final stage, the validation stage is carried out to analyse the performance of the framework with different base classifiers as well as compare the proposed methods (Figure 2) with benchmark works. This final stage would thus; the complete activity of the research activity at the same time discusses the achievements based on the research objectives.

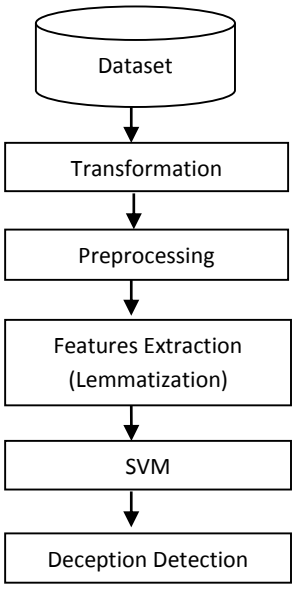


Fig. 2. The proposed deception detection approach

FINDINGS AND DISCUSSION

Several experiments have been carried out for this study where Fake-Real News (FRN) Dataset (RalucaChitic, 2018) and News Article (NA) Dataset ("Fake News | Kaggle," 2018). Dataset has been selected to validate the hypothesis. The experiment is to determine the type of fake or truth of online news. The experiments are conducted using 80% of the dataset for training and 20% for testing with five-fold cross-validation. Subsequently, Deception Classification Technique is tested under four different types of base classifiers, namely Support Vector Machine (SVM), Stochastic Gradient Descent (SGD), Logistic Regression (LR) and K-Nearest Neighbour (k-NN) are applied at the training stage. The parameter is set up under Linear Kernel SVM with probability TRUE and C=5. SGD used loss parameter 'hinge,' penalty l2, alpha, and tol = le-3, and maximum iteration 1000. LR used parameter solver 'lbgs' and maximum repetition of 1000. The initial steps are performed using the same parameter setting, and features extraction.

This experiment implemented to present the performance of the proposed method to retrieve specific types of deception for online news. The label set L to consist of 2 label types of Fake and True News. The features are basic text extracted from the real-world setting of the dataset.

The accuracy of false news is described in Table 1. Specifically, this method can define one-by-one efficiency per name of online news. It can address the type of report and which online news. This proposed method manages to send the particular two examples of a story between ranges of 0.001-0.990 of accuracy per label.

Table 1: Accuracies of deceptive news

Dataset	Features	SVM	SGD	LR	k-NN
Fake Real News	Base (TFIDF)	70.26%	62.55%	70.5%	71.54%
	Base + Lemmatization	93.67%	91.29%	91.38%	81.67%
	Base + Syntactic	67.03%	62.77%	67.32%	71.13%
	Base + Bigram	90.33%	86.62%	88.22%	77.74%
	Base + Lemma + Syntactic	93.41%	90.45%	90.29%	66.53%
	Base + Lemma + n-grams	90.33%	86.35%	88.22%	77.74%
	All	90.33%	81.29%	85.14%	74.86%
Fake News Article	Base (TFIDF)	94.29%	73.45%	78.73%	77.61%
	Base + Lemmatization	98.17%	95.50%	96.76%	86.36%
	Base + Syntactic	79.49%	62.85%	78.68%	77.20%
	Base + Bigram	99.35%	94.46%	96.54%	87.99%
	Base + Lemma + Syntactic	97.76%	91.59%	94.51%	81.49%
	Base + Lemma + n-grams	98.70%	94.54%	96.54%	87.99%
	All	98.52%	87.27%	92.90%	85.06%

Overall, using SVM, news article recorded more than 90% of the accuracy. Based on the different classifier above, SVM is the best classifier for this model.

Performance Measures

Table 2 presents the measures of precision, recall, and F1 with associated five-fold cross-validation results for our deception detection model. Each table represents different data. Precision was highest when lemmatization included. The bigram provided improvement for the outcome and showed the most top recall performance. The F-Score was maximized in the case when lemmatization and syntactic is used.

Table 2: Performance Measures

Datasets	Features	Precision	Recall	F-Score
Fake Real News	Base (TFIDF)	0.73	0.72	0.72
	Base + Lemmatization	**0.94**	0.94	0.94
	Base + Bigram	0.92	0.92	0.92
	Base + Syntactic	0.70	0.67	0.67
	Base +Lemma + Bigram	0.92	**0.92**	0.92
	Base + Lemma + Syntactic	0.95	0.95	**0.95**
	All	0.92	0.92	0.92
Fake News Articles	Base (TFIDF)	0.95	0.95	0.95
	Base + Lemmatization	**0.98**	0.98	0.98
	Base + Bigram	0.99	0.99	0.99
	Base + Syntactic	0.80	0.80	0.80
	Base +Lemma+ Bigram	0.99	**0.99**	0.99
	Base + Lemma + Syntactic	0.98	0.98	**0.98**
	All	0.99	0.99	0.99

CONCLUSION

From the research that we had done, we can solve the problem of misleading news and analyze the method of machine learning so that we can build our model and test it with any online news. Our result shows above 90% accuracy. In term of machine learning, Support Vector Machine is the best choices for our results. The result can further be improved with the help of other features. We compare our result with Rubin paper where their solution for satirical news; meanwhile, we choose online news as our problem solver.

REFERENCES

Anand, A., Chakraborty, T., & Park, N. (2017). We used neural networks to detect clickbaits: You won't believe what happened next! *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10193 LNCS*, 541–547. https://doi.org/10.1007/978-3-319-56608-5_46

Chakraborty, A., Paranjape, B., Kakarla, S., & Ganguly, N. (2016). Stop Clickbait: Detecting and preventing clickbaits in online news media. *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2016*, 9–16. https://doi.org/10.1109/ASONAM.2016.7752207

Fake News | Kaggle. (2018). Retrieved March 15, 2019, from https://www.kaggle.com/c/fake-news/data

Kinmont, K. (n.d.). Fake News Detection in Twitter. *Projects.Cs.Uct.Ac.Za*. Retrieved from http://projects.cs.uct.ac.za/honsproj/cgi-bin/view/2017/karodia_kinmont_lu.zip/resources/literatureReview-KNMKRI002.pdf

Pérez-Rosas, V., & Mihalcea, R. (2015). Experiments in Open Domain Deception Detection. *Proceedings Ofthe 2015 Conference on Empirical Methods in Natural Language Processing*, (September), 1120–1125.

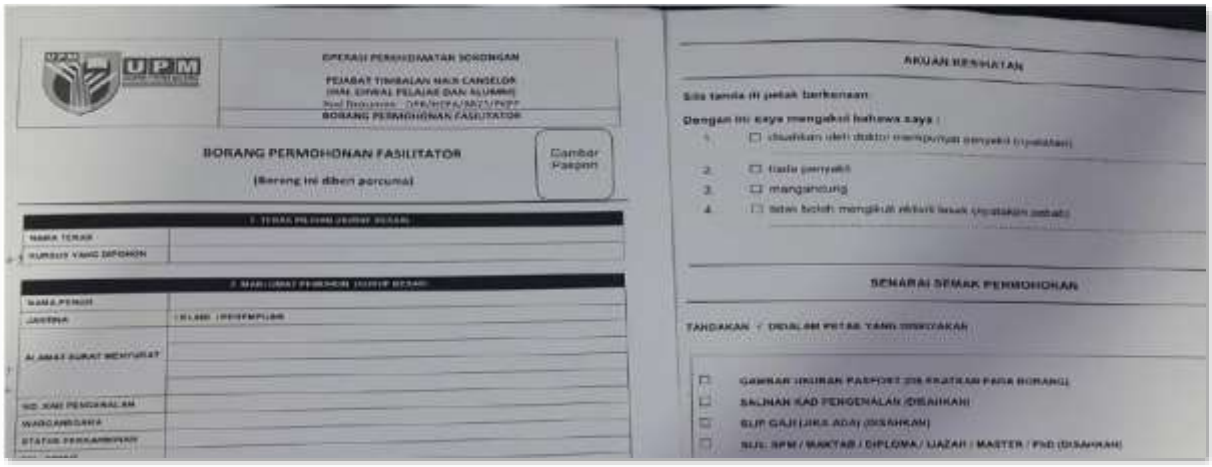
RalucaChitic. (2018). real_or_fake | Kaggle. Retrieved March 15, 2019, from https://www.kaggle.com/rchitic17/real-or-fake

FACILITATOR APPLICATION MANAGEMENT SYSTEM FOR STUDENT CO-CURRICULUM AND STUDENT DEVELOPMENT CENTER OF UPM

Iskandar Ishak[1]
[1]*Faculty of Computer Science and Information Technology, Universiti Putra Malaysia Selangor, Malaysia*
iskandar_i@upm.edu.my

INTRODUCTION

Facilitator job application in for Credited Co-curricular courses in UPM is handled by the UPM's Student Co-curriculum and Development Center (SCDC). Currently, the job application management process is using the manual and web-based platform. Paper-based approach tends to produce errors in terms of data quality, data validity as well as data restoration as it depends on hardcopy forms. Web-based facilitator application for SCDC can be the solution to have a better facilitator application management, reduce data error and can store data for future purposes.



METHODOLOGY

The methodology that chosen to be use in this project is Waterfall Model. It is a linear sequential model in which all the phase is organized in linear. Each phase has to be complete first before the next phase can start. This model is chosen because it is simple to understand and the requirement is clear and well known as the main objective is to automate the existing manual application process.

