1. **Title of project:** Graph Theory Driven Approach to Analyze UML Class Diagrams Relationships using XMI

**Objectives(s) of project:**

- To analyze the complexity of classes in a UML class model using XMI and graph theory.
- To identify the classes with high impact, i.e. classes being heavily reuse, besides the utility classes, interface classes, and design patterns.
- To measure the complexity and impact of classes in the UML model using graph theory metrics.

**Brief description:**

This research focuses on software design, specifically on UML class diagrams, to analyze the complexity of UML classes, and subsequently identify the classes with high impact, i.e. classes being heavily reuse, utility classes, interface classes, and design patterns. Analyzing and understanding the behavior of software systems can be associated with their modeling through graphs (as in the graph theory concept), where software components are represented as nodes, while inter-relationships among them are represented as edges. Graphs have been successfully applied on various domains such as the World Wide Web, social network, power grid, and scholarly citation network to provide a high-level graph abstraction view.

One way to analyze UML diagrams is by using the XMI (XML Metadata Interchange) representation of the analyzed UML class diagram. XMI is used to exchange metadata information of UML models using XML (Extensible Markup Language) representation. All major CASE (Computer Aided Software Engineering) tools, such as ArgoUML, Rational Rose, Enterprise Architect, MS Visio, Altova, Visual Paradigm, etc can export and import class diagrams in the format of XMI. With the support of XMI, it would ease the task of measuring the complexity and impact of classes using graph theory metrics.

**Expected Outcome:** An approach to analyze the complexity of UML class model using XMI and graph theory metrics.

**Tools/Programming languages to be used:** Suitable development tools
2. **Title of project:** Approach for Mapping Bug Reports and Change Commit to Relevant Classes in a Source Code Management System

**Objectives(s) of project:**

- To explore state-of-the-art techniques in mapping unstructured data (bug reports) to structured data (source code or classes in a source code management system (SCMS) such as GitHub) using Information Retrieval (IR) techniques.
- To propose an automatic approach that maps bug reports with respect to their relevant source code or classes in the selected source code management system.

**Brief description:**

This research focuses on proposing an automatic approach that maps bug reports with respect to their relevant source code or classes in a selected SCMS. The proposed approach can aid developers by minimizing the search space when trying to pinpoint software components that are responsible for causing software bugs. Since a bug report in GitHub is construed as a query (in verbal form) and the source code in the GitHub repository are viewed as a collection of documents, the problem of finding source code that corresponds to a bug report can be addressed using information retrieval (IR) technique. The candidate is expected to explore state-of-the-art techniques in mapping unstructured data (bug reports) to structured data (source code or classes) using IR techniques. The candidate is also expected to be familiar with object-oriented programming languages and possesses skills in scripting technologies.

GitHub is arguably one of the most used SCMSs with more than 10 million active projects in the GitHub repository (Kalliamvakou et al., 2014). GitHub provides a variety of services, among which the most important ones are the version control and bug tracking systems. The version control system (VCS) stores all versions of the source code, while the bug tracking system (BTS) provides a unified interface for reporting errors. Bug reports can be used to identify faulty and previously fixed source code. The VCS, on the other hand, can also indirectly highlight certain bug fixing commits that are otherwise not reported in the BTS. Hence, both the version control and bug tracking features in GitHub contain information that could help in fixing software bugs, with the overall aim of improving software quality.

A GitHub developer who is assigned to address a particular bug report usually needs to reproduce the software bugs and perform code reviews in order to highlight the classes or methods that are responsible for causing the anomaly. However, the diversity and uneven quality of bug reports in large projects can make this process tedious. Essential information is often missing from the bug reports, especially in GitHub where bug reports created by developers are usually in verbal form. Bug reports in GitHub are usually annotated in a discussion forum to allow developers to discuss and verify the reported bugs. While it is useful for discussion, the bug reporting system in GitHub does not contribute directly toward finding the software components that are responsible for causing a particular software bug or fault.

**Expected Outcome:** A software prototype that supports this research.
Tools/Programming languages to be used: Any suitable object-oriented development and scripting tools. The candidate is expected to be familiar with object-oriented programing languages and possesses skills in scripting technologies.

