

## BORANG PENGESAHAN STATUS TESIS

JUDUL: Text Processing of Student's Project Content Classification


SESI PENGAJIAN: 2006/2007

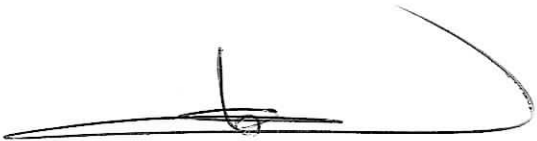
Saya Sultan Syed Ibrahim Bin Sahol Hamid  
(HURUF BESAR)

mengaku membenarkan tesis (PSM/Sarjana/Doktor Falsafah) ini disimpan di Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Kolej Universiti Teknikal Kebangsaan Malaysia.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. \*\* Sila tandakan (/)

<u>          </u>	SULIT	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)
<u>          </u>	TERHAD	(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)
<u>  /  </u>	TIDAK TERHAD	

  
(TANDATANGAN PENULIS)  
Alamat tetap : No. 25 Jalan 1, RRM PKNS,  
Batu 16 48000 Rawang, Selangor D.E.  
Tarikh : 27 Oktober 2006

  
(TANDATANGAN PENYELIA)  
En. Mohd. Sanusi Bin Azmi  
Nama Penyelia  
Tarikh : 22/11/2006,

CATATAN: \*\* Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa.  
Tesis dimaksudkan sebagai Laporan Projek Sarjana Muda (PSM)

raf

QA76.76.D47 .S94 2006



0000039094

Text processing of student's project content classification /  
Sultan Syed Ibrahim Sahol Hamid.

## TEXT PROCESSING OF STUDENT'S PROJECT CONTENT CLASSIFICATION

SULTAN SYED IBRAHIM BIN SAHOL HAMID

This report is submitted in partial fulfillment of the requirements for the  
Bachelor of Computer Science (Software Development)



FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY  
KOLEJ UNIVERSITI TEKNIKAL KEBANGSAAN MALAYSIA  
2006

## DECLARATION

I hereby declare that this project report entitled

### TEXT PROCESSING OF STUDENT'S PROJECT CONTENT CLASSIFICATION

is written by me and is my own effort and that no part has been plagiarized without citations.

STUDENT :   
(SULTAN SYED IBRAHIM SAHOL HAMID)  
SUPERVISOR:   
(MOHD SANUSI BIN AZMI)

Date : 20 November 2006

Date : 22/11/2006

## DEDICATION

First and foremost I would like to dedicate my heartiest appreciation to my family, especially to my beloved mother and sister, Khatijah Said Mohammed and Rahima Beevi Sahol Hamid for their support and guidance. My special thanks is also dedicated to En. Mohd. Sanusi for his dedicated supervising through this thesis completion.

This special dedication also goes to the entire person who supports me with their concern during completing this project. Thank you very much.

## ACKNOWLEDGEMENTS

First and foremost, I thanked Allah for blessing me to complete Projek Sarjana Muda 1 (BITU 3973). I would like to enlarge my appreciation to Mr. Mohammad Sanusi Azmi and Mrs. Rosmiza for their kindness to accept me as one of the student under their supervision. Special thank also dedicated to both of them for all comments, idea, and guidance for completing this thesis.

This appreciation also goes to all friends that always give support, opinion, and advices for me to complete this report especially my friend under Mr. Mohammad Sanusi Azmi supervision.

To my beloved family, I would like to forward my obliged to them for their continuous support during my study period, their patience and benevolence. Lastly, I would like to thank to everyone who has contributed during my Project Sarjana Muda. Your kindness and cooperation in completion of paper work is much appreciated.

