

**INFORMATION VISUALIZATION FOR UM LIBRARY DATABASE**

**SHEREENA MOHD ARIF**

**FACULTY OF COMPUTER SCIENCE & INFORMATION  
TECNOLOGY  
UNIVERSITI MALAYA  
KUALA LUMPUR  
2005**

**INFORMATION VISUALIZATION FOR UM LIBRARY DATABASE**

**SHEREENA MOHD ARIF**

**DISSERTATION SUBMITTED IN FULFILMENT FOR THE DEGREE OF  
MASTER OF COMPUTER SCIENCE (MULTIMEDIA)**

**FACULTY OF COMPUTER SCIENCE & INFORMATION TECNOLOGY  
UNIVERSITY OF MALAYA**

**2005**

## DECLARATIONS

I hereby declare that the work in this thesis is my own except for quotations and summaries which have been duly acknowledged.

4 JUNE 2005

SHEREENA MOHD ARIF  
WGAY 10138

Universiti Malaya

## ACKNOWLEDGEMENTS

This dissertation contains many of my own ideas and interpretations of the ideas of others. However, I would not be in a position to have ideas and interpretations had it not been for the opportunities I was given at the beginning of my career and the work of inspirational researches and writers in the field of Information Visualization and VRML.

I want to take this opportunity to thank Mrs Nurul Fazmidar Mohd Noor for being my supervisor while I'm writing this dissertation at University of Malaya. She has constantly give guidance as the manuscript came to her in several pieces. Her professionalism, creativity, and experience have taken this dissertation to a higher level than I imagined. I am also deeply indebted and grateful to my superiors and colleagues in the Information Science & Technology Faculty at the National University of Malaysia for being my mentors and give support in finishing the project.

I also want to express my gratitude and love to my husband; Mr Hamizi Hamdan and son for their patience and understanding that were shown along the period of my dissertation completion. Their support in all ways has given me strength and faith to make this dream come true. And to my dad, Mr Mohd Arif Abas; my mom, Mrs Paridah Harun; I am really grateful for their financial and mental support; as they has taken this work to a much higher level than I could ever do on my own.

Lastly, many thanks to individual who has contributed to the completion of the project which I haven't mention their names, may Allah bless you all.

## ABSTRACT

This system is an enhanced prototype, called “Doc-with-Cone” to visualize information for University of Malaya Library (UML). The visualization concept adopted “Cam Tree” and “Cat-a-Cone” interfaces to perform a search on UML database. Focus of the project is to introduce an easier-to-use information retrieval system by embedding Human-Computer Interaction (HCI) features with stress on usability aspect in the prototype. The system development is also to help users retrieve more accurate results from the database in less time; compared using traditional OPAC system. The development is carried out based on System Development Life Cycle (SDLC) methodology with prototyping approach. The “Doc-with-Cone” is developed using Virtual Reality Modeling Language (VRML) and 3D Studio Max. It can be viewed via any web browser pre-installed with Cortona® VRML Client plug-in. The proposed visualization works effectively on 3D plane for hierarchically structural information. Based on the assessment done with 20 participants at the end of the project, most users agree on positive aspects of usability factors and HCI features embedded in the prototype. Many of them see the prototype as a new positive concept to be introduced to Malaysian community in searching database.

## TABLE OF CONTENTS

<u>Title</u>	<u>Page</u>
DECLARATIONS	ii
ACKNOWLEDGEMENT	iii
ABSTRACT	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	vi
LIST OF TABLES	vii
CHAPTER 1 INTRODUCTION	
1.1 Overview	1
1.2 Problems of existing system	2
1.2.1 Difficulty in finding the correct word for search process	3
1.2.2 Difficulty to decide which information is more relevant	4
1.2.3 The working memory load	5
1.2.4 No alternative interface for different group of users	5
1.2.5 Time-consuming	6
1.3 Scope of research	6
1.4 Objectives of study	7
1.4.1 To build a better info retrieval system in usability aspect	7
1.4.2 To help the users find the info faster and more accurate	7
1.4.3 To introduce new concept and technique of information retrieval through the use of 'information visualization'	8
1.5 Methodology of study	8
1.5.1 Library research	8
1.5.2 Lab work	8
1.5.3 Field work	9

1.5.4	Interview	9
1.5.5	Survey	9
1.5.6	Evaluation	9
1.6	Significance of the research	10
1.7	Organization of the dissertation	10
1.7.1	Chapter 1 – Introduction	10
1.7.2	Chapter 2 – Information Visualization for Libraries	11
1.7.3	Chapter 3 – Research Methodology & Design Spec.	11
1.7.4	Chapter 4 – Prototype development & Implementation	11
1.7.5	Chapter 5 – Evaluation & Statistical analysis	11
1.7.6	Chapter 6 - Conclusion	12
1.8	Summary	12
CHAPTER 2	INFORMATION VISUALIZATION FOR LIBRARIES	
2.1	Introduction	14
2.1.1	Increasing amount of data and information	15
2.1.2	The current inability to have overall understanding of the information	16
2.1.3	Towards more usable and user-oriented information retrieval system	17
2.1.4	To exploit visual power in human-brain processing	18
2.2	Classification of information visualization	20
2.2.1	Data to be visualized	21
2.2.1.1	One-dimensional data	21
2.2.1.2	Two-dimensional data	22
2.2.1.3	Multidimensional data	23
2.2.1.4	Text and hypertext	23
2.2.1.5	Hierarchies and graphs	25
2.2.1.6	Algorithms and software	26
2.2.2	Visualization techniques	26
2.2.2.1	Geometrically Transformed Displays	27
2.2.2.2	Iconic Displays	28
2.2.2.3	Dense Pixel Displays	28

2.2.2.4	Stacked Displays	29
2.2.3	Interaction and Distortion Technique	29
2.2.3.1	Dynamic projections	30
2.2.3.2	Interactive Filtering	30
2.2.3.3	Interactive Zooming	30
2.2.3.4	Interactive Distortion	30
2.2.3.5	Interactive Linking and Brushing	30
2.3	Examples of suitable techniques for libraries	31
2.3.1	Hyperbolic Browser	31
2.3.2	Perspective Wall	32
2.3.3	Lighthouse	34
2.3.4	The Information Cube	36
2.4	Related issues	37
2.4.1	Provide informative feedback	38
2.4.2	Permit easy reversal of actions	38
2.4.3	Support an internal locus of control	38
2.4.4	Reduce working memory load	39
2.4.5	Provide alternative interfaces for novice and expert users	39
2.5	Summary	40
CHAPTER 3	RESEARCH METHODOLOGY & DESIGN SPECIFICATION	
3.1	Introduction	41
3.2	Project planning	42
3.3	Data analysis	44
3.3.1	Requirement determination	45
3.3.2	Requirement structuring	46
3.3.3	Alternative generation and selection	50
3.4	Designing	57
3.4.1	Dialogue and interfaces	58
3.4.1.1	Who will use the interfaces or dialogues?	58
3.4.1.2	What is the purpose of the dialogues or interfaces?	58
3.4.1.3	When and where is the dialogue or	59



	interfaces needed and used?	
	3.4.1.4 How many people need to use or view the dialogues or interfaces?	59
	3.4.2 Files and databases	59
	3.4.2.1 Laying out sample data	60
	3.4.2.2 Normalization process	62
	3.4.2.3 Transforming E-R diagram to normalized relations	63
	3.5 Summary	64
CHAPTER 4	PROTOTYPE DEVELOPMENT & IMPLEMENTATION	
	4.1 Introduction	66
	4.2 Prototype development	67
	4.2.1 Lay out the database structure	67
	4.2.2 Visualize the database structure in 3D using 3D Studio Max	68
	4.2.3 Export into VRML	69
	4.2.4 Insert text on each node	70
	4.2.5 Sensing viewer actions to make the cone rotation possible	71
	4.2.6 Apply the color coding	72
	4.2.7 Insert hyperlink for each node	74
	4.2.8 Publishing prototype layout online	74
	4.3 Minimum system requirement	75
	4.4 Overview of <i>Doc-with-Cone</i>	76
	4.5 Assessing usability factors	80
	4.5.1 Offering alternatives to novice and expert users	80
	4.5.2 Reduce short-term memory load	81
	4.5.3 Prompting cues	82
	4.5.4 Colors and balance	83
	4.5.5 Simplicity and clean	84
	4.6 Summary	84

CHAPTER 5	EVALUATION & STATISTICAL ANALYSIS	
5.1	Introduction	86
5.2	Objectives	86
5.3	Procedures	87
5.4	Demographic information	88
5.5	Findings	91
5.5.1	Assist users in finding information	91
5.5.2	Usability factors	92
5.5.3	Feedback and comments	101
5.5.3.1	Positive aspects	101
5.5.3.2	Negative aspects	103
5.6	Summary	104
CHAPTER 6	CONCLUSION	
6.1	Introduction	106
6.2	Discussion	106
6.3	Advantages of study	112
6.4	Weaknesses & future enhancement	113
6.5	Summary	114
BIBLIOGRAPHY		
APPENDIX		
A	Organization chart of UML database	115
B	Full content of help menu	122
C	Survey on Human-Computer Interaction elements in <i>Doc-In-Cone</i> system	126

## LIST OF FIGURES

<u>Figures No.</u>		<u>Page No.</u>
2.1	Lifelines project interface by University of Maryland	21
2.2	A zoomable Pad++	22
2.3	Parallel Coordinates	23
2.4	Tilebars search on medical diagnosis	24
2.5	WebToc to visualize the contents of a web site	25
2.6	Mars to assist mobile network usage analysis	26
2.7	Hyperbolic Browser	32
2.8	The Perspective Wall	33
2.9	Lighthouse visualization technique	35
2.10	The Information Cube	37
3.1	Research methodology flow chart	42
3.2	Systems development life cycle with analysis phase highlighted	44
3.3	Context diagram of University of Malaya's OPAC system	46
3.4	Level-0 DFD of UML's current logical OPAC system	47
3.5	Level-0 DFD of UML's new logical OPAC system	48
3.6	E-R diagrams for UML's OPAC system	49
3.7	Cam Tree of a directory hierarchy	52
3.8	The Cat-a-Cone interface for combining category and text search	53
3.9	The pop-up window in <i>Doc-with-Cone</i>	54
3.10	Systems development life cycle with designing phase highlighted	57
3.11	Sample records from UML database	60
3.12	Description of fields in book record	61
3.13	Relations of UML's database analysis phase	62
3.14	Normalized relations of UML's database	63
3.15	Final E-R diagram for UML's OPAC system	63
4.1	UML book holdings stored in Microsoft Access 2000	68
4.2	3D scene of the prototype built using 3D Studio MAX	69
4.3	Labeling code for nodes	70
4.4	Nodes with label	71
4.5	VRML code for <i>CylinderSensor</i> and the parent cone	71
4.6	VRML code for routing	72
4.7	The prototype with color coding application	73
4.8	Anchor node for hyperlink capability in VRML	74
4.9	Screen shot of <i>Doc-with-Cone</i>	75
4.10	Search results screen	76
4.11	Detail Information screen shot	77
4.12	Screen with 'Navigation cues/legend' hyperlink clicked	78
4.13	Overview map screen	78
4.14	Help screen on the left frame	79
4.15	Mock screen of the new interface	80
4.16	Use of cues in VRML world	83
4.17	Search results from <i>Doc-with-Cone</i>	84
5.1	Percentage of participants group range	88

5.2	Percentage of participants' profession group	89
5.3	Percentage of participants with referral to experience in using 3D software	90
5.4	Percentage of users' opinion on system simpleness	93
5.5	Percentage of opinion on easy reversal of actions	94
5.6	Exact figure on level of easiness to find information using the system	94
5.7	Numbers of participants clear with information organization	95
5.8	Percentage of users in preference against the interface	96
5.9	Participants expectation on system capabilities	97
5.10	Fraction of users opinion on accomplishing task quickly	98
5.11	Percentage of users' understanding of system's interaction	98
5.12	Users' opinion on learnability aspect of the system	99
5.13	Numbers of fraction on users acceptance on system speed	100

Universiti Malaysia

## LIST OF TABLES

<u>Tables No.</u>		<u>Page No.</u>
4.1	Color and its application	73
5.1	Summary of participant demographic analysis	90
5.2	Summary of the questionnaire results	101

Universiti Malaya

## CHAPTER 1

### INTRODUCTION

#### 1.1 Overview

The University of Malaya Library (UML) has been established over 40 years ago at the centre of the University of Malaya's (UM) main campus. Its mission is to be a premier academic library providing excellent collections and information services in support of the University of Malaya's mission (University of Malaya Library, 2002).

The UML consists of the main library, branch libraries and special libraries. The main library provides library facilities to the whole campus, while the branch and special libraries hold the subject collections pertaining to their respective faculties such as Law, Medicine, Dentistry, Engineering, Economics & Public Administration, Education, Language & Linguistics, Institute of Postgraduate Studies & Research, Center of Foundation Studies in Science and the Academy of Islamic Studies.

During the years, UML holdings have reached to 1,239,749 volumes of monographs, bound journals and manuscripts. It also has 3,631 titles of journal subscriptions and 39 subscriptions of index/abstract and full-text databases in CD-ROM or online access (University of Malaya Library, 2002).

Apart from providing basic facilities for library's users, UML also provides the following services:

- Information Searches Service
- Library Orientation Programmes
- Information Skills Programmes
- Computer Labs
- Professional Consultancy
- Training for library professionals
- Interlibrary loans
- Document Delivery Services

## **1.2 Problems of existing system**

Libraries use a standard system to catalog and classify their holdings. They usually use Anglo-American Cataloging Rules to describe the materials. They also use organizational scheme such as Library of Congress or Dewey Decimal Classification to assign subject codes and subject headings. Thus, library cataloging relies on the same information shared and centered through bibliographic utilities such as Online Computer Library Center (OCLC).

Structure that consists behind the cooperation among libraries, which use different online catalog, is Machine Readable Cataloging Record (MARC) (University of Malaya, 2002). MARC is the data format, which uses national and international standards such as Information Interchange Format (ANSI Z39.2) and the Format for

Information Exchange (ISO 2709). The MARC consists of several types, for example USMARC and UKMARC and used all over the world.

The list of a library's holdings contained in library catalogs. At the beginning, the card catalog served as a tool to find collection(s) in a library.

UML first implemented the Online Public Access Catalog (OPAC) in 1994 as a finding tool, searchable by author, title, keyword and subject headings. Thus, OPAC is the card catalog in automated form.

Because the OPAC is just an automated form of the traditional catalog, the users still encounter problems with the existing system; which are:

### **1.2.1 Difficulty in finding the correct word for search process**

Many success or failure of the search process depends on the word chosen by the searcher compared to the cataloger (Theimer, 2002). In OPAC, the search process can be conducted using keyword in 'Author', 'Title', 'Subject' and 'Keyword'. It is quite straightforward when we use the term 'Author' and 'Title' but it may be confusing when we want to use the other two terms.

'Subject' means that users have to enter the correct word to find the right collection as Library of Congress coded it; the cataloging system used by many libraries throughout the world, including UML. For example, if the user types the word 'computer' to find books about Macintosh machine, the Library of Congress may code it under the term 'technology'. Thus, the user would not get the maximum search results.



The term 'Keyword' may also be misleading. Users may not understand what exactly the term refers to. Some users might think that anything typed using 'Keyword' will refer to the contents of the entire document, and some might assume it is referring to anything cited to the keyword.

Word spelling may also contribute to this difficulty. OPAC users, especially students have different ways in spelling. For example, to find information about catalog, users may spell it as 'catalogue' or enter the Malay word 'katalog'. Different word spelling will give different results and this portrays how word selection may interfere with the search process.

### **1.2.2 Difficulty to decide which information is more relevant**

Before information visualization era, the most largely adopted model by commercial libraries, including UM library is Boolean model; which based on set theory and Boolean algebra (Baeza-Yates & Ribeiro-Neto, 1999). The drawback of the model is that it is based on binary criteria; which a document is predicted to be either relevant or irrelevant. There is no grading scale and this will prevents it from good retrieval performance. The major advantage of this model is its simplicity and clean formalism behind the model. To retrieve the document's title, users can easily type any keyword relevant to the information needed. Apart from that, the exact matching behind the Boolean model may lead to too many or too few documents retrieved; and this will give difficulties to users to select which documents are important and relevant.

### **1.2.3 The working memory load**

Another disadvantage of using the current OPAC system is the users have to remember how to use the system in order to find the required document(s). This heavy memory load is not suitable for novice users and will create trauma among those who use the system for the first time. It may also discourage new users to visit the library and gain benefits from the OPAC system.

Thus, to make full use of OPAC, the administration of UML has designed a program to train the users to use OPAC. The program is compulsory to all undergraduate students; but optional to other users. Ariyapala & Edzan (2002) indicated that majority of foreign postgraduate students learn to use OPAC by themselves, and this includes on-screen instructions available in OPAC. Since most users learnt the OPAC by themselves, this makes the process of using the current system more complicated and may lead to waste resources due to minimal usage of the system.

### **1.2.4 No alternative interface for different group of users**

The users of an information retrieval system such as OPAC may vary from their expertise and background. Novice users may want a simple interface to use the system, while expert users may need a quick shortcut in retrieving documents. The current OPAC system has not provided users with alternative interface to make them feel easier and more comfortable.

### **1.2.5 Time-consuming**

The listed difficulties and problems as above have contributed to time-consuming factor in using OPAC. The misleading use of the correct word in search process may delay the precise result needed by the user. The long list of text containing titles of collection in UML holdings may also make users need more time to decide which collection to choose; the possibility of missing the right information is not included. Apart from that, the non-user-centered interface may take some time for users to become competent with the system.

### **1.3 Scope of research**

This research will focus on existing function of OPAC in finding collection within UML holdings. Added features such as borrower's information are not the main concern.

The prototype will be limited to documents from Computer Science subject headings and title.

The subject that will be used to test and evaluate the system will only involve various users with or without experience in using OPAC. At the end, the evaluation will determine whether the proposed system is more effective and usable than the existing OPAC system used in UML.

## **1.4 Objectives of study**

The main purpose of the study is to transform the presentation of information in textual form within OPAC environment into visual form using information visualization. However, there are three objectives that influence the research; which are as follows:

### **1.4.1 To build a better information retrieval system in usability aspects**

The users will find the system is user-friendly and learn to manipulate information presented in colorful 3-dimensional (3D) images. Thus, the interface used in information visualization can be easily deployed among non-technical community.

### **1.4.2 To help the users find the information faster and more accurate**

An effective information retrieval system such as OPAC will enable users to get information fast, making sense from the information retrieved and reach decisions in a relatively short time (Gershon, 1998). One of the objectives of the study is to help users find the books faster by having both query function and manipulating the 3D images that will lead them to the required relevant information.

### **1.4.3 To introduce new concept and technique of information retrieval through the use of 'information visualization'**

The main objective of information visualization is to provide the users with overall understanding of information in one screen. The proposed system is built to enable users to manipulate the information directly, by zooming in or

out to get the details or general context of the whole database presentation. This will help them to understand more about the information structure and help them feel more convenient about the system. The research outcome will introduce new technique for information retrieval based on information visualization.

With reference to the objectives, the prototype system will be more usable compared to the current OPAC system. It will give more satisfaction to the users and less hassle for maintenance.

## **1.5 Methodology of study**

There are few methods used in the study, which will include:

### **1.5.1 Library research**

Examples of recent techniques or technologies used in the industry will be searched in the library. Current development and advancement relevant to the topic will be also revised using sources from library.

### **1.5.2 Lab work**

Lab work will only include the development of the system prototype, which involves VRML and Java programming using personal computer.

### **1.5.3 Fieldwork**

Fieldwork will be necessary to investigate further about current UM library databases and the use of OPAC system for information retrieval.

### **1.5.4 Interview**

Interviews may be necessary if relevant information about OPAC system and UM library databases is not enough.

### **1.5.5 Survey**

Few surveys will be conducted at the end of the prototype development to evaluate the effectiveness and efficiency of the developed system. The surveys may include expertise in OPAC usage such as librarians and end-users such as students. These people will be asked to use the developed prototype and compare them to the current OPAC system. The surveys will include answering questionnaires to evaluate the usability criteria for both OPAC and the prototype system.

### **1.5.6 Evaluation**

Evaluation will be based on the survey conducted earlier. This evaluation will be in statistical form and some important conclusion for further enhancement of the prototype will be suggested.

## **1.6 Significance of the research**

The main interest of this study is to transform the traditional way of presenting large information in textual form into visual form using information visualization. The transformation will decrease the complexity in interpreting information and presenting them with more user-centered approach.

This research is vital especially to people who involve with administration of large hierarchical dataset. The communities who get the benefit from this research are staff in database development and automation division, users of large set of data and those who are interested to use this approach in presenting information.

## **1.7 Organization of the dissertation**

The dissertation consists of 6 chapters and each chapter will discuss the following topics:

### **1.7.1 Chapter 1 - Introduction**

This chapter gives a brief description of the research. It discusses on the overview of existing OPAC system used in UML and the problems that influence the research. It also gives the main idea of the whole study such as scope of the research, objectives and the methodology.

### **1.7.2 Chapter 2 – Information Visualization for Libraries**

Chapter 2 explains the main concept of the research background; which is the information visualization. It also explains what is information visualization and the main types of information visualization.

This chapter gives several examples of suitable systems for libraries and reason why “Doc-with-Cone” has been chosen to represent information for UML database. At the end, it discusses some issues for design proposal.

### **1.7.3 Chapter 3 – Research Methodology & Design Specification**

This chapter discusses the requirement analysis and design specification that includes Data Flow Diagram, interface design and tools to represent the specification. It explains the concept prototype at the beginning of the chapter to guide readers to get the general idea about the final prototype. At the end, this chapter discusses the steps taken to integrate the UML database into our prototype.

### **1.7.4 Chapter 4 – Prototype development & Implementation**

Chapter 4 explains the technical and programming details involved. It also gives the explanation on interface description of the proposed system and some usability factors considered in building the prototype.

### **1.7.5 Chapter 5 – Evaluation & Statistical Analysis**

This chapter discusses about the testing procedures in evaluating the proposed system. It also presents the analysis of the test data from system



testing process that has been conducted at the end of the research. This chapter also discusses on how the prototype finds information in UML.

### **1.7.6 Chapter 6 – Conclusion**

Chapter 6 is the final chapter in this dissertation, which concludes the contents of this study. It consists of lessons learnt and objectives that have been achieved until the end of the research.

This chapter outlines the research constraints and future enhancement needed to develop the system into better information retrieval tool. This chapter concludes which system is more suitable to represent data in UML based on the comparison from other researchers.

## **1.8 Summary**

This chapter discusses about the nature of the research domain. Beginning of the chapter gives a brief overview about University of Malaya Library (UML) and its services.

The next section presents the problems encountered by the users in using the current OPAC system; while the third section elaborates further on the scope of research.

To make sure that the research meets the target, four objectives have been determined; as discussed in the fourth section. The following section explains

about methodology of study that will be used to carry out the research. At the end, we lay out the significance of the research and organization of the dissertation.

Next chapter will introduce the whole research domain and comparison between current research and research by other researchers.

Universiti Malaya

## CHAPTER 2

### INFORMATION VISUALIZATION FOR LIBRARIES

#### 2.1 Introduction

Before we understand the nature of information visualization, we will discuss about the visualization itself. According to Tegarden(1999), visualization technologies fall into three general classes; which are scientific visualization, data/information visualization and virtual reality.