References


3. **Title of project:** Automated Approach for Recovering Unreported Software Bugs in Source Code Management Systems with the Aid of Weighted Complex Networks

**Objectives(s) of project:**

- To propose an approach inspired from weighted complex network for automatically recover unreported bugs in a source code management system (SCMS).
- To propose a method to correlate software metrics and graph theory metrics to fault proneness of a software system for revealing fault-prone components.
- To automate the proposed approach.

**Brief description:**

When there is a software change request, the decision-making process is often driven by prior experience and intuition of software maintainers. To rely less on the experience and intuition of specific maintainers, researchers have proposed techniques to analyse the characteristics of previously identified bugs in software repositories, such as source code version control and bug repositories, to predict and isolate bug-prone software components (Gyimesi, Gyimesi, Tóth, & Ferenc, 2015). To fully utilize these repositories, software developers have started to adopt a SCMS as a medium of collaboration to track, audit, and report software bugs and defects. However usually not all bugs are reported because SCMSs solely depend on maintainers to manually report and annotate bugs in commit logs (Romo & Capiluppi, 2015). If bug reports are out of sync due to wrong reporting, the data mined from such bias and incomplete repositories might deliver false negative bug prediction. Hence, it is important to identify the nature and extent of the bias, and develop an approach to aid in ensuring the integrity and validity of data mined from a SCMS.

In this research, a foundation inspired from weighted complex network to automatically recover unreported bugs in a SCMS is proposed. First, a weighted complex network is created to illustrate the interactions of all software components in a software system from a graph theory point-of-view. Subsequently, software metrics and graph theory metrics are correlated to fault proneness of the software are applied to reveal fault-prone components. Finally, an automated reporting tool will annotate any unreported bugs to the SCMS based on the findings. The proposed approach can aid in minimising the gap between the reported and actual bugs to reduce false negative bug prediction and thus minimising maintenance costs.

**Expected Outcome:** An approach and method inspired from weighted complex network for automatically recover unreported bugs in a source code management system.

**Tools/Programming languages to be used:** Suitable development tools
References


1. **Title of project:** Analysis, Classification and Extraction of Share Market News Releases

**Objectives of project:**

1. To analyse and classify the different types of share market news releases according to its main focus which include release of dividend, change of board of directors of a company, launching of new product, news about price drop of petroleum, etc., by companies listed in the main market of BURSA Malaysia. Then, develop a share market news classification database.

2. Based on the news releases by companies listed in the main market of BURSA Malaysia, develop a news extraction system that incorporates text analytic techniques to extract the essence of the news, and classified them according to the classes defined in No. 1 above, together with the details of the news.

   For example, for news about release of dividend of a company, the details that should be recorded include the date the dividend news is released, the company name and stock code, the amount or percent (%) of dividend, government tax (if dividend is not tax exempted, and the percent of tax imposed), entitlement date, Ex-date, payment date of dividend, etc.

3. Analyse 200 or more news releases using the news extraction system to evaluate the accuracy of share market news extraction and classification results.

**Brief description:**

The share market prices of all companies are affected by news releases. The news could produce a positive or negative impact on the market price of a particular company in a country or create a global impact on the market prices of all companies worldwide. This research aims to develop a classification scheme to catalog the different types of news releases according to its main focus so that its impact could be analysed after the news have been released.
Students shall first analyse all the different types of share market news to develop a comprehensive classification scheme to categorise the news. A news (text) extraction system shall be developed to retrieve the main focus/issue of the news. The details of the news will be recorded systematically for the use of prediction of share market prices based on news releases.

Expected outputs:
1. A comprehensive classification scheme of share market news releases.
2. A news extraction system to classify the share market news according to the classification scheme developed.

Tools/Programming languages to be used: Suitable programming language(s), database, development tools, graphical tools, and any other related tools and technologies. However, students are required to discuss with project supervisor and all other project team members who embarked on this research to decide on the programming languages, tools and technologies to be used.