## ABSTRACT

This project papers is discuss about “*Text Processing of Student’s Projects of Contents Classification*”. The project will be implemented specifically for lecturers to search information about student’s projects for their reference purpose. The focus of this project will be on searching the student’s projects and categorized text according to project abstract. This project will be developed by using Java platform and TCP/IP programming for some of the module. Besides that, the project aims is to use a client/server to develop this application framework. Thus, it provides users a search engine function and a display the required results. Paper begins by discussion how the system flow and the problem face nowadays on current system which is manual searching process of student’s projects information. A few case studies also include in this paper to gather more exact information. Paper continues in discussing analysis and design for this project. Object oriented Analysis & Design (OOAD) is chosen as a methodology to develop this project. Unified Modeling Language (UML) uses to apply in Object-Oriented Approach and Rational Rose 2002 is a tool to develop the diagram. The users of this project are admin and KUTKM lecturers. This project is hope to bring and adapt a high quality technical solution in KUTKM lecturers and provides them with better services toward computing world.

## ABSTRAK

Kertas projek ini adalah khas untuk *Pemprosesan teks projek pelajar klasifikasi kandungan*. Project ini akan dilaksanakan terutamanya untuk pensyarah bagi tujuan pencarian maklumat mengenai projek pelajar untuk rujukan mereka. Projek ini akan memberi tumpuan dalam pencarian projek pelajar and mengklasifikasikan teks berdasarkan abstract projek tersebut. Projek ini akan dibangunkan menggunakan platform Java dan Pengaturcaraan TCP/IP. untuk beberapa modul. Selain itu, tujuan projek ini adalah untuk menggunakan *client/server* untuk membangunkan rangka aplikasi. Dengan ini, ia membekalkan fungsi enjin pencarian memaparkan maklumat yang dikehendaki. untuk pengguna. Kertas ini bermula dengan membincangkan bagaimana aliran sistem tersebut dan masalah yang dihadapi oleh sistem yang sedia ada pada masa kini yang mana merupakan proses pencarian secara manual. Beberapa kajian kes telah dikaji bagi memperolehi maklumat yang lebih tepat dan bernas. Seterusnya, bab-bab seperti analisa dan rekabentuk akan dibincangkan bagi memperolehi konsep dan idea yang lebih tepat mengenai projek ini. Analisa dan rekabentuk berorientasikan objek (OOAD) telah dipilih sebagai metodologi bagi membangunkan aplikasi ini. *Unified Modeling Language (UML)* digunakan untuk diaplikasikan dalam pendekatan Orientasi-Objek dan Rational Rose merupakan peralatan yang digunakan untuk membangunkan gambarajah. Pengguna aplikasi ini adalah *admin* dan pensyarah KUTKM. Dengan adanya projek ini, adalah diharapkan ianya dapat membawa dan menanam satu penyelesaian teknikal yang berkualiti tinggi kepada pensyarah KUTKM dan memberi perkhidmatan yang lebih baik dalam zaman pengkomputeran

## TABLE OF CONTENTS

CHAPTER	SUBJECT	PAGE
	<b>DECLARATION</b>	<b>i</b>
	<b>DEDICATION</b>	<b>ii</b>
	<b>ACKNOWLEDGEMENTS</b>	<b>iii</b>
	<b>ABSTRACT</b>	<b>iv</b>
	<b>ABSTRAK</b>	<b>v</b>
	<b>TABLE OF CONTENTS</b>	<b>vi</b>
	<b>LIST OF TABLES</b>	<b>x</b>
	<b>LIST OF FIGURES</b>	<b>xii</b>
	<b>LIST OF ABBREVIATIONS</b>	<b>xiii</b>
<b>CHAPTER 1</b>	<b>INTRODUCTION</b>	
	1.1 Project Background	1
	1.2 Problem Statement	2
	1.3 Objective	3
	1.4 Scope	3
	1.5 Project Significance	4
	1.6 Conclusion	5
<b>CHAPTER 2</b>	<b>LITERATURE REVIEW AND PROJECT METHODOLOGY</b>	
	2.1 Introduction	6
	2.2 Fact and Findings	6
	2.3 Project Methodology	7
	2.4 Project Requirements	9