- **Scientific visualization** was developed much earlier than information visualization due to the need of scientists and engineers to view data in visual and graph format (Herman et al., 2000). It deals with the transformation of data produced through scientific or engineering calculations or experiments into images.
- **Data/Information visualization** emerges when many graphical displays of data evolved and the applications that were developed no longer fit into scientific visualization category. As quoted from Tegarden,  
*“Data/Information visualization addresses the transformation of non-spatial and or behavioral data into visual images that represents an analogy or metaphor of the problem space.”*

(Tegarden, 1999, p. 8)

- **Virtual Reality** is a 3D computer-generated with simulated environment that is rendered in real time according to behavior of the user. The interface is more to explain to the user the real environment of the display or to imitate an environment; which are unavailable to the user such as being on the moon or under the sea. The display is basically not about how to represent the data to make it understandable and usable.

Referring to the above discussion, information visualization is about creating tools that exploit human visual system in helping users to effectively digest the information been presented. In other words, visualization is the process of transforming data, information and knowledge into visual form to exploit human's capabilities (Gershon, 1998). Because of the capabilities of modern computers and increasing amount of data, information visualization is becoming a vital research topic especially in multimedia and human-computer interaction fields.

There are few reasons why information visualization is becoming popular; which are as follows:

### **2.1.1 Increasing amount of data and information**

The emerging technology of World Wide Web and the availability of fast digital network access have contributed to demand of large amount of data and information to the users. This has led to advances in storage technology and thus created new possibilities for information navigation, retrieval and access in which visualization could play a major role (Bushell et al., 1991). The large sets of data or information can be reduced in graphic form to represent the whole structure of information in one screen.

Information visualization can also make use of 3D environment because it offers better aspect ratio and perspective when using limited screen space to display large information structures (Robertson et al., 1989)

### **2.1.2 The current inability to have overall understanding of the information**

Before the era of large data and information available in the information highway environment, the need of having overall understanding of the information is not as crucial as nowadays. There is usually no way for the user to know if they have missed out some information. The user often has to go back to previous spot to backtrack the relationship between information.

However, with information visualization, the user can have an overall understanding of the information in one screen, allowing the users to identify interesting subsets. Information visualization also includes methods to interact directly with the information displayed for navigation, which will provide more powerful and useful information environment.

An experiment has been carried out by Schaffer et al. (1996) to compare two methods for viewing hierarchically clustered networks; using both a standard full-screen zoom view and a fisheye view. In this case, the fisheye view is representing the information visualization system and the full-screen zoom view is representing traditional information retrieval system.

Twenty subjects that consist of senior undergraduate students and graduate students were selected from University of Calgary. All of them are familiar with

graphical user interfaces and general data structures. The experiment found that subjects using fisheye view were more efficient at performing the task given. They took less time to complete the task and the amount of navigation was reduced. More than that, they were able to focus directly on the task and not distracted by the need to visualize the network mentally. Most of the subjects found that the context provided by the fisheye view as a valuable resource to complete the task because they can keep track on where they are in the network by using the overview.

### **2.1.3 Towards more usable and user-oriented information retrieval system**

Boolean and Exact Query Matching methods that are currently dominating information retrieval system (Bushell et al., 1991), especially in libraries; are difficult to use. Most users found it was difficult to formulate the effective Boolean queries. However, with good design of information visualization system, users can explore and filter information in a flexible, dynamic and interactive way. Users can use the new system as a visualization tool or a visual query language to help them to search the information easily.

Hornbæk and Frøkjær (2001) has investigated if interfaces using information visualization techniques can support reading of electronic documents by comparing 20 subjects' reading activity in a linear, fisheye and overview+detail interfaces. Linear interface is similar to documents presented on paper and in most interfaces for electronic documents in practical use. The fisheye interface shows certain parts of the document are more important than the remaining parts. The overview+ detail interface shows the documents as a

linear sequence of text and pictures together with a tightly coupled overview of the document. If the user clicks on certain area in overview pane, the detail pane will show readable part linking to the activated overview and vice versa.

The results from the above experiment showed that the overview+detail interface supports reading electronic documents better than the linear and fisheye interface. The experiment also tested on the ability of subjects understanding the main content of the document. It has shown that subjects' answers to this test are graded higher when the overview+detail interface is used. In evaluating the interface, subjects strongly prefer the overview+detail interface than the remaining two interfaces, pointing out that it supports navigations and helps to gain an overview of the structure of the document. Although the user prefer overview+detail interface, they use less time to complete the question-answering tasks using fisheye view. The results show that the usability aspects of information visualization using fisheye view and overview+detail are graded higher than the normal display of information.

#### **2.1.4 To exploit visual power in human-brain processing**

In information visualization, large sets of data are transformed to graphic form, which makes the human perception able to detect the patterns of information structure easier than using direct analysis of numbers. Thus, by using visual form to represent large sets of information or data, users easily understand complex structural relationships of information. For example, if a user uses traditional information retrieval system to search for "Java programming language", he may insert "Java" or "java, programming" as keyword in the query field. This will end up with a long list of documents that contains the

word “Java”, including culture and language of Java race in Indonesia. A user who is not from computer science field will find this very confusing and time-consuming to find the relevant document. Nevertheless, by using visual form of representation, the user can easily track the cluster of the information. “Java programming language” documents will contain in one particular subgroup and other kinds of Java will contain in other subgroups. The user can also see other programming languages that may be used instead of Java.

Hornbæk and Frøkjær (1999) have carried out another experiment to investigate the usefulness of an information visualization system; called the Thematic Map. The experiment used subjects who were master thesis students in computer science to study differences in the interaction process between a non-visual information retrieval interface and a visual information retrieval interface. The Boolean interface is used to represent the non-visual information retrieval interface and the Thematic Map represented the visual information retrieval interface.

The results from the experiment showed that four out of six subjects preferred the Thematic Map interface to the non-visual interface. They found that the graphics are pleasing and they liked the overview gained from browsing the map. They also found ideas to formulate queries from the terms they saw on the map.

Generally, the current situation of information around human makes information visualization an important field to be explored and studied. Visualization applications



can enhance human mind capability to process vast amount of information in faster and more convenient way to be benefited by many categories of users.

## **2.2 Classification of information visualizations**

In everyday life, there are many data being stored and many more information being retrieved and analyzed. Almost all transaction done by human will be recorded with assumption that it will be a potential source of valuable information. However, without having a good technique to visualize the information, the data that are recorded earlier would be meaningless. Thus, there are many researches and developments have been done in many areas of expertise to best visualize their data and information.

With many types of data, there are many approaches and techniques. To make the existing techniques available to the user, there are some efforts to classify the varieties of visualizations, as will be discussed here.

According to Keim (2002), information visualization techniques can be classified into three criteria; which are **data to be visualized**, the **visualization technique** and the **interaction distortion technique**.

### **2.2.1 Data to be visualized**

The data to be visualized in information visualization usually consists of large number of records. Each record consists of some variables and dimensions, which make them multidimensional. For example, there are five criteria to determine if a person is physically fit. Each criteria may be represented by one

variable. We call the number of these variables as the dimensionality of data set. Data sets are as follows:

### 2.2.1.1 One-dimensional data

One-dimensional data is known as temporal data where it usually has one dense dimension. Basically, this type of data set deals with timelines. Before information visualization age, the timelines is represented as simple lists but the fact that temporal events can be simultaneous and overlapping could not best presented in this way.

An example of project done in this area is Lifelines. The Lifelines project has been carried out by University of Maryland to visualize medical patient record. Lifelines has one screen overview of the record using timelines; which provide direct access to the data. Medical records, hospitalization and medications represent horizontal lines, while icons represent discrete events such as physician consultations, progress notes or test, as shown in figure 2.1 below.



Figure 2.1 Lifelines project interface by University of Maryland

### 2.2.1.2 Two-dimensional data

A two-dimensional data has two distinct dimensions and usually used in geographical data; where the two variables are longitude and latitude. The variables may also be width and height and other dimensions that can be considered.

An example of projects develop to visualize two-dimensional data is Pad++ as shown in figure 2.2. This project has been developed at University of New Mexico and NYU Media Research Laboratory. The visualization technique provides zoomable display; where in one research a prototype zooming WWW browser using Pad++ has been developed. In the project, a multiple-scaling environment has been established. Instead of having a single page visible at one time, multiple pages and the links between them are depicted on a large zoomable information surface (OLIVE, 1997).

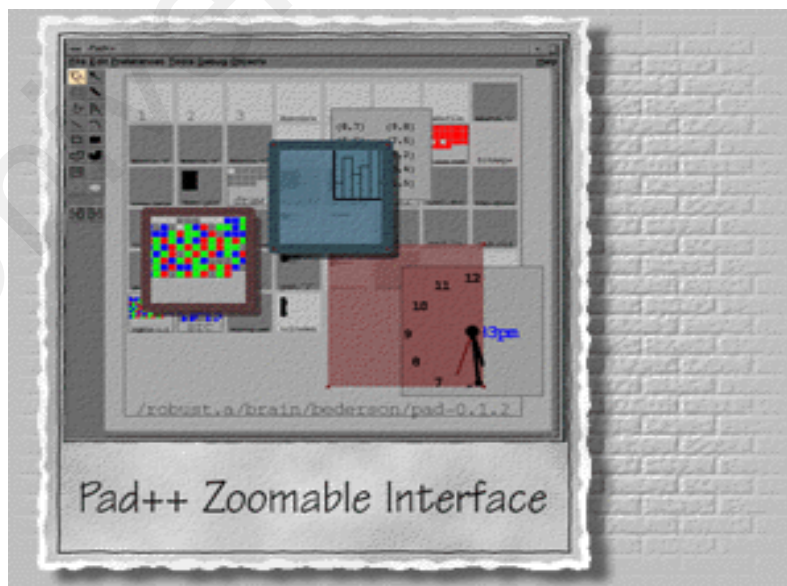


Figure 2.2 A zoomable Pad++

### 2.2.1.3 Multidimensional data

Many data sets consist of more than two variables, which makes them multidimensional. Thus, more sophisticated visualization techniques are needed. One example of visualization technique to represent multidimensional data is Parallel Coordinates, as shown in figure 2.3.

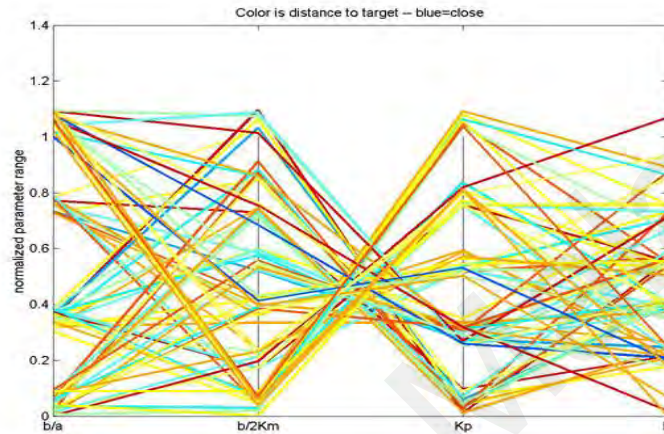


Figure 2.3 Parallel Coordinates

Parallel Coordinates have axes, which are parallel to each other and equally spaced. By drawing the axes parallel, the visualization can display points, lines and planes more than 3 dimensions. Each data dimension is represented as a horizontal or vertical axis. A data element in N-dimensional space is mapped to a polyline that traverses across all of the horizontal or vertical axes.

### 2.2.1.4 Text and Hypertext

Some data could not be presented in numbers, but text. Typical examples include program listings, documents with many lines and document search result. Users who deal with such data set might

want to further search for more specific results or compare between lists.

The common approach to deal with such data sets is to allow users to scroll down through long lists until the desired element is achieved. One of the popular researches to overcome this problem is Tilebars. Marti A. Hearst from Xerox Palo Research Center (OLIVE, 1997) has proposed this visualization technique. Tilebars provides iconic representation of the documents content according to the query terms. It allows users to make decisions about which document to view and passages of the document; based on the behavior of the query terms in the document. figure 2.4 below shows the Tilebars paradigm used in medical diagnosis.

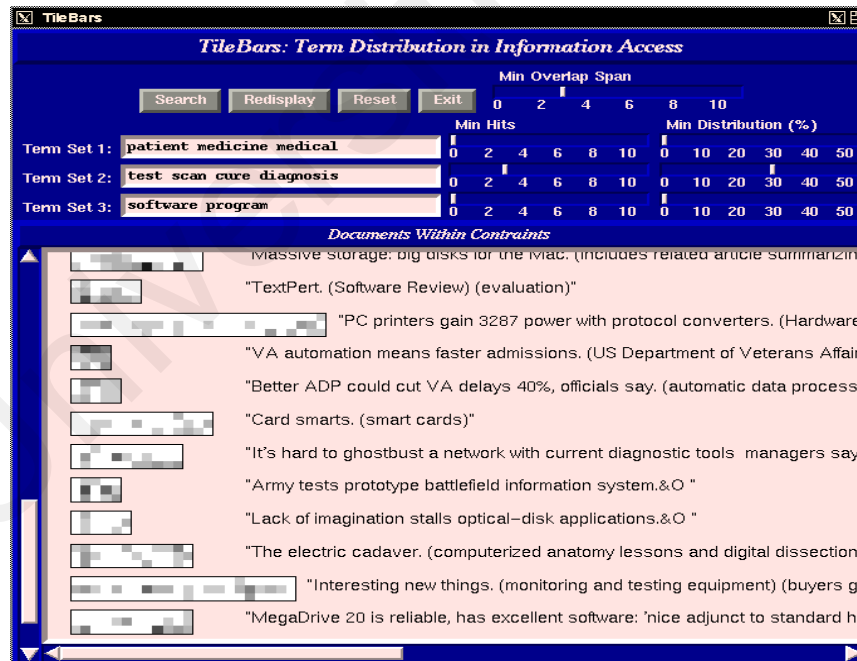


Figure 2.4 Tilebars search on medical diagnosis

### 2.2.1.5 Hierarchies and Graphs

Most of the time, data sets have connection to other information. Graphs and hierarchies always represent such data. They consist of several set of objects, known as nodes. There are varieties of hierarchical data such as structures of organizations, disk space management, genealogies and file systems.

An example of visualization that manipulates hierarchical data is WebToc, as shown in figure 2.5. The visualization technique has been developed in University of Maryland. By following local links, WebToc generates and displays a hierarchical representation of the documents local to the site in the standard web-browsing environment (University of Maryland, 1997).



Figure 2.5 WebToc to visualize the contents of a web site

### 2.2.1.6 Algorithms and Software

The advancement in computers processing power leads to increasing need of visualizing data for large software projects. The goal of the visualization is to assist software developer to debug the program by providing help such as visualized the errors.

The visualization can also be used to represent network system in mobile network as used in Mars, as shown in figure 2.6. The prototype was developed to help analysts to study mobile network usage patterns. Current research in mobile networking relies heavily on simulation; therefore, researches need models of user movement based on actual observation using the Mars visualization technique.

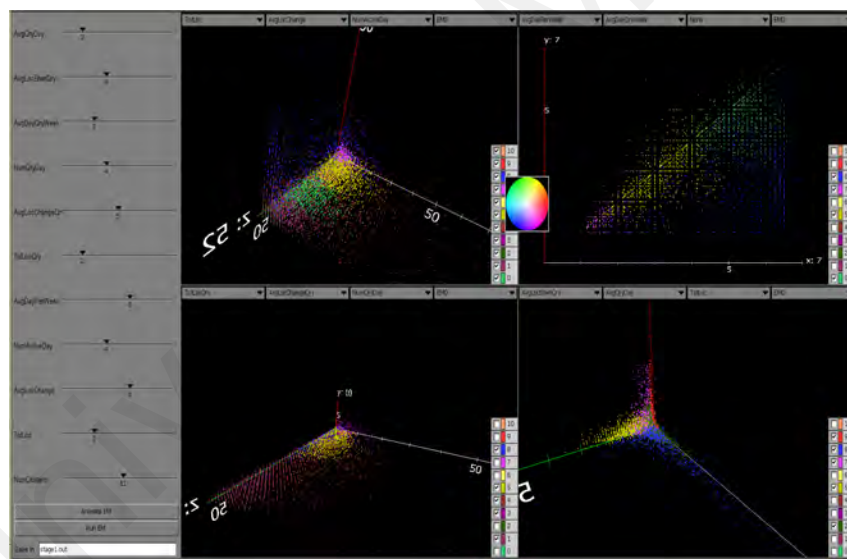


Figure 2.6 Mars to assist mobile network usage analysis

## 2.2.2 Visualization techniques

Apart from the standard 2-D techniques and bar charts, there are many visualization techniques to visualize the data. This section discusses the

classes of visualization techniques based on basic visualization principles as following:

#### **2.2.2.1 Geometrically Transformed Displays**

This technique will transform multidimensional data sets into an interesting display. A popular example of this technique is Parallel Coordinates, as shown in figure 2.3; where it maps the  $k$ -dimensional space onto the two dimensional display by using  $k$  axes which are parallel to one of the display axes. The axes represent the dimensions or variables and are linearly scale from the minimum to the maximum value of the corresponding dimension. Each data presents as a polygonal line; intersect with each of the axes. The intersection represents the value of the considered dimensions.

#### **2.2.2.2 Iconic Displays**

Another visualization technique is Iconic display. The technique always represents values of multidimensional data sets by icons. These icons may vary from little faces, needle icons, star icons and Tilebars as shown in previous figure 2.4.

#### **2.2.2.3 Dense Pixel Displays**

Dense pixel techniques map each attribute value to a colored pixel and group the nearby pixels into a cluster. Since the technique uses only one pixel to represent each data, it is probably the largest amount of data displayed in visualization technique. This technique is popular in



giving information on hot spots, local correlations and dependencies as shown by Mars technique in figure 2.6.

#### 2.2.2.4 Stacked Displays

Stacked display techniques present hierarchical data sets. The basic idea behind the technique is to stack the representation of next level of hierarchy data onto the previous level data. A famous technique using stacked display is Information Cube as shown in figure 2.10. A brief description of Information Cube can be obtained in section 2.2.4.

### 2.2.3 Interaction and Distortion Technique

Besides visualization technique, effective information visualization also needs interaction and distortion technique. The difference between interaction and distortion lies on how developers use them to help users understand the information better.

- **Interaction techniques** enable the user to interact directly with the visualization and able to relate and combine possible visualizations.
- **Distortion techniques** provide users with detail information and an overview of the whole structure of information.

The following section will discuss interaction and distortion techniques outlined by Keim (2002). The term *dynamic* means that direct user interaction can change visualization automatically while the term *interactive* means users change the visualization manually.

### **2.2.3.1 Dynamic Projections**

The basic idea of dynamic projection is to transform the real environment data into one particular or a series of visualizations. This will help users to explore the data sets, especially when dealing with large multidimensional data sets. Example of this technique is Mars as shown in figure 2.6

### **2.2.3.2 Interactive Filtering**

Filtering is about focusing on interesting items and filter out uninteresting items. There are two ways in focusing on interesting items, which are browsing and querying. However, these methods are difficult to use with large data sets since browsing will take some time and the user might miss some important points while browsing. Querying often does not produce the desired results as it depends on retrieval system's querying method.

An example of a technique using interactive filtering is Magic Lenses. Magic Lenses uses a tool like a magnifying glass to filter the data directly in the visualization. The results are displayed differently from the other data set. It will show a modified view of the selected region and the rest of the visualization remains the same as before.

### **2.2.3.3 Interactive Zooming**

Users of information retrieval system typically have an interest in some portion of a collection and they need to zoom in the interesting part. This is especially important when the visualization at the same time

preserve the overview or context of the information. A user can zoom on one dimension at a time by moving the zoom bar controls or by adjusting the field-of-view box. One popular method is by pointing at the interesting point and issues a zoom command by double click on the mouse. Pad++ visualization technique as shown in figure 2.2 also uses this method.

#### **2.2.3.4 Interactive Distortion**

This technique helps users understand and process the information visualized by preserving the context of the whole information structure. The basic idea is to show portions of the data with high-level details and data with low-level details. It usually used on hierarchies and graphs but can also be applied to the other visualization techniques. An example of distortion technique is Perspective Wall as shown in figure 2.8 in later subsections.

#### **2.2.3.5 Interactive Linking and Brushing**

The objective of this technique is to combine different visualization methods to enhance the usefulness of a single technique. For example, interactive linking and brushing can also be used in Mars technique where *select* gesture can link a data to its details and users are able to use brushing to differentiate sets of data with others depend on their objective. Connecting linking and brushing to multiple visualization can also provide more information compared to using a single visualization independently.

## **2.3 Examples of suitable techniques for libraries**

There are many visualization techniques that have been developed by researchers in the market. This section will give brief descriptions on some existing visualization techniques that are suitable to be used in libraries for information retrieval.

### **2.3.1 Hyperbolic Browser**

This technique has been proposed by Lamping and Rao (1996) to display and manipulate large hierarchies and graphs, which have been converted to trees (Lamping & Rao, 1996). The developers have implemented versions of the hyperbolic browser in Common Lisp, C++ and Java to visualize hierarchies. The technique is suitable for browsing organization charts, file system hierarchies and link structure of World Wide Web (WWW) servers and spaces.

The hyperbolic browser uses both focus + context and fisheye view interaction models to enable the users visualizing the information. The visualization includes several parents, siblings and children to make the users able to explore the hierarchy easily. According to the developers, a conventional 2-D browser is capable to display 100 nodes with assumption that each node will have 3 text strings and using a 600 pixel by 600 pixels window. However, the hyperbolic browser can display 1000 nodes with the nearest 50 nodes to the focus, with each node can have 3 to dozens of characters of text (Lamping & Rao, 1996). This shows that the hyperbolic browser can display up to 10 times as many nodes and thus make it an

effective tool to display a hierarchy of information and suitable for displaying collections in libraries.

Figure 2.7 shows the hyperbolic plane is a non-Euclidean geometry where parallel lines diverge from one another. By default, the root of the hierarchy will be placed at the center but the display can bring other nodes into focus. This can be done when the user clicks any visible point to bring it to the center of focus or by dragging any visible point to other position. The rest of the display will transform accordingly to maintain the arrangement of the hierarchy.

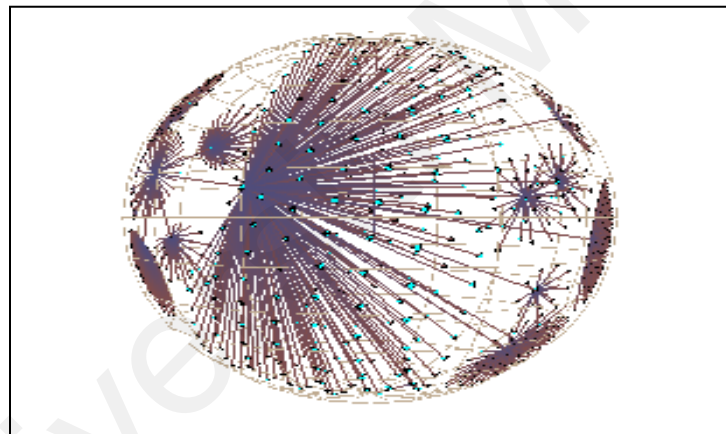


Figure 2.7 Hyperbolic browser in 3D

### 2.3.2 Perspective Wall

Mackinlay et al.(1991) has proposed a technique named Perspective Wall to visualize linear information by combining detailed and contextual views. It uses hardware support for 3D interactive animation to fold wide 2D layouts into 3D visualizations that have a center panel for detail and two perspective panels for context.

