

**UNIVERSITI TEKNOLOGI MARA**

**DESIGN AND IMPLEMENTATION OF  
MALAYSIAN MICROBIAL  
STRAIN DATABASE**

**HASDIANTY ABDULLAH**

Thesis submitted in fulfillment of the requirements  
for the degree of  
**Master of Science (Information Technology)**

**Faculty of Information Technology & Quantitative Sciences**

**November 2008**

## **ABSTRACT**

Malaysian Microbial Strain Database (MMSD) is a database of microorganisms collection from many places in Malaysia. It is developed specially to organize microbial data generated from two projects granted by MOSTI. Relational database modeling has been applied during the design phase in order to organize the microbial data. MMSD is also assisted with graphical user interface (GUI) query and retrieval system to facilitate its users in organizing and regaining the data stored. The system is implemented on three-tier architecture with MySQL as the DBMS at the bottom layer, PHP as server side script in the middle layer and the GUI as the first layer. The relational database system enables three main functions, one is to provide the researchers with web interface to perform data management task for data generated (sampling record, culture information, identified microbes information and 16S rRNA sequence). Secondly, the database system allows the users to issue query to obtain information from the database and third, the system allows users to view results in a user friendly interface.

## **ACKNOWLEDGEMENTS**

**Alhamdulillah, I thank Allah for giving me strength to finish this research, though I had to go through many difficulties in many phases of this research.**

**My foremost gratitude goes to my wonderful supervisor, Puan Fauziah Redzuan for her patience and her valuable guidance in completion of this thesis. Because of your guidance and advice has made it possible for me to finish this thesis in a very limited time I have. My foremost gratitude also goes to my co-supervisor, Puan Norehan Abdul Manaf for her guidance and suggestions during the completion of this thesis writing.**

**My deepest gratitude for Prof. Dr. (E) Abdul Latif Ibrahim, the director of IIALSB for his moral support and also to Ministry of Science, Technology and Innovation (MOSTI) for the financial support of this master thesis.**

**I would like to express my gratefulness to my friend, Encik Ahmad Fuad Hilmi Muhamad for his technical support during the development phase of this project and for introducing me the world of database application and SQL scripting. The development of this database would have been impossible without the trainings he conducted as I have no knowledge in databases and SQL before.**

**My utmost sincere gratitude goes to my husband and son, for the patience and sacrifice in all the time I dedicated to this research. Your support, love and attention have inspired me to keep going to the end of this thesis process. My utmost sincere gratitude also goes to both my parents. Ayah and mama, thank you for everything, for always be there whenever I need you. Not to forget, my eldest sister, for all the help during the completion of this thesis.**

**I also would like to express my appreciation to my fellow friends, Nor Suhaila Yaacob, Norhatiah Md Lias and Suhaiza Ahmad Jamhor for a wonderful company and advice anytime I need, making the process of finishing this thesis easier than it should have been.**

**Thank you.**

# TABLE OF CONTENTS

| <b>CONTENT</b>                 | <b>PAGE</b> |
|--------------------------------|-------------|
| <b>ABSTRACT</b>                | <b>ii</b>   |
| <b>CANDIDATE'S DECLARATION</b> |             |
| <b>ACKNOWLEDGEMENTS</b>        | <b>iii</b>  |
| <b>TABLE OF CONTENTS</b>       | <b>iv</b>   |
| <b>LIST OF TABLES</b>          | <b>viii</b> |
| <b>LIST OF FIGURES</b>         | <b>ix</b>   |
| <b>LIST OF ABBREVIATION</b>    | <b>xi</b>   |
| <b>DEFINITION PAGE</b>         | <b>xii</b>  |
| <br>                           |             |
| <b>CHAPTER 1 INTRODUCTION</b>  |             |
| 1.0 Introduction               | 1           |
| 1.1 Research Background        | 2           |
| 1.2 Problem Statement          | 3           |
| 1.3 Research Questions         | 3           |
| 1.4 Research Objectives        | 4           |
| 1.5 Research Scope             | 4           |
| 1.6 Research Significance      | 4           |
| 1.7 Research Limitation        | 5           |
| 1.8 Chapter Overview           | 6           |
| 1.9 Summary                    | 7           |

# CHAPTER 1

## INTRODUCTION

### 1.0 Introduction

One of the hallmarks of modern genomics research is the generation of enormous amounts of raw sequence data (Xiong, 2006). Biologists today are swimming in a rapidly rising sea of data. Efficient experimental techniques, primarily DNA sequencing and microarrays, have led to an influx of basic biological data. The exploitation of these biological data creates the need for databases that are easy to maintain and can be easily managed by users who are most likely to be biologist. The current research in biology depends heavily on the effective exploitation of huge amounts of data (Bornberg & Paton, 2002). However, biologists have been dealing with problems of information management since 17<sup>th</sup> century (Gibas & Jambeck, 2001). Thus, the very first challenge in the genomics era is to store and handle the staggering volume of information through the establishment and use of computer databases. Biological databases have been considered as such a tool to assist scientists in data management (Chen, 2006). The development of databases to handle the vast amount of molecular biological data is thus a fundamental task of bioinformatics.

Bioinformatics is a new discipline of science that incorporate biology, computer science and information technology (Barnum, 2005). Bioinformatics represents a new field at the interface of the twentieth-century revolutions in molecular biology and computers (Pevsner, 2003). In its broadest sense, the term bioinformatics can be considered to mean information technology applied to the management and analysis of biological data (Attwood & Parry-Smith, 1999).