2.4.1	Software Requirements	10
2.4.2	Hardware Requirements	11
2.4.3	Others Requirements	12
2.5	Project Schedule and Milestones	13
2.6	Conclusion	13
<b>CHAPTER 3</b>	<b>ANALYSIS</b>	
3.1	Introduction	14
3.2	Problem Analysis	15
3.2.1	Background of the Current System	15
3.2.2	Problem Statement	16
3.3	Requirement Analysis	17
3.3.1	Functional Requirement	17
3.3.2	Business Flow	19
3.3.3	Use case View	19
3.3.4	Actors	20
3.3.5	Use Case Description	21
3.3.6	Interaction Diagram	26
3.3.7	Software Requirements	27
3.3.8	Hardware Requirements	28
3.3.9	Other Requirements	29
3.4	Conclusion	29
<b>CHAPTER 4</b>	<b>DESIGN</b>	
4.1	Introduction	30
4.2	High Level Design	31
4.2.1	System Architecture	31
4.2.2	Static Organization	33
4.2.3	User Interface Design	35
4.2.3.1	Navigation Design	38
4.2.3.2	Input Design	40
4.2.3.3	Output Design	41
4.2.4	Database Design	42
4.2.4.1	Logical Database Design	42

4.2.4.2	Business Rules	43
4.2.4.3	Data Dictionary	43
4.2.4.3	Deployment view	43
4.3	Detailed Design	44
4.3.1	Software Specification	45
4.3.1.1	Class Detail	45
4.3.1.2	Methods/Operations Detail for Interface	46
4.3.2	Physical Database Design	49
4.3.2.1	Data Dictionary	50
4.4	Conclusion	51
<b>CHAPTER 5</b>	<b>IMPLEMENTATION</b>	
5.1	Introduction	52
5.2	Software Development Environment Setup	53
5.3	Software Configuration Management	53
5.3.1	Configuration Environment Setup	53
5.3.2	Version Control Procedure	58
5.4	Implementation Status	58
5.5	Conclusion	61
<b>CHAPTER 6</b>	<b>TESTING</b>	
6.1	Introduction	63
6.2	Test Plan	63
6.2.1	Test Organization	63
6.2.2	Test Environment	63
6.2.3	Test Schedule	64
6.3	Test Strategy	66
6.3.1	Classes of Test	67
6.4	Test Design	68
6.4.1	Test Description	68
6.4.2	Test Data	70
6.5	Test Results and Analysis	72
6.6	Conclusion	75

<b>CHAPTER 7</b>	<b>PROJECT CONCLUSION</b>	
	7.1 Observation on Weaknesses and Strength	76
	7.1.1 Strength	77
	7.1.2 Weaknesses	77
	7.2 Propositions for Improvement	77
	7.3 Contribution	78
	7.4 Conclusion	78
<b>REFERENCES</b>		80
<b>BIBLIOGRAPHY</b>		81
<b>APPENDICES</b>		82

## LIST OF TABLES

<b>TABLE</b>	<b>TITLE</b>	<b>PAGE</b>
2.1	Software Requirements	10
2.2	PC Requirements for Server Side and Developer	11
2.3	PC Requirements for Client Side and End User	11
2.4	Other Requirements	12
3.1	Login Use Case Description	21
3.2	Select Student's Projects (Admin) Use Case Description	22
3.3	Generate Process Use Case Description	23
3.4	Search Student's Projects (Client) Use Case Description	24
3.5	View Results Use Case Description	25
3.6	Save Results Use Case Description	26
3.7	Software Requirements	27
3.8	PC Requirements for Server Side and Developer	28
3.9	PC Requirements for Client Side and End User	28
3.10	Other Requirements	29
4.1	Input Design	41
4.2	Output Design	41