This visualization technique is designed from an early idea of Bifocal Display that was developed by Spence and Apperly (Mackinlay, 1991). The Bifocal Display is a 2-D version of the Perspective Wall that is used to visualize a hierarchy of journals, volumes, issues and articles. The visualization technique is designed to enable users to see the context of a single piece of information placed in a large strip (Werner, 2002) is in 2-D presentation brings out one main problem which is it cannot integrate context and detail, leading to invention of the Perspective Wall.

The Perspective Wall has a panel in the center to view details and two panels on each side of the center panel to view context as shown in figure 2.8. The side panels are shaded to give perception of 3D.

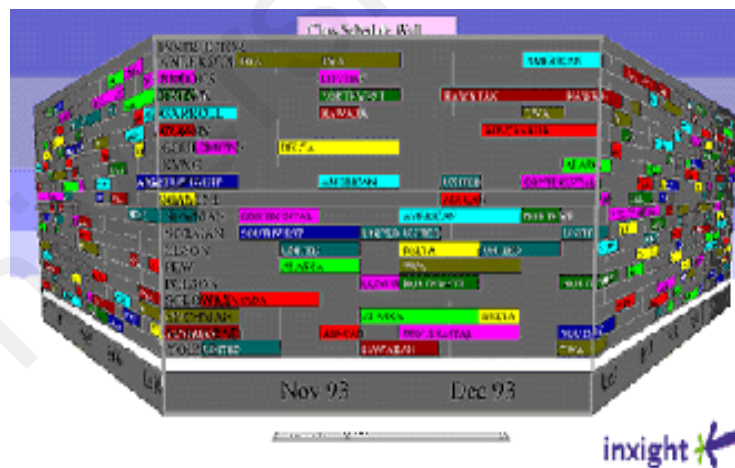


Figure 2.8 The Perspective Wall

The view transition can be accomplished by adjusting the wall to make selected portion of the information be the center of user's view. The

animation helps the user perceive object constancy; which is required by the cognitive system in human to understand the view after it had changed.

The technique has been used to visualize some temporally structured information, for example file systems and corporate memoranda that often involves spanning properties such as time. Some adjustment can also be made to the visualization to represent information used in libraries.

### **2.3.3 Lighthouse**

Lighthouse is a visualization technique developed by Leuski and Allan (2000) for a Web-based information retrieval system. The concept of this technique is drawn from many studies, which proposed that document clustering is a better way of organizing the retrieval results.

This technique accepts queries from a user, collects the retrieved documents from the search engine, organizes and presents them to the user (Lamping & Rao, 1996). The Lighthouse presents the document as spheres floating in the space and positions them according to their inter-document similarity, as shown in figure 2.9. If two documents are similar, the spheres that represent the documents will be closely located. The visualization also display ranked list to show how similar the retrieved documents are compared to the original query.

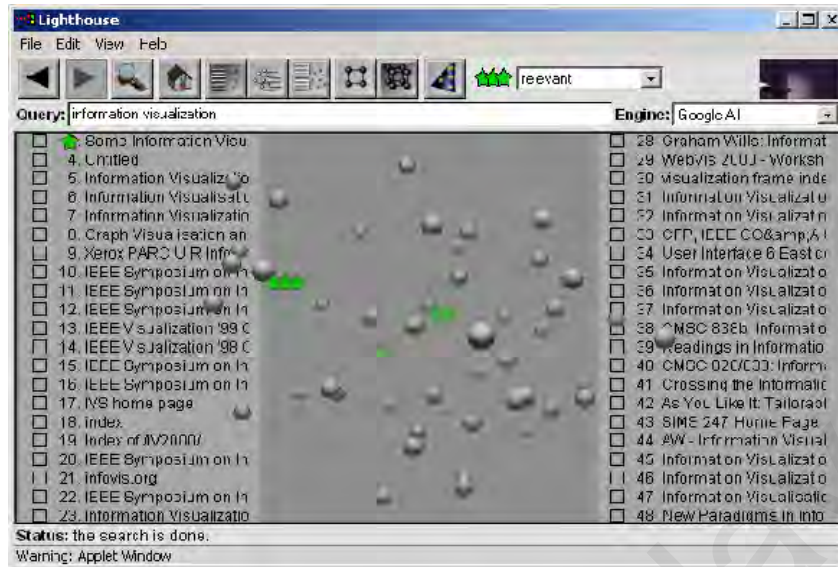


Figure 2.9 Lighthouse visualization technique

The retrieved list is positioned according to the rank with the most relevant document is on the top and the least relevant is at the bottom. On each page there will be 25 documents title on each left and right of the screen. The fifty spheres that represent the document titles are floating at the centre of the screen to preserve some screen space and stress the integration of the ranked list and the visualization. Each sphere is linked to the corresponding document title in the ranked list. The corresponding document title will be highlighted when a user click on a sphere. The user has an option to view the sphere structure either by rotating, zooming and sliding the structure by dragging the mouse pointer. At the same time, the ranked list will remain in place.

A document description can be viewed in a small window when the user point to corresponding sphere or document title while pressing the control key.



The user can double-click on the document title or sphere to open the document in the web browser.

Since Lighthouse is mainly developed for Web-based application, it can also be implemented to visualize libraries collections using internal and wide-area network system.

#### **2.3.4 The Information Cube**

The Information Cube is a 3D visualization technique to represent hierarchical information. This technique has been proposed by Rekimoto and Green (1993) based on the nested metaphor. The concept of using the 3D presentation comes from a common problem where the screen space is too limited to display a large tree structure.

Based on example of the technique as shown in figure 2.10, the outermost cube corresponds to the top-level data, whereas the innermost cube correspond to bottom level data with title on each cube surface. The user can manipulate and interact with the data by using a DataGlove to rotate and move the cubes with his hand. The user can also select a cube and navigate through it to do further inspection. Each cube is displayed in a semi-transparent color that allows the user to see through inside of the cubes.



Figure 2.10 The Information Cube

As collections in libraries are naturally in hierarchical form, thus the Information Cube can also be used to represent the information needed.

#### **2.4 Related issues**

Before information visualization era, the most adopted model by commercial libraries, including UM Library is Boolean model, which based on set theory and Boolean algebra (Baeza-Yates & Ribeiro-Neto, 1999). The drawback of this model is it is based on binary criteria, which a document is predicted to be either relevant or irrelevant. There is no grading scale and this will prevents it from good retrieval performance. The major advantage of this model is its simplicity and clean formalism. The user can easily type any keyword relevant to information they want to retrieve the documents title. Apart from that, the exact matching using the Boolean model may lead to too many or too few documents retrieved. This will give difficulties to users to select which documents are important and relevant to them.

In Modern Information Retrieval (Baeza-Yates & Ribeiro-Neto, 1999), quoted by Marti A. Hearst, Ben Shneiderman has listed principles to design the user interfaces. Since Boolean model has being used for a long time ago, it lacks of these five major criteria in a good information retrieval system. These issues shall be considered to be implemented in the new system are:

#### **2.4.1 Provide informative feedback**

A good information retrieval system shall provide users with usable informative feedback. Information visualization system can provide the feedback by giving users the grading scale of relevant documents in order of most relevant to less relevant titles. It can also visualize clusters of documents to show which documents belong to certain subgroups, so that the users can easily select which subgroup they interested in.

#### **2.4.2 Permit easy reversal of actions**

An information retrieval system shall allow the users to reverse several repeating actions easily to minimize the total time of documents retrieval.

#### **2.4.3 Support an internal locus of control**

Users who are using an information retrieval system shall want to feel in control while using it. Users' will have high satisfaction when they feel that they are in control.

The well-organized hierarchical structure that will be used in the new system shall promote the sense of natural behavior in user for finding information. The

user shall clearly see what will happen after each action they perform in the information visualization system.

#### **2.4.4 Reduce working memory load**

The main disadvantage of current Boolean model used in the libraries is that the users have to remember how to use the system in order to find the documents they want. Information visualization shall reduce the working memory load by using visual in representing information and this is believed will help users to use the system naturally without difficulties in remembering additional steps to be taken in retrieving documents.

The interactive animation that will be used in the developing prototype can also shift a user's task from cognitive to perceptual activity (Robertson et al., 1989). For example, when a user rotates an object, the user can easily perceive the relationship of the information been displayed. Without animation, the user would need more cognitive effort to understand the relationship between different displays.

#### **2.4.5 Provide alternative interfaces for novice and expert users**

The users of an information retrieval system have different level of expertise and background. Novice users may want a simple interface to use the system, while expert users may want to have a quick shortcut in retrieving documents. The wide range of user's expertise shall be considered in developing an effective and more user-oriented system.

## **2.5 Summary**

Chapter 2 generally introduces the readers to the information visualization and its nature. As an introduction, we shall discuss the origins of information visualization and the types of visualizations as outlined by Daniel A. Keim. The section also discusses about the importance of the information visualization.

In the following section, the discussion shall focus on the classification of information visualizations. We shall describe some existing techniques to make us better understand about the nature of each technique.

During the study of the information visualization for UM Library databases, we shall look through some techniques that can be considered as an alternative to OPAC system. This is discussed in section 2.3.

Before we discuss the design proposal of the information visualization, we shall evaluate some related issues on an effective information retrieval system. Section 2.4 has discussed the human-computer interaction issues.

Next chapter will discuss further about the analysis and research methodology followed for the project. Design specification will also be outlined in the following chapter to give readers some pictures on steps taken to build the prototype.

## CHAPTER 3

### RESEARCH METHODOLOGY & DESIGN SPECIFICATION

#### 3.1 Introduction

We developed the system based on traditional System Development Life Cycle (SDLC), with some refinement on the steps taken to reduce time taken for analysis and design specification.

The first phase of the system development is project planning, during which a study is being taken to identify problems in the current OPAC system and several solutions are proposed. Next, data analysis and design are carried out to determine current procedures and technical requirement in developing the system. Building a prototype to match the design into working system then follows the crucial part. Various subjects have tested the prototype with different background in evaluation phase to find out the effectiveness of the proposed navigation aids incorporated with it.

The prototyping phase will be iterated if evaluation result shows that prototype has to be modified to satisfy user's requirement. Figure 3.1 below shows the flow of research methodology used.

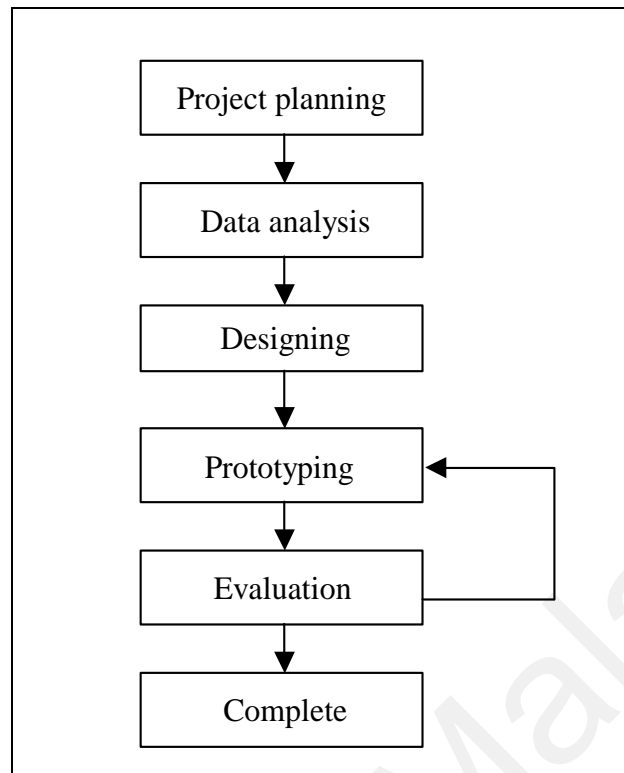


Figure 3.1 Research methodology flow chart

### **3.2 Project planning**

The main purpose of this phase is to identify the need for a new or enhanced system; compared to the existing system. First, the library user's requirements are examined. The information system needs are the result from the requests to deal with problems in current procedures and also from the desire to perform additional tasks. Most of information gathering in this phase is conducted based on our own experience as a user of UM Library, and also through journal and proceedings publications reading, mainly from the Association for Computing Machinery (ACM) and Institute of Electrical and Electronics Engineers (IEEE).

To get some ideas on general domain of the dissertation, net surfing has been carried out using the term '*information visualization*', '*3d information visualization*' and '*library information retrieval*'. The term 'information visualization' is briefly discussed below and detail discussion can be found in previous chapters.

- Information visualization is the process of transforming data, information and knowledge into visual form to exploit capabilities of human nature.
- Information visualization applies human-computer interaction (HCI) elements to make it more usable and effective.
- Research in the area of information visualization for libraries is quite new in Malaysia and the outcome might trigger various feedbacks from UML users to accept the alternative solution to the existing OPAC system.

The search on '*library information retrieval*' has led to common structure of information presentation used in libraries that is usually presented hierarchical structure. The next research in '*hierarchical information visualization*' domain is conducted to gather some information on previous work related to the same domain. This activity has been completed to gain some ideas on techniques used to overcome existing problems and to enhance the current work in delivering better and simplified outcome.

Problems in existing OPAC system and techniques used by previous researches in the same domain have been discussed in Chapter 2, '*Information Visualization for Libraries*'.



### 3.3 Data analysis

The main purpose of analysis phase is to thoroughly study the UML current procedures and the information system used to perform book searching using UML database.

There are three sub-phases in the analysis phase; which include:

- Requirement determination
- Requirement structuring
- Alternative generation and selection.

The three sub-phases are shown in the following figure 3.2.

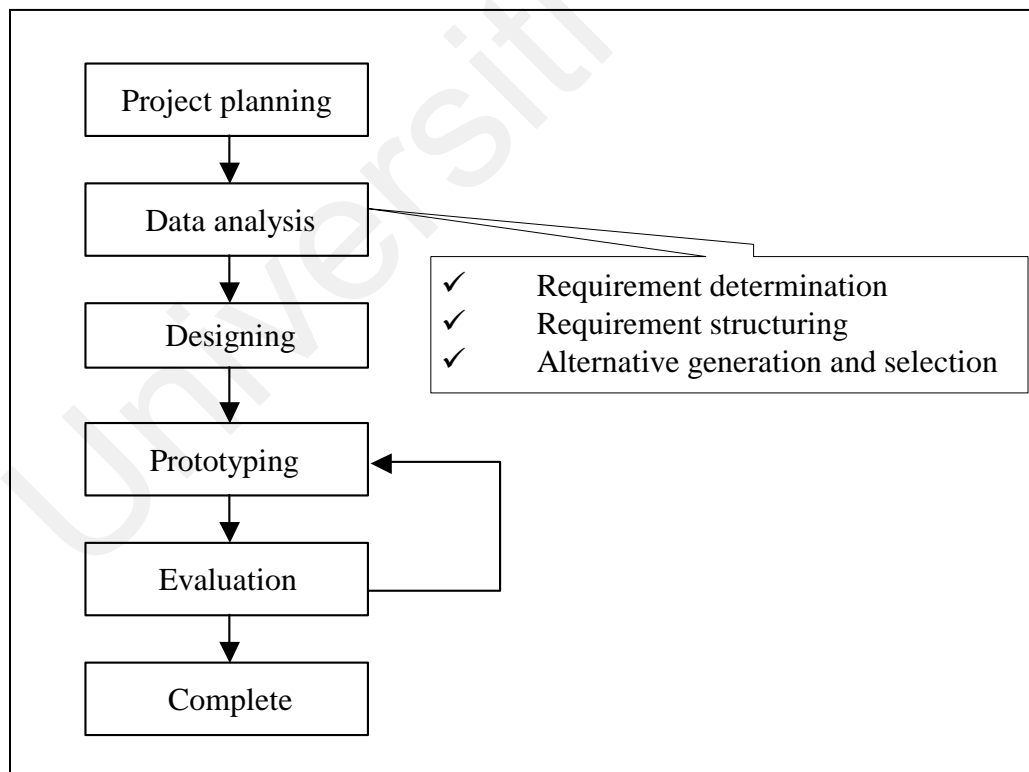


Figure 3.2 Systems development life cycle with analysis phase highlighted

### 3.3.1 Requirement determination

The purpose of this sub-phase is to determine what the new proposed system should do. During requirement determination, we have gathered information on what the new system should do from users of the current OPAC system, from observing users and from procedures taken to complete the search process.

The first step taken to deliver the outcome is to make an appointment with the appropriate person in-charge of UML database development. We have conducted an interview with Ms Bathmawathi Krishnan from Database Development Division of UML to know the procedures taken in storing books information in the database.

Each new book arrived in UML will go through cataloging process which requires a librarian to read through the details and content of the book; and the process will take a day or two. Then the book will be given a unique call number. The call number will follow subject headings as developed by Library of Congress in the LC Classification Schedules (Library of Congress 1998). This call number will be stored in the UML database and all the details of the book such as name of author, publisher and year published will refer to the same call number.

From the meeting, we have developed a general knowledge of cataloging process and database structure of UML books. Since our own domain is information visualization; which differs from library information retrieval domain of interest, we have decided not to alter the current information retrieval

technique used to access a book. In contrast, we started from the OPAC system and developed a way to visualize the whole database structure into one main page.

### 3.3.2 Requirement structuring

After information systems requirements have been determined, we presented the information gathered using data flow diagrams (DFD). To deliver a DFD showing the flow of information inside the proposed system, we presented a context diagram to show the scope of the system, indicating which elements are inside and outside the system. Figure 3.3 below shows the context diagram developed.

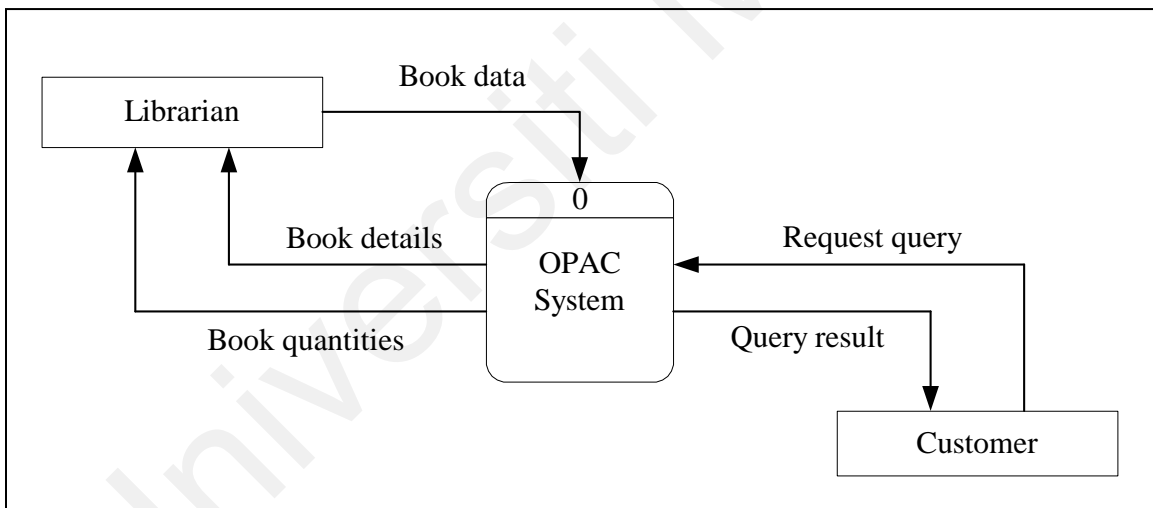


Figure 3.3 Context diagram of University of Malaya's OPAC System

The context diagram shows that only two external entities are involved in the OPAC system, which are *customer* and *librarian*. The *librarian* will store book data in the OPAC system and in return, the system will provide book details and book quantities to the *librarian*. Meanwhile, *customer* may use the OPAC

system to request for book query and the system will send back query result to the *customer* as a feedback. However, the dissertation will only focus on information transformation involved when *customer* query for a book; without amending the current information link between the *librarian* and the *OPAC system*.

Using the context diagram, we have identified two different processes shown in the following DFD as figure 3.4. Process 1.0 is receiving book data from the *librarian*. This process will then produce added information to be stored in the UML's database *book inventory*. The *book inventory* database produces book details and book quantities information to the *librarian*. The database also supplies book details information to be used by process 2.0 also known as book query process. *Customer* will perform query process to search for a book; and the process will send the query back to the database. Lastly, query result will be given back to *customer* as the final outcome of the system.

