**MULTIMEDIA UNIVERSITY**

**FACULTY OF COMPUTING AND INFORMATICS**

**FINAL YEAR PROJECT**

**Title Change Requisition Form:**

Note: (i) Title change has to be approved by the supervisor and FYP Committee (ii) Students cannot change the title during FYP2 (under normal circumstances).

**Student ID:** 1181203334

**Student Name:** Mak Yen Wei

**Project ID:**

**Student and Project Specialisation (Note: Student and Project Specialisation should match)**

**(Select one)**

Data Science

**Approved project Title:** Time Series Based Text Summarization

**New Project Title:** Forum Text Processing and Summarization

**Justification for Changing the Title (Give specific reasons, after consultation with your supervisor)**

The reason for the title change is based on the overall scope we’re working on, previously the title was Time Series Based Text Summarization, and was only limited to time series and summarization. Since then, upon discussions after discussions with Dr Goh Hui Ngo, my Supervisor, we agreed that it is wise to also explore other contexts of the data from the forums, such as sentiment analysis, topic modeling, scoring and so on. In addition, the focus in this work is on forums, therefore the title is changed to “Forum Text Processing and Summarization”.

**Project Type: Research-based**

**Project Description: (Note: If there is no change, state ‘NIL’ , Highlight the change, if any)**

**Previous Description**

Background: Applicable to the tech industry, specifically in software. Bugs and issues are inevitable in the software world. Stackoverflow provides a platform for developers around the world to ask and answer questions related to problems that engineers face while developing software.

Problem Statement: Stackoverflow answers can be crowded with answers that are too old, or no longer applicable due to a new release of a package or language. Thus, developers spent hours and hours debugging the bug or issue with no versioning context of the solution.

Methodology: In this research, we attempt to develop an application to analyze time series based texts found in stackoverflow and merge the data with the github issues tab to provide versioning context of the package or language used. This brings features such as language filters, versioning filters and much more to help developers to find the solution way faster. Datas from both domains will be merged, analyzed and summarized into a Q&A format.

Expected Output: A web application for users to find solutions with ease. Data will be aggregated from both stackoverflow and github issues and will be analyzed based on the context of time and topic.

**New Description**

Background: Applicable to the tech industry, specifically in software. Bugs and issues are inevitable in the software world. Online forums such as Stackoverflow, Github, Reddit and Twitter provide a platform for developers around the world to ask and answer questions related to problems that engineers face while developing software.

Problem Statement: Forums can be crowded with loads of un-related, incorrect and outdated discussions. It is particularly hard to locate the right answers, discussions in the software engineering domain.

Methodology: In this work, we attempt to propose a framework for gathering, processing and summarizing these discussions in online forums in the hopes that it helps software engineers to better understand and locate the right discussions and answers easily in such crowded forums.

Expected Output: A framework that helps users to find solutions with ease. A frontend application integrated with the framework.

**Project Objectives: (Note: If there is no change, state ‘NIL’, Highlight the change, if any)**

**Previous Objective**

Developing an application to analyze time series based texts and using string similarization to summarize a combination of answers coming from stackoverflow and github issues. Datas from both domains will be merged, analyzed and summarized into a Q&A format

**New Objective**

Proposing a framework for gathering, processing and summarizing discussions in online forums.

**Project Scope (Note: If there is no change, state ‘NIL’, Highlight the change, if any)**

**Previous Scope**

Expected output: A web application for users to find solutions with ease. Data will be aggregated from both stackoverflow and github issues and will be analyzed based on the context of time and topic.

Limits: Stackoverflow and github issues are the only domains that we considered for data aggregation.

**New Scope**

Expected Output: A framework for gathering, processing and summarizing discussions in online forums.

Platforms: Stackoverflow, Github, Reddit and Twitter.

**Number of Students in the project (Select one):**

One

**Two Student Project Details:**

**Student 1 Subtitle (Note: If there is no change, state ‘NIL’)**

**Student 1 Work Distribution (Note: If there is no change, state ‘NIL’, Highlight the change, if any)**

**Student 2 Subtitle (Note: If there is no change, state ‘NIL’)**

**Student 2 Work Distribution (Note: If there is no change, state ‘NIL’, Highlight the change, if any)**

**Recommendation from the Supervisor:**

**Comments from the Moderator (Not applicable during FYP1):**

**Received by the FYP Committee on**

**Comments from FYP Committee:**