4.3	Data Dictionary	50
5.1	Computer Requirements for Server Side and Developer	57
5.2	Computer Requirements for Client Side and End User	57
5.3	List of Procedure and Control	58
5.4	Implement Status for User Interface	59
5.5	Implement Status for Database	59
5.6	Authentication User	60
5.7	Implement Status for Select Student's Projects	60
5.8	Implementation Status for Searching Student's Projects	61
6.1	System Configuration and Specification	64
6.2	Test Schedule according Tasks, Activities and Duration to carry out Testing Activities.	64
6.3	Test Description of Login menu	68
6.4	Test Description of Admin Menu	69
6.5	Test Description of Client Menu	69
6.6	Test Data of Login Menu	70
6.7	Test Data of Admin Menu	71
6.8	Test Data of Client Menu	72
6.9	Test Result of Login Menu	73
6.10	Test Result of Admin Menu	74
6.11	Test Result of Client Menu	74

## LIST OF FIGURES

FIGURE	TITLE	PAGE
3.1	Activity Diagram of Current System	16
3.2	Overview of Text Processing of Student's Projects of Contents Classification	18
3.3	Activity Diagram for To-be-System	19
3.4	Global view of use case model	20
4.1	Overview of System Software Architecture based on 3-Tier Architecture.	32
4.2	Overview of To-be-System based on Layering Architecture	33
4.3	To-be-System Packages	34
4.4	Class Diagram for Searching package	34
4.5	User Login	35
4.6	Select Student's Projects Interface (Admin)	36
4.7	Simple Search Interface (Client)	37
4.8	Advanced Search Interface (Client)	38
4.9	Navigation Design for Admin Menu	39
4.10	Navigation Design for Client Menu	40
4.11	Entity Relationship Diagram (ERD) for To-be-System.	42
4.12	Deployment view of To-Be-System	44
4.13	Class LoginForm	45
4.14	Class AdminForm	45
4.15	Class ClientForm	46
5.1	System Properties	54
5.2	Environment Variables	55
5.3	Edit System Variable	56

## LIST OF ABBREVIATIONS

JDK	Java Development Kit
KUTKM	Kolej Universiti Teknikal Kebangsaan Malaysia
OOAD	Object Oriented Analysis and Design
PSM	Projek Sarjana Muda
RMI	Remote Method Invocation
TCP/IP	Transfer Control Protocol/Internet Protocol
TPSPCC	Text Processing of Student's Project Content Classification
UML	Unified Modeling Language

# CHAPTER 1

## INTRODUCTION

### 1.1 Project Background

Generally, classification refers to the automated grouping of textual or partial textual entities for document retrieval, categorization, domain specific information extraction processing and others. The popularity of text processing is easy to view. Classification is also associating a body of text with a category. The category is often of a thematic or topical nature that involves '*understanding*' of natural language.

Therefore, the system will carried out text processing of student's projects of contents classification which will be implemented specifically for lecturers to search information about student's projects for their reference purpose. The focus of this project will be on searching student's projects and categorized text according to field of project.

The developed system used Java platform and TCP/IP programming for some of the module. Besides that, the system project aims is to use a client/server to develop this application framework. The client/server model is a form to distributed computing where one program (the client) communicates with another program (the server) for the purpose of exchanging information.

In addition, the project also applies the Remote Method Invocation (RMI). RMI is a core package of the JDK1.1 and above that can be used to develop distributed applications.



Finally, developing the system in Java look efficiently and hopefully can help to those lecturers who are handled the student's past year project in a simplest way.

## 1.2 Problem Statements

The aim of this system is to provide an efficient system and solve existing problem. The problems identified are:

- **Manual searching process**  
Lecturers need to search manually to find information about student's projects. It might be takes time to find the needed data seems they have no time and busy with their teachings.
- **Difficulties in accessing needed file for reference purpose by lecturers.**  
Lecturers difficult to access needed data from student's projects because there are plenty of student's projects. Thus, it makes the lecturers hard to find the desired data and it is also time consuming.
- **Lost or in bad condition**  
The entire project documents which is not stored or kept in specific place may risk disclosing to unauthorized persons or lost. In other way, it may also make all the project documents untidy due to dusty condition or others at the particular space.
- **Redundancy of data**  
Redundancy of data might be happen with current student's project title because lecturers are difficult to detect the past year student's projects by manual searching process.

### 1.3 Objectives

The objectives of Text Processing of Student's Projects of Contents Classifications are as following:

- To develop a searching engine for student's projects information retrieval
- To enable searching accessible to lecturers at anywhere
- To avoid redundancy of data

### 1.4 Scopes

This system will be implemented specifically for lecturers to search information about student's projects for their reference purpose. The focus of this system will be on searching and categorized text according to field of project. The scopes of this project are as follow:

#### i. Multiple User Access with System

- Admin  
Have a privilege on this system to select the file of student's projects and store the searching results into database.
- Clients  
Will perform search process of student's projects. Only lecturers can be client.

#### ii. Searching Function

- Admin  
Able to make a search of student's projects by browse the file name in order to searching results.
- Client  
Make a search of student's projects which is simple search by project title while advanced search by author, matrix number, year and supervisor.

### iii. Classification Function

- This function will automatically categorize text of student's projects based on project abstract.

## 1.5 Project Significance

The system will benefit Kolej Universiti Teknikal Kebangsaan (KUTKM) Malaysia and also its all lecturers in many ways to ease the process flow through paperless and efficient system.

Besides that, the system is able to develop a searching engine for student's projects information retrieval. It also enables searching accessible of student's projects to lecturers at anywhere. So, lecturers can search the needed data easily, efficient and save the time.

In addition, the system also can avoid redundancy of data from having a same project title, contents and other information between past year student's project with currents. Furthermore, it will be able to perform classification which helps the process of searching method.

The important of each function will overcome the problems that occurred in the manual searching process which used for finding information about student's reports. The approaches that are proposed will apply the new technologies which are computerized information and systematic data handling.

## 1.6 Conclusion

As the world enters new millennium, many revolutionary changes are emerging throughout our lives especially the technology that enrolling computerized field. Thus, this system has been proposed to overcome the problem to improve manual searching process by the adoption of information technology.

This system has been proposed to ease the process flow and it's related with system objectives which are to replace the current approach and decrease the probabilities of human error while doing the searching process and classification to provide software qualities such as correctness, reliability, efficiency, integrity and usability.

In addition, the objective of this system is to enable searching accessible to lecturers at anywhere and to find a better searching engine for student's projects information retrieval. Hence, with added features in this system, more systematic and effective services can be provided and thus benefit to lecturers. Nevertheless, this system is useful if lecturers take advantage from the added features using new technology.

Henceforward, literature review and project methodology will be continued as a next stage. This part will review the previous project and make a comparison with project that has been proposed. Besides, this chapter explains about the software methodology that will be used in this project.

## CHAPTER 2

### LITERATURE VIEW AND PROJECT METHODOLOGY

#### 2.1 Introduction

This chapter focuses literature review and project methodology that will be opted to optimized development of to-be-system. Literature review includes study and research of published materials like journal, thesis and technical documents. In project methodology section, selected approach or methodology will be described the activities.

The project requirements that are requisite in this system will be explained in detail and followed with project schedule and milestones. This chapter will be continued with conclusion whereby it will conclude about this chapter and also gives an overview about the next chapter that is Analysis.

#### 2.2 Fact and findings

According to the free dictionary by Farlex, classification system is an organized structure for arranging or classifying; “he changed the arrangement of the topics”; “the facts were familiar but it was in the organization of them that he was original”; “he tried to understand their system of classification”.

Norbert Fuhr and Gerhard Weikum (2001), in their journal said that classification is extracting and analyzing information from a collection of linked documents at a locality to enable categorization of documents and prediction of documents relevant to a focus document.