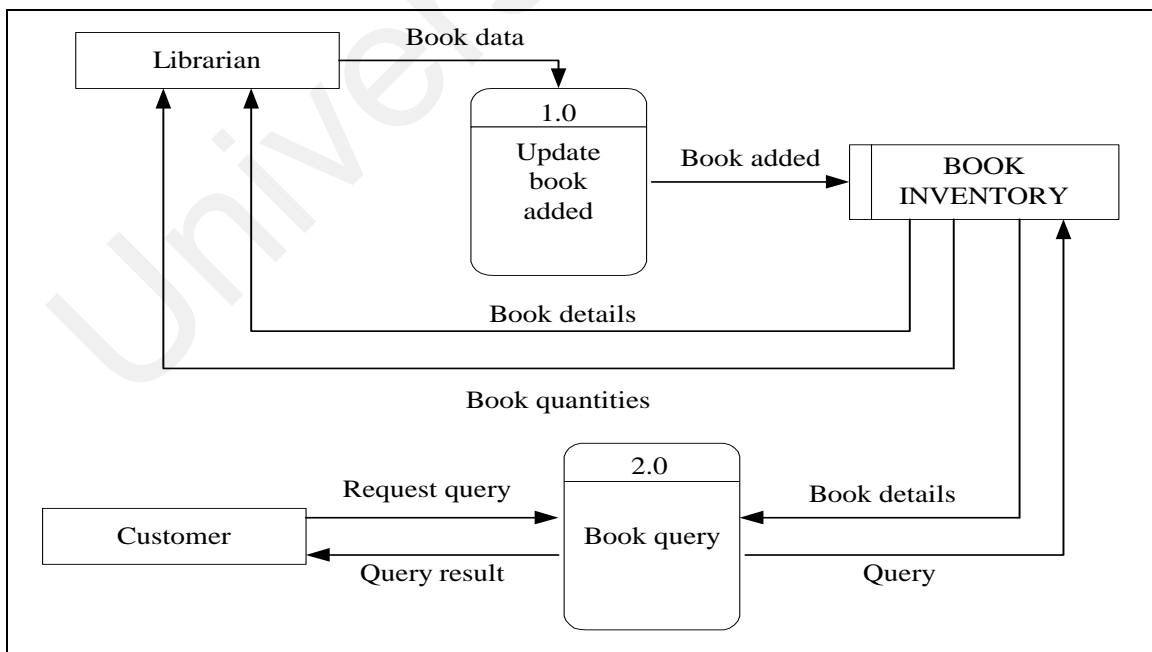


Figure 3.4 Level-0 DFD of UML's current logical OPAC system

Based on the above current logical OPAC system, we proposed the new logical OPAC system as shown in figure 3.5

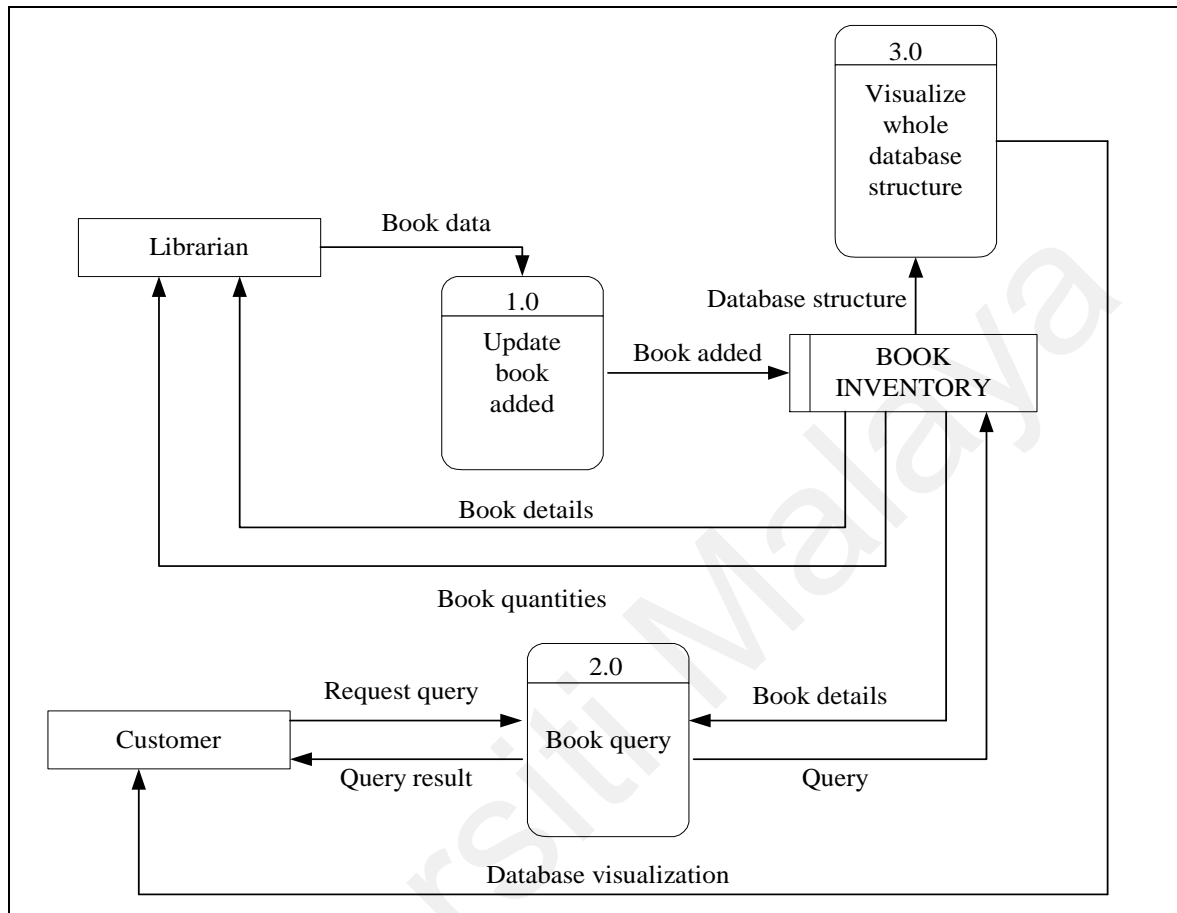


Figure 3.5 Level-0 DFD for UML's new logical OPAC system

Process 3.0 is the new proposed logical OPAC system. The process uses the database structure of the same book inventory database to visualize whole database structure. The database visualization is then presented to the customer to help them making more precise selection of the book based on the query they have done before.

The previous DFD only shows the flow of data between manual and automated steps in the current and proposed OPAC system. The next section

will explain about conceptual data model; which is a detailed model that captures the overall structure of data involved in the OPAC system. We present the conceptual data model using Entity-Relationship Diagram (E-R diagram) using crow's foot notation.

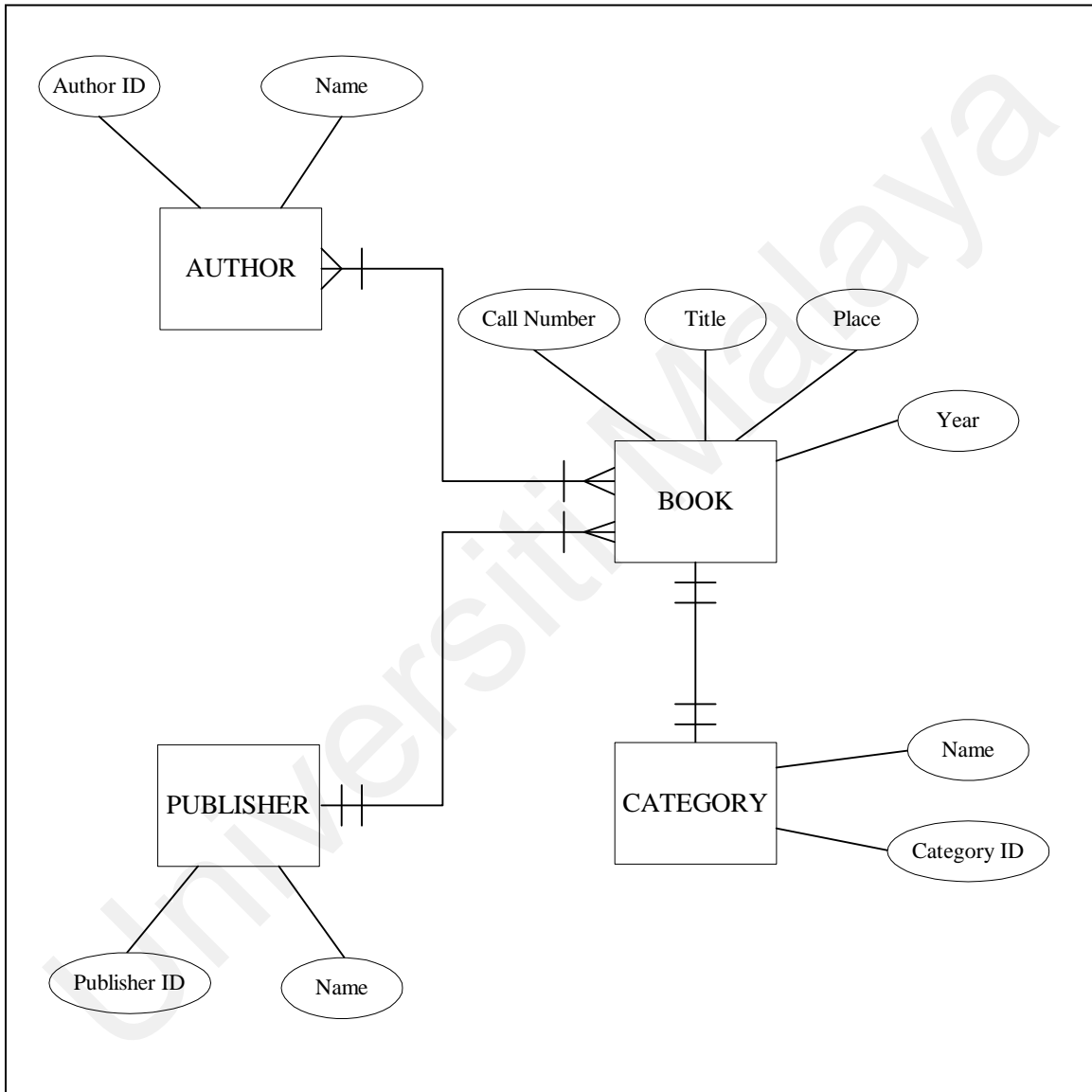


Figure 3.6 E-R diagrams for UML's OPAC system

Figure 3.6 above shows the E-R diagram for the proposed UML's OPAC system. Only important attributes will be shown to users and less important

attributes such as librarian's ID who keyed in the book's data will be left out. Only the listed attributes will be used in the development of the prototype at the end of the research. Further normalization process of the database will be done during designing phase.

There are four entities involved in the OPAC system, which are *Author*, *Book*, *Publisher* and *Category*. The entity *Author* holds two attributes; *Author ID* as the primary key and *Name* as the other attribute. A particular author may be an author of one or many books and a particular book may has one or many authors. The entity *Book* holds four attributes; *Call Number* as the primary key, while *Title*, *Place* and *Year* as the other attributes. The next relationship is between *Book* and *Category* entities; one entity *Book* may falls under one or many *Category* and a particular entity *Category* may holds one or many entity *Book*. The entity *Category* holds *Category ID* as the primary key and *Name* as the other attribute. Lastly, there is relationship between entity *Book* and entity *Publisher* where only one *Publisher* may publish a *Book* but a *Publisher* may publishes one or many *Book*. The entity *Publisher* holds *Publisher ID* as the primary key and *Name* as its other attributes.

### **3.3.3 Alternative generation and selection**

We now have reached to the point in the analysis phase where we are ready to transform all the information gathered and structured into some concrete ideas about the nature of design for the new OPAC system. This sub-phase generates several alternatives to the current OPAC system and make selection of the best alternative.

Referring to the information we gathered so far, we come out with a concept prototype as an alternative to the current OPAC system. The early prototype has been built to gather an initial set of basic requirements and later will allow us to quickly convert the requirements into a working version of the proposed information visualization. This dissertation will propose a visualization technique to provide alternative to the existing information retrieval system called OPAC in UM Library. The proposed visualization technique is a combination of two hierarchy visualization techniques named “Cam Tree” and “Cat-a-Cone” as shown in figure 3.7 and figure 3.8. Throughout the discussion, the visualization technique will be referred as “Doc-with-Cone”.

The *Doc-with-Cone* name derives from “*Document within Cones*”. It is suitable to represent information in hierarchies form. The information in *Doc-with-Cone* is presented in 3D animated form. There are some advantages on using 3D to represent information;

*The effect of 3D/Rooms is to make the screen space for immediate storage of information effectively larger (in the sense that the user can get to a larger amount of ready-to-use information in a short time). The effect of rapid zooming, animation and 3D is to make the screen space effectively denser, (in the sense that the same amount of screen can hold more objects) which the user can zoom into or animate into view in a short time.*

*(Robertson, Card & Mackinlay, 1993, p.65)*

Nodes are drawn like 3 x 5 index cards. The beginning of hierarchy is placed at the most left of the 3D room and is the apex of a horizontal cone with its





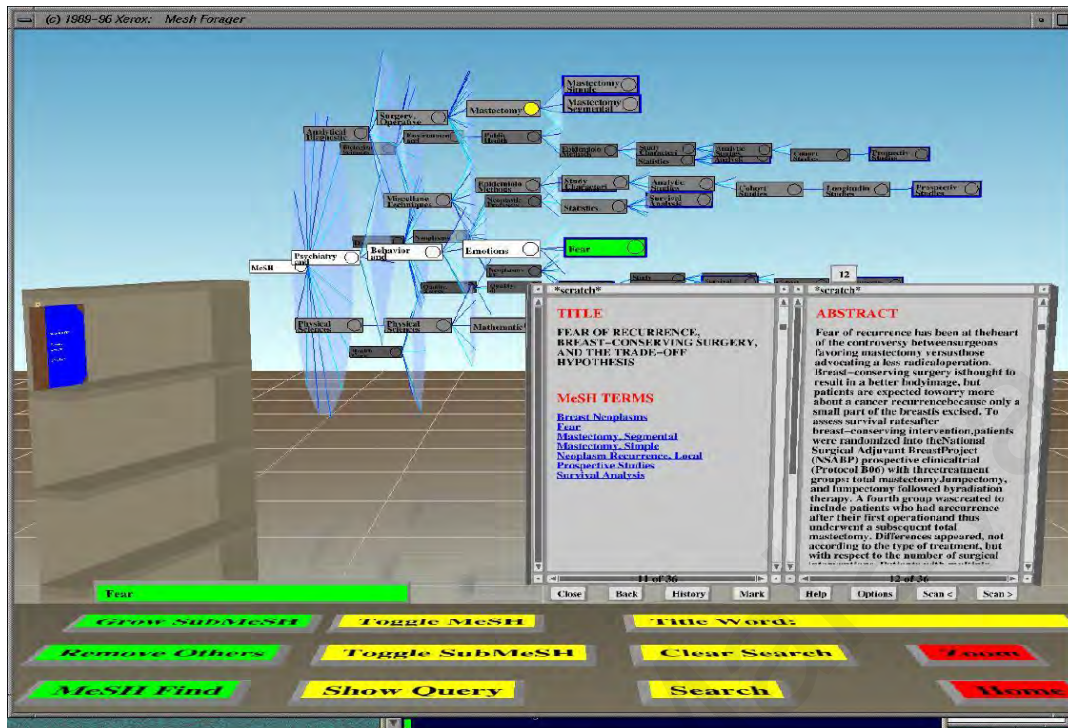


Figure 3.8 The Cat-a-Cone interface for combining category and text search.

Any cone within the structure can be rotated to bring up any node within a selected tree structure. The rotations of each substructure are done in parallel, and are animated so the user can track the transformation.

For a complex hierarchy, users may need ways of hiding selected parts to focus on a particular substructure. This operation has been suggested in Cone Trees where it can be done by menu or by gestures directed at a node. If a user clicks a node toward the top node, all of its descendants are hidden from view. Growing the children node back into view can be done by clicking the node away from the top node or by clicking the last node in current view (Robertson et al., 1993).

However, based on our preliminary analysis, users decline the use of show-and-hide sub nodes because it is confusing and may introduce panic among new users. Based on the suggestion, we omit the operation to avoid disorientation while navigating the 3D information visualization.

Query operation is initiated by clicking the button labeled 'Search' on the top of the room. A pop-up window will appear showing 'Titles in this subgroup' and 'Selected title particulars'. The window is necessary to help users deal with long document titles and various titles of the same query parameter. For example, if a user keys in 'Java, programming' as the search parameter, the window may be shown as figure 3.9.

Titles in this subgroup	Selected title particulars
<ol style="list-style-type: none"> <li>1. Java by Dissection</li> <li>2. Java for beginners</li> <li>3. Java programming for engineers and scientists</li> <li>4. Java for experts</li> <li>5. Learning Java in 21 days</li> </ol>	<p>Title: Java by Dissection</p> <p>Author: Ira Pohl and Charlie McDowell</p> <p>Call No: QA76.73</p> <p>Location: Main Library</p> <p>Abstract: provides a comprehensive, example-based coverage of Java language.</p>

Figure 3.9 The pop-up window in *Doc-with-Cone*

By default, the 'Selected title particulars' will represent the first title in 'Titles in this subgroup' partition. Selecting any title in the left partition in the pop-up window will make the relevant path nodes in the hierarchy highlighted and the right partition of the window will show the selected title particulars.

With reference to the selected visualization technique proposed, the proposed environments and software to develop the prototype are OpenGL, VRML and Macromedia Director. Further details on the specification and capabilities will be discussed in the following subsections:

a) OpenGL

OpenGL stands for “Open Graphics Library”. The word *Open* is significant because OpenGL specification is an independent operating system and windowing system. It is a graphic application programming interface (API), which was developed by Silicon Graphics (OpenGL 2004).

While it can produce magnificent 3D graphics, it needs many programming skills to come out with direct user interaction to the graphics produced.

b) VRML

VRML or Virtual Reality Modeling Language is an open, extensible, industry standard scene description language for 3D scenes on the internet (Ames et al., 1997). With VRML, we can author and view distributed interactive 3D worlds that are rich with text, images, animation, sound, music and even video. VRML 2.0 supports complex 3D animation, simulations and behaviors by allowing Java and JavaScript programmers to write scripts that act upon VRML objects.

To use VRML, we only need to download free VRML browser from any website that offers them. The download process does not take too much

time and easy to conduct. The developer may also embed the VRML world in the Hypertext Markup Language (HTML) document to make it available online and easy to interact with database using Active Server Pages (ASP) technology.

c) Macromedia Director

Macromedia Director is an advanced environment for authoring multimedia applications (Persidsky, 2000). We can use Director's built-in visual development capabilities where media or sprites can be dragged and dropped over a timeline and then the sprites and media can have behaviors copied onto them from a library imparting interactivity. We can also use the object-oriented language Lingo to script behaviors and then control the program when a more sophisticated form of interactivity or control is desired.

The Director needs large resources in developing the 3D world. A fast processor is required to have it run at an acceptable rate. This will make transferring files under development a hassle and difficult to be done. Direct interaction between the database and Director may not be feasible. There is almost no website discussing the possibility of integrating 3D world developed using Director with a database. The main reason we do not use this software is the cost of the software itself. Macromedia Director has to be purchased at around USD1, 000.

With the above issues discussed, we have decided to build the prototype with VRML and integration to Ms Access database will be done using ASP technology.

### 3.4 Designing

After we have completed all three sub-phases in data analysis, we have carried out designing phase as shown in the following figure 3.10.

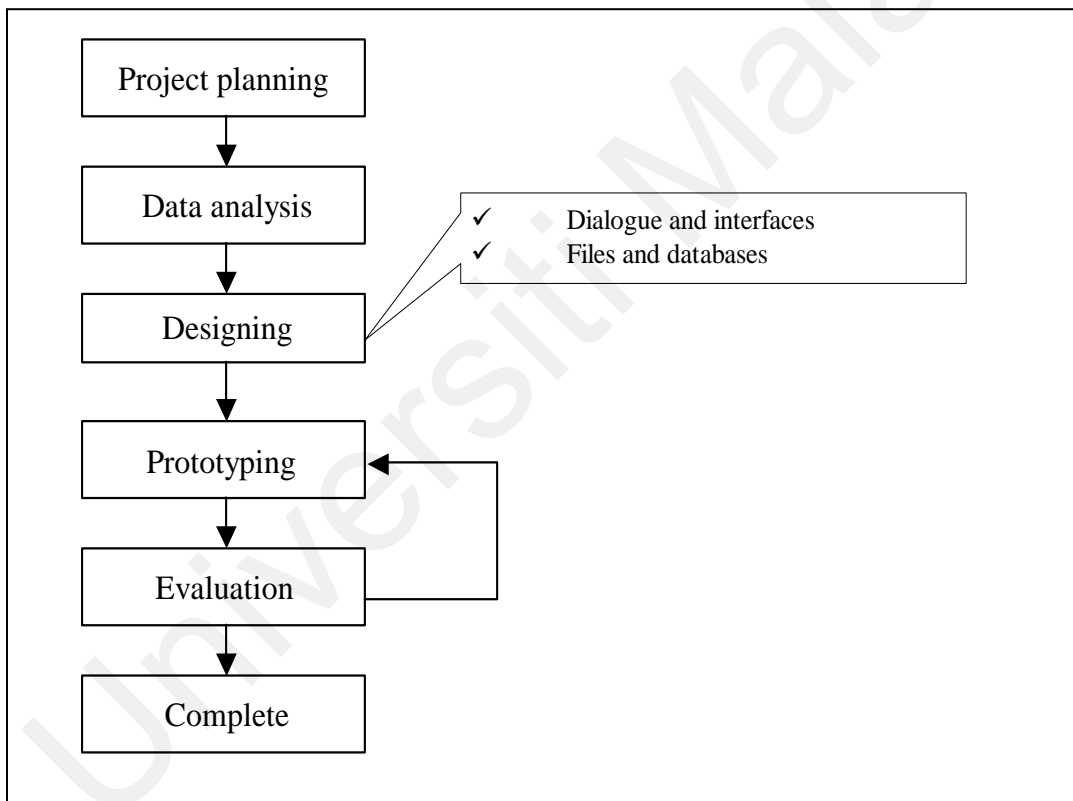


Fig 3.10 Systems development life cycle with designing phase highlighted

As shown in figure 3.10, there are two sub-phases in the designing phase consist of designing dialogues and interfaces, and designing files and databases.

### **3.4.1 Dialogue and interfaces**

We followed user-centered approach in designing dialogues and interfaces. The “who, what, when, where and how” related to the creation of all dialogues and interfaces need to be answered to have a good prototype design. Following subsections will discuss relevant issues in designing dialogues and interfaces:

#### **3.4.1.1 Who will use the interfaces or dialogues?**

UM Library’s users, mainly students and academicians, will use the OPAC system interfaces. Only new students may not be very familiar with the interfaces while the rest has been using OPAC since they have been UM Library’s users. UM Library provides short-term courses for new users on how to use the OPAC system. Thus, the flow of dialogues and ways to use the proposed system will not be too different from the previous interfaces. It is also important that the new interfaces are easy to learn and easy to use by users.

#### **3.4.1.2 What is the purpose of the dialogues or interfaces?**

The purpose of the dialogues or interfaces is to search for books and review its details. The interface that will be developed shall be suitable to display textual information of book information and availability.

#### **3.4.1.3 When and where are the dialogues or interfaces needed and used?**

The dialogues and interfaces are available all the time and not limited to UM Library's operating hours only. Users may also want to access the interfaces even if they are not in the library. The solution is to make the system available online. With regard to network problem that may occur, the interface will be installed manually in the terminals available in UM Library.

#### **3.4.1.4 How many people need to use or view the dialogues or interfaces?**

All UM Library's users need to use the interfaces at any time during the day. It is important under heavy system usage does not slow down the respond time to generate the interface.

### **3.4.2 Files and databases**

The next step after designing dialogues and interfaces is to design files and databases. Currently, UML uses more than one type of database system such as Oracle, MySQL, DB2 and many others. UML requires larger database system to support more data as time goes by. They decided to integrate all the databases to the same OPAC system, rather than going through database conversion to avoid data lost during the process. Because of this factor, we need to build a new database using Microsoft Access 2000 to store all the data that will be used by the prototype.



There are three steps taken in designing the database which will be discussed in the following subsections:

### 3.4.2.1 Laying out sample data

Laying out sample data is important for us to list out all the attributes for the entities. The diagram shown in figure 3.5 consists of brief expectation of developer to the current OPAC system. The entities and attributes used in the proposed prototype are accurate because it is based on the real data sample from the library.

Figure 3.11 shows three sample book records taken from UML database. Each record has at least nine data, which are *Call Number*, *Title*, *Place*, *Year*, *Author*, *Publisher*, *Location*, *Status* and *Due*. Explanation of each field is depicted in the following figure 3.12.