Number of Students: 1

Status: New
2. **Title of project:** Share Market Price Prediction Model based on Company News Analyses

**Objectives of project:**

1. To analyse the trend of share market prices of a company listed in the BURSA Malaysia after the news about dividend or other share-related news have been released by companies from a specific sector, using data mining technique.
2. To establish different share market price prediction models for different sectors based on the findings from the share market price analyses obtained from No. 1 above.
3. To develop a share market price predicting system that incorporates the share market price predicting model (from No. 2 above).
4. To evaluate the accuracy of the share market price predicting model using the share market prices of the companies listed in the BURSA Malaysia website.

**Brief description:**

This research aims to formulate different share market price prediction models (it can be a mathematical formula) by analysing the trends of the share market prices after the dividend news or any other share-related news have been released by the companies from different sectors. Students must choose companies listed in the main market of BURSA Malaysia from one of the sectors such as finance, properties, hotels, trading/services, etc. Based on the trends (findings) of the past 10-20 years of share market prices, establish different share market price prediction models to predict the increase or decrease in the share market prices of the companies after the companies have released their next dividend news or any other share-related news such as share offer for sale, or news about rights issue.

Students shall develop a mobile application that incorporates the different prediction models. The system shall record details about the companies (company name and stock code), the profits gained by the companies for the financial reporting period, and details about the dividend (% of dividend, government tax (if dividend is not tax exempted, the % of tax imposed), payment date of dividend, etc.). The system shall show the predicted trends in tabular and graphical formats. Both formats shall show the predicted percent of increase or decrease in the share market prices from the next day that the dividend news or share-related news have been announced until the payment date of the dividend or the related event (e.g.: rights issue date). The accuracy of the different prediction models shall be evaluated using the data from the BURSA Malaysia Website or related sources (i.e. using the past as well as the current year's share market prices, dividend data, etc.).
Expected outputs:

1. Different share market price prediction models to predict and show the trends of the share market prices of companies from different sectors for the period after the next company dividend news or any share-related news have been released until the dividend payment date (period DAD-DPD) or related event date (period SRAD-RED).

2. A share market price prediction system (mobile application) to illustrate and predict the trends of share market prices of the companies from different sectors for the period DAD-DPD or SRAD-RED.

Note: DAD-DPD: Dividend Announcement Date-Dividend Payment Date; SRAD-RED: Share-Related Announcement Date-Related-Event Date

Tools/Programming languages to be used: Mobile development tools, graphical tools, and any other related tools and technologies. Students are required to discuss with project supervisor and all other project team members who embarked on this research to decide on the programming languages, tools and technologies to be used.

Number of Students: 1

Status: Old
Lecturer: Associate Professor Dr. Zarina Mohd Kasirun

Email: zarinahmk@um.edu.my

1. **Title of project**: A multi-criteria decision making approach for legacy systems modernization

**Objectives(s) of project:**

- To investigate approaches and functions in multi-criteria decision making tools and systems in supporting legacy systems modernization.
- To formulate a multi-criteria decision making approach for modernizing legacy systems.
- To develop a proof-of-concept to evaluate the approach.

**Brief description:**

This project focuses on investigating multi-criteria decision making approaches for modernizing legacy systems. The candidate is required to review existing approaches that typically consider useful functions. Among them are genetic algorithm and desirability function. Desirability Function for example is able to provide a unified metric representative of the suitability between the complete set of ranking-factors set available and organization weighted-needs for the systems. This way, decision makers can quantitatively assign needs for the systems even when the most desirable factors are not available from the existing systems.

**Expected Output:**

A multi-criteria decision making approach for modernization legacy system

**Tools/Programming languages to be used:** Any suitable tools.

**Number of students:** 1

**Status:** New
Lecturer: Dr. Mumtaz Begum Peer Mustafa
Email: mumtaz@um.edu.my

1. Title of project: English to Malay speech translation system as a communication aid for tourist.

Objectives(s) of project: This research will identify and develop suitable techniques for converting English speech to Malay speech with acceptable conversion accuracy.