According to Rich, as a Co-Founder and Vice President of Corporate Strategy Butterfield from Information lifecycle management (ILM) at Outer Bay Technologies (1998), classifying data is a prerequisite which means it is necessary to identify the data which want to work with and articulate the business, regulatory and technical requirements for that data. Besides, it also identifies and classifies the data according to those requirements and automates data storage and movement according to those classifications. Accurate data classification is perhaps the most important and challenging part of the process.

According to Heidi Biggar (2005), classification is preparing data (determining who, what, when, where, importantly, why by scanning and analyzing data patterns, history, building a rich index and categorizing and classification that data), taking action against data (search and retrieval, set retention periods, move and delete data, encrypt data, set and enforce policies). It's also taking actions with a specific process goal in mind.

### **2.3 Project Methodology**

A growing body of software development organizations implements process methodologies. “*ISO 12207*” is a standard for describing the method of selecting, implementing and monitoring a lifecycle for a project.

After doing the research and reviewing study, Object-Oriented Analysis and Design (OOAD) is the most suitable recommended method that can be used in development of this project. Besides, Unified Modeling Language (UML) which suitable to OOAD approaches also referred as Rational Unified Process (RUP).

OOAD methods are becoming the most widely used methods for system design. The UML has become the standard language used in OOAD. It is widely used for modeling software systems and is increasingly used for designing non-software systems and organizations.

OOAD applies object modeling techniques to analyze the requirements for a context (e.g., a system, system modules, organization, or business unit), and to design a solution. Most modern OOAD methodologies are use-case driven across requirements, design, implementing, testing and deployment.

The main reason choosing OOAD methodologies for this project are it is consistency between analysis and design models which deal with entities from the problem domain that carry over into the solution domain. Besides that, OOAD method allows more aspects of the system to be modeled than do structured techniques. Further more, the encapsulation offered by Object-Oriented methods allows changing the implementation of an object without having to make changes to the rest of the system. Objects that have the same interface can be swapped with each other as needed without affecting the rest of the system.

In Addition, its ability to easily add and make changes to Object-Oriented systems that allows being scalable. It also allows to write modular code while encapsulation that is more reusable, inheritance takes reuse one step further by allowing to write new classes that inherit most of their functionality from previously written classes.

Besides that, all objects can be concurrent. In object-oriented modeling, objects send messages to each other. These messages can be mapped to the implementation using simple sequential function calls.

According to Dr. James Rumbaugh (1997), Object Oriented Analysis & Design (OOAD) is the *art* of developing systems based on a set of cooperating objects. Success depends upon systems analysis and software development teams understanding OOAD and available tools. Smooth and natural transition occurs between analysis and design teams due to nature of defined processes.

According to Donald Firesmith from Dictionary of Object Technology (SIGS Books, 1995), analysis is the “development activity consisting of the discovery, modeling, specification and evaluation of requirements”, while Object Oriented analysis is the “discovery, analysis and specification of requirements in terms of objects with identity that encapsulate properties and operations, message passing, classes, inheritance, polymorphism and dynamic binding”. Firesmith also states that Object- Oriented design is the “design of an application in terms of objects, classes, clusters, frameworks and their interactions.”

Unified Modeling Language (UML) is the standard language for visualizing, specifying, constructing, and documenting the artifacts of a software-intensive system. The UML combines the best of the best data that is modeling concepts (Entity Relationship Diagrams), business modeling (work flow), object modeling and component modeling. It can be used with all processes, throughout the development life cycle and across different implementation technologies.

Thus, UML is most suitable for this project because its produce products and services that address user needs and requirements. Requirements maybe considered the problem and the product or services maybe considered the solution. The problem and solution occur within some domain.

To solve problems, the appropriate knowledge of the problem and solution must be captured (modeled), organized (architecture), and depicted (diagrams) using some mechanism that enables communication and leverage of our knowledge. So, UML is the choice of this project to develop an efficient system.

## **2.4 Project Requirements**

The project requirements are important to identify in order to select an appropriate software and hardware that are required in this project. The topic will focused about the projects requirements and facilities in detail.