	<b>Record 1</b>	<b>Record 2</b>	<b>Record 3</b>
<b>Call Number</b>	QA763 Tittel	QA7628 COM	QA7628 J4
<b>Title</b>	MCSE NT Server 4 in the enterprise exam cram	Computer concepts problems and exercises	How to pass computer programmer aptitude tests
<b>Place</b>	New York	Chicago	New York
<b>Year</b>	1998	1970	1967
<b>Author</b>	Ed Tittel	Science Research Associates	John J. Jensen
<b>Publisher</b>	Coriolis Group	Science Research Associates	Cowles Education Corp
<b>Location</b>	Engineering	Main Library	Main Library
<b>Status</b>	Missing	Available	On hold
<b>Due</b>			11/ 1/2004

Figure 3.11 Sample records from UML database

Name of field	Description
<b>Call Number</b>	Unique key assigned by librarian with reference to LC Classification Schedules
<b>Title</b>	The book's title
<b>Place</b>	Name of place where the book has been published
<b>Year</b>	The book's publication year
<b>Author</b>	The book's author name
<b>Publisher</b>	The book's publisher name
<b>Location</b>	The location where the book has been kept
<b>Status</b>	Status of availability: Missing/Available/On hold
<b>Due</b>	Date the book will be available if the is borrowed by a customer

Figure 3.12 Description of fields in book record

The *Call Number* field is assigned by librarians to show the book is under which category. We obtained the category structure in Computer Science field from the LC Classification Schedules (1998) and transformed it to organization chart as shown in appendix A. The prototype will include books within this category structure only; which starts with Call Number QA75.5 until QA76.5. To understand how to use the category classification, we compare the *Call Number* of the first record in figure 3.11 with the organization chart in appendix A. The *Call Number* "QA763 Title" falls under the parent category "Study & Teaching. Research" which consists of Call Number "QA7627-7633". The parent category has three subcategories including the Call Number "QA763" for "Certification of electronic data processing personnel".

Thus, the book with Call Number "QA763 Title" will falls under three categories; "Electronic Computer. Computer Science" as the main

category, “Study & teaching. Research.” and “Certification of electronic data processing personnel” as the sub-categories. The first phrase in the Call Number (QA763) always shows the book’s category classification; while the phrases that follow (Title) may refer to author’s name or year of publishing. These combinations will make the Call Number unique.

### 3.4.2.2 Normalization process

Normalization is a process of converting a data structure into simple and stable data structures (Kroenke, 1995). We used normalization as a guideline for checking the desirability and correctness of relations. The following figure 3.13 shows four entities with their attributes from sample data lay out previously.

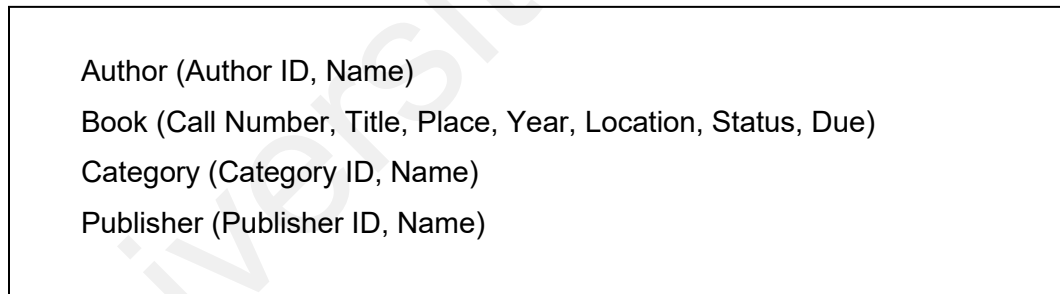


Figure 3.13 Relations of UML’s database from analysis phase

After the normalization process, we have come out with the normalized relations as shown in figure 3.14. Attributes underlined are the primary keys and attributes with dash line are the foreign keys.

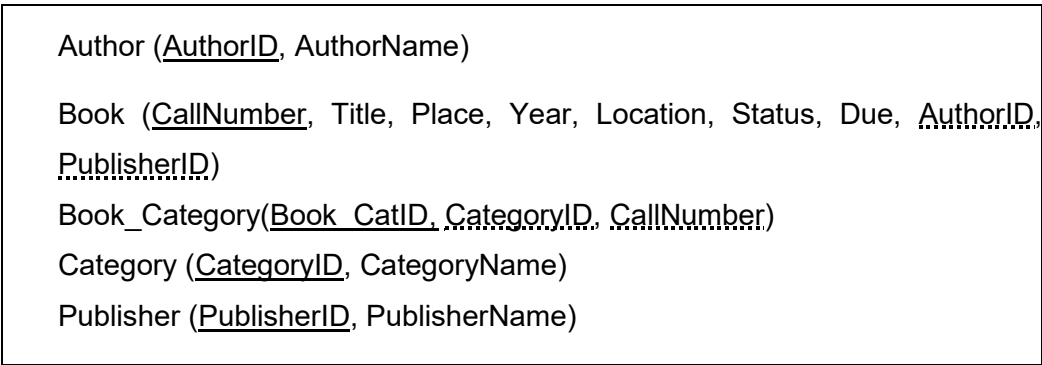


Figure 3.14 Normalized relations of UML's database

**3.4.2.3 Transforming E-R diagram to normalized relations**

After the normalization process, we transform the previous E-R diagrams to normalized relations shown in the following E-R diagram as in figure 3.15. The final E-R diagram is developed to make it easier in developing the database system in future.

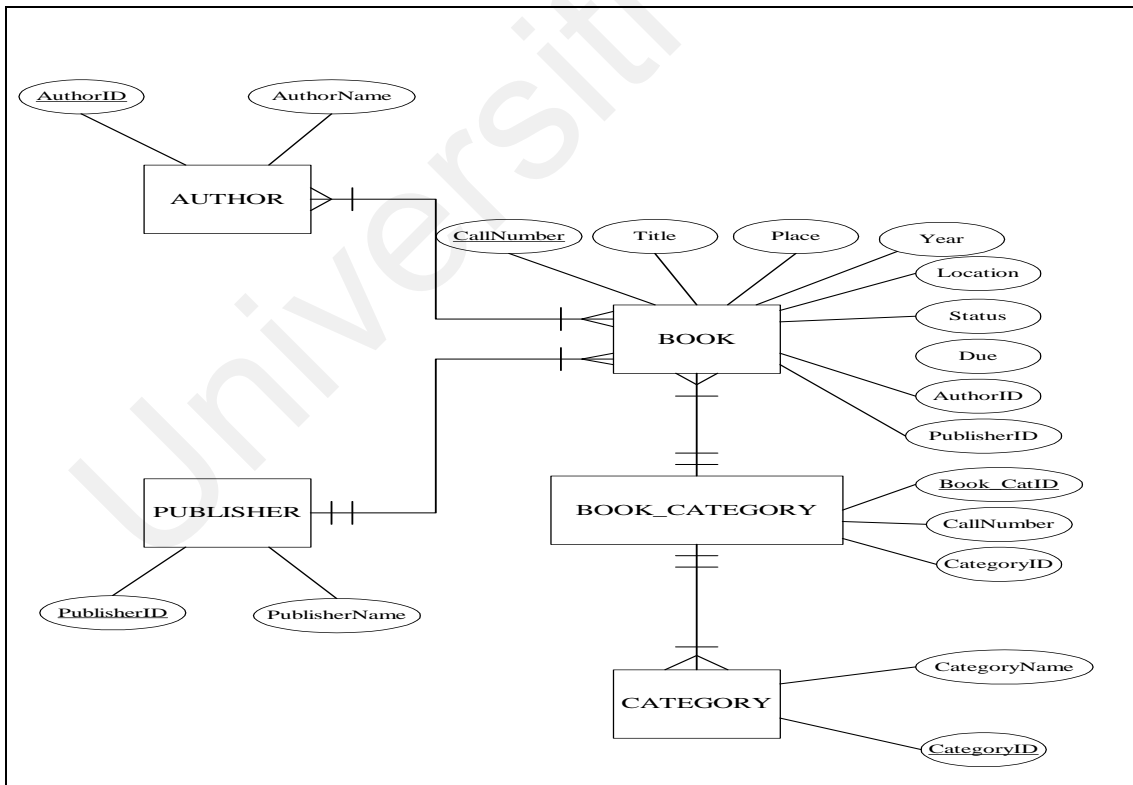


Figure 3.15 Final E-R diagram for UML's OPAC system

### **3.5 Summary**

This chapter discusses about analysis and design aspect of the information visualization that will be developed.

Section 3.1 discusses about the steps taken in completing the research. There are five phases carried out which are project planning, data analysis, designing, prototyping and evaluation phase.

Section 3.2 discusses project planning phase in detail. The main purpose of this phase is to identify the need for a new or enhanced system; compared to the existing OPAC system. One of the subsections outlines the preliminary design proposal for the prototype.

Section 3.3 discusses the analysis process phase. The purpose of this phase is to thoroughly study the UML current procedures and the information system used to perform book searching using UML database. DFD and E-R diagrams have been used to help in completing the analysis phase.

The designing process is discussed in section 3.4. There are two sub-phases in the designing phase consist of designing dialogues and interfaces, and designing files and databases.

Generally the visualization will combine two existing visualization named "Cat-A-Cone" and "Cam Tree". Some may think that this system is more to enhance the

“Cat-A-Cone” technique. However, the author decided to start from “Cam Tree” and build an additional detail window as in “Cat-A-Cone” without imitating the remaining features.

Universiti Malaya

## CHAPTER 4

### PROTOTYPE DEVELOPMENT & IMPLEMENTATION

#### 4.1 Introduction

With analysis and design issues discussed in Chapter 3, this chapter discusses about prototyping and implementation issues.

We used a prototyping methodology to iteratively collecting information, constructing a prototype, assessing usability and making refinements to embed usable interfaces and dialogues in our prototype.

Based on our pilot test, we developed the working prototype as an alternative to the current OPAC system. Later sub section will give brief descriptions on screens included in the prototype.

Following sections will discuss the major steps taken in completing the research.

## **4.2 Prototype development**

To produce a usable visualization technique, we have presented the prototype of our preliminary visualization technique proposal to some potential users as pilot testing. The proposed visualization technique has been outlined in section 3.3.3; Alternative generation and selection.

Using information gathered so far; we have constructed a prototype to visualize UML databases. In general, these are the steps taken to develop the prototype:

### **4.2.1 Lay out the database structure**

With reference to organization chart of the database as shown in appendix A, we lay out the database structure to insert record of book holdings in UML. The development of the database structure has been discussed in earlier chapter.

There are total of 248 records involved in developing the prototype. All these records are taken from current OPAC to make sure users can evaluate the prototype more effectively. The comparison between the proposed system and the current OPAC can be done more significantly. All the records are stored in Microsoft Access 2000 as the software can be used easily and its use is sufficient within the research scope. Figure 4.1 below shows screen shot from the database we built.



CallNumber	Title	Place	Year	Author	Publisher	Location	Status	Due
QA764 D6	Analogue computers	London	1969	K. N Dodd	English Un	Main Library	Available	
QA764 E753	Analogue computers	New York	1960	G. Segal	Pergamonr	Main Library	Available	
QA763 Hawhee	Microsoft certified syste	Indianapc	1998	Howard Ha	Que	Main Library	Available	
QA763 Tittel	MCSE NT Server 4 in thi	New York	1998	Ed Tittel	Coriolis Gr	Engineering	Available	
QA763 Majors	MCSE Windows NT Ser	California	1998	Ken Majors	IDG Books	Engineering	On hold	11/ 1/2004
QA763 Lamml	MCSE: TCP/IP for NT S	San Fran	1998	Todd Lamr	SYBEX Inte	Main Library	Available	
QA763 Strebe	MCSE Internet informati	San Fran	1998	Matthew S	Network Pr	Engineering	Available	
QA7628 COM	Computer concepts prot	Chicago	1970	Science R	Science Re	Engineering	Available	
QA7628 J4	How to pass computer p	New York	1967	John J. Jer	Cowles Edi	Engineering	On hold	11/ 1/2004
QA7628 L67	1001 questions & answe	New York	1977	Kensington	Petrocelli/C	Main Library	Available	
QA7628 TOD	CLAIT(Computer literacy	Essex	2000	Carole Tod	Pearson Ec	Main Library	Available	
QA7627 Heck	Guidelines for evaluating	Va	1981	William P.	National Cc	Engineering	Available	
QA7627 Gilchri	Private data processing	New Jers	1974	Bruce Gilc	N/A	Engineering	Missing	
QA7627 IFIP W	Working conference on t	Oxford	1982	Conference	North Oxfo	Engineering	Available	
QA7625 French	Up the EDP pyramid: thi	New York	1981	Jack Frenc	Wiley	Main Library	Available	
QA7625 Greer	A career in computing	London	1986	Paul Greer	Batsford	Engineering	On hold	11/ 1/2004

Figure 4.1 UML book holdings stored in Microsoft Access 2000

#### 4.2.2 Visualize the database structure in 3D using 3D Studio Max

3D Studio MAX (3ds max) is used to transform the database structure into 3D environment because it can export the 3D scene to VRML format. Unlike VRML, which requires developer to create the 3D scene with programming; 3ds max offers editor to build the 3D scene by only using mouse stroke.

Since every sub nodes will be placed evenly spaced along its parent node, we need the capability of cloning objects in circular arrays in 3ds max to do the job. All the transformation of cloning objects will be done relatively to a center point, which is where the parent node resides.

For example, for “Computer Science” category, there are 20 sub categories reside under it. Thus, we put a node with label “Computer Science” as the parent node and a cone; which can be rotated at its right. There are 20 sub nodes connected to the cone and labeled with the sub categories name.

At the end of the process, we have managed to transform seven pages of database structure as in Appendix A into one screen; as shown in figure 4.2. There are 180 nodes in the screen; with maximum of 56 sub nodes reside under a category.

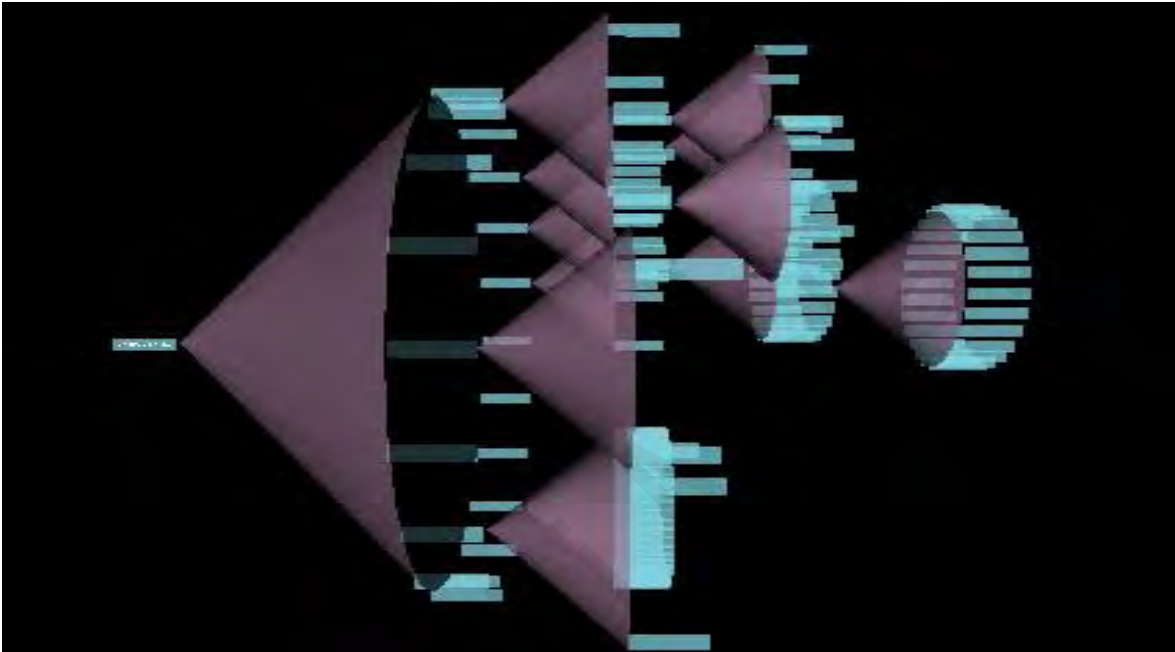


Figure 4.2 3D scene of the prototype built using 3D Studio MAX

### 4.2.3 Export into VRML

The skeleton of the prototype built using 3ds max is then exported into VRML. Refinement has been made to all the exported VRML codes that come together with the scene from 3ds max. This is to make sure that the VRML browser will be able to support all the nodes used in the code that has been generated. The VRML browser that we will use is Cortona® VRML Client because the browser is free and easy to download from the net.

The refinement has helped us in reducing the file size and eliminate some of the codes to build the boxes by using instances capability in VRML. Instances refer to a condition where a node can be duplicated to make other nodes. In VRML we used the node 'DEF' and 'USE' to make instances possible.

#### 4.2.4 Insert text on each node

Each node need to be labeled for further navigation by the users. The labeling process is done using VRML code with node 'string'. The example shown in figure 4.3 below is used to label the parent node (the most left node in the prototype) with text 'Computer Science'.

```
DEF ParentText Transform {  
  translation -28.9 -0.58 -68.9  
  children [  
    Shape {  
      appearance DEF WhiteText Appearance {  
        material Material {  
          diffuseColor 1.0 1.0 1.0  
        }  
      }  
      geometry Text {  
        string "Computer Science"  
        fontStyle FontStyle {  
          size 0.5  
          justify "left"  
        }  
      }  
    }  
  ]  
}, #ParentText
```

Figure 4.3 Labeling code for nodes

As the default background color is black, we use white color label to make it easier to read by users. The following figure 4.4 shows some of the nodes with label.

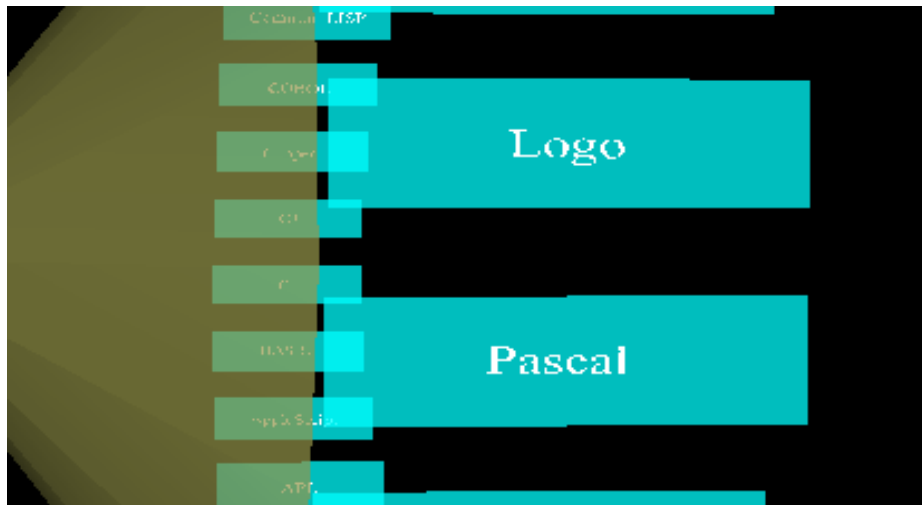


Figure 4.4 Nodes with label

#### 4.2.5 Sensing viewer actions to make the cone rotation possible

To enable the users to rotate the cones, we have inserted sensor in the VRML code. The capability of rotating the node is important; so that the users are able to bring the nodes hidden at the back to their front view. We are using *CylinderSensor* node to sense the viewer action and 'ROUTE' component to wire the sensor and the cone. The following figure 4.5 shows *CylinderSensor* node named 'Cone01Sensor' and part of the parent cone shape named 'Cone01Set'.

```

DEF Cone01Sensor CylinderSensor {}
DEF Cone01Set Transform {
  children[
    Shape {
      appearance DEF ConeColor Appearance {
        material Material {
          diffuseColor 1 1 0.5
          ambientIntensity 1.0
          specularColor 0 0 0
          shininess 0.145
          transparency 0.3
        }
      }
      geometry DEF Cone_FACES Cone {bottomRadius 20 height
    }
  ]
}

```

Figure 4.5 VRML code for *CylinderSensor* and the parent cone

Figure 4.6 below shows the VRML code to wire node 'Cone01Sensor' and 'Cone01Set' together.

```
#routing for cones rotation  
ROUTE Cone01Sensor.rotation_changed TO Cone01Set.set_rotation
```

Figure 4.6 VRML code for routing

The cone can be rotated 360 degrees in clockwise or counter clockwise; as users pleased. This is to make the users feel more comfortable and have sense that they are controlling the scene.









#### 4.2.6 Apply the color coding

There are total of 180 boxes of nodes involved in the prototype with minimum of 1 and maximum of 56 sub nodes. To make it easier for the users to differentiate each group, we apply color-coding method.

Color is an effective way of conveying information and it has important uses in tasks where identification, coding and response times are important. We use colors to code; by identifying categories and by showing trends and relationships in information.

For example, for subject heading "Digital Computers" which have 4 layers of sub nodes, we have applied light blue color to the entire nodes box; including the parent and child nodes. There are 8 colors used in the prototype. The following table 4.1 shows the colors used and its application.

Table 4.1 Colors and its application

Color	Description of subject headings
	Parent node & nodes without sub nodes
	Digital computers and sub nodes
	Communication of info and sub nodes
	Research and sub nodes
	Biography and sub nodes
	Computer software and sub nodes
	Special computers and sub nodes
	Other topics and sub nodes

In a nutshell, we use different colors for different group as shown in the following figure 4.7.

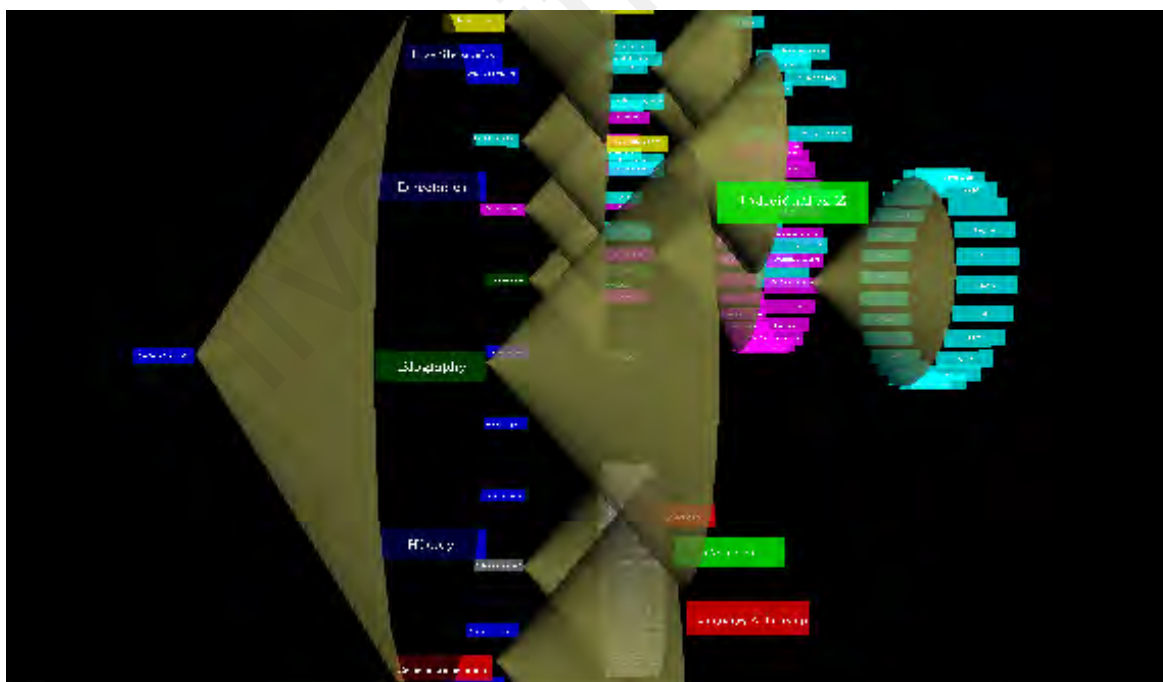


Figure 4.7 The prototype with color coding application

#### 4.2.7 Insert hyperlink for each node

Each parent and sub nodes in the prototype are linked to a new window that will show detail information of the node. The new window will display list of books under the category that the users have chosen. We used *Anchor* node in VRML code to make the hyperlink capability possible. In the example shown in Figure 4.8, the node 'Computer Science' can be clicked and linked to the address specified in *url* element.

```
Anchor {  
  url ["http://localhost/remote_NewFinal/VRML_QComScience.asp"]  
  parameter ["target=_blank"]  
  description "Computer Science"  
  children [  
    DEF ParentPlane Transform {  
      ]} #ParentPlane  
  ]} #Anchor ParentPlane  
]} #Group
```

Figure 4.8 Anchor node for hyperlink capability in VRML

#### 4.2.8 Publishing prototype layout online

The real task in this project development is to put the VRML scene together with search engine online. We used Dreamweaver MX (DMX) application to build the search engine for its stability. We chose DMX because the technical help is available within reach among academician in UM itself.