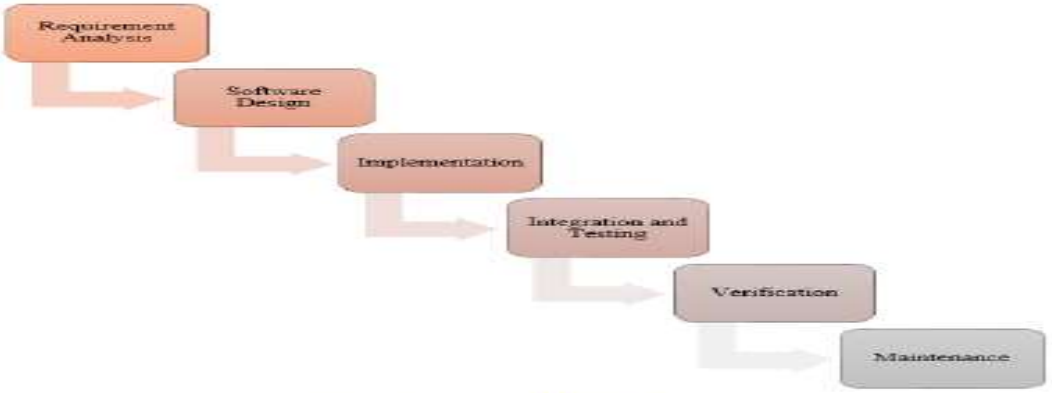


Figure 8: Waterfall Model

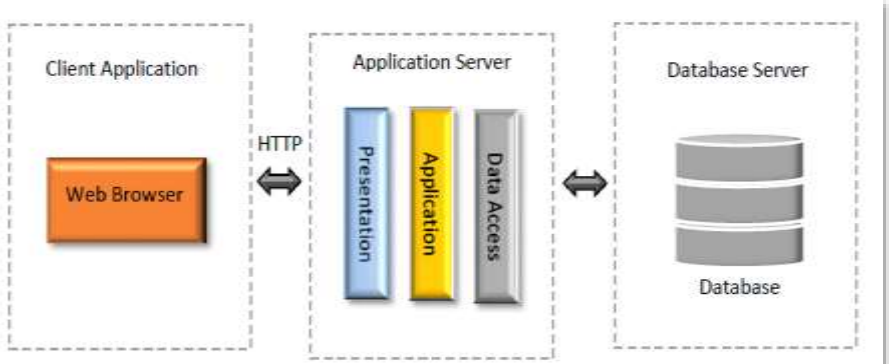
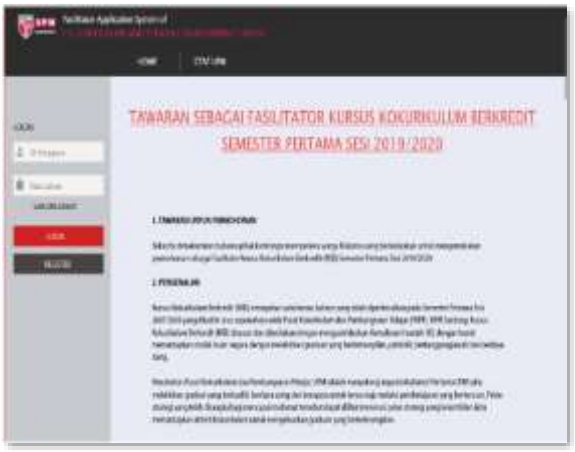
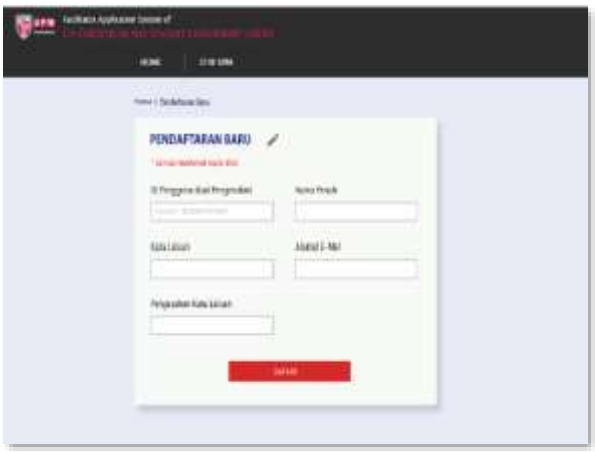


Figure 9: System Architecture for Facilitator Application Management System

Facilitator Application Management System for Co-Curriculum and Student Development Centre (PKPP) is a web-based application that follow the three-tier architecture. It consists of three different tier such as presentation tier, application tier and data tier that have different type of functionalities. Presentation tier is the top most level which handle the interaction with user by providing the user interface. The application tier contains set of rule for information processing and business logic while the data tier consists of database and database management system.





CONTRIBUTIONS

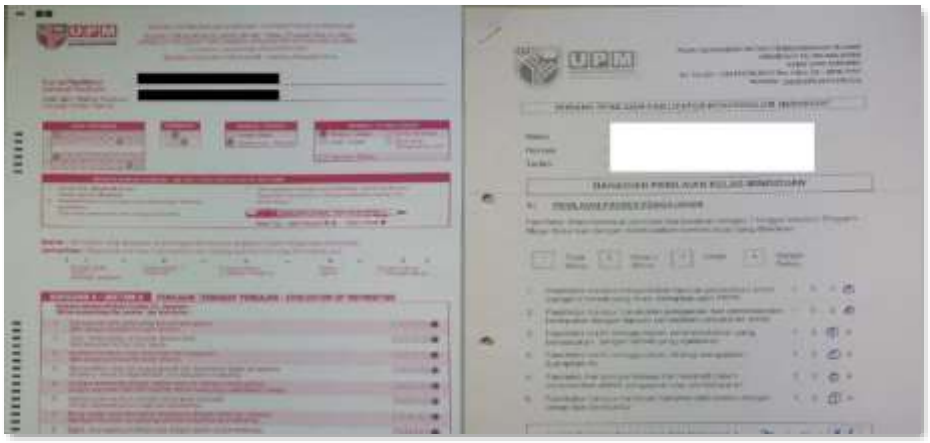
This projects aim to improve the facilitator application process through paperless approach in which all information is stored in a database. Easier reporting and data management tasks can reduce time for CSDC UPM to manage data related to managing facilitator application of credited co-curricular teachings in UPM**.**

FACILITATOR EVALUATION SYSTEM FOR STUDENT CO-CURRICULUM AND STUDENT DEVELOPMENT CENTER, UNIVERSITI PUTRA MALAYSIA

Iskandar Ishak[1]
[1]*Faculty of Computer Science and Information Technology, Universiti Putra Malaysia*
Selangor, Malaysia
iskandar_i@upm.edu.my

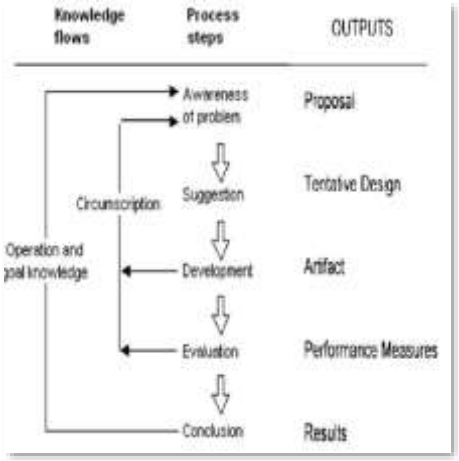
INTRODUCTION

In Universiti Putra Malaysia (UPM), the performance of a facilitator is always being monitored under Co-Curriculum and Student Development Centre to ensure that an optimal standard of facilitating is achieved by eliminating those facilitators who failed to comply with the standards. The current evaluation is done through manually form-filling by both supervisor and student based on different set of questionnaires prepared and printed by CSDC. the benefits of having a web-based facilitator evaluation system outshine its disadvantages in hoping to deal with the troublesome manual form filing and storing of files. As such, this project is here to study and develop a reliable alternative solution for the staffs of CSDC, supervisor and student.



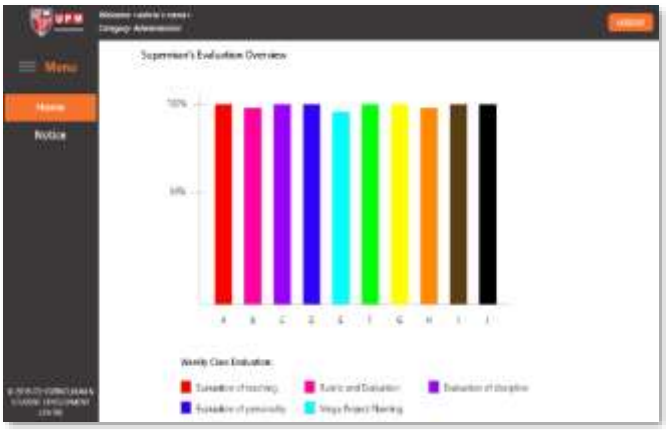
METHODOLOGY

A multitier or client-server architecture with three layers will be used to develop the proposed system. As illustrated in the diagram above, the system consists of user interface layer, application logic layer and database layer. The separation of tiers allows easier modification and addition of layers in future development without the need to rebuild the entire system**.**



RESULTS AND SUMMARY

In summary, the web-based facilitator evaluation system should provide convenient way for both supervisor and students to evaluate the facilitator based on each class respectively. On the other hand, the staff at Co-Curriculum and Student Development Centre should be able to review and analyse the evaluation form submitted by both supervisor and student easily.



CONTRIBUTIONS

1. The system is able to generate questionnaires based on the type of users.
2. The system is able to compute the scores for each facilitator based on the questionnaires completed by the users.
3. The system is able to store the evaluation record for each facilitator.
4. The system is able to generate overview of the result of the evaluation based on each facilitator.

MANAGING DYNAMIC AND ADAPTIVE CHARACTERISTICS IN SOFTWARE PROJECT MANAGEMENT

Marzanah A. Jabar[1], Norhayati Mohd. Ali[1], Yusmadi Yah Jusoh[1], Salfarina Abdullah[1] and S.Mohanarajah[2]
[1]Faculty of Computer Science & Information Technology,
University Putra Malaysia, 43400, Serdang, Selangor, Malaysia
Faculty of Computer, Engineering and Technology,
[2]Asia Pacific University, Technology Park, Bukit Jalil, 57000, Kuala Lumpur, Malaysia
marzanah@upm.edu.my, smohan@apu.edu.my
corresponding author: marzanah@upm.edu.my

Abstract: There has been a growing interest in the usage and adaptive ramifications within the broad spectrum of agile methodologies and its current trend as an alternative to traditional methodologies has not been widely accepted. Whilst a lot of hype has been seen since its introduction in 2001 with the Agile manifesto where delivery success was promised, software practitioners still appear to be weary as unfavorable usage perceptions dominate the software industry and the transition appears to be a formidable task. Project success and failure scenarios have been researched but the factors that establish these are still widely inconclusive and elusive creating more interest and emphasis for embarking on further research. This paper highlights and investigates the relationship between agile software development methodologies and the success and failure scenarios prevalent in the software industry.

Keywords: Dynamic And Adaptive Characteristics, Software Project Management, Traditional method, Agile method

INTRODUCTION

The software industry initially had its roots with the use of traditional methodologies for ensuring successful deliveries. Towards the latter half of the 20th century the concept of agility as being a flexible approach made its appearance although its roots date far back as the 1930s (Tins Karrbom Gustavsson et al., 2013). This was interesting as agility, although caught on much later, was a concept that linked closely to quality (used by IBM and NASA) while the traditional methodologies still focused and the triple constraints (Scope, time and cost – PMBOK Body of Knowledge) to ensure quality all this while. It could be that agile methods were a culture shock as it passed control to the operative team and project managers felt left out or in fact lost an important role in managing projects. Tremendous amount of research has been made on critical success and failure factors but not many provide the interconnection between these factors (Timo O, A. Lehtinen et al., 2014). Our goal is to provide a set of critical success and failure factors that where the applicability of the factors are constrained and limited within certain boundaries which can take inputs to tailor the processes to facilitate the needs of the project, building interconnect bridges and deeming it to achieve success.

METHODS

The methodology used is the Evidence-Based Software Engineering (EBSE) research and practice. Additionally, the Systematic Literature Review (SLR) method was also incorporated as it provided a rigorous review and synthesis of research results. In analyzing the reviewed data, we have used the Analytical Hierarchy Process (AHP) method as it provided a clear approach in attaching priorities to recommended solutions. In developing the theoretical and conceptual framework for the analysis of the research material, which was essentially linear, as our concern was that our framework was described and motivated by existing literature. In total 82 academic papers were reviewed and about 54 were found to be useful in developing a theoretical and conceptual framework.