Brief description:
Malaysia is one of the well-known tourist destination in the world, with people from many different countries visit Malaysia annually. One of the great attraction is our hospitality where we give our foreign visitors a warm welcome the Malaysian way. However, one of the problem faced by foreign visitors in Malaysia is communication with local community particularly the suburban. While English can be considered as second language in Malaysia, not all Malaysian has the luxury to understand and speak in English. As such a speech-to-speech translation engine can be an effective tool of cross lingual communication particularly for English speaking foreigners to communicate with the locals in Malay. The speech-to-speech translation is highly demanded as it can benefit billions of people around the world. There have been very few attempts to develop the translation engine for Malay locally with acceptable conversion accuracy. This research aims at identifying suitable techniques that can be applied in converting English speech to Malay speech with acceptable conversion accuracy.

Expected Output: English-Malay speech-to-speech translation system, translation techniques

Tools/Programming languages to be used: Speech Technology toolkits: HTK, HTS

Number of students: 1

Status: New
2. **Title of project:** An Automated Academic Performance Prediction System for University students using Academic Factors

**Objectives(s) of project:** The main objective of this project is to design an automated academic performance prediction system for university students using the identified academic factors.

**Brief description:**

The aim of every university is to produce students with acceptable academic performance so that the graduating students have better employability. However, students, on their own may not be able to plan their academic progress, and different student has different capability. An automated academic performance prediction system for university students can assist the management in identifying potential students that can excel in education as well as students that require special attention of the faculty. This is because a good university will not neglect its weakest students. The proposed system will predict a student’s performance in term of GPA and CGPA based on certain academic factors such as past performance, sociological background, past courses as well future courses to be taken. The system will also allow the management to keep track of the performance of a student over a period of time (semester, yearly and so on).

**Expected Output:** Academic Performance Prediction System, Prediction techniques, Academic Factors

**Tools/Programming languages to be used:** Any suitable programming language and data mining tools

**Number of students:** 1

**Status:** New
1. **Title of project:** Improving code quality of End-User Programming through design of tool support.

**Objectives(s) of project:**

1. To identify Software Engineering Knowledge (SEK) that should be taught to end-user programmers to increase the quality of code they produced.
2. To propose design of tool support that teaches SEK to end-user programmers.
3. To implement the proposed design by developing a prototype.
4. To evaluate the proposed design and prototype.

**Brief description:**

End-user programmers differ from professional developers in that end users' programs are not the end in itself, but rather a means to perform their own tasks [1]. Examples of end users developing software are when they use spreadsheet applications, web authoring tools, and when they develop educational simulations, spreadsheets, and e-business web applications [2]. Many end-user programmers lack formal computer science training and the software they developed contains many errors [3]. This research aims to improve the quality of code produced by end-user programmers by designing tool support that teaches SEK to these non-professional developers.

**Expected Output:**

1. A model showing the relationships of SEK and quality of code.
2. Design of tool support that teaches SEK to end-user programmers.
3. A prototype tool.
4. Evaluation results of the proposed design and prototype.

**Tools/Programming languages to be used:** To be determined. This might involve customising or extending or integrating suitable existing tools.

**Number of students:** 1
Status: New

References:


2. **Title of project:** An Approach for Collaborative Chunking of Software Design Information

**Objectives(s) of project:**

1. To investigate how information related to the software design of a system can be chunked or grouped together based on the collaborative usage of the information.
2. To propose an approach for usage-driven collaborative chunking of software design information.
3. To develop a prototype tool that supports the proposed approach.
4. To evaluate the approach and prototype.

**Brief description:** During the development of a software system, project team members frequently access information related to the software design of the system and evaluate its usefulness for the task-at-hand. The evaluations happen implicitly within the minds of the team members, are often forgotten as time passes and remain non-transferable between team members. The knowledge of the usefulness of the information can be re-used if team members provide evaluation information explicitly, for example, by providing ratings, comments, and tags, on the evaluated information. This evaluation data as well as the access data (for examples, frequency of access, and time spent on information) can be analysed and subsequently used to chunk or group information, for easier retrieval of the set of information required for a particular task.

**Expected output:**

1. An approach for usage-driven collaborative chunking of software design information.
2. A prototype tool.
3. Evaluation results of the proposed approach and prototype.

**Tools/Programming languages to be used:** To be determined. This might involve customising or extending or integrating suitable existing tools.