The search engine is connected to the database that we have developed earlier. To make this possible, we set up a Personal Computer (PC) as a server using ASP technology. It enables the users to search for certain books by keyword. The search method is imitated from the current OPAC to reduce the effort of learning new method among users.

The VRML scene is embedded next to the search engine, which shows the whole UML holdings in one screen. Users can interact directly with the VRML world from the same screen. They can rotate and click to open a hyperlink on the nodes. Figure 4.9 shows the whole prototype; named “Doc-with-Cone” online.

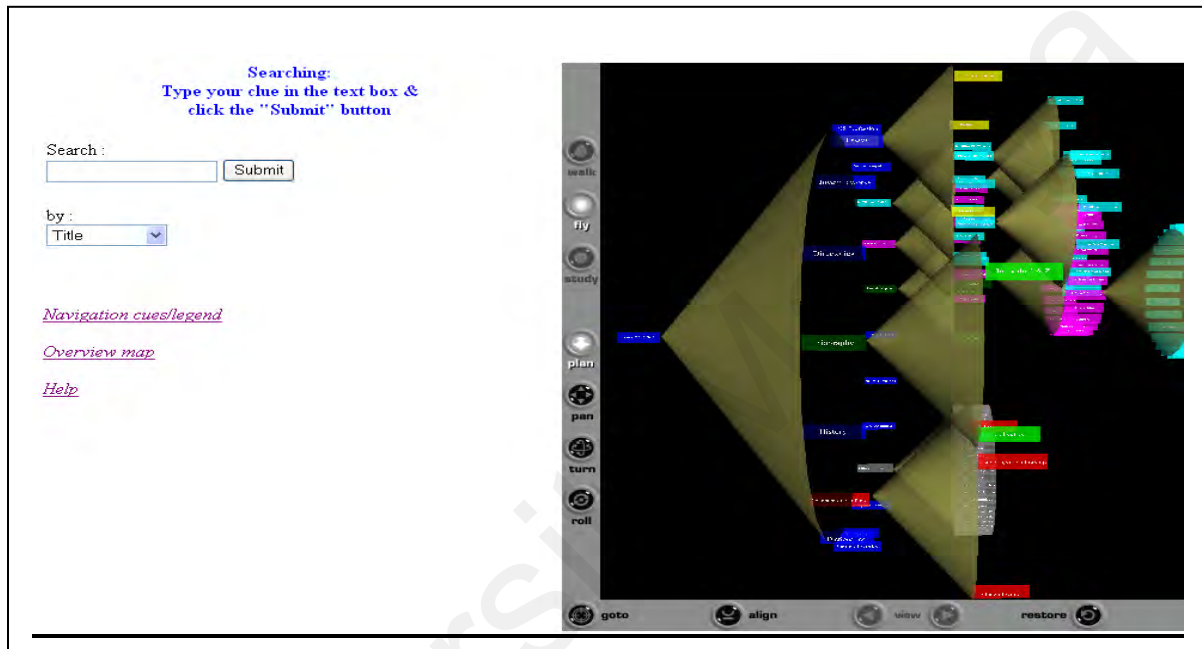


Figure 4.9 Screen shot of “Doc-with-Cone”

### 4.3 Minimum system requirement

The proposed visualization technique is built on PC environment because the majority of Malaysian are more accustomed to PC; rather than Macintosh. Thus, the minimum requirement for this system to run smoothly is to use Windows 2000, Windows NT or Windows XP.



Users also need to use at least Intel Pentium III processor with 600 megahertz (MHz) with Netscape Navigator or Internet Explorer Version 5.0 or later. The minimum requirement for Random Access Memory (RAM) is 256 megabyte(MB).

#### 4.4 Overview of Doc-with-Cone

The main page of “Doc-with-Cone” looks like figure 4.9 above. The screen is divided into two frames, which are the current 2D OPAC interface on the left and the 3D visualization technique interface on the right.

When a user keys in the word ‘Java’ in the search text field, selects ‘Title’ option in the following list box, and clicks the ‘Submit’ button; a separate window will pop-up and display the search results. Following figure 4.10 shows the search results screen.

You searched on : java

Call Number	Title	Year	Author	Publisher
<a href="#">QA7673J38 Afergan</a>	Java quick reference	1996	Michael M. Afergan	Que
<a href="#">QA7673J38 Liang</a>	Introduction to Java programming	1999	Y. Daniel Liang	Que
<a href="#">QA7673J38 Morrison</a>	Teach yourself more Java in 21 days	1997	Michael Morrison	Sams net
<a href="#">QA7673J38 Scneider</a>	Using enterprise Java	1997	Jeff Scneider	Que
<a href="#">QA7673J39 Holman</a>	Instant JavaScript	1997	Brian Holman	Prentice-Hall International
<a href="#">QA7673J39 N44</a>	JavaScript for the World Wide Web	2001	Tom Negrino	Peachpit Press
<a href="#">QA7673J39 P69</a>	JavaScript: a complete reference	2001	Thomas Powell	McGraw-Hill
<a href="#">QA7673J39 Wyke</a>	Pure JavaScript	1999	R. Allen Wyke	Sams net

Records 1 to 8 of 8

Figure 4.10 Search results screen

As can be noted in the figure 4.10 above, each Call Number is clickable to display further details of a particular book. The details of the book included in the screen are call number, title, place published, name of publisher, the location where the book is held currently and borrowing status of the book with the due date; if applicable. Figure 4.11 shows the screen shot when the third book in the above list is selected.

<b>Detail Information</b>	
<b>Call Number</b>	: QA7673J38 Morrison
<b>Title</b>	: Teach yourself more Java in 21 days
<b>Place</b>	: Indiana
<b>Year</b>	: 1997
<b>Author</b>	: Michael Morrison
<b>Publisher</b>	: Sarns net
<b>Location</b>	: Main Library
<b>Status</b>	: Available
<b>Due</b>	:

Figure 4.11 Detail Information screen shot

There are three hyperlinks on the left frame; which are 'Navigation cues/legend', 'Overview map' and 'Help'. A click on 'Navigation cues/legend' hyperlink will open a page that shows legend of colors; which are used in the 3D visualization interface. The details of the navigation cues and the description have been explained in earlier subsection about color-coding method; as can be seen in table 4.1. The screen is shown in the following figure 4.12.

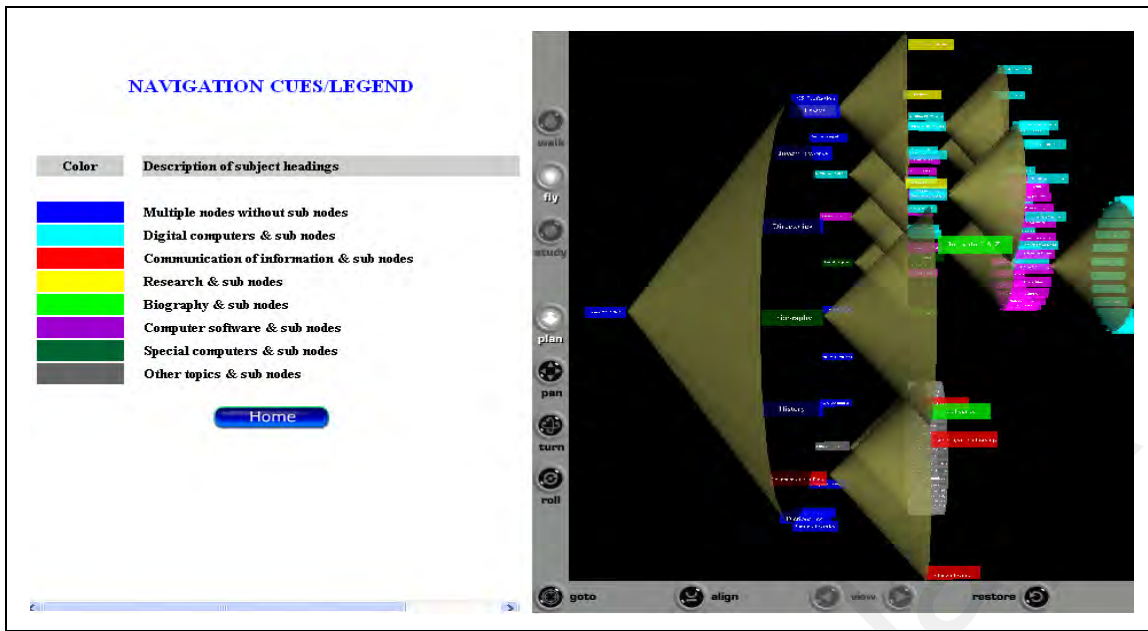


Figure 4.12 Screen with 'Navigation cues/legend' hyperlink clicked

To go back to the main page, users can click on 'Home' button at the bottom of the page. When users click on 'Overview map' hyperlink from the main page, the same 3D visualization technique interface will be opened in a separate window. This overview map can be used as guidance for navigation in the right screen. Figure 4.13 shows the 'Overview map' screen.

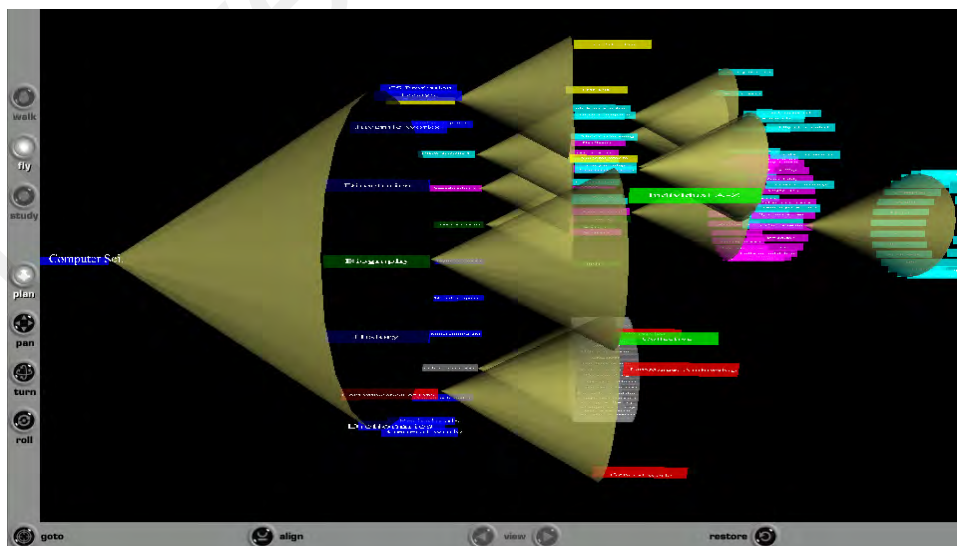


Figure 4.13 Overview map screen

Another hyperlink in the main screen is 'Help'. The 'Help' hyperlink will open a page at the left frame as shown in figure 4.14. This screen explains to users on how to perform a search process and how to navigate in Cortona browser. The full content of this window is as depicted in Appendix B.

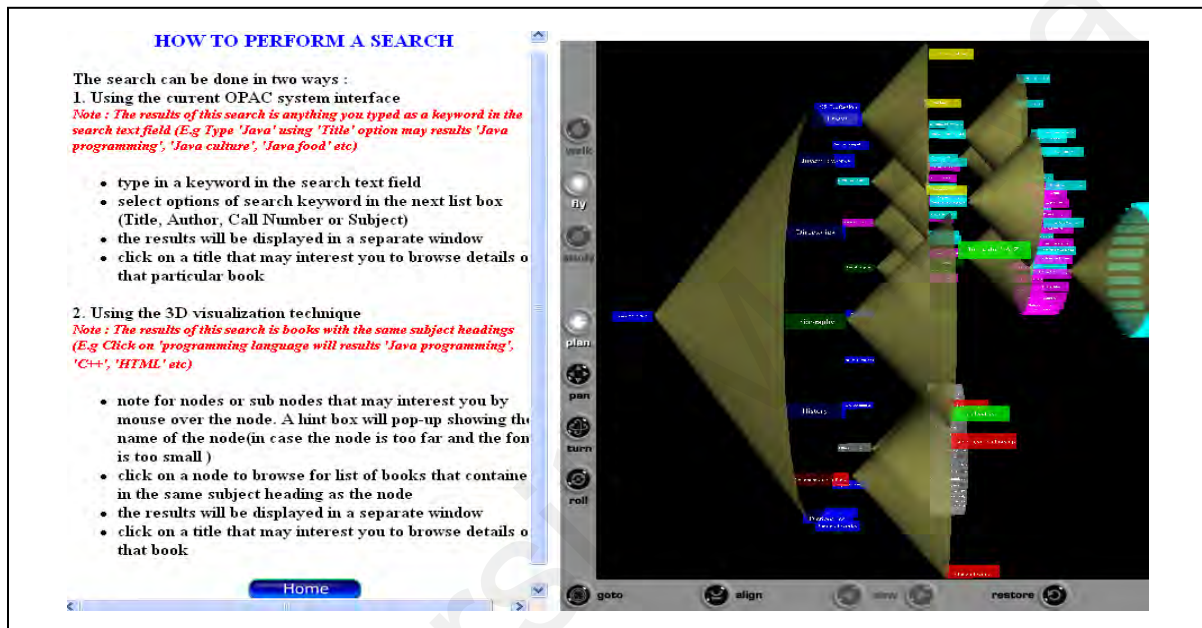


Figure 4.14 Help screen on the left frame

So far, we have discussed on the use of the left screen. The right screen and the overview map can be used interchangeably. If a user prefers to use the right screen as the overview map, the 3D visualization technique interface opens in a separate window as the searching tool and vice versa. An expert user who has used the system frequently might not need the overview map, since the user already knows where the node is. The user also might need the navigation cues from 'Navigation cues' hyperlink. Each node represents a subject heading.

When a user mouse over a node, a hint box will appear to show the name of the node, in case if the node is too far and the font is too small to read. A click on a sub node will obtain the listing books in the subject heading. The same screen as shown in previous figure 4.10 and 4.11 will display listing books and detail information.

#### **4.5 Assessing usability factors**

Before we proceed in evaluating the prototype, we take into account to consider usability factors. This is important because we realized that over the past several years, industry and academic researches have spent considerable effort investigating how information formatting influences individual task performance and perceptions of usability. We have gathered several guidelines from previous research work and we will discuss them in the following subsections:

##### **4.5.1 Offering alternatives to novice and expert users**

The following figure 4.15 shows initial mock screen on how the propose system's interface will look like.

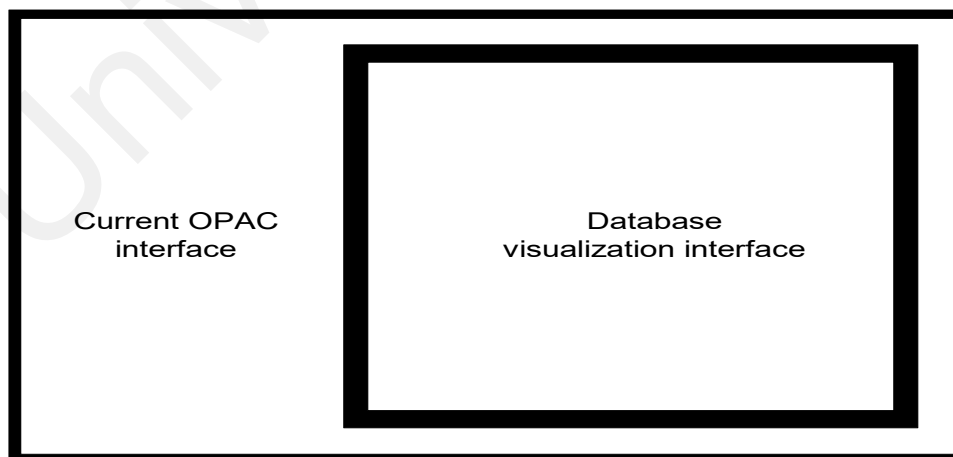


Fig 4.15 Mock screen of the new interface

The new proposed system consists of two interfaces, side by side. The right screen is for database visualization interface, which will be developed in 3D environment; while the left screen will use OPAC interface for novice users who are not familiar with 3D elements and more comfortable with current interface. The OPAC interface that we use here is only a basic interface, not the current retrieval system used by UML; where the OPAC system is embedded in the new retrieval system modified by the library.

The objective of combining both interfaces is to offer several choices of interfaces to novice and expert users. We hope the proposed visualization technique will attract more users in the future; rather than forcing them to learn the whole new interface in a short period of time.

While the novice users get the benefit from 'early' introduction to the database visualization interface just by understanding the organization structure of the database; the expert users can jump into using the visualization database to find information as explained in previous section 4.4.

#### **4.5.2 Reduce short-term memory load**

This guideline has been adopted from Shneiderman(1997). The objective is to simplify the screen and at the same time offers effective information to the users. We put together the VRML screen and the current OPAC system into the same window; as shown in previous figure 4.15. The following hyperlink will be shown in a new window to avoid loss of navigation among users. For a particular search; maximum number of windows that will appear are two. We

assumed two windows are optimum for a lot of searching process that will be used.

#### **4.5.3 Prompting cues**

Prompting cues is one type of providing feedback; as mentioned by Hoffer et al.(1996). We adopt the usage of tool tips to provide feedback; as suggested in many HCI discussions. A tool tip is a method for embedding small snippets of information into a user interface. A tool tip usually displays an object's name or function whenever the user's pointer rests over that object on the screen. Referring to our prototype, a tool tip or a box with text will appear when a user points the cursor over a node. The text will show the node's name. This is to avoid users clicking wrong nodes and in advance; inform the users which node they will click soon.

The use of *Anchor* node in VRML provides cues automatically. When a user points to a node in the VRML scene, the node that has hyperlink will turn the mouse pointer to anchor shape. It will prompt a sign to the user that the node can be clicked and will bring the user to detail information of the node. Figure 4.16 below shows the use of cues for node "Individual A-Z" (the mouse pointer could not be captured in view).

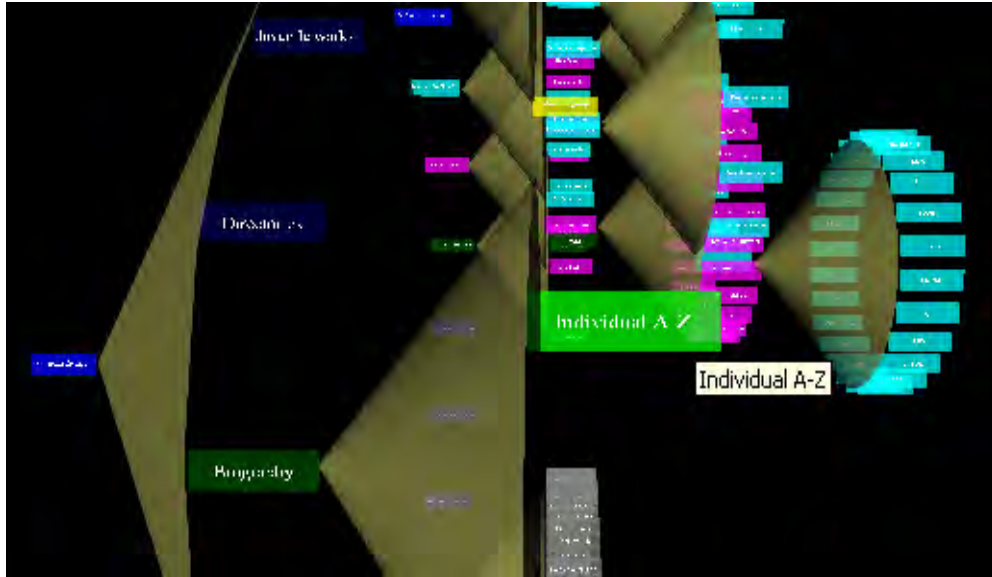


Figure 4.16 Use of cues in VRML world

#### 4.5.4 Colors and balance

Colors play an important role in designing user interface (Morton, 2004). Color-coded has been used to visualize network configuration by Shneiderman et al.(1994); where compact color-coded treemaps with dynamic queries allowing user-controlled filtering and animation of the data display proved well-suited for representing the multiple containment hierarchies in networks. We use colors on nodes that will look significant different from each other. We apply the same color for sub nodes within the same group; and the color will be different to other groups. The use of color-coding method and details on colors used in this prototype has been explained in subsection “4.2.6 Apply Color Coding”. In other words, colors are the key point to help the users navigate the space in VRML.

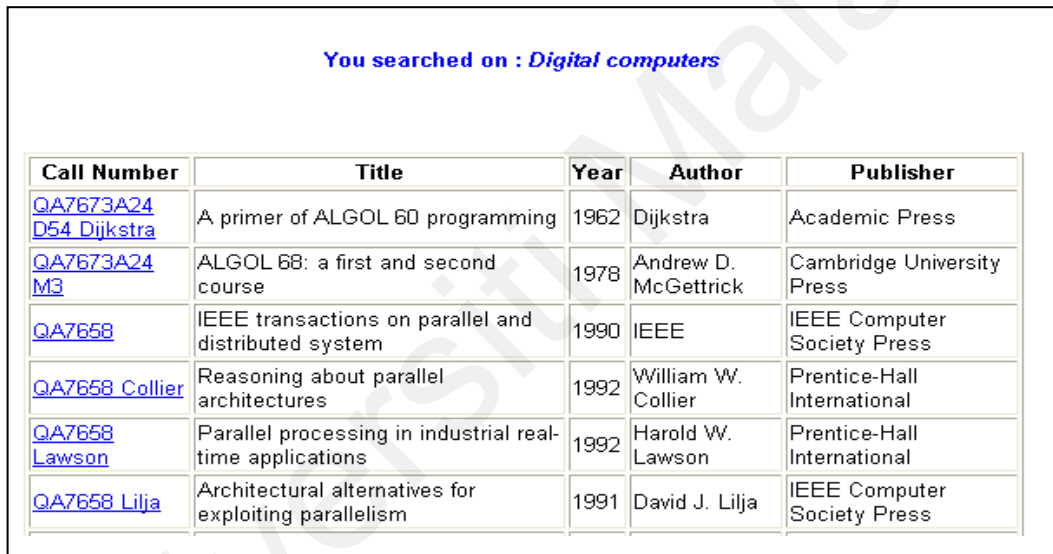
All the colors used are balanced with the black background of the VRML world. This is vital to produce an attractive user interface; but not cluttered.