RESULTS

As the discussion on the research questions and the analysis covers a wide area of motivation factors and objectives a summary is provided to consolidate the various analysis, findings and the outcomes that could be used as further research.

Research Questions	**Results of Analysis of the Literature**	**Outcomes and new proposed research areas**
RQ1. Have the problematic definitions of software project failures affected agility as the perceptions of success have not been uniformly viewed and interpreted by different stakeholders?	The definition of success amongst stakeholders being satisfied appears to be a universal derivative of success and to be satisfied could involve a multitude of perceptive inclinations. The idea of mental models (Xiaodan et al., 2014) with teams brings in an important perspective. Teamwork does not automatically arise and to use this as a decisive input to success could relate back to the original work done on the principles of the agile manifesto where research on the relevance of teamwork to agility was not explored (Stephen Wood et al., 2013).	A study of mental models with a link to the agile manifesto as an important new perspective to be used to streamline the decisive input to formulate a uniform definition of success.
RQ2. Are current identified success and failure causes grouped into typical broad areas of categorization and standardization and do they work a set of ranked priorities (AHP)?	Major success factors are leadership, perception, methodologies, KPIs, strategic policies, effective communication and leadership with the most important being stakeholder management. The evidence also suggests that we should try to develop new critical success factors or try to model the success of agile projects with different methods, (e.g. the Analytical Hierarchy Process – AHP) and provide a ranking perspective in terms of importance and criticality by associating weights to these factors. Establishing a clear set of critical success causes using the AHP method would require a standard use of the terms in identifying and developing the critical factors with an established hierarchical structure.	To use categories, factors, variables, attributes and performance indictors in a hierarchical structure as key distinguishing terms or elements to arrive at the success and failure causes to facilitate the effort. This should be used for traditional and agile methods. AHP and RCA should be used to determine the ranking.

RQ3. What potential failure causes identified from previous studies still exist that need to be transformed to success causes within the spectrum of agile methodologies to make it more development and implementation centric?	Studies suggest that failure causes related to customer collaboration, business core values and development principles need further exploration and research while integrating the advance knowledge on themes beyond what we already know (Sun-Wen Chuang et at., 2014). Other studies indicate that the nature of agile teams (small teams and its self-organizing nature - a feature of the first of the agile values in the manifesto) to function effectively encounter barriers as changes to corporate policies and procedures are not accepted easily (Jean Binder et al., 2014).	Focus should be in identifying new failure causes and not in improving causes already researched. Barriers with corporate cultures on the acceptance and use of agile would be another crucial area.
RQ4. What are the most commonly identified success and failure causes failures within the pool of agile methodologies and are there any similarities when compared to traditional methodologies?	The importance of agile methodologies is being recognised as the Project Management Institute (PMI) has included the bridging of traditional and iterative approaches. Evidence of stakeholder's satisfaction always seems to be the highest ranked in terms of success and/or failure causes in most studies which is beyond the golden triangle (Meghann L. Drury-Grogan, 2014). Iteration pressure (Rashinda Hoda et al., 2013) is gaining momentum and with agile is replacing waterfall, increased adoption levels, reduced inhibiting factors and improved benefit levels the maturation of Agile Software development Processes is seen (David Bustard et al., 2013).	New research using the 'golden triangle' from traditional methodologies to be included with the stakeholder's management process in the hierarchical philosophy for agility. This further supports and reinforces outcomes from research objectives 1, 2 and 3.
RQ5. What traditional and agile processes currently support the mapping of the agile manifesto with its values and principles to embrace current success and failure causes?	The use and direct link of the critical success and failure factors to the 12 principles appears to be relatively non-existent (only 1 research paper in 2014) in recent publications of research articles although they focused on agile methodologies.	New studies and research to be done not just to map the agile values and principles to agile and traditional methodologies, but also to include specific success and failure causes.
RQ6. How do we use the mapping to propose a framework to develop a dynamic and adaptive hybrid version of an agile methodology?	Support for the development and use of hybrid methodologies as new area of research (Sun-Wen Chuang et al., 2014). Further support is also seen to combine both traditional and agile methodologies in one project and take advantage of the relevant components and tailor it (Mario Spundak, 2014).	Research on bridging techniques that incorporates the traditional and agile methodologies should be done and investigated to develop the agile hybrid framework and methodology.

Table 1: Mapping research questions with objectives and motivation factors.

The new areas proposed for research should be looked as new pilot studies where hypothesis are formulated and tested through empirical research.

CONCLUSION

Success and failure scenarios are the mainstream derivative constructs of project delivery studies in terms of the moving forward in achieving higher project success rate. Issues with stakeholders were the major areas of concern as success and failure contributors. Other important factors were organizational culture and methodology, which needed equal attention as gaps still prevailed and at times festered to its own disadvantage. The most interesting find was that these contributors were equally important whether we used traditionally formal methods or moved to agilitic methods. This also explains the flat success rate for software projects over the last 5 years as we probably reached a point where we have been addressing areas in terms of causes and were not specifically trying to weed out the areas independent and applicable to the software industry. We recommend that a dynamic approach both to the causes and the methodology is the answer as the software industry does not sit still in todays' ever changing user demands and expectations. Environments with a dynamic methodology, which encompasses a dynamic set of critical success causes as its inception based on a set of dynamic success indicators, is the key answer. The dynamic success (and failure) factors, indicators and the derived methodology are recommended from our study. This we feel should be the approach for all new and future research.

The new areas of research in this study relate closely to dynamic and adaptive agile approaches and this will be investigated further.

BENEFITS

The study has contributed in proposing a new set of success and failure causes and the roots to a new agile methodology from the analysis and findings. These would be useful for researchers and software practitioners who are interested to do studies on the further adoption and use of existing agile methodologies or to tailor agile methodologies as hybrids versions in the future. Like all new contributions in any field of study, improvements will be identified and new areas for research will be the focus. Some might go further to complete the initial aims and objectives identified at the start of the study while others might additionally also branch to new concepts and theories. This current research has done both and in the six areas it has contributed.

First, an overall trend analysis over the last 18 years based on Standish report on IT project success and failure rates and the usage of agile methodologies over the last decade. Definitions of project success has been varies due to different perceptions by project stakeholders.

Second, a review of the recent articles published from 2010-2014 has provided a trend and the cause of project success and failures. A compilation of an exhaustive list of the success and failures causes has been provided from the analysis, these were grouped in various broad categories to facilitate a root analysis perspective.

Third, a theoretical and conceptual framework used for the study.

Fourth, a hierarchical structure (AHP) of project elements presented as categories, factors, variables, attributes and indicators to support a root cause analysis (RCA) of success and failure causes to improve our understanding on the successful adoption, use and implementation of Agile in software projects.

Fifth, a new set of success and failure causes with an appreciation of the barriers and obstacles that has been overcome and its link to the agile values and principles.

Sixth, the tools that can be used to develop a new agile methodology or a hybrid version

Seventh, a proposed framework has been developed from the gaps that have been seen as new research areas from the analysis and findings of the 6 research questions. The new research problem areas will be formulated as a set of hypotheses to be investigated and tested. Agile characteristics has not assisted software development methodologies to reach a dynamic set of critical success causes that can be used to fit a broad range of projects (simple to complex) for successful deliveries. Thus these views lead back to the questions on the definition of success, the need for critical failure causes and the development of an adaptive and dynamic methodology which was the original basis and roots of our research.

REFERENCES

Alas Martakis, Maya Daneva, 2013. Handling Requirements Dependencies in Agile Projects: A Focus Group with Agile Software Development Practitioners, IEEE 7th International Conference on Research Challenges in Information Science (RCIS)

Antonio G. Sanjuan, Thomas Froese, 2013, The Application of Project Management Standards and Success Factors of a Project Management Assessmsnt Tool, Procedia, Social and Behavioural Sciences.

Barbara Kitchenham, Pearl Brereton, David Budgen, Mark Turner, John Bailey, Stephen Linkham, 2009, Systematic literature review in software engineering - A systematic literature Review, 2013, Software Engineering, Information and Software Technology journal, ELSERVIER.

Barbara Kitchenham, Pearl Brereton, 2013, A Systematic Review of Systematic Review process research in Software Engineering, Information and Software Technology journal, ELSERVIER.

Brendan Murphy, Christian Bird, Thomas Zimmermann, Laurie Williams, Nachiappan Nagappan, Andrew Begel, 2013, ACM / IEEE International Symposium on Empirical Software Engineering and Measurement.

David Bustard, George Wilkie, Des Greer, 2013, The Maturation of Agile Software Development Principles and Practice:Observations on Successive Industrial Studies in 2010 and 2012, 20th Annual IEEE International Conference and Workshops on the Engineering of Computer Based Systems (ECBS).

Dragan Stankovik, Vesna Nikolic, Miodrag Djordjevic, Dac-Buu-Cao, 2013, A Survey study of critical success factors in agile software projects in former Yugoslavia IT companies, The Journal of System Software, ELSERVIER.

Farzana Asad Mir, Ashly H Pinnington, 2014, Exploring the value of Project Management: Linking Project Management Performance and Project Success, ScienceDirect, ELSERVIER.

Frank K. Y. Chan, James Y.L. Thong, 2008, Acceptance of agile methodologies: A critical review and conceptual framework, Decision Support Systems, ELSERVIER)

Gilles Garel, 2013, A History of Project Management Models: From re-models to the standard models, SciVerse ScienceDirect, ELSERVIER.

Jean Binder, Leon IV Aillaud, Lionel Schilli, 2014, The project management cocktail model: An approach for balancing agile and ISO 21500, Procedia - Social and Behavioral Sciences, ELSEVIER.

Jim Sheffield, Jlien Lemetayer, 2013, Factors associated with software development adility of successful projects, International Journal of Project Management. ELSEVIER.

Jonas Soderland, Sylvain Lenfle, 2013, Making Project History: Revisiting the Past, Creating the Future, International Journal of Project Management.

Kate Cavis, 2014, Different Stakeholder groups and theor perspective of project success, ScienceDirect, International Journal of Project Management, ScienceDirect, ELSERVIER.

Lavagon A Ika, Amadon Diallo, Denis Thuillier, 2012, Critical Success factors for World Bank projects: An empirical investigation, ScienceDirect, International Journal of Project Management, ScienceDirect, ELSEVIER.

Leo Vijayasarathy, Dan Turk, 2012, Drivers of agile software development use: Dialectic interplay between benefits and hindrances, Information and Software Journal, ELSEVIER.

Linda G. Wallace, Steven D. Sheetz, 2014, The Adoption of software measures: A technology acceptance model (TAM) perspective, Information & Management, ELSEVIER..

Luis Rodriquez, Rossitza Setchi, Jose L. Salmeron, 2007, Modelling IT project success: Emerging methodologies reviewed, ScienceDirect, ELSEVIER.

Maarit Laanti, Outi Salo, Pekka Abrahamsson, 2011, Agile methods rapidly replacing traditionally methods at Nokia: A Survey of opinions on agile transformation, Information and Software Technology, ELSEVIER.

Magne Jorgensen, 2014, The Failure Factors of Small Software Projects at a Global Outsourcing Marketplace, Journal of Systems Software, ELSEVIER

Mario Spundak, 2014, Mixed agile/traditional project management methodology – reality or illusion?. Procedia-Social and Behavioral Sciences, ELSEVIER.