**Number of Students:** 1

**Status:** Old
1. **Title of project:** Evaluation framework/method that compare and select a combination of software systems for data analytics pipeline

**Objectives(s) of project:**

1. To review, investigate and validate existing software tools for supporting the tasks in a data analytics pipeline.
2. To identify selection criteria from desirable features and quality attributes to compare and evaluate software systems for data analytics pipeline.
3. To develop a decision-making evaluation framework/method to compare and select a combination of software tools to support all the stages in a data analytics pipeline.
4. To evaluate the framework/method by conducting a case study for cardiac simulation and modelling data analytic pipeline.

**Brief description:**

Data analytics is a broader term that includes data analysis as necessary subcomponent. Analytics defines cognitive processes an analyst uses to understand problems and analyse data in meaningful ways. Specific tools, techniques, and methods are used to perform analytics and communicate results successfully. Many real-world data analytics scenarios involve an integration of multiple data analytics stages, which are often carried out using different software tools. In this research, we call the integration of multiple data analytic stages as “Data Analytics Pipeline”. A data analytics pipeline is a set of data analytics processing stages connected in series, where the stages reconnected one to the next to form a pipe. The output of one stage is the input of the next stage. The stages of a data analytic pipeline are often executed in parallel or overlapped in execution. Selecting a combination of data analytics tools is complex. Problems arise when various tools available to support each stage of a data analytics pipeline. Each tool provides a set of features and can process certain data. Some tools are able to handle more than one task. Researchers are facing problems in selecting a combination of software tools to be used in a series of stages throughout the data analytics process.

Techniques such as decision trees, fuzzy logic, case base reasoning, hybrid knowledge-based system and artificial neural network are well known approaches in software selection. However,
most of the existing studies focus on selecting one software tool at a time, and not a combination of software tools. In a data analytic pipeline, different software tools are used in acquiring data, importing data (supported format) into the tool, conducting pre-processing, performing the analysis, visualize and reporting results, or automating repeatable tasks. A tradeoff between features and quality attributes need to be analyzed to select a combination of tools that allows the development teams to work with their project needs and constraints.

In this research, an evaluation framework/method will be proposed by applying decision making methods such as decision tree, case base reasoning and fuzzy logic to discover the hidden patterns, associations, and anomalies of the characteristics of different kinds of data analytics software systems. The discovered knowledge can be used for assisting the software development teams and researchers who are new to data analytics to optimize the selection of a combination of data analytics software tools to meet specific project or domain needs.

![Figure 1: A Generic Data Analytics Pipeline](image)

**Expected Output:**

In this research, an evaluation framework/method will be proposed to compare and select a combination of data analytics software tools that can be used to handle a series of tasks in a data analytics pipeline based on the identified criteria or constraints.

**Tools/Programming languages to be used:** Any suitable tool/Programming language to develop the method used to compare and select a combination of data analytics software tools.

**Number of students:** 1

**Status:** Old
2. **Title of project:** Process reference model/framework to develop machine learning (ML) based systems

**Objectives(s) of project:**
1. To identify the differences between ML based systems development and generic software development.
2. To analyse the best practices or steps used to ensure the statistical validity of the ML components in the ML based systems.
3. To develop a process reference model/framework to integrate ML component development process into generic software development process to ensure the statistical validity of the ML components.

**Brief description:**

Machine learning (ML) will enable cognitive systems to assist us in making good decisions by bringing the right recommendation to us in a more natural and personalized way. Furthermore, many ML algorithms have been applied to analyse big data quickly and automatically in recent software systems development. However, many experienced software engineers are experts in generic software development but they are a novice at ML-based system development and unfamiliar with the methodology of implementing ML components. On the other hand, a novel ML method is typically proposed by a ML researcher or team of researchers. A problem is that the code is written by researchers that may or may not be trained in the discipline of software development to ensure the statistical validity of the ML components.