#### 4.5.5 Simplicity and clean

Nielsen(1999) has always emphasized on simplicity when building a web site or online system. Figure 4.9 shows the clean screen shot of “Doc-with-Cone” which consists of two main screens in one window.

The use of buttons has been minimized and the following hyperlink can be retrieved in another window. Figure 4.17 shows search result after the users click on node named “Digital Computers”.



*You searched on : Digital computers*

Call Number	Title	Year	Author	Publisher
<a href="#">QA7673A24 D54 Dijkstra</a>	A primer of ALGOL 60 programming	1962	Dijkstra	Academic Press
<a href="#">QA7673A24 M3</a>	ALGOL 68: a first and second course	1978	Andrew D. McGettrick	Cambridge University Press
<a href="#">QA7658</a>	IEEE transactions on parallel and distributed system	1990	IEEE	IEEE Computer Society Press
<a href="#">QA7658 Collier</a>	Reasoning about parallel architectures	1992	William W. Collier	Prentice-Hall International
<a href="#">QA7658 Lawson</a>	Parallel processing in industrial real-time applications	1992	Harold W. Lawson	Prentice-Hall International
<a href="#">QA7658 Lilja</a>	Architectural alternatives for exploiting parallelism	1991	David J. Lilja	IEEE Computer Society Press

Figure 4.17 Search result from “Doc-with-Cone”

#### 4.6 Summary

This chapter describes in detail the steps taken in developing the prototype. The implementation of the prototype also takes into account usability factors to ensure users' satisfaction.

We also lay out minimum system requirement for the visualization technique to run smoothly. The overview of the system is also described to relate the results of users' evaluation upon the prototype; that will be discussed in the following chapter.

Next chapter will discuss the procedures taken in evaluating the prototype together with the results and feedback from the participants.

Universiti Malaya

## CHAPTER 5

### EVALUATION & STATISTICAL ANALYSIS

#### 5.1 Introduction

This chapter discusses about the method used and processes involved in evaluating the prototype. Discussion of the results and feedback from the participants are also outlined.

#### 5.2 Objectives

The main objective of the evaluation is to test the usability factors from the users' point of view. The usability factors assessment are important in delivering a usable system that can be used comfortably by potential users. Another objective for conducting the evaluation test is to note positive and negative aspects of the prototype for further enhancement.

The evaluation is to determine whether the two objectives mentioned above are successfully achieved. In later subsection will discuss the two evaluation methods that have been conducted; which are the "speak-out loud" and questionnaire technique.

### **5.3 Procedures**

We installed the prototype to a laptop with 40 gigabytes of storage to make it portable for testing requirement. The laptop is used as a server using ASP technology with Windows XP Professional operating system.

The methods of evaluation are 'speak-out loud' and questionnaire technique. All participants are given their own time to test the prototype individually. First, they are given a briefing on what are the objectives of the test.

In 'speak-out loud' method, they are not shown how to handle the system, but they can ask if they have doubt on how to use the prototype in searching for UML holdings. This method is to evaluate the system's learnability. Then, they are asked to find certain book using the prototype. They are encouraged to speak out their ideas and comments loudly without interference from others. The researcher notes participant's movement and comments.

Immediately after the test, they have been asked to answer a questionnaire given. 10 significant questions are asked to the participants to assess usability factors of the prototype. All these questions derived from usability testing conducted earlier by Perlman (1997) and Lewis (1995). The main goal is to come out with a simple questionnaire but capable to obtain feedbacks. At the end, the result of the questionnaires are recorded and analyzed.

#### 5.4 Demographic information

We focus to target respondents from a small but represent sample of potential UML users. According to a paper published by Virzi (1992), 4 or 5 users can detect 80% of all usability problems. Additionally, the first few users will most likely detect the biggest usability problems. Nielsen and Landauer (1993) provided further support for this theory when they found that the first five users will uncover about 70% of the major usability problems and the next few users will find nearly all the remaining problems, up to 85% or so. However, we followed suggestion by Perfetti and Landesman (2001), which stated that more number of respondents might bring in more accurate feedback from the potential users.

There were 20 volunteers participated in testing the prototype. The range of their age is between 18 to 35 years old. This range of age is selected because most UML users will belong to this group. In detail, 90% of the participants range 18 to 25 years old, and 10% are between 26 and 35 years old; as shown in figure 5.1 below.

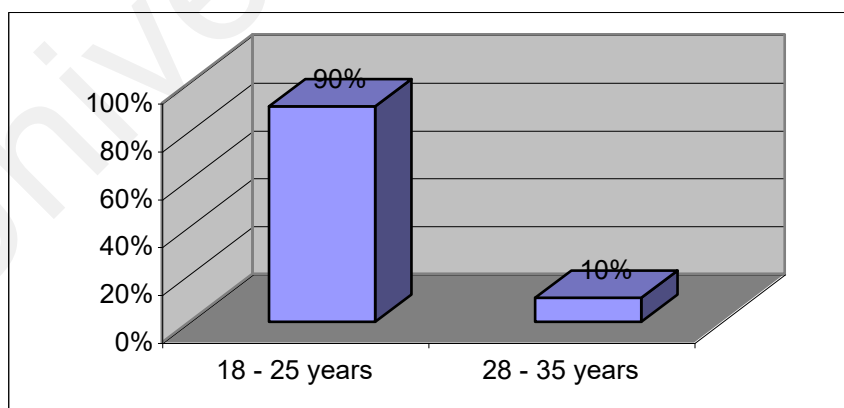


Figure 5.1 Percentage of participants' group range

60% of the participants are female and 40% of them are male. They have different background from academicians to private workers. 13 out of 20 participants are students, 4 are research assistant, 1 is a web programmer and the other 2 are working in private sectors. The percentage of this profession group is shown in the following figure 5.2.

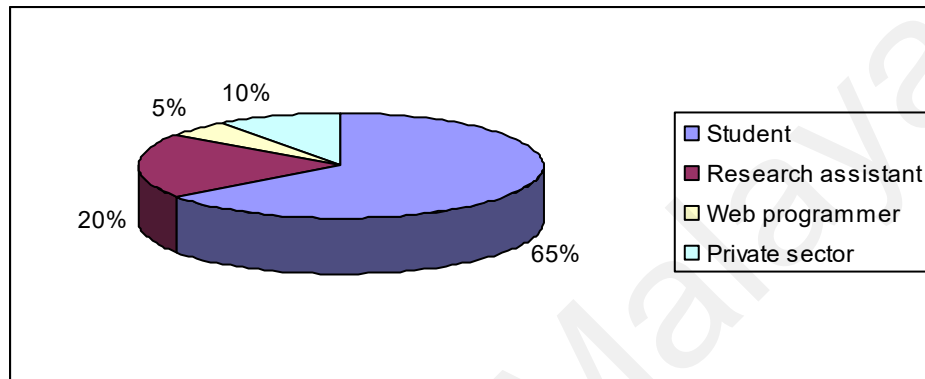


Figure 5.2 Percentage of participants' profession group

All of them have experience in using computer. This can be expected because most of Klang Valley's populations are computer literate and 90% of them are familiar with OPAC system.

To test on 3D usage in the prototype; 85% participants have experience in using 3D software and only 15% of them have no experience at all. The figures are as shown in figure 5.3 below.

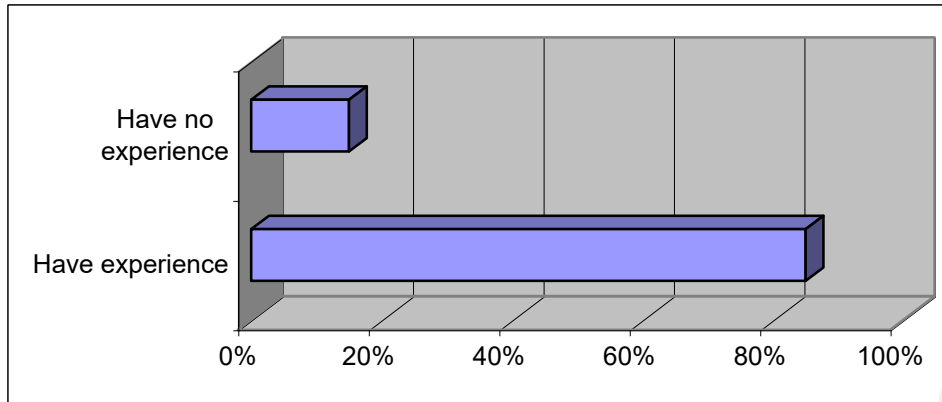


Figure 5.3 Percentage of participants with referral to experience in using 3D software

In summary, the following table 5.1 shows figures for group of participants involved.

Table 5.1 Summary of participant demographic analysis

Description	Group	Number of participant
1. Age	18 – 25 years	18
	26 – 35 years	2
2. Sex	Female	12
	Male	8
3. Profession	Student	13
	Others	7
4. Computer experience	Yes	20
	No	0
5. OPAC experience	Yes	18
	No	2
6. 3D software experience	Yes	17
	No	3

## **5.5 Findings**

We will discuss in detail the results of the experiment we carried out in the following subsections.

### **5.5.1 Assist users in finding information**

The main result of the experiment is the evaluation of the visualization method capability to assist users in finding information from UML database. From the observation we made, most users who have experience using 3D software will directly try to use the VRML screen. They feel that they can handle the rotation and excited to take control of the prototype. Since the system is new to them, they have taken few minutes to understand how the prototype works by trying each function on the prototype.

We also observed that some of the participants who have been using VRML browser quickly understand that nodes are clickable and the cones can be rotated. They found that the node clicked will pop-up search details based on the node's title. They also have no problem on using shortcuts and buttons to pan, zoom, rotate or move up, down, left and right in Cortona VRML browser. However, they feel that the 3D scene will take longer time to search for a specific book title.

For the rest of the group who do not have experience in using 3D software will try the familiar screen of OPAC to search for books. When asked to try the 3D scene, they feel faltered because the scene has no steady plane and uses 3D space to the maximum.



However, most of them do not realize the connection between current OPAC screen and the 3D screen, which stood next to each other. Thus, they are not using the prototype to the maximum. None of them knew that the OPAC screen can be used to search for a particular book, and the 3D scene can be used to know what books laid in a certain subject. For example, if they type 'programming language' in OPAC system using 'title' method, they will only retrieve books, which have 'programming language' as its title. However, by using the 3D scene, they can search for node named 'programming language' and when they clicked on it will give the user books title which lay under 'programming language' subject headings. In other words, the Cat-a-Cone will give users more accurate search results.

### **5.5.2 Usability factors**

The usability factors evaluation is based on the questions asked in the questionnaire as follows:

- **It was simple to use the system**

When asked the above question, 50% of participants disagree with the statement. 15% of respondents agree and 35% strongly agree that it is simple to use the system. It can be concluded that half from the users would find the prototype is easy to use and another half will feel the other way.

This figure derived from the situation that most of the users have no experience on using Cortona VRML browser, though 80% of respondents

have experience in using 3D application including 3D games, which are popular among computer games lovers nowadays. The following figure 5.4 shows fraction of the figures.

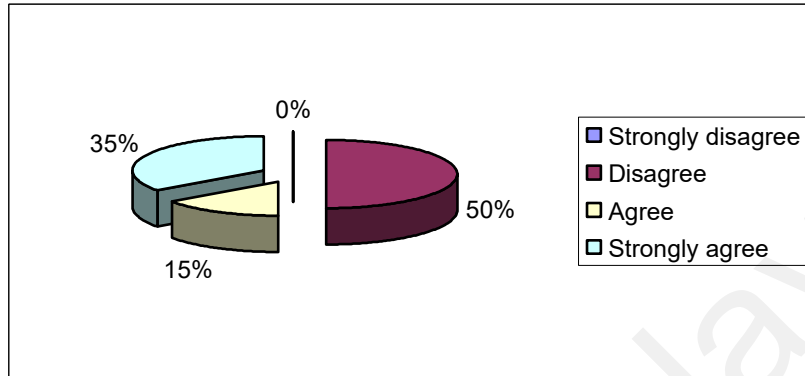


Figure 5.4 Percentage of users' opinion on system simplicity

- **Whenever I make a mistake using the system, I recover easily and quickly**

This question is to test the ability of error recovery and easy reversal of actions among users. One participant strongly feels that this system prohibits easy reversal of action, while 4 out of 20 participants disagree to the above statement.

On contrary, 7 participants agree that users can recover easily when made mistakes; the remainder 8 users strongly feels positive about the statement. In summary, many users would find that the prototype permits easy reversal of actions and they can recover quickly when they made mistakes. Figure 5.5 below shows the figure in percentage.

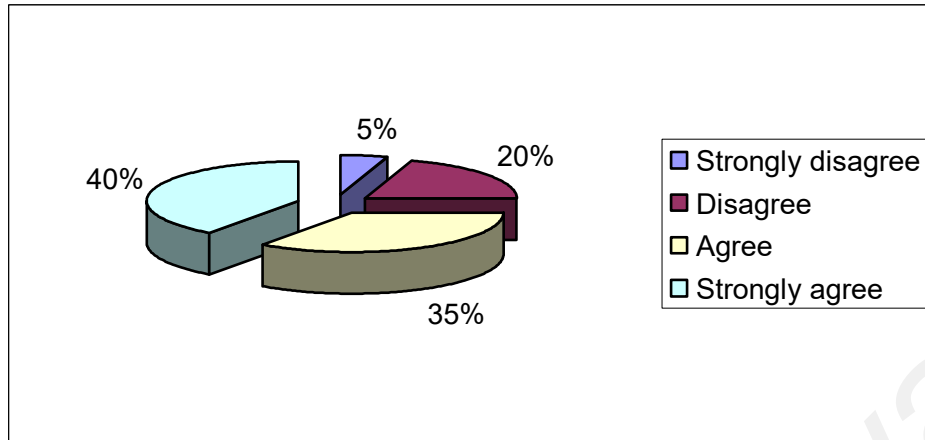


Figure 5.5 Percentage of opinion on easy reversal of actions

- **It is easy to find the information I needed**

Basically, 80% of participants agree that this system really help them in finding information required. In detail, 60% of participants strongly agree to the above statement and 20% agree. However, another 20% feels the opposite way.

It is shown that more than half of participants found it is easy to look for information needed when they use the prototype. The following figure 5.6 shows fraction of figures in exact.

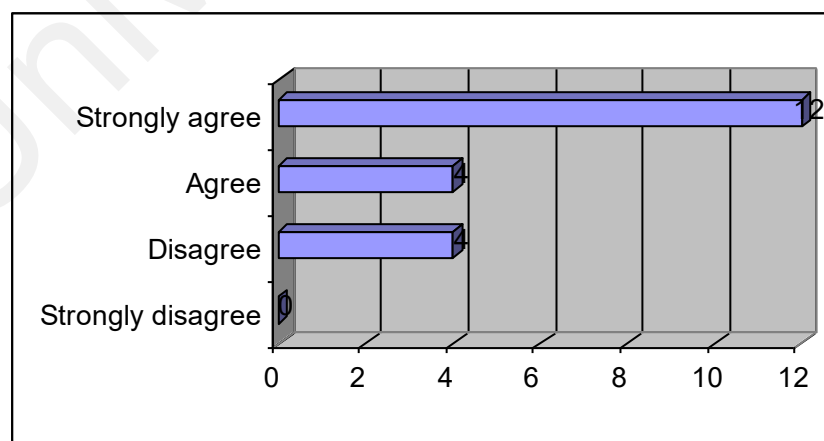


Figure 5.6 Exact figure on level of easiness to find information using the system

- **The organization of information on the system is clear**

This question is to evaluate how far the participants realized about the hierarchical structure of information in the visualization method introduced to them. Referring to figure 5.7 below, 30% of participants strongly agree that they understand the organization of information on the screen, 15% only agree to the statement. 35% of participants disagree with the statement while the remainder 20% strongly disagrees.

The figure 5.7 shows that more than half of users do not understand about information structure in the 3D environment. This explains why users do not take full advantage on search method using both OPAC and the 3D visualization.

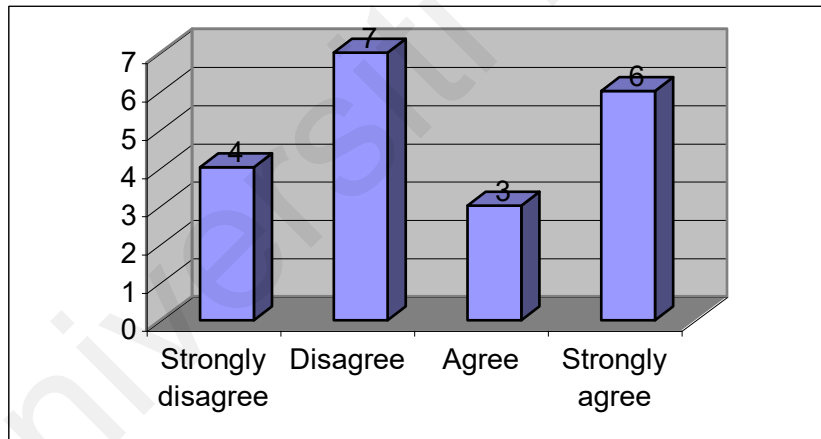


Figure 5.7 Numbers of participants clear with information organization

- **I like using the interface of this system**

This statement is to assess users' preference against the system's interface. 35% of them like to use the interface of this system, with 57% of

them strongly prefer the interface and the remainder 43% only agree to the above statement.

65% of 20 participants dislike the interface. 92% of them disagree with the interface and 8% strongly disagree. In general, most respondents found the system's interface is not attractive. Figure 5.8 shows the fraction in visual form.

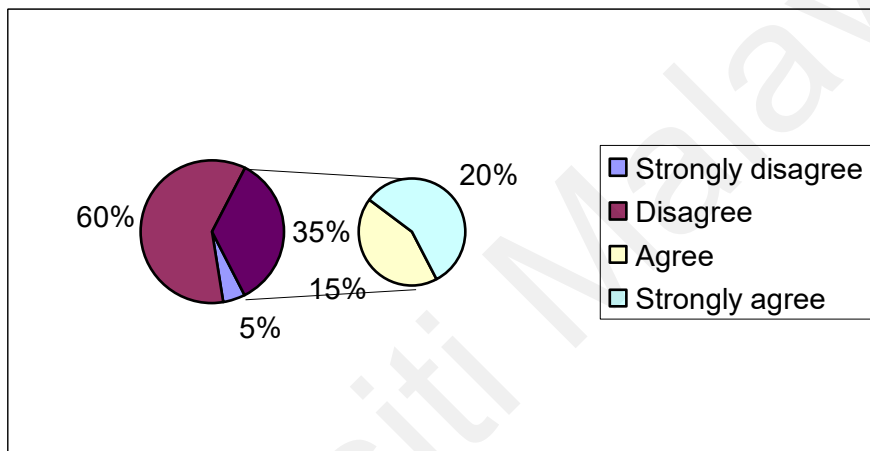


Figure 5.8 Percentage of users in preference against the interface

- **This system has all the function and capabilities I expect it to have**

Figure 5.9 below shows the fraction of participants' opinion on the above statement. This statement is to assess how far the system fulfills users' expectation and need.

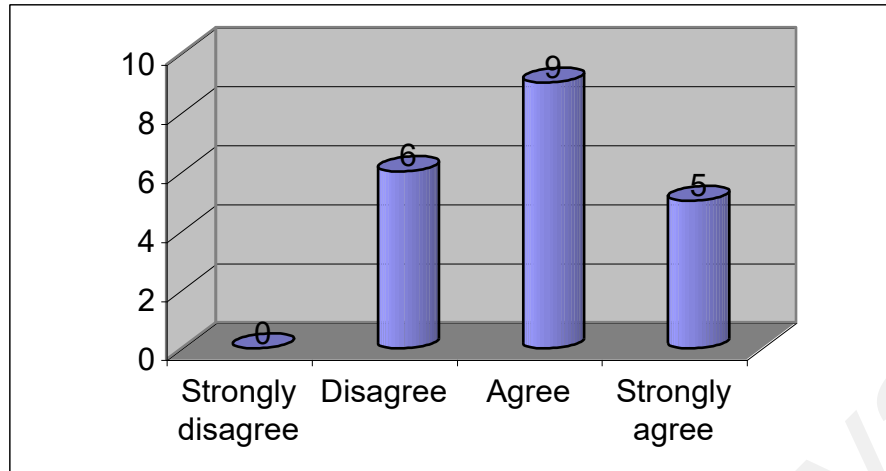


Figure 5.9 Participants expectation on system capabilities

It was found that 70% of participants agree that the system has all the function and capability they expected, 36% of them strongly agree and 64% of them agree with the statement. However, 30% of the participants feel that the system does not achieve their expectation. This shows that most users would find that the system is reliable to carry out the task they wanted it to accomplish.

- **Using this system in my job would enable me to accomplish tasks more quickly**

This question is to obtain participants' opinion on whether the prototype assists them in finding information more quickly. 75% of participants agree that this system assists them to accomplish task more quickly; while 25% feels the other way.

In detail, 60% of participants strongly feel the prototype helps them to accomplish task in less time and 15% only agree to the above statement.

Figure 5.10 explains these in pie chart with exact number (not in percentage).

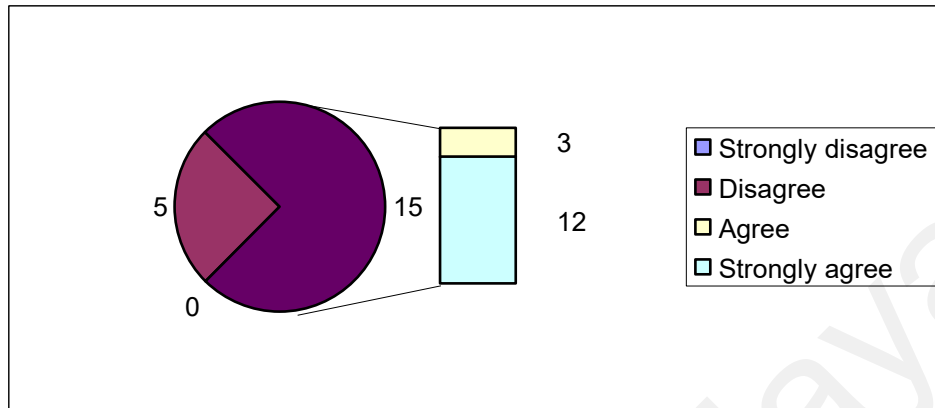


Figure 5.10 Fraction of users opinion on accomplishing task quickly

- **My interaction with the system would be clear and understandable**

Figure 5.11 below shows how participants in this experiment understand what is going on along their interaction with the system.