Meghann L. Drury-Grogan, 2014, Performance on agile teams: Relating interation objectives and critical decisions in project management success factors, Information and Software Technology, ELSEVIER.

Osama Sohaib, Khalid Khan, 2010, Integrating Usability Engineering and Agile Software Development, Int. Conference on Computer Design and Application (ICCDA 2010).

Peerasit Patanakul, 2014, Managing large-scale IS/IT projects in the public sector: Probles and causes leading to poor performance, Journal of High Technology Management Research, ELSEVIER.

Philip Achimugu, Ali Selamat, Roliana Ibrahim, Mohd Naz'ri Mahrin, 2014, A systematic literature review of software requirement prioritization research, Information and Software Technology, ELSERVIER.

Rashinda Hoda et al., 2013, Towards Learning Teams, IEEE Software

Rakshith H. P., Annapurna P Paul, An Impoverished Perspective To Agile Methodology, International Conference on Systems Engineering and Technology, 2013.

Stephan Wood, George Michaelides, Chris Thompson, 2013, Successful extreme programming: Fidelity to the methodology or good teamworking, Information and Software Technology journal.

Sun-Wen Chuang, Tainyi Luor, His-Peng Lu, 2014, Assessment of institutions, scholars and contributors on agile software development (2001-2012), The Journal of Systems and Software, ELSEVIER.

Sven Overhage, Sebastian Schlauderer, 2012, Investigating the Long-Term Acceptance of Agile Methodologies: An Empirical Study of Developer Perceptions in Scrum Projects.

Timo O.A. Lehtinen, Mika V. Mantyla, Jari Vanhanen, Juha Itkonen, Casper Lassenis, 2014, Perceived causes of software failures – An analysis of their relationships, Information and Software Technology, ELSEVIER.

Tina Karrbom Gusavsson, Anette Hallin, 2013, Rethinking dichotomization: A critical perspective on the use of the 'hard' and 'soft' in project management research, International Journal of Project Management, ELSEVIER.

Torgeir Dingsoyr, Sridhar Nerur, VenuGopal Baijepally, Nils Brede Moe, 2012, A decade of Agile Methodologies: Towards explaining agile software development. The Journal of Systems and Software, ELSEVIER.

Tory Dyba, 2011, Special Section on Best Papers from XP2010, Infromations and Software, ELSEVIER.

Tsun Chow, Dac-Buu Cao, 2010, A survey study of critical success factors in agile software projects, ScientDirect, ELSERVIER.

Victor Escobar-Sarmiento, Mario Linares-Vasquez, 2012, A Model for Measuring Agilityin Small and Medium Software Development Enterprise, IEEE

Wei Ding, Peng Liang, Anthony Ting, Hans van Vliet, 2014, Knowledge-based approaches in software documentation A systematic literature review, Information and Software Technology Journal.

William Chavres de Souza Carvalho, Pedro Frosi Rosa, Michel dos Santos Soares, Marcos Antonio Teixeira da Cunba Jr. and Luiz Crlos Buiatte, 2012, A comparative Analysis of the Agile and Traditional Software Development Processes Productivity, 30th International Conference of the Chilean Computer Science Society.

Xiaodan Yu, Stacie Petter, 2014, Understanding agile software development practices using shared mental models theory, Journal of Information and Software Technology, ELSEVIER.

CRISIS INFORMATICS IN SMART CAMPUS: OPPORTUNITIES, CHALLENGES, AND FUTURE DIRECTIONS

Umar Ali Bukar[1,2], Marzanah A. Jabar[1] and Fatimah Sidi[3]
[1]Department of Software Engineering and Information System, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (UPM),
[2]Department of Mathematical Sciences, Computer Science Unit, Taraba State University Jalingo
[3]Department of Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (UPM)
marzanah@upm.edu.my, fatimah@upm.edu.my
corresponding author: marzanah@upm.edu.my

Abstract: Crisis Informatics is a term that describes the use of social media for emergency. This paper select articles on schools crisis communication from an ongoing research on assessment of crisis communication theories and models. The focus is to present state of the art opportunities and challenges facing universities and educational establishment in crisis. This review helps stakeholders to see the significance of social media as part of the crisis communication plan. The findings recommend ways for future research and for emergency correspondence to comprehend the significance of the universities utilization of social media during an emergency, why that is significant, and why specialists should make viable social media communication strategies a need during an emergency.

***Keywords*:** Social media, crisis communication, campus,

INTRODUCTION

Organizations are still susceptible to crises despite their best efforts to avoid one (Omilion-Hodges & McClain, 2016). Ulmer et al. (2011) as cited in (Soehner, et al., 2017) emphasized that "No community, no organization, public or private, is immune from crises". While Social media have come with a new set of communication issues that schools need to consider during and after school crises (Mazer, et al., 2015). It is evident that publics are increasingly relying on social media during crises. It is imperative for organizations and so with education bodies to understand how to effectively incorporate these new technologies into crisis management (Mazer, et al., 2015). Crises are unexpected, and crisis management focuses on how stakeholders (organizations, actors, and institutions) cope with surprising negative pre-events, during-event, and post-events. Crises induce a high degree of uncertainty (Nijkrake, et al., 2015) and anxiety among most actors involved. Every emergency circumstance is associated with unique factors. The study of crisis management and crisis leadership is advanced by examining the distinctive emergency circumstances and their novel conditions.

OVERVIEW OF CRISIS COMMUNICATION LITERATURE

Graham et al. (2015) provides a brief overview of theoretical foundation of crisis communication theories, particularly situational crisis communication theory (SCCT) and the social-mediated crisis communication model (SMCC) are presented. Coombs', SCCT, is verifiably the most frequently utilized theory to examine crisis management and communication (Coombs, 2004, 2007). In a brief, SCCT recommends that an organizational crisis response ought to compare to the degree of its obligation regarding the emergency and the reputational danger presented by the emergency (Coombs, 2007). Jin and Liu (2010) proposed an alternative of the SCCT that consolidates online networking sites users, the social-mediated crisis communication model (SMCC) to guide emergency managers in their social media efforts following an emergency. Specifically, SMCC plots the communications between an organization engaged with a crisis and the various sorts of publics who produce and use data about the emergency by means of social networking sites (Liu et al., 2013). The principal public portrayed in SMCC includes individuals who make crisis data for others to use

and are known as influential social media creators. The second public in SMCC are known as online followers and incorporates the individuals who used the information provided by social media influencers. The third public distinguished by SMCC incorporates people who used information indirectly by means of implication, social media inactive. By recognizing various publics utilizing SMCC, crisis managers, through monitoring social media, can know how and when to react on the online (Graham et al., 2015). Crisis communication is incomplete without bringing the dominant and alternative form of communication like the social media (Soehner, et al., 2017).

CRISIS INFORMATICS IN THE CAMPUS

Dabner (2012) investigated a Facebook page created purposely for the disaster that occurred at a university. Their findings suggest how the university uses the page for information, communication, and collaboration during crisis from both staffs and students. Omilion-Hodges & McClain (2016) conducted research on how first information respondents (FIRs) react to crisis messages, reframing messages, and their dissimination patterns. The paper concludes that FIRs accurately communicate news of crisis and are willing to participate in crisis discussion.

Table 1: Summary of Related Work on Crisis Communication in University

Reference	Theory/Model	Methods	Applications	Crisis type	Remarks
(Heverin & Zach, 2012)	B. Dervin's (1983) theory of sense making	Content, complex time series, and discourse analysis	twitter	Campus shooting	Microblogging and collective sense making
(Dabner, 2012)		Content analysis	Facebook	Earthquake	Information, communication and collaboration on Facebook during crisis
(Mazer, et al., 2015)	Dependency theory	Social media analysis	Facebook, twitter, blogs	Campus shooting	Evidence of lacking crisis communication in universities
(Omilion-Hodges & McClain, 2016)	Channel specification theory, and gratifications theory	Content comparative analysis, and thematic analysis	Facebook	Campus shooting	Information dissemination, meaning of content
(Soehner et al., 2017)	Crisis Communication Theory	Descriptive		Bed bug crisis in library	Reduce rumour and misinformation
(Formentin et al., 2017)	SCCT	Content analysis	Facebook	Sex scandal	Response strategy and public response to crisis
(Fortunato et al., 2018)	SCCT and Image restoration theory	Descriptive	twitter	Racial tensions at the campus	Organisational vulnerabilities
(Hong & Kim, 2019)		Descriptive	Facebook		Impact of using social media by university during crisis

Most of these people used their phones and text messages to share information. Soehner et al. (2017) stated in their paper one of the reasons universities goes public during crisis (bed bug) was the potential for social media to be the bearer of the news. While Mazer et al., (2015) analyses the information generated from social media and provide the implication of educational sector not communicating to the demand of users in social media that includes the student and the parents. Formentin et al. (2017) further conducted a study that examines the types of responses strategy a university were engage and how users responded based on SCCT. Summary of sample studies of crisis informatics found in educational sector can be found in table 1. Moreover, four-phase model for evaluating crisis-management content curricula for teaching was proposed (Alajmi & Al-Qallaf, 2019).

In table 1, few studies presented contribution of research on the area of crisis informatics in educational settings.

OPEN ISSUES

Communication is the most important element in crisis management (Palttala et al., 2012). Social Media as a communication tool for crisis communication is facing significant challenges by schools and is an area that schools least prepared to manage (Mazer et al., 2015). Crisis informatics is highly lacking research on school crisis communication which can leave schools underprepared for school crises and their lasting impact on students, parents, and the larger community.

The network and semantic analysis is also area that needs more research efforts. For example, Omilion-Hodges & McClain (2016) analysed content of users generated to understand the the nature of the messages toward the crisis. However, the authors do not investigate how the messages are written (punctuation) like by using exclamation marks and capitalised words.

In addition, social media progressively turned out to be part of students' lives, higher education must consider if, when and how their potential can best be utilize (Dabner, 2012). Microblogging is yet another term that describes public participation in an online discussion. Haverin and Zach (2012) study demonstrated that microblogging communication found that information sharing practices was used in the early reaction period of the crisis, and opinion sharing expanded after some time, mostly during the recovery period (post crisis). The investigation of individual microblogging text recognized different subjects in the discussion strings that not just helped individual givers comprehend the circumstance yet in addition helped other people who pursued the discussion. The results of this study show that microblogging can play a vital role in collective sense-making during crises.

CONCLUSIONS

This study aim to inform the researchers willing to explore the potential of social media for higher education crisis communication and management. The major crisis type found in these studies is campus shooting. Facebook and twitter are the only social networking sites reported from the study. Majority of the studies conducted content analysis of social media content. Moreover, very few studies examine the crises based on the dominant crisis communication theories (SCCT). Different application of studies in this area proved to the fact that research from the academia and industry is lacking to address issues and challenges currently facing crisis informatics in the educational sector.

BENEFITS FOR HIGHER EDUCATION

Crisis is inevitable and crisis management is becoming an integral part of organisations, and communication is the most important element in crisis management. Crisis communication cannot be complete without taking into consideration the impending form of communication through social media.