There are various risks that can be identified and need to be addressed when we integrate ML-based component(s) into a software system. Sculley et al. (2014) have identified several ML specific risk factors that need to be avoided or refactored during ML based system development. These include boundary erosion, entanglement, hidden feedback loops, undeclared consumers, data dependencies, changes in the external world, and a variety of system-level anti-patterns (Sculley et al, 2014). There is a change of software engineering role in the ML based system development. The development teams need to ensure that the requirements that we expect from ML-based systems can still be met when the ML-based components keep involving. The outputs produced by ML components should always meet the statistical validity specified in the requirements. For example, ML-based system may deal with big business/personal data which is different from dealing with scientific data. There are unchanging scientific laws underlying scientific data while the learned laws using big business/personal data are not only more fluid but also the produced predictive results that may "change" the end users and the underlying learned rules over time.
This research aims to address the following research questions:

- What are the software development practices that could make a big difference when experimenting and testing ML algorithms?
- What are the good ML development practices that are applied by ML researchers to ensure the statistical validity of ML components?
- How to integrate the practices or steps used in ML components development into generic software development processes?

This research aims to propose a conceptual process reference model/framework for ML based systems, that support software engineers, ML researchers, statisticians and/or data analyst in developing and maintaining a ML based systems to ensure the statistical validity of the ML components. The model/framework should consist of a holistic life cycle model for ML based systems, product quality criteria and metrics for ML components, best practices for different stakeholders, and recommendations for action, as well as tools that support stakeholders in developing and maintaining ML-based systems.

**Expected Output:**

In this research, a conceptual process reference model/framework will be proposed for supporting ML based systems development to manage the ML component development and ensure the statistical validity of the ML components.

**Tools/Programming languages to be used:** Any suitable tool that can be used to demonstrate the new process framework.

**Number of students:** 1

**Status:** Old

**Reference:**

3. **Title of project:** Analysing and Predicting Cardiovascular Diseases in Malaysia

**Objectives(s) of project:**

1. To identify challenges in analysing and predicting risks of cardiovascular diseases in Malaysia
2. To review, investigate and validate existing data mining techniques for cardiac-based clinical data
3. To propose a model for analysing and predicting the risks of cardiovascular diseases in Malaysia
4. To develop an analytic dashboard to present retrospective and prospective analysis to show the trends of cardiovascular diseases of Malaysian in the past and predict the future trend in the future.

**Brief description:**

Cardiovascular disease is one of the main causes of death in Malaysia. Early diagnosis, detection and treatment of cardiovascular diseases can help to reduce the rate of mortality in Malaysian. It is important to address the risk factors such as tobacco use, unhealthy diet and obesity, physical inactivity, high blood pressure, diabetes and raised lipids among the Malaysian population. It is important to monitor the Malaysian population characteristics by age, gender, state, symptoms and other demographics to identify the movement and concentration of cardiac patients population in Malaysia with certain type of cardiovascular diseases or condition. Clinical intelligence is a new way to improve patient care outcomes, quality of health care and patient population health by applying a set of computer-enabled methods, processes and discipline to transform clinical data into meaningful clinical that gives new insight, discovery and knowledge.

National cardiovascular disease databases have collected large amount of clinical data related to cardiovascular diseases in Malaysia. These cardiac-based clinical data can be analysed by applying data analytics methods to extract the hidden knowledge and new discovery. The aim of this research is to propose a model based on data mining techniques to classify patients into various cardiac disease categories (e.g. by age, gender, race, state/postcode, type of cardiac disease, risk profiles and other demographics). Also, retrospective and prospective analysis will be conducted to show the trends of cardiovascular diseases of Malaysian in the past and predict the trend in the future. The real-life cardiac-based clinical data analytics will contribute to detect abnormal and outlier cardiac patients, identify the effectiveness of the existing clinical practice in treating the cardiovascular diseases and predict the trend of the cardiovascular diseases based on Malaysian cardiac patient population characteristics. Hence, this leads to the potential of generating a knowledge-rich environment that can help to rapidly improve the quality of clinical decisions.

This research also contributes to provide interactive analytics dashboard present the results of retrospective and prospective analysis (e.g. a geographic map dashboard to show the Malaysian
population with myocardial infarction, cardiac and pre-cardiac condition). The analytics dashboard allows Ministry of Health, cardiac health professionals, healthcare providers and other clinical staffs to improve make more effective clinical and healthcare decisions, improve population health management and monitor cardiac patient population by proposing new strategies and policies.