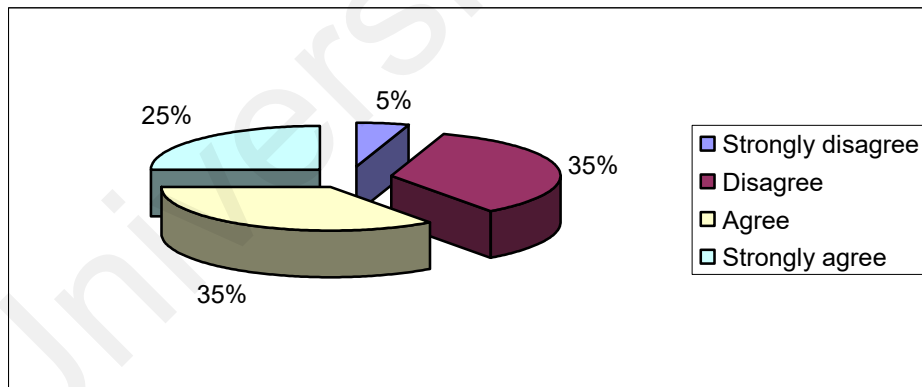


Figure 5.11 Percentage of users' understanding of system's interaction

The experiment found that 12 participants feel they can interact with the system. They also find that the interaction is clear. However, 8 out of 12 participants are against this fact. It shows that more than half of

respondents believe their interaction with the system would be clear and understandable.

In detail, 5 participants strongly disagree with the above statement and 7 participants disagree. Another 7 participants agree on the fact that the interaction is clear and understandable and the remainder 5 participants strongly agree to the statement.

- **The system is easy to use**

This question is to obtain the learnability aspect of the prototype. 60% of participants feel that the system is easy to use; while 40% of them disagree to this statement. The following figure 5.12 shows fraction of users' opinion on the above matter.

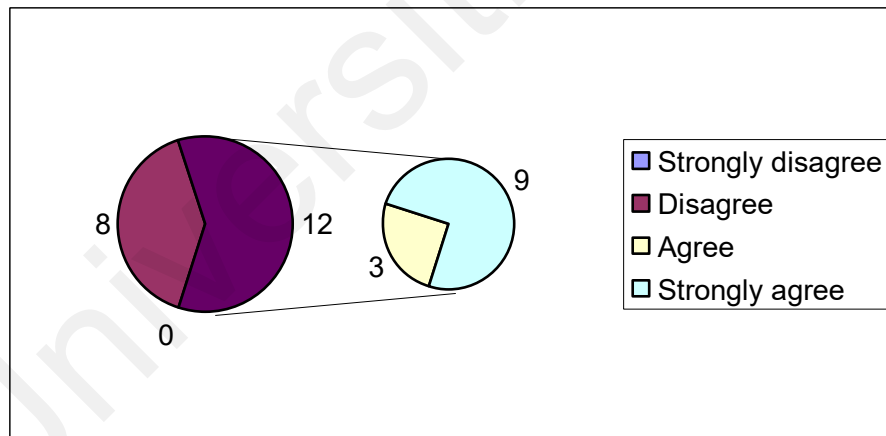


Figure 5.12 Users' opinion on learnability aspect of the system

- **System speed is fast enough**

This question is vital because 3D scene rendering always take amount of time to be completed. However, 65% of participants feel they can accept



the system speed while the remainder 35% cannot compromise with the system speed. This shows that the system speed is acceptable by most users.

To be exact, 60% of participants strongly agree with the above statement, 5% agree, 20% disagree and 15% strongly disagree. The following figure 5.13 shows the fraction in detail.

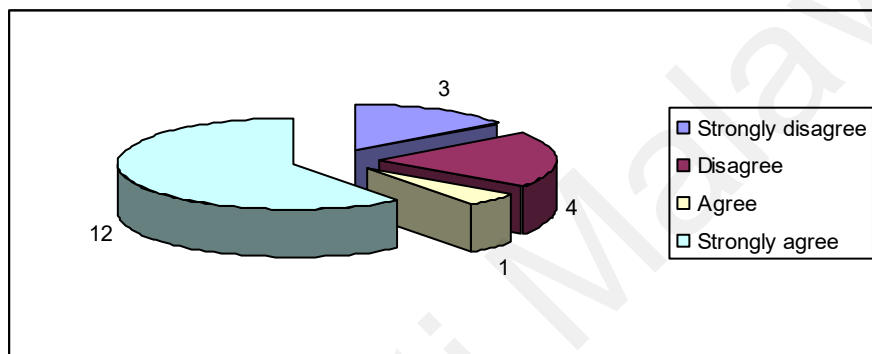


Figure 5.13 Numbers of fraction on users acceptance on system speed

As we can see, most of the questions asked bring positive feedback from the users. This shows that the prototype is acceptable when judged on usability issues.

Table 5.2 below is the summary of the questionnaire results taken after the evaluation process. The ranking 1 is for strongly disagree, 2 for disagree, 3 for agree and 4 for strongly agree.

Table 5.2 Summary of the questionnaire results

No	Questions	Ranking			
		1	2	3	4
1	It was simple to use this system	0	10	3	7
2	Whenever I make a mistake using the system, I recover easily and quickly	1	4	7	8
3	It is easy to find the information I needed	0	4	4	12
4	The organization of information on the system screens is clear	4	7	3	6
5	I like using the interface of this system	1	12	3	4
6	This system has all the functions and capabilities I expect it to have	0	6	9	5
7	Using this system in my job would enable me to accomplish tasks more quickly	0	5	3	12
8	My interaction with the system would be clear and understandable	1	7	7	5
9	The system is easy to use	0	8	3	9
10	System speed is fast enough	3	4	1	12

### 5.5.3 Feedback and comments

At the end of the questionnaire, we enclosed open-ended questions to obtain users' perception on positive and negative aspects of the prototype. The following sub sections will discuss further about this matter:

#### 5.5.3.1 Positive aspects

Some users commented the visualization technique as a new good concept to introduce in Malaysia. The technique is not meant to be used to display a database structure but also have potential in presenting large structure of other information; such as in World Wide Web.

Some stated that the visualization technique is able to display overall structure of categorized search result very well. They feel that

organization of subject headings gives much new information that could not be accessed before using OPAC. It helps in displaying overall structure of the database and parent-child relationship for each node. This is important as it enables the user to broaden or deepen the search process.

The capability of the visualization technique to display the whole database on one screen has led to another positive aspect, which is saving time. These users feel that the visualization technique has helped them in understanding relationship between nodes and the time taken to decide which category is relevant to the users' interest can be shorten.

Since the visualization technique introduces new ways of finding information on database, many are attracted to the colors that are used in the system. They state that the colors are interesting, and the search process is fun and interactive.

Other positive comments from the users are good information flow. Our consideration in deploying simple and clean design in the prototype has gain appreciation among the participants. The interface is straightforward and the results of interactions with the prototype are predictable.

Most users find that the prototype provides acceptable rate of system speed on opening the web site for this prototype. However, the following screens provide fast system speed for online application.

#### **5.4.3.2 Negative aspects**

Contrary from positive aspects discussed above, some users complained about the size of font; which they feel too small and difficult for them to read. Leading from the same problem, some users feel uncomfortable to see the topics that are listed on the screen. This is true because the main idea of this visualization technique is to show overall description of the whole database. To put everything on one screen, we have to sacrifice the font size and hope that the users will take advantage on the capability of the database structure to search for UML holdings.

There are also possibilities that users might be lost in 3D space while perform the search. This situation that has been pointed out by a user is significant with previous findings by other research in the past about using 3D environment to visualize information. However, users can always press the 'restore' button to bring the screen back to initial state before any manipulation made by users.

Few users also show that it is very hard for them to find specific book using the visualization technique proposed. As mentioned in earlier section, the visualization technique is proposed to help users in finding information. They cannot be used solely to find the

information; without the use of OPAC system. In other words, the OPAC system and the 3D visualization technique have been merged and used together to gain maximum benefit to users in finding information.

Apart from the above aspects pointed out, another negative aspect is the technology constraint. Different VRML browser may support different perspective of nodes in VRML. For example, VRML browser named Cortona® VRML Client able to support Cone Sensor node, while another VRML browser known as Community Place maintained by Sony Ltd does not support this same node. To overcome the difficulties, we have made the effort to let the users download Cortona® browser if they do not have them when they first open *url* of this application.

However, all the negative aspects discussed above will not interfere with the efficiency of the system to visualize UML database.

## **5.6 Summary**

This chapter discusses about the evaluation process and feedbacks received from the prototype evaluation.

Early section in this chapter explains the two objectives of this evaluation process; which are to assess usability factors in the prototype and to note positive and negative comments from the users. It is clear that the two objectives are achieved; which has been discussed in the later subsections. The following section describes how the evaluation process has been carried out; by using two methods. The methods are 'speak-out loud' and questionnaires answering session.

In section 5.4, we present demographic information of the participants involved in the study. This is crucial to make sure that we have selected only participants, which are likely to represent UML users, as our domain of interest.

The last section discusses about the results of the evaluation process; which are divided into three subsections. The first subsection shows how the habit of specific users helps them in finding UML holdings. The later subsection lay out the assessment of usability factors through results from questionnaires answered. And the last subsection notes the positive and negative aspects of the prototype as pointed out by the participants as their feedbacks.

The next chapter will discuss further the results of this research and future enhancement to be made for this study.

## CHAPTER 6

### CONCLUSION

#### 6.1 Introduction

Information visualization is still a new concept to Malaysian academic community even though the research in this field has been conducted since mid 1980s in developed countries. The research has contributed to many research in Malaysia to introduce 3D information visualization tool and it is used in our research.

This chapter will conclude all the findings cited along the research progression. The subsequent subsections will describe the strengths and weaknesses of the study to be enhanced in the future.

#### 6.2 Discussion

The research begins with an overview of the study domain. Existing problems in the OPAC system have been addressed in the first chapter of this dissertation. We present *Doc-with-Cone* to overcome these problems as describe below:

- **Difficulty in finding correct word for searching**

Using subject heading as the nodes' name has many benefit as a result it has been used to catalogue the vast number of books produced everyday all around the world. Using subject heading as a node, users will not need any right keyword to make sure they will be displayed with details to the correct books anymore. Users can select a node to start their search and follow the path of the node's hierarchy in order to find the correct subject heading that may interest them. For example, to find book with C++ book, users just have to start with 'Computer Science' node and follow the hierarchy until they see 'Programming languages' node and they will find 'C++'.

Compared with the OPAC system, a book with title 'Bible for computer programmer in banking sector' may discuss in detail about programming using C++ in banking sector. But a search with 'C++' as the title keyword will only retrieve books with 'C++' as their title.

- **Difficulty to decide which information is relevant**

The ability of seeing the whole hierarchy of the database gives another benefit to users in deciding relevant information. A user who seeks for computer programming book may enter keyword such as 'programming' or 'computer programming'. The keyword search may provide the user with too many or too few books titles. The user will have difficulty to decide which book that may contain information that they are looking for.

However, using the visualization technique described previously, the user can decide which node is relevant to their interest and which are not.



- **The working memory load**

There is an epigram says, “A picture is worth thousand words”. Based on this, many people use diagram and graphs to represent data in order to simplify the textual description. A view on the visualization tool in representing UML holdings can reduce working memory load to understand the basic organization of the database.

- **No alternative for different group of users**

As stated in earlier chapter, the novice users and the experts has their own choice in selecting ways to perform their search for UML holdings. Novice users can still use the current OPAC interface even after they were introduced to the 3D information visualization technique tool. There are also help menus and overview map to guide them in using the 3D interface. For expert users, they can use the 3D interface screen to maximize the capabilities offered by the system, such as direct interaction, pop-up hint box and color coding.

- **Time consuming**

All the problems in using OPAC system will contribute to time-consuming factor. The solutions provided by *Doc-with-Cone* will save time in performing the search. This is supported by findings in the evaluation process, where 75% of participants found searching performance using *Doc-with-Cone* is faster compared to OPAC system.

There are three objectives of this study as described in Chapter 1. All these objectives are achieved based on the evaluation process conducted and

discussed in Chapter 5. Only the first objective that is to build a usable information retrieval system is agreed by half of the participants. While the other two objectives which are to help users find the information faster and more accurate and also to introduce new concept of information retrieval using 'information visualization' has been achieved with an agreement by the majority of the participants.

Information visualization is a transformation of non-spatial or behavioral data into visual images that represents metaphor of the information domain. Meanwhile 3D is becoming popular and common among computer applications including computer games, educational software, entertainment and application tools.

This research is conducted with stress on 3D information visualization application in UML. A preliminary research has been done earlier as a guidance to clear the small differences between classes of visualization techniques. The research has been discussed in Chapter 2; which states that this research has been completed in the domain of information visualization.

The various benefits of information visualization technique that has been described in Chapter 2 have been proven based on the results and feedback received from the participants on the prototype evaluation. The proposed prototype is able to overcome the problem of increasing amount of data and information, help the users to have overall understanding of the information, contribute in development of more usable information retrieval system and able to exploit visual power in human-brain processing.

The literature review conducted at the early stage of the research has shown the strengths and enhancements needed from previous research in the same domain. Based on these, we have adopted an information visualization technique according to criteria classification suggested by Keim(2002). The information visualization technique, which contains the structural organization of UML databases, is displayed to the users. We adapt both iconic displays and stacked displays as its visualization technique and some interaction and distortion techniques. The interaction and distortion technique used in the prototype is interactive zooming; where users can zoom in the interesting part of the visualization. The interactive distortion helps users understand and process visualized information by preserving the context of the whole information structure. The capability of interactive distortion is offered by the 'Overview map' in the prototype. Another visualization technique embedded in the proposed tool is interactive linking and brushing; where select gesture can link a data to its details using hyperlink capability in both HTML and VRML environment.

Some related issues that have been considered in developing the prototype; as discussed in Chapter 2 are described below:

- **Provide informative feedback**

The *Doc-with-Cone* provides informative feedback by using cluster of information to show groups of UML holdings subsiding in different subject headings. Users can easily find information faster and effectively.

- **Permit easy reversal of actions**

Anytime the users need to perform a new search, the users can click on the 'Restore' button provided in VRML browser to start over the 3D scene. To start a new search using the OPAC interface, users can click on 'Refresh' button of the internet browser.

- **Support an internal locus of control**

The feel of being in control while using a system is a crucial part in developing a usable system. All the direct interactions in the 3D scene are understandable using common interaction while surfing the net. The changes on mouse pointer show that an object can be manipulated; moved, clicked or rotated in performing the search.

The following issues as discussed earlier are the problems of existing OPAC system:

- **Reduce working memory load**

- **Provide alternative interface for novice and expert users**

In chapter 3; which discusses research methodology and design specification, we provide the concept prototype of *Doc-with-Cone* in subsection 'Alternative generation and selection'. This concept prototype has been introduced to some potential users in preliminary evaluation process to receive their opinion in developing a usable prototype.

Based on users' feedback and steps taken as discussed in Chapter 3, we have developed the *Doc-with-Cone* as a working prototype to perform search process for UML holdings. The details of prototype development and specification have been discussed in Chapter 4. We also have described the minimum system requirement and overview of *Doc-with-Cone* screen shots in the following sub sections. Later in Chapter 4, we discussed the usability factors embedded in the prototype.

To evaluate whether the objectives of this research has been achieved, we have conducted an evaluation and statistical analysis; as described in Chapter 5. The negative and positive feedbacks from the participants will be considered in future research in the same domain.

### **6.3 Advantages of study**

Apart from the advantages of the prototype discussed in earlier chapters, perhaps the most positive comment received from the participants of evaluation process is the developed prototype has introduced a new way to perform a search in book retrieval. The same basic idea can be embedded to represent hierarchical information online.

The experience in developing the prototype shows that a modeling language such as VRML and application tools such as Dreamweaver MX can produce a usable and powerful tool to visualize information. The fact that we are new in

using all the technology and the research field is not an obstacle to develop a working prototype that will be considered in future.

The system proposed indicates that information retrieval system used in libraries should be updated to accommodate new technologies. The current monochrome screen applied in many OPAC systems should be replaced with more interesting interface, such as applied in *Doc-with-Cone*.

This research has contributed to many research done in 3D information visualization. However, there are not many research done in Malaysia as an attempt to visualize information for databases; particularly large databases that reside in libraries.

The significance of this study is that a new technique for information retrieval based on “information visualization” has been studied. This technique is an alternative to OPAC system, which is currently used worldwide. Even though the technique involved limited the number of databases, further enhancement could be done to accommodate bigger collection of databases that are suitable for libraries. This study has contributed to the new technique of information visualization based on 3D graphical format.

#### **6.4 Weaknesses & future enhancement**

There is a major concern when we visualize a very large database that holds millions of holdings reside in many subject headings. The working prototype only

holds 248 records with 180 nodes visualized. A node may have maximum 56 sub nodes. These figures are only applicable under 'Computer Science' subject headings. In total of all UML holdings especially in bigger libraries such as Library of Congress and even for Malaysia's National Library; the 3D information visualization may be cluttered.

To overcome this problem, a new way of visualization technique will be introduced in the future to enhance research in information visualization domain.

Referring to the evaluation process conducted at the end of the research, more than half of participants commented that the prototype is difficult to use and a bit complicated. This shows that Malaysian academic community is still new to this concept and it may take some time for us to apply a 3D visualization tool for use in public.

## **6.5 Summary**

Chapter 6 concludes the whole dissertation in order to evaluate how far the main objective of the research has been achieved.

In a nutshell, we hope that this study may trigger further interest in information visualization technique to other academicians and scientist; mainly in Malaysia.

## REFERENCES

- Ames, A. L. et al (1997). VRML 2.0 Sourcebook, 2<sup>nd</sup> ed. Canada: John Wiley & Sons.
- Ariyapa, P.G. and Edzan, N.N. (2002). Foreign postgraduate students and the online catalogue at the University of Malaya Library. *In Malaysian Journal of Library & Information Science*. 7(1). P. 57-67.
- Baeza-Yates, R. and Ribeiro-Neto, B. (1999). Modern information retrieval. New York : ACM Press.
- Bushell, C. et al. (1991). Information Visualization : The next frontier. *In Proc. Of the 21<sup>st</sup> annual conference on Computer Graphics and Interactive Techniques*. p. 485-486. New York : ACM Press.
- Gershon, N., Eick, S. G. and Card, S. (1998). Information Visualization. *In Interactions*. 5(2). p. 9-15.
- Herman , I., Melancon, G. and Marshal, M.S. (2000). Graph Visualization and Navigation in Information Visualization. In *IEEE Transactions on Visualization and Computer Graphics*, 6(1), p. 24-43.
- Hoffer, J. A. et. al (1996). Modern Systems Analysis and Design. California: Benjamin-Cummings Publishing.
- Hornbæk, K. and Frøkjær, E. (2001). Reading of Electronic Documents: The Usability of Linear, Fisheye, and Overview+Detail Interfaces . *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Washington, United States. p. 293-300. New York: ACM Press.
- Hornbæk, K. and Frøkjær, E. (1999). Do Thematic Maps Improve Information Retrieval? *In International Conference on Human-Computer Interaction (INTERACT '99)*, Edinburgh, Scotland. p. 179-186. Scotland: IOS Press.
- Keim, D.A. (2002). Information Visualization and Visual Data Mining. In *IEEE Transactions on Visualization and Computer Graphics*, 8(1), p. 1-8.
- Kroenke, D.M. (1995). Database processing: Fundamental, design and implementation. New Jersey: Prentice Hall.
- Lamping, J. and Rao, R. (1996). Visualizing Large Trees Using the Hyperbolic Browser. *In Proceedings of the CHI '96 Conference Companion on Human Factors in Computing Systems*, Vancouver, Canada. p. 388-389. New York: ACM Press.



- Leuski , A. and Allan, J. (2000). Lighthouse: Showing the Way to Relevant Information. In *IEEE Symposium on Information Visualization 2000*, Salt Lake City, USA, 9-10 October 2000. p. 125-165.
- Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. In the *International Journal of Human-Computer Interaction*. (7), (1). P. 57-78.
- Library of Congress, Cataloging Policy & Support Office (1998). Subject Headings, LC Classification Schedules. Detroit : Gale Research Inc.
- Mackinlay, J.D , Robertson, G.G, and Card, S.K (1991). The Perspective Wall: Detail and Context Smoothly Integrated. In *Conference on Human Factors and Computing Systems*. p. 173-176.
- Morton, J. L. Color and usability matters. Date retrieved: 18 November 2004. <http://www.colormatters.com/usability.html>.
- Munzner, T.(2002). Information visualization. In *IEEE Computer Graphics and Applications*. 22(1). p. 20 – 21.
- Myers, B.(1996). Strategic Directions in Human-Computer Interactions. In *ACM Computing Surveys*. 28(4). p. 794-809.
- Nielsen, J.(1999). Designing Web Usability. New York: New Riders Press.
- Nielsen, J., & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. In the *Proceedings of INTERCHI 93*, p. 206-213. New York, New York: ACM Press.
- OLIVE homepage. <http://www.otal.umd.edu/Olive>. Last updated: October 7, 1997. University of Maryland.
- OpenGL's main homepage. <http://www.opengl.org>. Retrieved March 20, 2004. Last updated: 2004.
- ParallelGraphics main homepage. <http://www.parallelgraphics.com>. Retrieved January 10, 2005. Last updated: January 2005.
- Perlman, G. (1997). Practical Usability Evaluation. In *CHI Extended Abstracts 1997*. p. 168-169.

- Perfetti, C., & Landerman, L. (2002). Eight is not enough. [http://www.uie.com/articles/eight\\_is\\_not\\_enough.htm](http://www.uie.com/articles/eight_is_not_enough.htm). Retrieved April 14, 2003.
- Persidsky, A. (2000). Director 8 for Macintosh and Windows: Visual QuickStart Guide, 5<sup>th</sup> ed. Berkeley: Peachpit Press.
- Rekimoto, J. and Green, M. (1993). The Information Cube: Using Transparency in 3D Information Visualization. In *Proceedings of the Third Annual Workshop on Information Technology and Systems (WITS '93)*, p. 125-132.
- Robertson, G.G, Card, S.K and Mackinlay, J.D (1993). Information Visualization using 3D Interactive Animation. In *Communications of the ACM*. 36(4). p. 57-71.
- Robertson, G.G, Card, S.K and Mackinlay, J.D (1989). The Cognitive Coprocessor Architecture for Interactive User Interfaces. In *Proceedings of the 2<sup>nd</sup> annual ACM SIGGRAPH Symposium on User Interface Software and Technology*. Williamsburg, Virginia, 1989. p. 10-18. New York : ACM Press.
- Schaffer, D. et al (1996). Navigating Hierarchically Clustered Networks through Fisheye and Full-Zoom Methods. In *ACM Transactions in Computer-Human Interactions*, 3(2), p. 162-188. New York : ACM Press.
- Shneiderman, B. (1997). Designing the User Interfaces. Massachusetts : Addison-Wesley.
- Tegarden, D. P. (1999). Business Information Visualization. In *Communication of the AIS*, 1(4), p. 1-38. Atlanta : Association for Information Systems.
- Theimer, S.(2002). When a 21<sup>st</sup> Century User Meets a 20<sup>th</sup> Century OPAC: How Word Choice Impacts Search Success. In *PNLA Quarterly*. 66(3). p. 23 – 24.
- University of Malaya Library's homepage. <http://www.umlib.um.edu.my>. Last updated: November 12, 2002. University of Malaya Library.
- University of Maryland Human-Computer Interaction Laboratory homepage. <http://www.cs.umd.edu/hcil/webtoc>. Last updated: December 22, 1997. University of Maryland.

Virzi, R. A. (1992). Refining the test phase of usability evaluation: How many subjects is enough? In *Human Factors*, 34(4), p. 457-468.

Werner , K. (2002). A Comparison of Information Visualization Methods. In *Conference on Human Factors and Computing Systems*, Minneapolis, USA, p. 872-873.

Universiti Malaya