REFERENCES

Alajmi, B.M. & Al-Qallaf, C.L., 2019. Crisis-Management Content in LIS Curricula: Developing a Model for Future Improvement. *Journal of Library Administration*, Vol. 58(Issue 7), pp.Pp. 645-673.

Coombs, T., 2018. Crisis Communication. *Crisis Communication*, 22 August.

Coombs, W.T. & Holladay, J.S., 2012. The paracrisis: The challenges created by publicly managing crisis prevention. *Public Relations Review*, volume 38, pp.pp. 408 - 415.

Dabner, N., 2012. 'Breaking Ground' in the use of social media: A case study of a university earthquake response to inform educational design with Facebook. Internet and Higher Education, Volume 15, pp. pages 69-78.

Formentin, M., Bortree, D.S. & Fraustino, J.D., 2017. Navigating anger in Happy Valley: Analyzing Penn State's Facebook-based crisis responses to the Sandusky scandal. *Public Relations Review*, 43, pp.pp. 671-679.

Fortunato, J.A., Gigliotti, R.A. & Ruben, B.D., 2018. Analysing the dynamics of crisis leadership in higher education: A study of racial incidents at the University of Missouri. *J Contingencies and Crisis Management*, 26, pp.pages 510-518.

Graham, M.W., Avery, E.J. & Park, S., 2015. The role of social media in local government crisis communications. *Public Relations Review*, 41, pp.pp. 386-394.

Heverin, T. & Zach, L., 2012. Use of Microblogging for Collective Sense-Making During Violent Crises: A Study of Three Campus Shootings. *Journal of the American Society for Information Science and Technology,* Volume 63, pp. pp. 34-47.

Hong, S. & Kim, B., 2019. Exploring social media use in university crisis communication: An experiment to measure impact on perceived crisis severity and attitudes of key publics. *J Contingencies and Crisis Management,* Volume 27, pp. pp. 61-71.

Liu, B.F. & Kim, S., 2011. How organizations framed the 2009 H1N1 pandemic via social and traditional media: Implications for U.S. health communicators. *Public Relations Review*, 37, pp.pp. 233-244.

Mazer, J. P. et al., 2015. Communication in the face of a school crisis: Examining the volume and content of social media mentions during active shooter incidents. Computers in Human Behavior, Volume 53, pp. pages 238-248.

Nijkrake, J., Gosselt, J. F. & Gutteling, J. M., 2015. Competing frames and tone in corporate communication versus media coverage during a crisis. Public Relations Review, Volume 41, pp. pages 80-88.

Omilion-Hodges, L. M. & McClain, K. L., 2016. University use of social media and the crisis lifecycle: Organizational messages,first information responders' reactions, reframed messages and dissemination patterns. Computers in Human Behavior, Volume 54, pp. pp. 630-638.

Palttala, P., Boano, C., Lund, R. & Vos, M., 2012. Communication Gaps in Disaster Management: Perceptions by Experts from Governmental and Non-Governmental Organisations. *Journal of Contingencies and Crisis Management*, 20(1), pp.pages 2-12.

Soehner, C., Godfrey, I. & Bigler, G. S., 2017. Crisis Communication in Libraries: Opportunity for New Roles in Public Relations. The Journal of Academic Librarianship, Volume 43, pp. pages 268-273.

TRUST, PURPOSE, AND ROLE-BASED ACCESS CONTROL MODEL FOR PRIVACY PROTECTION

Mohd Rafiz Salji[1,2], Nur Izura Udzir[1], Mohd Izuan Hafez Ninggal[1], Nor Fazlida Mohd. Sani[1] and Hamidah Ibrahim[1]
[1]Faculty of Computer Science and Information Technology, Universiti Putra Malaysia
[2]Faculty of Information Management, Universiti Teknologi MARA, Malaysia
mohdrafiz@sarawak.uitm.edu.my, {izura, mohdizuan, fazlida, hamidah}@upm.edu.my
corresponding author: izura@upm.edu.my

***Abstract*:** Data privacy is one of the fundamental needs of the people. In a computing environment, there are various issues of data privacy protection in the enterprise. To enforce the automation of privacy policies and law, access control has been one of the most devoted subjects. Role-based access control model has been proposed to protect customer's data. However, relying on role only is insufficient and inefficient to protect data especially sensitive attributes, and this may cause risks of privacy disclosure to unauthorized and untrusted users. This paper presents a finer-grained access control called Trust, Purpose, and Role-Based Access Control (TPRBAC) model to efficiently protect data particularly sensitive attributes. In the proposed model, purpose and role is applied to permit access to data, while trust is applied to control access to sensitive attributes. A prototype system is developed and tested, and the result shows sensitive attributes are protected. Experiments are conducted to validate the proposed model. The results show that the proposed work is efficient and improved privacy protection. Therefore, this research solves the issue of insufficient and inefficient access control mechanism in protecting data especially sensitive attributes.

***Keywords*:** Trust, purpose, role, privacy protection, sensitive attributes.

INTRODUCTION

Data privacy is increasingly becoming one of the main concerns in data management. People are now more conscious about how their data are being protected by the organization. This awareness is increasingly highlighted when sharing and collecting data become seamless and prevalent by the omnipresence of Internet connection. There are many approaches to preserve data privacy, for example, encryption and digital signatures (Abdul Ghani, 2013). However, the scope of this study is focused on access control to protect data as this mechanism is the most common approach to protect data from unauthorized user (Bertolissi and Fern´andez, 2014). Many studies related to access control have been proposed to protect privacy, however, there are still issues that impede the development of efficient access control models.

A fine-grained access control known as role-based access control (RBAC) model has been proposed (Yousafi, 2013) to permit user access to the data based on role, i.e., job title or job function. However, in this model all users with the same role can be allowed to access customers' data without considering their purpose. For example, lecturers can access students' data; but, not all lecturers are authorized to access all students' data. Purpose means "for what reason data are accessed or used" (Agrawal et al., 2002). To access data, user needs to be evaluated based on the purpose of the usage. Purpose-based access control (PBAC) is a common access control model, which considers the purpose of access as an essential factor in deciding whether to permit or deny access to the resources. Many PBAC models (Kabir et al., 2012, 2011; Sun and Wang, 2012; Sun et al., 2012; Abdul Ghani, 2013; Wang et al., 2014b; Elgendy et al., 2017) have been proposed to preserve data privacy. In these models, purpose is considered in order to allow certain users to access certain data to avoid privacy violation. There are three options which the customer can set the level of privacy in the PBAC model before permitting or denying user access to the data: to allow, conditionally allow, or prohibit access (Kabir et al., 2012, 2011). For example, a customer Alice allows her age, i.e., 37 to be accessed by users for admin purpose; but she prohibits users to access it for marketing purpose. Subsequently, she may conditionally allow her age, i.e., 35-40 to be accessed by users for purchase purpose. Based on this example, customer data should be accessed by the user based on role and

purpose, not based on role only to protect customer privacy. Therefore, a new access control model needs to be proposed that considers the user's purpose to protect data.

Data in nature is a sensitive information, but sensitive attributes must remain safe (Maheshwarkar et al., 2012). In general, data are divided into three types of attributes, namely, de-identified, quasi identifier, and sensitive (Sweeney, 2002b). De-identified data are the obvious identifying records that need to be concealed, for instance social security number. In contrast, quasi identifier such as race, age, and zip code is a non-key attribute that needs to be generalized before it can be released. Meanwhile, sensitive attributes such as medical condition and income are classified data which privately belong to a customer. Therefore, sensitive attributes require critical restricted access in the system and access to this attribute is limited to trusted users only. However, existing access control models (Yousafi, 2013; Kabir et al., 2012, 2011; Sun and Wang, 2012; Sun et al., 2012; Abdul Ghani, 2013; Wang et al., 2014b; Elgendy et al., 2017) do not focus on protecting sensitive attributes. This may lead to the risk of inappropriate access and use of sensitive attributes. Therefore, a mechanism is needed to permit only trusted users to access sensitive attributes. In access control model, one of the common types of access control called trust-based access control (TBAC) is applied to protect the resources of the system. TBAC is inspired by an important aspect in human life, which is trust. In this study, trust refers to firm believe to a user in an organization. By this concept, a user that is highly trusted will be granted access to more resources. However, trust is mutable in response to the changing situations. Therefore, it is paramount to design an efficient access control model that can capture the dynamic nature of human trustworthiness. Based on previous literature, access control models based on trust (Toahchoodee et al., 2009; Li et al., 2009b, 2012) have been proposed to protect data, but not specifically protecting sensitive attributes. Therefore, a new access control model needs to be proposed to consider trust to protect sensitive attributes.

In this paper, a Trust, Purpose, and Role-Based Access Control (TPRBAC) model is proposed to protect data especially sensitive attributes. Test and validate the TPRBAC model by using the prototype system and compare it with the previous model (Yousafi2013) is conducted. The result shows that the TPRBAC model is efficient and privacy protection is improved as compared to the previous model (Yousafi, 2013).

METHODS

The TPRBAC model is developed to test whether it is applicable to protect data particularly sensitive attributes. Tests are conducted using a prototype system named XYZ Bank Information System to identify which user level of granularity is permitted to access data especially sensitive attributes. After completing the test of the TPRBAC model, then the TPRBAC model needs to be validated by comparing the efficiency in terms of response time between the proposed and the previous model (Yousafi, 2013) in accessing sensitive attributes. Next, both models are compared in terms of privacy protection by determining the number of users whom are permitted and not permitted to access sensitive attributes. Finally, to determine the improvement of privacy protection in the proposed model, the number of sensitive attributes that cannot be accessed by the untrusted user is calculated.

FINDINGS

Efficiency: Based on Figure 1, the result shows that the TPRBAC model is efficient where untrusted users faster in accessing data without sensitive attributes as compared to the previous model which permits user to access sensitive attributes without considering user's trust.

Figure 1: Performance of an Untrusted User (Proposed Model) and an Authorized User (Previous Model) Accessing Data

Privacy Protection: In order to compare the privacy protection between the proposed model and the previous model (Yousafi, 2013), both models are tested to access customer data to determine which user is permitted or authorized to access to data. In order to compare both models, ten customer data are used in this test, which are nine data and one sensitive attribute, i.e., *earned income*. A total of 1000 customer data is applied in this study. In this test, we assume users who are working in the Marketing Department with the role as a Marketing executive, and the purpose of Service-Update is allowed to access customer data in the proposed model, while in the previous model, only role is considered to access the data.

Based on Tables 1 and 2, User 2 in Table 1 is the same as User 2 in Table 2 that using the role to access data. However, User 2 in the RBAC model is permitted to access data, but User 2 in the proposed model is prohibited to access it because the proposed model requires both role and purpose to access data. However, the previous model requires role only to access it. Unlike the previous model, without trust, sensitive attributes are prohibited to be accessed by the user in the proposed model such as User 7. The result shows that the privacy is protected in the proposed model as compared to the previous model.

Table 1: Access Control Mechanism in the RBAC Model (Yousafi, 2013)

	User Authentication	Data Authorization
User	**Role**	**Data**
1	**X**	**X**
2	✓	✓

Table 2: Access Control Mechanism in the TPRBAC Model

	User Authentication			Data Authorization	
User	**Role**	**Purpose**	**Trust**	**With Sensitive**	**Without Sensitive**
1	**X**	**X**	**X**	**X**	**X**
2	✓	**X**	**X**	**X**	**X**
3	**X**	✓	**X**	**X**	**X**
4	**X**	**X**	✓	**X**	**X**
5	✓	**X**	✓	**X**	**X**
6	**X**	✓	✓	**X**	**X**
7	✓	✓	**X**	**X**	✓
8	✓	✓	✓	✓	✓

Next, to measure the improvement of privacy protection, the number of the privacy that are prohibited to be accessed by untrusted users in the proposed work is calculated. Based on calculation of an untrusted user who is prohibited to access one sensitive attribute, it means that 1000 data elements or one data out of 10000 data elements or 10 data are prohibited to be accessed by an untrusted user. The calculation is as follows:

Privacy protection = $\frac{1000}{10000}$ * 100

The result of the calculation shows that the privacy is protected by 10% in the TPRBAC model. This test shows that the privacy protection is improved by protecting sensitive attributes from an untrusted user in the proposed model as compared to the previous work (Yousafi, 2013).

DISCUSSION

The TPRBAC model is tested and validated by using a prototype system and experiment. Based on the test of the TPRBAC model, it shows that the TPRBAC model can be used to protect data especially sensitive attributes. Based on the validation of the TPRBAC model, the result shows that the TPRBAC model is efficient and improve privacy protection as compared to the previous model (Yousafi, 2013).

CONCLUSION AND BENEFITS FOR HIGHER EDUCATION

This study proposes a finer-grained access control called trust, purpose, and role-based access control (TPRBAC) model to protect data especially sensitive attributes. The TPRBAC model was tested and validated to prove that it can be used to protect data especially sensitive attributes. As a result of the test, it shows that the data especially sensitive attributes can be protected. Based on the validation of the TPRBAC model by comparing with the existing model (Yousafi, 2013), experiment proves that the TPRBAC model is efficient and privacy protection is improved as compared to the previous model. The issue of the previous model that is insufficient and inefficient which provides the role only in protecting data especially sensitive attributes was solved by proposing an efficient TPRBAC model to protect data especially sensitive attributes.

The benefit of this study to the higher education is that the proposed model can be applied to protect data especially sensitive attributes of students, educators, and non-academic staff.

REFERENCES

Abdul Ghani, N. 2013. Credential Purpose-Based Access Control for Personal Data Protection in Web-Based Applications. PhD thesis, Universiti Teknologi Malaysia, Faculty of Computing.

Agrawal, R., Kiernan, J., Srikant, R. and Xu, Y. 2002. Hippocratic Databases. In VLDB'02: Proceedings of the 28th International Conference on Very Large Databases, 143-154. Elsevier.

Bertolissi, C. and Fernandez, M. 2014. A Metamodel of Access Control for Distributed Environments: Applications and Properties. Information and Computation.

Elgendy, R., Morad, A., Elmongui, H. G., Khalafallah, A. and Abougabal, M. S.2017. Role-Task Conditional-Purpose Policy Model for Privacy Preserving Data Publishing. Alexandria Engineering Journal 56 (4): 459-468.

Kabir, M. E., Wang, H. and Bertino, E. 2011. A Conditional Purpose-Based Access Control Model with Dynamic Roles. Expert Syst. Appl. 1482-1489.

Kabir, M., Wang, H. and Bertino, E. 2012. A Role-Involved Purpose-Based Access Control Model. Information Systems Frontiers 14 (3): 809-822.

Li, M., Sun, X., Wang, H. and Zhang, Y. 2012. Multi-Level Delegations with Trust Management in Access Control Systems. Journal of Intelligent Information Systems 39 (3): 611-626.

Li, M., Wang, H. and Ross, D. 2009b. Trust-Based Access Control for Privacy Protection in Collaborative Environment. In e-Business Engineering, 2009. ICEBE'09. IEEE International Conference on, 425-430. IEEE.

Maheshwarkar, N., Pathak, K. and Choudhari, N. S. 2012. K-anonymity model for multiple sensitive attributes. International Journal of Computer Applications (IJCA).

Sun, L. and Wang, H. 2012. A Purpose-Based Access Control in Native XML Databases. Concurrency and Computation: Practice and Experience 24 (10): 1154-1166.

Sun, L., Wang, H., Soar, J. and Rong, C. 2012. Purpose Based Access Control for Privacy Protection in E-Healthcare Services. Journal of Software 7 (11).

Sweeney, L. 2002b. K-Anonymity: A Model for Protecting Privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10 (05): 557-570.

Toahchoodee, M., Abdunabi, R., Ray, I. and Ray, I. 2009. A Trust-Based Access Control Model for Pervasive Computing Applications. In Data and Applications Security XXIII, 307-314. Springer.

Wang, Y., Zhou, Z. and Li, J. 2014b. A Purpose-Involved Role-Based Access Control Model, 1119-1131. Springer Berlin Heidelberg.

Yousafi, H. 2013. A Role-Based Access Control Schema for Materialized Views. Master's thesis, University of Windsor.

WATERMARK BLENDING TECHNIQUE FOR SECURING USER AUTHENTICATION DATA IN MOBILE APPLICATION SYSTEMS

Kartini Mohamed[1], Fatimah Sidi [2], Marzanah A. Jabar[3] and Iskandar Ishak[4]
[1]Group Human Resource, SIRIM Berhad, No.1, Persiaran Dato' Menteri, Peti Surat 7035, Section 2, 40700 Shah Alam, Selangor.
[2,3,4] Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 Serdang Selangor
kartinim@sirim.my, fatimah@upm.edu.my, marzanah@upm.edu.my, iskandar_i@upm.edu.my
corresponding authors: fatimah@upm.edu.my*, kartinim@sirim.my*

Abstract: Watermark is traditionally used for protecting copyrighted materials such as scripts, movies, images or sounds. Nowadays, the watermark is still widely used but more in a digitized environment. There are many ways of how the watermarks are being applied. In this paper, the watermark is not applied for the purpose of protecting copyrighted materials but more on digitally securing user authentication data used for accessing mobile application systems (mobile apps). A user authentication model is developed to improve the access control to mobile apps. The model introduces three protection techniques namely multi-factoring, ciphering and watermarking techniques. In multi-factoring technique, the number of user authentication factors is increased from having just username and password to raise the difficulty of guessing the authenticating data. The added factors are mobile IMEI number, mobile SIM card number, random number and time. In ciphering technique, the user authentication factors are not only encrypted but also hashed to ensure the encrypted data cannot be easily decrypted. Finally, in the watermarking technique, the authenticating data are concealed by blending them in such a way that they cannot be un-blended without knowing the blending formula. Although there are three protection techniques introduced, only watermarking technique is described in more details in this paper. The effectiveness of the introduced techniques is measured based on user acceptance level using a statistical analysis. The analysis shows that although ciphering technique is the highest contributor among other techniques, watermarking technique not only has the strongest relationship but also becomes the dominant factor in making them acceptable by users of mobile apps.

***Keywords*:** watermark, mobile application system, security, user authentication, data transmission

INTRODUCTION

The use of mobile apps involves data transmissions in a wireless environment which is very vulnerable to data theft by irresponsible hackers (Petal et al., 2016). Communication using mobile apps may involve transmitting of highly confidential data such as personal particulars, financial information, medical records, high profile business deals as well as certain governmental affairs. These types of data require stronger data protection compared to those of non-confidential data. Securing these data before, during and after transmission is very crucial to ensure users are confident that their personal information and other confidential data are well protected.

Section 40 in the Malaysian Personal Data Protection Act 2010, requires all data administrators to get concents from data owners before any sensitive personal data can be processed (AGC, 2016). This implies that strict control and strong protection of data are necessary to ensure all sensitive personal data are not simply released to unintended personnel.

Normal usage of watermark is to indicate the correct owners of the copyright materials (Parhi & Nishitani, 1999). However, the irresponsible persons can still misuse the sensitive materials without the knowledge of the owners. Nevertheless, if these materials are in digital form, stronger protections can easily be applied to stop the irresponsible users from obtaining these data. The blending of watermark is a technique that can be used to achieve this purpose.

This paper is organized into several sections as follows: introduction to watermarking, the blending technique, summary of statistical analysis, results and discussion, and conclusion.

WATERMARK

Watermarks have become useful tools since many years ago. Their applications become overwhelming when digital watermarks are discovered. Digital watermarks are digital data that can be made hidden in a media content which are invisible to human eyes but detectable by computers. Due to this capability, it is suitable to be applied in numerous ways such as signatures to identify content owners, fingerprinting to identify digital content buyers, broadcast and publication monitoring for determining royalty payments, authentication to filter out the falsified content, copy control to ensure only controlled content are validated and secret communication to ensure no information are leaked out to unauthorized personnel (Parhi & Nishitani, 1999). Another advantage of applying watermarks is when imposing them with digital signatures, the data security and integrity can be increased since they can no longer be tempered by illegitimate users (Shukla et al., 2012). Watermark is a better and suitable tool to be applied compared to public key cryptography or digital signatures which are too heavy for Wireless Sensor Networks (WSN). Due to this characteristic of watermark, Kamel et al. (2009) have introduced a Distortion-Free Watermarking Scheme to increase robustness in preventing various attacks such as data injection, deletion, modification or node replacement. Since watermarks can enhance digital data protections in many ways, next section of this paper describes how watermarks can improve the access control to mobile apps using a blending technique.

WATERMARK BLENDING TECHNIQUE

The data used in this blending technique are the multi-factor user authentications consisting of username, password, mobile IMEi number, mobile SIM card number and random number which are then clasified into known and unknown data. The known data are username and password since the users key-in these items on their own. The unknown onces are the additional factors introduced in this user authentication model except time. These additional factors are ‘unknown’ by users since they are auto retrieved by the system which is also considered as the secret information or unseen ‘watemark’.

These watermark data are blended with username dan password using the following steps which are depicted in Fig.1:

Step 1: Obtained all the five user authentication as raw data
Step 2: Hash each of these data using a hash function, preferably SHA-512
Step 3: Split each of the hash data into two parts. The length of hashed data become 128 each. After spiting, the lenght is only 64 each.
Step 4: Blend them by randomly switched the split data, the random ‘formula’ is memorized by the system because it be used for un-blending process
Step 5: Encrypt the blended data using AES encryption of AES256 formula. After encryption, the length of each encrypted data will result in 108. These 5 encrypted data will be attached together into a string of 5x108 equivalent to 540 characters.

These blended user authentication data can be transmitted between mobile phone and the server which will be compared with the stored authentication data in database. Before the comparison can be done, the transmitted data need to be un-blended. The steps for this un-blending process are as shown in Fig.2:

Step 1: Obtained all the encrypted data with 540 characters. Split them into 5 parts with equal lenght to obtain 108 characters for each part.
Step 2: Decrypt each part using AES256 decryption formula. The decrypted data is actually in hashed form at this stage.
Step 3: Un-blend the data using the based on memorized ‘formula’ identified earlier. Split each of the un-blended data into two parts.

Step 4: Switch back the hashed data by pairing back each of the two parts.
Step 5: De-hash each of the paired data to get back the original raw data.

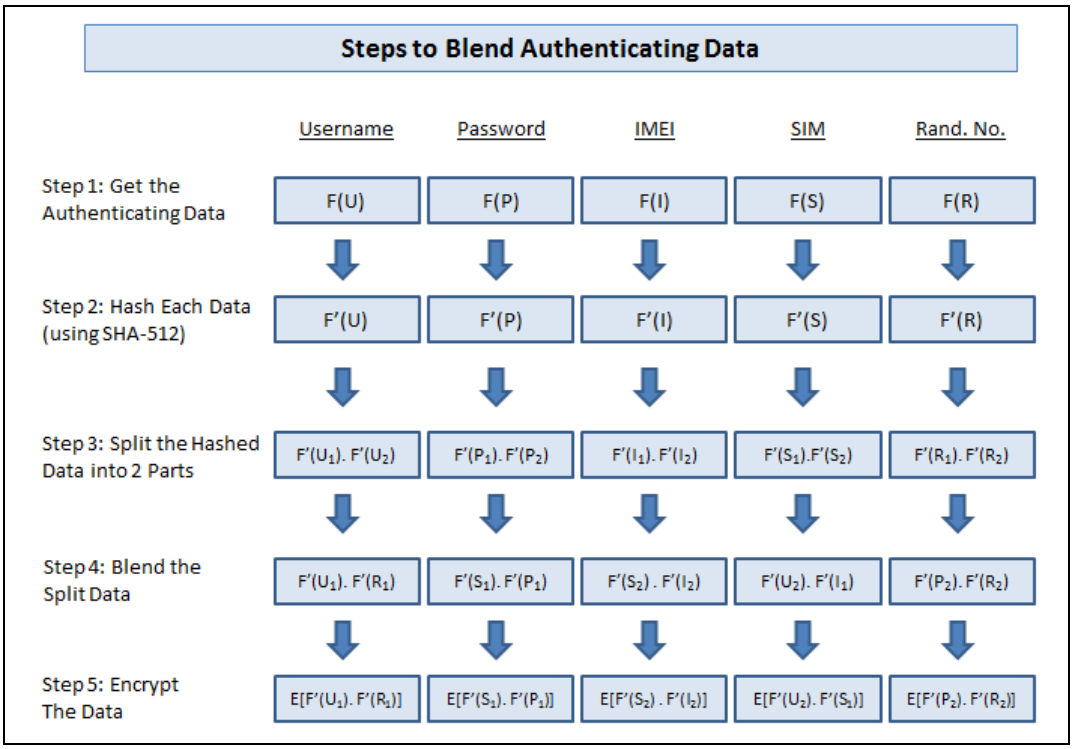
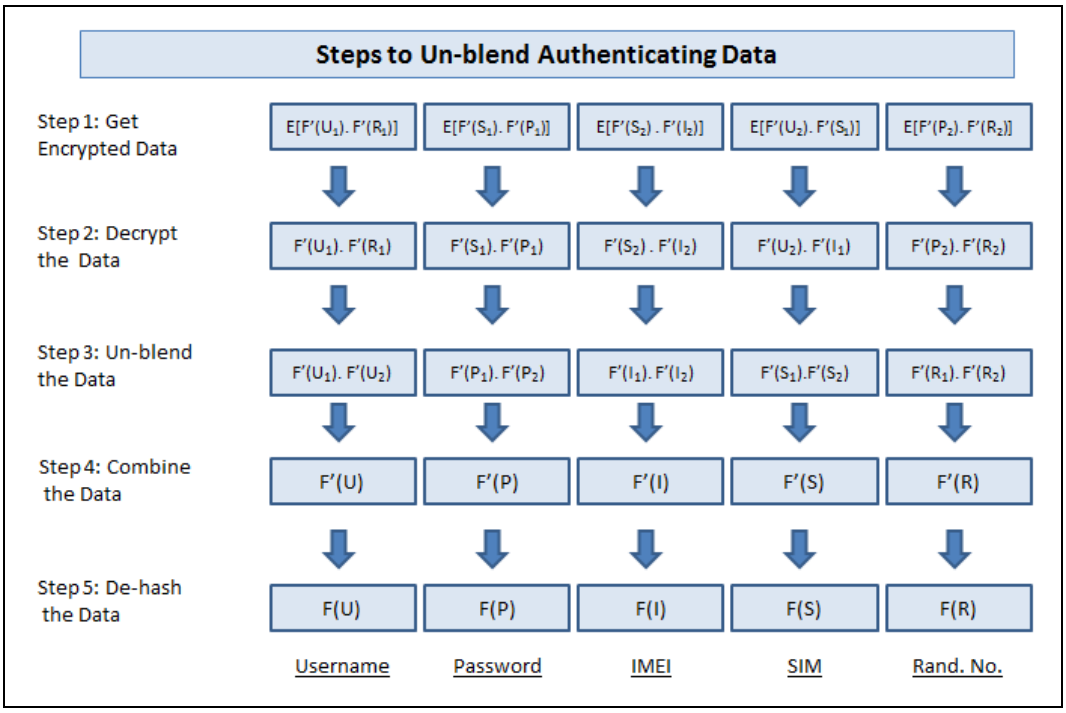


Fig. 1: Steps in Blending Technique

Fig. 2:



Steps in Un-blending Technique

By completing all the above processes, the protection techniques of multi-factoring, ciphering and watermarking are all applied. To measure the effectiveness of these techniques, a statistical analysis is carried out to measure the acceptance level by users of mobile apps.

STATISTICAL ANALYSIS

The user acceptance level is measured using a statistical analysis which consists of three tools namely Desriptive Analysis, Pearson Correlation, and Multiple Linear Regression. The analysis requires the model to be divided into two variables – independant and dependent variables. Independent variable consists of multi-factoring, ciphering and watermarking techniques while dependent variable refers to acceptablity of user authentication as shown in Fig.3.

Descriptive Analysis is performed to undertand the level of agreement between the independent and dependent variables while Pearson Correlation assists in finding the relationship between the two vaiables and Multiple Linear Regression helps in identifying the most dominant factor in the independant variables that contributes to the achievement of dependent variable.

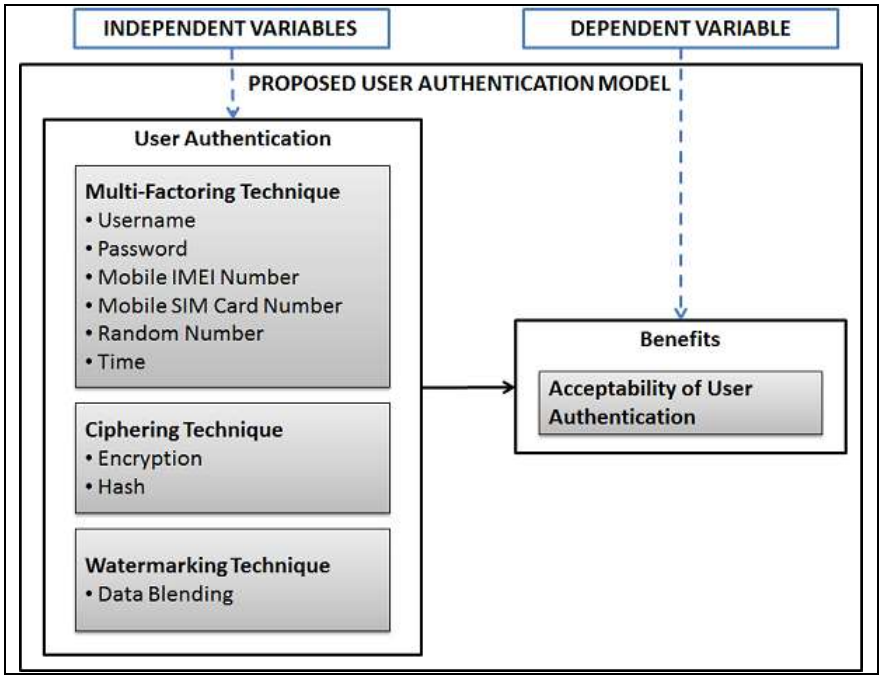


Fig.3: Independent and Dependent Variables

Survey forms are constructed and distributed to 250 smart phone users and 175 of them have responded. From this number of responses, results are analysed and discussed.

RESULTS AND DISCUSSION

In Descriptive Analysis, level of agreement is determined using a formula adopted by George & Mallery (2013) which limits the Low level between 1~2.33, Moderate between 2.34 ~ 3.66 and High between 3.57 ~ 5. Survey results indicate Multi-factoring scores 3.89, Ciphering 4.16 and Watermarking 4.05. This indicates that although all three techniques in independent variables score at High level or agreement, ciphering technique scores the highest.

In Pearson Correlation, the relationships between independent and dependent variables are interpreted using the Rule of Thumb by Guildford (1973) where the Pearson (r) value is divided into 5 categories of Negligible (0.0~0.29), Low (0.3~0.49), Moderate (0.5~0.69), High (0.7~0.89) and Very High (0.9~1.0) relatioships. Survey results show that multi-factoring and ciphering have the correlation

value of 0.556 each and watermarking has 0.641 correlation value. This means watermarking has the highest correlation value.

In Multiple Linear Regressions, the dominant factor is identified using the method introduced by Bluman (2012. The formula of simple linear regression model is

$$Y_1 = B_0 + B_1X_1 + B_2X_2 + B_3X_3$$

where,

Y_1	= Dependent variables
B_0	= Constant (Intercept)
B_{1-3}	= Estimates (Regression Coefficients)
X_1, X_2, X_3	= Independent variables

The regression coefficients from the survey results indicates that the constant value is 0.97, multi-factoring is 0.21, ciphering is 0.13 and watermarking is 0.40. Unfortunately, the significant value for ciphering is 0.08 which is above 0.05 which means ciphering is not significant and not considered in the linear regression. Therefore, Multiple Linear Regression formula obtained from this survey is $Y_1 = 0.97 + 0.21X_1 + 0.40X_3$ where X_1 is multi-factoring and X_3 is watermarking. Since watermarking has a higher coefficient value, it indicates that watermarking is the dominant factor.

CONCLUSIONS

The study has proposed a model for better control of access to mobile apps using three protection techniques of multi-factoring, ciphering and watermarking. The effectiveness of the proposed protection is measured using a statistical analysis consisting of Descriptive Analysis, Pearson Correlation and Multiple Linear Regressions. Surveys form are distrbuted to smart phone users and 175 of them responded. Based on the analysis, even though ciphering technique is the highest contributor to the acceptability of mobile apps users, watermarking technique is found to have the strongest relationships and become the dominant factor compared to the other protection techniques. This also implies that watermarking is a protection technique acceptable by mobile apps users in making the authentication data secured.

ACKNOWLEDGEMENT

This research is funded by Ministry of Higher Education of Malaysia under the Fundamental Research Grant Scheme (FRGS) - FRGS/2/2014/ICT07/UPM/02/1.

REFERENCES

V. Patel, R. Chellappa ve D. B. Chandra. 2016. Continuoues User Authentication on Mobile Devices, *IEEE Signal Processing Magazine,* no. July, pp. 49-61.

AGC, agcportal, 15 June 2016. [Çevrimiçi]. Available: http://www.agc.gov.my/agcportal/uploads/files/Publications/LOM/EN/Act%20709%2014%206%20 2016.pdf. [Accessed: 7 July 2019].

K. K. Parhi ve T. Nishitani. 1999. A Review of Watermarking Principles and Practices, *Digital Signal Processing in Multimedia Systems,* pp. 461-485.

S. S. P. Shukla, S. P. Singh, K. Shah ve A. Kumar. 2012. Enhancing Security & Integrity of Data Using Watermarking & Digital Signature, *IEEE*.

I. Kamel, O. Al-Koky ve A. Al-Dakkak. 2009. Distortion-Free Watermarking Scheme for Wireless Sensor Networks, *International Conference on Intelligent Networking and Collaborative Systems*.

D. George and P. Mallery. 2013. IBM SPSS Statistics 21 Step By Step: A Guide and Reference, 13th Edition., Pearson.

J. P. Guildford. 1973. Foundamental Statistics in Psychology and Education, 5th Edition., New York, USA: McGraw-Hill.

A. G. Bluman. 2012. Elementary Statistics, A Step by Step Approach, 7th dü., New York: McGraw Hill.

LIST OF e-PROCEEDING POSTERS

Facilitator Application Management System for Student Co-curriculum and Student Development Center of UPM

Introduction

Facilitator job application in for Credited Co-curricular courses in UPM is handled by the UPM's Student Co-curriculum and Development Center (SCDC). Currently, the job application management process is using the manual and web-based platform. Paper-based approach tends to produce errors in terms of data quality, data validity as well as data restoration as it depends on hardcopy forms. Web-based facilitator application for SCDC can be the solution to have a better facilitator application management, reduce data error and can store data for future purposes.

Methodology

Figure 8: Waterfall Model

The methodology that chosen to be use in this project is Waterfall Model. It is a linear sequential model in which all the phase is organized in linear. Each phase has to be complete first before the next phase can start. This model is chosen because it is simple to understand and the requirement is clear and well known as the main objective is to automate the existing manual application process

Results and Summary

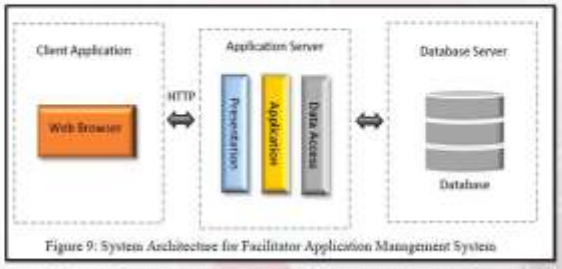


Figure 9: System Architecture for Facilitator Application Management System

Facilitator Application Management System for Co-Curriculum and Student Development Centre (PKPP) is a web-based application that follow the three-tier architecture. It consists of three different tier such as presentation tier, application tier and data tier that have different type of functionalities. Presentation tier is the top most level which handle the interaction with user by providing the user interface. The application tier contains set of rule for information processing and business logic while the data tier consists of database and database management system

Contributions

This projects aim to improve the facilitator application process through paperless approach in which all information are stored in a database. Easier reporting and data management tasks can reduce time for CSDC UPM to manage data related to managing facilitator application of credited co-curricular teachings in UPM.

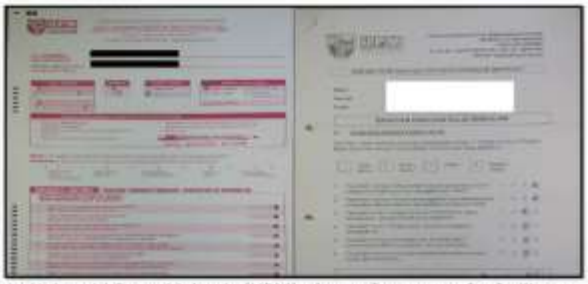
Presenter:
Dr Iskandar Ishak

Project Team member:
Tan Yong Lin, Paramasivam Subramaniam, Nur Izura Udzir, Rosiha Abdul Razak, Mohd Syawal Sauti, Ahmad Zikri Zainol, Ahmad Azim Kamaruddin, Hairul Nizam Mohd Saroji

Collaboration between:
Co-curriculum and Student Development Center and Faculty of Computer Science & Information Technology, Universiti Putra Malaysia

Facilitator Evaluation System for Co-Curriculum and Student Development Centre of UPM

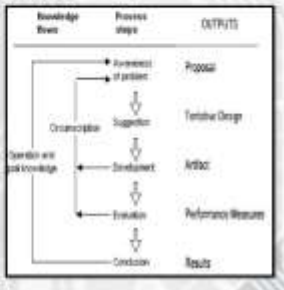
Introduction



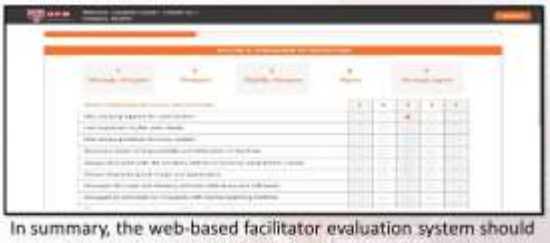
In Universiti Putra Malaysia (UPM), the performance of a facilitator is always being monitored under Co-Curriculum and Student Development Centre to ensure that an optimal standard of facilitating is achieved by eliminating those facilitators who failed to comply with the standards. The current evaluation is done through manually form-filling by both supervisor and student based on different set of questionnaires prepared and printed by CSDC. the benefits of having a web-based facilitator evaluation system outshine its disadvantages in hoping to deal with the troublesome manual form filing and storing of files. As such, this project is here to study and develop a reliable alternative solution for the staffs of CSDC, supervisor and student.

Methodology

A multitier or client-server architecture with three layers will be used to develop the proposed system. As illustrated in the diagram above, the system consists of user interface layer, application logic layer and database layer. The separation of tiers allows easier modification and addition of layers in future development without the need to rebuild the entire system



Results and Summary



In summary, the web-based facilitator evaluation system should provide convenient way for both supervisor and students to evaluate the facilitator based on each class respectively. On the other hand, the staff at Co-Curriculum and Student Development Centre should be able to review and analyse the evaluation form submitted by both supervisor and student easily.

Contributions

1. The system is able to generate questionnaires based on the type of users.
2. The system is able to compute the scores for each facilitator based on the questionnaires completed by the users.
3. The system is able to store the evaluation record for each facilitator.
4. The system is able to generate overview of the result of the evaluation based on each facilitator.

Presenter:
Dr Iskandar Ishak

Project Team member:
Lee Jin Wee, Paramasivam Subramaniam, Nur Izura Udzir, Rosiha Abdul Razak, Mohd Syawal Sauti, Ahmad Zikri Zainol Ahmad Azim Kamaruddin, Mohd Hairul Nizam Saroji

Collaboration between:
Co-curriculum and Student Development Center and
Faculty of Computer Science & Information Technology,
Universiti Putra Malaysia

GLOBAL INVESTMENT SYSTEM (GLINSYS)

Together We Eradicate Poverty!

INTRODUCTION

Global Investment System (GLINSYS), is a decision support system to assist investors in determining the investment risk in a country.

The system will calculate the investment risk based on selected determinant factors using a set of rules of fuzzy logic.

Utilizing GLINSYS will encourage investors to invest in non-developing country, this will indirectly boost the country's economy by creating more job opportunity.

Hence, by this it can in another way reduce poverty in those non-developing countries.

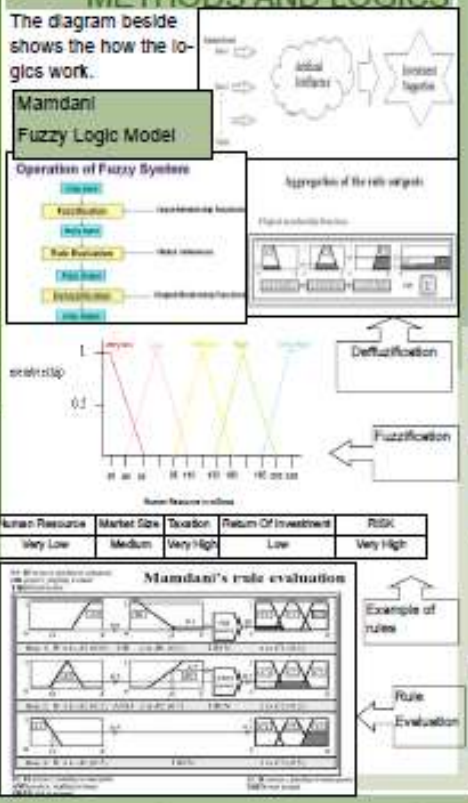
MAIN OBJECTIVES

Utilizing the factors collected deduce rules of fuzzy which will then be used for the calculation of the risk of investment.

Calculate the risk of investment based on selected determinants by using fuzzy logic.

METHODS AND LOGICS

The diagram beside shows the how the logics work.



Human Resource	Market Size	Taxation	Return Of Investment	Risk
Very Low	Medium	Very High	Low	Very High

RESULT

	Malaysia	Philipines	Brazil	New Zealand	Paraguay	South Africa	United States
Rubber	53.07	52.23	N/A	N/A	53.04	42.69	N/A
Oil Palm	N/A						
Gold Mine	N/A						
Petroluem	24.8						
Hotel	N/A						
Medicine	N/A						
Housing	N/A						
Cocoa	N/A						
Poultry	N/A						

Legends:
Very Low Low Medium High Very High

The risk table of investing in different countries based on selected determinants.

ADVANTAGES/BENIFITS

GLINSYS can calculate the risk of investment in various countries which will encourage investor to invest in non-developing countries, indirectly creating more job opportunities.

CONCLUSION

In conclusion, GLINSYS not only encouraging investment to non-developing countries which will indirectly affects the job opportunities, it also boost the economics growth of the world, creating a balance investment between developed and non-developing countries.

With these, comes to solving the problem of poverty in those poor countries slowly over the time.

Student Name	TAN WUI PENG
Matric number	148378
Department	INFORMATION SYSTEM
Contact	+6016-4770315
	friendly_negg@hotmail.com

Supervisor Name	Dr. MARZANAH A. JABAR
Department	INFORMATION SYSTEM
Contact	+6017-3314754
	marzanah@fsktm.upm.edu.my

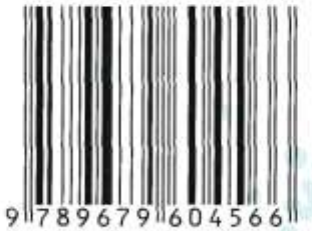
Faculty of Science Computer and Information System University Putra Malaysia

©Copyright Reserved

UPM
UNIVERSITI PUTRA MALAYSIA
BERILMU BERBAKTI

Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
www.fsktm.upm.edu.my

ISBN 978-967-960-456-6



powered by IDEC