**Expected Output:**

A model for analysing and predicting the risks of cardiovascular diseases in Malaysia and an analytic dashboard to present retrospective and prospective analysis.

**Tools/Programming languages to be used:** Any suitable data analytic tool for data processing and programming language to develop a

**Number of students:** 2

**Status:** New
4. **Title of project:** Mining Mobile App Reviews: Mobile App Quality Feedback Analysis

**Objectives(s) of project:**

1. To analyse the way how users express the quality issues through app reviews.
2. To identify the keywords and topics to extract mobile app quality information from app reviews.
3. To develop a quality feedback model with a set of pre-defined linguistic rule and classification to group the reviews according to software quality attributes.
4. To evaluate the proposed model using the sample of mobile app reviews from different categories of mobile apps.

**Brief description:**

User reviews are valuable feedback, requests and ideas getting directly from mobile application (app) users. Mobile app reviews are text-based and contain information and various topics (e.g. user experience, bug reports and user ratings). User feedback are crucial to reasons why users like or dislike a mobile app in the market. However, the huge number of user reviews are unstructured data that consists of a lot of text-based content that need to be analysed. It is challenging to extract useful information and provide constructive feedback for app companies and their development teams to improve the quality of their apps.

Most of the existing studies focus on analyzing the positive and negative feedback from users to identify the user preferences, user behaviours and new feature requests from the users. However, the feedback content also has an impact on software quality. Besides enhancing the functionality and usability of mobile apps, more attention should be paid on improving other software quality aspects (e.g. reliability, performance, security, maintainability) of the mobile apps. To the best of our knowledge, there is no work that have been done to do further analysis on the user reviews to extract specific quality information for each quality attribute. It is important to make use of the user feedback and extract useful quality feedback to provide insights of the app quality.

In this research, text mining techniques will be used to analyse the user reviews identify the ways how users express their feedback on software quality. Keywords and topics will be identified to retrieve quality information and match them to specific quality attribute (i.e. functional suitability, reliability, security, performance, compatibility, portability and usability). A quality feedback model will be proposed to (a) extract quality information from user reviews using a set of pre-defined linguistic rule; (b) classify the quality issues into specific quality attribute; and (c) provide feedback regarding the users’ major concerns on software quality of a mobile app. A tool will be developed
to perform the data analytics from raw user reviews automatically and to evaluate the accuracy of the model using sample of mobile apps reviews from different types of mobile apps.

**Expected Output:** A conceptual quality feedback model and a tool to demonstrate the model.

**Tools/Programming languages to be used:** Any suitable text mining Tool or any suitable programming language to develop a tool for mining mobile app reviews

**Number of students:** 1

**Status:** New
Lecturer: Dr. Raja Jamilah Raja Yusof

Email: rjry@um.edu.my

1. Title of project: Automatic assessment of programming lessons through physical games

Objectives(s) of project:

1. To identify fundamental concepts of programming to be used for physical games module
2. To design and develop physical games module to enhance the understanding of fundamental concept of programming
3. To implement and evaluate the transfer of knowledge of programming using physical games through automatic assessment system

Brief description:

(This project is funded by University Malaya through UM-LiTer grant 2016 and therefore candidate will be paid for working through this project)

Many application software has been developed to teach people to program such as BlueJ (Barnes & Kölling, 2016; Kölling, Quig, Patterson, & Rosenberg, 2003) developed in Australia and Scratch (Resnick, et al., 2009) in MIT, United States of America. Some are available online and some requires us to buy the licence to access the features and functionalities. These are software that covers fundamental topics of programming such as variables, if statements, loops, arrays, pointers and data structures (Farrell, 2014). These softwares are useful although the associated users did not claim any significant difference between those students using the software and not using them because of reasons associated to ethics and difficulties of setting the learning environment in using these softwares in their courses for the purpose of experimental studies (Kundle & Allen, 2016).

We propose playing physical games using learner's body movement or some part of their motor control (such as the hands, body and legs) in contrast to purely computer-based or board games as a tool to teach programming. The motivation behind this approach is firstly based on learning style theories that many students are of the kinaesthetic typed compared to the visual and auditory
(Klement, 2014) and this is also true in the context of computer science students (Ates & Altun, 2008). Kinaesthetic students process new knowledge (or understanding) with the involvement of their body movement. In context of teaching programming, we feel that most of the teaching approaches only benefit the visual and auditory students even though many are among the kinaesthetic typed.

Therefore, this project aims to teach programming using physical games modules and then evaluate students’ understanding (of the programming concepts) using an automated system to assess their understanding after playing the designed physical games.

Expected Output:

1. A module to teach programming concepts through physical games
2. An automated assessment system to evaluate students’ understanding of the above module

Tools/Programming languages to be used: Java/ Web-based platform languages

Number of students: 1

Status: New

References:


2. **Title of project:** Heuristics to evaluate interactivity effects on text memorization retention after using different medium of writing: the pen, the laptop and the smart phone.

**Objectives(s) of project:**

1. To investigate the effect of text memorization retention:
   - After writing the text using the pen
   - After typing the text using the laptop
   - After typing the text using the smart phone

2. To study the above using written text in Malay, English and Qur’anic Arabic.

3. To develop and evaluate a prototype to measure memorization effectiveness base on the above input

**Brief description:**

*(This project is funded by finishing HIR grant on Quranic education for human development through ICT year 2014-16)*

One of the usability goals in Human Computer Interaction is efficiency, how quickly one is able to perform certain tasks (Nielsen, 2012). In context of educational settings using technology, related memorization study are very minimal although the result can give a major impact in setting up guidelines on supporting memorization using technology. In the context of text memorization, such as historical facts, language rules such as in the Arabic language, formulas and others, many memorization approaches can be used to achieve memorization of these items. Some example are using Connect & Link (The Link Method), Make a Story (The Story Method), Associate Objects with Familiar Locations (The Loci Method) and others (Howes, 2015). In the context of Muslim culture, the memorization of the Qura’nic Text is also done through writing. According to a recent study (Mueller, 2014) using pen and paper, NOT laptops, to take notes boosts memory and the ability to retain and understand concepts. However, this study concentrate on notes taking while in a lecture setting. This study however, propose to investigate memory retention ability of a given text (in Malay, English and Qur’anic Arabic languages) after the text is written using pen and typed using laptop and smart phone. A Qualitative research study is proposed. A prototype is expected to be developed and evaluated for measuring the effectiveness of the three input styles for memorization through writing.
Expected Output:

1. The report on a paragraph of text (in Malay, English and Qur’anic Arabic) retention after using the pen, laptop and smart phone as a medium to support memorization.
2. A prototype for measuring memorization effectiveness
3. A heuristic guideline for developing memorization software through written text

Tools/Programming languages to be used: Java/ Web-based platform languages, touch screen display devices

Number of students: 1

Status: New

References:


3. **Title of project:** Case study on software development methodology and its effect to the design and development of user interfaces.

**Objectives(s) of project:**

1. To conduct a case study on software project in an identified company
2. To identify criteria of choosing a methodology for a given project and its effect to designing and developing user interfaces
3. To develop and evaluate prototype that helps software project managers to select a specific methodology

**Brief description:**

(This project is possibly being funded by Minister of Higher Education (MOHE) under the Case Writing Grant Scheme (CWGS) 2016, therefore candidate would possibly be paid for working through this project)

Survey results indicate that although agile methodologies are more prevalent than 10 years ago, traditional methodologies are still popular. Organizations also use multiple methodologies on projects. Furthermore, their choice of methodologies is associated with certain organizational, project, and team characteristics[1]. Theoretically, design and development of user interfaces should be affected according to the methodology chosen for example, rapid development methodology would result in brief interfaces compared to development methodologies which are more iterative and longer cycle periods. Sometimes it is also the case that there are too many request of change while the software development process is in progress. Such cases may be difficult to handle and certain methodology may be more advantage to be adopted. The effect of user interface development is to be studied based on the choice of methodology chosen.

**Expected Output:**

1. A case study report
2. A prototype for project managers to select a methodology which include predicted user interface development issues

**Tools/Programming languages to be used:** Java/ Web-based platform languages

**Number of students:** 1
Status: New

References: