

**UNIVERSITI TEKNOLOGI MARA**

**FACILITATING RESOURCE  
ALLOCATION DECISION  
THROUGH BIBLIOMINING: THE  
CASE OF UTM'S LIBRARY**

**MD RAZIB BIN KARNO**

Thesis submitted in fulfillment  
of the requirements for the degree of  
**Master of Science**

**Faculty of Information Management**

January 2015

## ABSTRACT

Library has vastly developed and demand from the users, institutions, international organization needs and technology advancement has changed the library planning and decision making approach in many ways including library budgeting, human resource and infrastructure allocations. This research described (a) the investigation undertaken to examine the characteristics of data from data reservoirs regarding user/patron information and circulation information. (b) The information seeking to explore the patterns and trends among these data reservoirs using data mining analysis with about 957,224 borrowing history and overall 31,052 registered readers and 139,195 title author of books from the Universiti Teknologi Malaysia library since 2008 to 2010. (c) To study how constructed patterns and trends generate informed decisions on resource allocation for circulation function by using cluster analysis, frequency statistics, averages and aggregates and market basket analysis algorithm. This thesis highlights the finding of a research using data mining technique (CRISP-DM) to explore the potentials of the bibliographic data of an academic library. With nearly 1 million records of collection in various formats, the Library of Universiti Teknologi Malaysia has been chosen as the case study for the research. The data mining technique was adopted to explore the relationship among statistically patterned and clustered bibliographic data. Bibliomining are tools that can visualize how libraries manage their costs, staff activity, customer service, user needs, marketing, popular books, circulation, reference transaction, quality of collection, educational programs etc. Similar data mining techniques are suggested to be employed in different library settings and even enterprises as to make more effective use of organizational resources.

## ACKNOWLEDGEMENT

Firstly, I wish to thank Allah s.w.t for giving me the opportunity to embark on my MsC and for completing this long and challenging journey successfully. My gratitude and thanks go to my supervisor Dr. Siti Arpah Noordin. Thank you for the support, patience and idea in assisting me with this project. I also would like to express my gratitude to the staff and lecturer of the Faculty of Information Management, UiTM Shah Alam, especially Mr Safawi Abd Rahman for providing the knowledge and assistance.

My appreciation goes to the Chief Librarian, Head of Department and staff of UTM Library, Skudai who provided the facilities and assistance during sampling. Special thanks to my colleagues and friends for helping me with this project.

The completion of this study would have been not possible if not dependent on the steadfast support and encouragement of my mother Khamsiah Hj Sulaiman, my very dear late father Karno Hj Hashim, my lovely wife Sa'idah Hashim, my two amazing daughter Nafisah and Ayuni. They hence paid equal contribution to the study for which I always feel profound gratitude in my heart.

# TABLE OF CONTENTS

	<b>Page</b>
<b>CONFIRMATION BY PANEL OF EXAMINERS</b>	ii
<b>AUTHOR'S DECLARATION</b>	iii
<b>ABSTRACT</b>	iv
<b>ACKNOWLEDGEMENT</b>	v
<b>TABLE OF CONTENTS</b>	vi
<b>LIST OF TABLES</b>	xi
<b>LIST OF FIGURES</b>	xii
<b>LIST OF GRAPHS</b>	xiv
<b>CHAPTER ONE: INTRODUCTION</b>	<b>1</b>
1.1 Introduction	1
1.2 Context of the Study	2
1.3 Problem Statement	6
1.3.1 Unorganised and Hidden Data Scattered Within the Library Bibliographic Data Reservoirs	6
1.3.2 Misinformation of Decision Making	6
1.3.3 Lack of Expertise, Experience and Interest in Data Mining	7
1.4 Research Questions and Objectives	7
1.5 Background Information of the Organization Where the Study Was Conducted	9
1.6 Theoretical Framework	19
1.7 Significance Of The Study	21
1.8 Conclusion	22
<b>CHAPTER TWO: THEORITICAL AND CONCEPTUAL FRAMEWORK</b>	<b>23</b>
2.1 Introduction	23
2.2 Evidence Based Librarian	24
2.3 Library Decision Making	26
2.3.1 Traditional Decision Making	26

2.3.2	Decision Making In A Libraries	28
2.3.3	Bibliomining As A Tool for Decision-Making In The Library Context	30
2.3.4	Strategic Planning In UTM	31
2.3.5	Sultanah Zanariah Library, UTM Strategic Planning	32
2.4	Bibliometrics vs Bibliomining	36
2.4.1	Bibliometrics	36
2.4.2	Bibliomining	37
2.4.3	Bibliographic Data Format	40
2.4.4	Variable fields	41
2.4.5	Tags	41
2.4.6	Indicators	42
2.4.7	Subfields	42
2.5	Library Circulation Transaction Data	43
2.6	Data Warehouse	49
2.7	Exploration of Data Sources	51
2.7.1	Data Sources From The Creation Of The Library Information System	51
2.7.2	Data Sources From The Usage Of Collection	52
2.7.3	External Data Sources	53
2.8	Data Mining Procedure	53
2.8.1	Association Rules (Market Basket Analysis)	56
2.9	Conceptual Framework Of The Research	57
	<b>CHAPTER THREE RESEARCH METHODOLOGY</b>	61
3.1	Introduction	61
3.2	Single Case Study	61
3.3	Research Activities	62
3.3.1	Research Tools	62
3.4	Cross-Industry Standard Process For Data Mining (CRISP-DM)	62
3.4.1	Business Understanding	65

3.4.2	Data Understanding	70
3.4.3	Data Preparation	82
3.4.4	Modelling	89
3.4.5	Evaluation	91
3.4.6	Deployment	92
3.5	Data Analysis And Interpretation	93
<b>CHAPTER FOUR: DATA ANALYSIS AND FINDINGS</b>		94
4.1	Introduction	94
4.2	The library system under study	94
4.2.1	Overview	94
4.2.2	System Architecture	96
4.3	Table characteristic	100
4.4	Knowledge Discovery Overview	103
4.5	Exploring data	105
4.5.1	Cluster Analysis	105
4.5.2	Summary Statistics	105
4.5.3	Frequency Statistics	109
4.5.3.1	Example Of Cumulative Frequency	109
4.5.4	Averages And Aggregates	113
4.5.5	Association Rules (Market Basket Analysis)	113
4.6	Bibliomining Result	114
4.6.1	Association Rule (Market Basket Analysis)	119
4.6.2	Cluster Analysis	139
4.7	Borrowing Trend	143
4.7.1	Types Of Item Borrowed	143
4.8	Popular works based On Subject	149
4.8.1	Borrowing Trends And Location	152
4.9	Borrowing And Types Of Degrees And Schools/Faculties	154
4.9.1	Faculty	154
4.9.2	Academics And Non-academics	154

4.9.3 External Members	154
4.10 Conclusion	157
<b>CHAPTER FIVE: DISCUSSION</b>	158
5.1 Library Bibliomining And PSZ Strategic Decision	158
5.2 Human Capital	160
5.3 Bibliomining And Library	161
5.3.1 Effective Promotion	163
5.3.2 Reading Stations	164
5.3.3 Library Liaison	164
5.3.4 Faculty Library	165
5.4 Collection Management	165
5.4.1 Cost Effective Analysis	166
5.4.2 Collection Development	168
5.4.3 Return On Investment (ROI)	168
<b>CHAPTER SIX: CONCLUSIONS AND RECOMMENDATION</b>	170
6.1 Introduction	170
6.2 Research Objective Revisited	170
6.2.1 Research Objective Revisited: New Set Of Data	170
6.2.2 Research Objective Revisited: New Knowledge Discovery Using Data Mining	171
6.2.3 Research Objective Revisited: Development of Library Bibliomining Process	172
6.3 Bibliomining Challenges	175
6.4 Future Research	176
6.5 Significance of Research	177
6.5.1 Theoretical contributions	177
6.5.2 Empirical Contributions	177
6.5.3 Practical And Professional Contributions	178
6.6 Conclusion	180

**REFERENCES**

181

**APPENDICES**

189



## LIST OF TABLES

<b>Tables</b>	<b>Title</b>	<b>Page</b>
Table 1.1	Ordering and Receiving Statistics by Faculty Year 2008 (Resource Development Dept. Annual Report 2008)	3
Table 2.1	Overall Percentage of Customer Satisfaction Index 2010	35
Table 2.2	MARC Tags	42
Table 2.3	Data Mining Operations, Features and Benefits	46
Table 3.1	Usage of LMS Modules: Report and Study by UTM Library From 2001-2010	66
Table 3.2	Previous Study in Doing Library Bibliomining	69
Table 3.3	Different Attribute Types	71
Table 3.4	Sample Data Attributes From CIRC.CHKOUT.STATS Filename	75
Table 3.5	Sample Types of Variables	76
Table 3.6	Nine Random Records in Library Bibliographic Data	76
Table 3.7	List of Dataset	77
Table 3.8	Attribute Definition of Patron Table	78
Table 3.9	Attribute Definition of Circulation Table	78
Table 3.10	Attribute Definition of Bibliographic Record Table	79
Table 3.11	Attribute Definition of Pieces Table	85
Table 3.12	Extracted Record From Data Warehouse	89
Table 4.1	Previous Studies on Bibliomining	101
Table 4.2	Usage of LMS Modules: Report and Study by UTM Library From 2001-2010	103
Table 4.3	Cluster by Patron ID Result	116
Table 4.4	Extracted Record From Data Warehouse	121
Table 4.5	Student “Basket” in Borrowing “Technology” Books	124
Table 5.1	Library Strategic Planning and Data Analysis	162
Table 5.2	Return on investment	169

## LIST OF FIGURES

<b>Figures</b>	<b>Title</b>	<b>Page</b>
Figure 1.1	UTM Library Strategy Map (2009-2013)	5
Figure 1.2	UTM Digital Library Systems	11
Figure 1.3	UTM Digital Library System: Integration Overview	13
Figure 1.4	Theoretical Framework	20
Figure 2.1	Potential Opportunities and Constraints	27
Figure 2.2	Benefits of a Business Intelligence System	29
Figure 2.3	UTM's Library Weakness and Strength	33
Figure 2.4	The Application Flow of Bibliomining	39
Figure 2.5	Data Mining Flow	48
Figure 2.6	Data Mining Models and Tasks	55
Figure 2.7	Research Framework	59
Figure 3.1	CRISP-DM Data Mining Process	64
Figure 3.2	CRISP DM Life Cycle Model	65
Figure 3.3	Data Collection Process	72
Figure 3.4	Patron Data Record	74
Figure 3.5	Sample Data From Circulation Filename	75
Figure 3.6	LMS Validation Data Model	80
Figure 3.7	Sample of Classifying PCODE Variables	81
Figure 3.8	Workflow of Data Extraction and Cleansing	83
Figure 3.9	List of Data Cleaning Using Excel	84
Figure 3.10	Integration Between Library Management System Modules	86
Figure 3.11	LMS Training Data Model	87
Figure 3.12	Data Transformation for Association Rules	88
Figure 3.13	Cluster Model	90
Figure 4.1	UTM's Digital Library System Integration	95
Figure 4.2	UTMs Digital Library System's Pillars	96

Figure 4.3	Architecture of Bibliomining Process	98
Figure 4.4	Telnet Interface	99
Figure 4.5	Sample of Telnet Command-line	100
Figure 4.6	UTMs Library Web Application Workstation	100
Figure 4.7	Knowledge Discovery Overview	104
Figure 4.8	Sample of Frequency Statistics	111
Figure 4.9	Sample of Frequency Statistics Graph Table	112
Figure 4.10	Patron Type Code	118
Figure 4.11	Cluster Model	118
Figure 4.12	Association Rule Procedure Using SPSS Clementine	120
Figure 4.13	Data Transformation for Association Rules	122
Figure 4.14	Technology Classification and Subclasses	126
Figure 4.15	Student Faculty of Education with Technology Borrowing Relation	127
Figure 4.16	Library Data Initial Stream	129
Figure 4.17	Selected Fields of Bibliomining	131
Figure 4.18	Bibliomining Data	132
Figure 4.19	Library Data Stream Model 1	132
Figure 4.20	Association Rule Model	133
Figure 4.21	Summary of Association Rule	134
Figure 4.22	Web Node Attached to Library Data Stream	135
Figure 4.23	Web Node Select Options	136
Figure 4.24	Web Node Display of Strong, Medium and Weak Link	138
Figure 4.25	Cluster and Call Number Similarity	139
Figure 4.26	Typical Data Flow for Bibliomining Applications	142
Figure 5.1	Research Contributions and Implications	159
Figure 6.1	Bibliomining Development Timeline Phases	174

## LIST OF GRAPH

<b>Graphs</b>	<b>Title</b>	<b>Page</b>
Graph 1.1	Number of Borrowing Transactions from 2006 – 2010	15
Graph 1.2	Number of Collection of Printed Titles (Cumulative)	17
Graph 1.3	Library Usage (2010)	18
Graph 4.1	Borrowing volume 2008 – 2010	107
Graph 4.2	User Borrowing (2008-2010)	108
Graph 4.3	Patron Divided by Cluster	117
Graph 4.4	Call Number by Cluster	141
Graph 4.5	Item Type Borrowed 2008-2010	143
Graph 4.6	Borrowing Location 2008 – 2010	144
Graph 4.7	Call Number Borrowed 2008 – 2010	145
Graph 4.8	Classification T Borrowing Statistics	147
Graph 4.9	Borrowing Record by L Call Number	148
Graph 4.10	Popular Subject	150
Graph 4.11	P Classification (Literature and language)	151
Graph 4.12	User Borrowing Records	153
Graph 4.13	Faculty of Education (I3) Borrowing Records	156
Graph 5.1	Total Stocks and Use Material	163
Graph 5.2	Material and Borrowed by Call Number 2008 – 2010	167

# CHAPTER ONE

## INTRODUCTION

### 1.1 INTRODUCTION

Academic libraries in Malaysia have vastly developed from traditional toward digital services in these few decades. The demands and needs of library users, local and international organizations well as the current technologies have affected the libraries in so many ways. As most libraries exist to serve the information needs of users, understanding the needs of individuals or groups is crucial with the increasing access to the Internet and the World Wide Web. As a result, digital libraries have burgeoned, thus serves a huge variety of different audiences.

The advent of low-cost data storage technologies and the wide availability of the Internet connections have made it easier for individuals and organizations to access large amounts of data. On the other hand, the accumulation of large amounts of data on the operations of the organizations and systems has become easier through the use of information technology. These data can be used to support decision making or risk management in the organizations and their systems (Tsumoto & Hong, 2011).

Data mining has emerged since 1960's. Calculation on analytical data was done manually in the form of statistic, regression, relationship and pattern in a simple and small group of data. Nowadays, the broad range of functionality on gaining new knowledge has brought this process to become more specific and detailed (Han & Kamber, 2006). The researches on mining patterns from traditional transaction databases have been developed for many years and rich kinds of patterns are discovered for different purposes (Shie, Yu & Tseng, 2012).

In complex organizations, public or private, decisions have been made on a continual short term basis (Vercellis, 2009). Furthermore, bibliometric records have been widely used in the information field and to be more precise it is a colony of bibliographic data in any database (Vercellis, 2009). Digital library for most is a warehouse of those precious data that integrate the internal or external databases. Meanwhile, the use of data mining to examine library data might be competently termed bibliomining as in the techniques of data mining for instance classification, prediction,

clustering, association, genetic algorithms and neural network help to achieve the goals of data mining which are to extract the hidden and unknown patterns from the database.

However, bibliomining in information and library areas are still need to be promoted. Small group with expertise in mathematical and computational study has scored lack of interest among this colony (Nicholson & Stanton, 2003). Huge amount of data has been stored in the library data warehouse especially in this library digital era. However, even though they have been gathered and stored in a systematic and structured way, it has to be processed by appropriate extraction tools and analytical methods capable of transforming them into information. Therefore, the relations between different category of library users and the works (books, journals, thesis, etc.) are visualized. As a result, these relations and patterns of the knowledge outcomes could help the library in their next phase of activities whilst dealing with decision making.

## **1.2 CONTEXT OF THE STUDY**

Librarians' imprecise perceptions about digital libraries need to be corrected. Rather than perceiving digital libraries as providing user friendly services, fast access to information and no boundaries data retrieval at a cheaper cost, the main development of digital libraries was in fact to preserve precious data stored in the libraries' database data warehouse (Chen and Chen, 2007). Many of these data have been unorganized and hidden in various library databases due to the lack of expertise among librarians to allocate, evaluate, extract and process these data into a new knowledge (Gold, 2007). Thus, the process of mining the data has involved other parties who did not possess any librarianship background resulted in the lack of understanding and enthusiasm in the work line which has affected the overall process and outcome.

Statistics and visual graphs for short term period of library data can be accessed by implementing a manual calculation within the OPAC (Online Public Access Catalogue). This can barely helpful in stating for incoming cost demand and user needs in a short term of the planning period, but in having a look for long term future plan it needs a more widely source of data that can support the plan.

The budgetary decision of a library is based on the percentage of purchase library items being referred by the users or high volume of usage. Libraries have to provide evidence for the top management on how well their collections are being utilized and keep track with the research, teaching and learning development.

The increase on the book title and copy of the subject will be divided by the increase of student in which faculty have the highest enrollment and also a suggestion by academician proposed by their student and research done by the circulation and acquisition librarian (see table 1.1: Resource Development Department Annual Report 2008).

Table 1.1:  
*Ordering and Receiving Statistics by Faculty Year 2008 (Resource Development Dept. Annual Report 2008)*

<b>ORDERING AND RECEIVING STATISTIC BY FACULTY FOR YEAR 2008</b>						
<b>FACULTY</b>	<b>PESANAN</b>			<b>PENERIMAAN</b>		
	<b>TITLE</b>	<b>COPY</b>	<b>TOTAL (RM)</b>	<b>TITLE</b>	<b>COPY</b>	<b>TOTAL (RM)</b>
FAB	921	969	266.551,37	987	1.096	293.686,25
FKA	518	519	226.169,67	650	660	293.949,44
FKE	948	952	423.340,84	881	937	405.717,79
FKM	956	958	459.144,24	1.007	1.033	526.475,76
FKKSA	701	602	320.246,69	626	649	333.865,66
FKSG	767	769	391.227,24	915	979	478.752,39
FP	1.962	1.978	409.115,69	2.011	2.182	444.212,65
FPPSM	1.907	2.031	435.778,71	2.016	2.265	466.530,89
FS	1.237	1.238	529.577,53	1.344	1.404	516.231,09
FSKSM	650	664	184.159,19	640	686	197.226,41
PPIPS	409	415	92.672,15	393	539	96.876,24
PSFAB	124	125	29.780,54	88	121	31.902,82
KST	1.121	3.016	495.005,36	1.458	2.738	477.191,51
BATC	508	533	170.654,93	496	515	170.650,32
IBS	280	280	77.028,36	277	306	77.893,39
SPS	13	15	5.823,76	41	42	3.604,67
FKBSK	277	277	102.847,29	261	261	96.566,96
FBB	323	323	168.823,52	351	359	180.552,82
AA	824	824	34.866,40	60	99	14.418,46
AR	72	75	15.866,17	546	565	39.297,67
PSZ	120	125	70.596,21	92	135	69.898,72
<b>TOTAL</b>	<b>14.638</b>	<b>16.688</b>	<b>4.909.275,86</b>	<b>15.140</b>	<b>17.571</b>	<b>5.215.501,91</b>

Moreover, since budgetary decisions are based on users' high volume of library items usage, it has become evident that all strategic considerations of libraries must consider or include the customers' satisfaction as well as the economical consumption of resources. Therefore, the effectiveness and efficiency of services are becoming a ubiquitous challenge in the library service planning (Decker & Hoppner, 2006).

One of the Universiti Teknologi Malaysia Library's (UTM Library) strategic plans (Figure 1.1) as well as its vision and mission statement is to provide excellent

services to the users as stated in its internal processes prospective strategic thrust whereby providing an effective decision making process and empowerment is crucial. Therefore, relevant information and knowledge are needed to facilitate effective decision making.



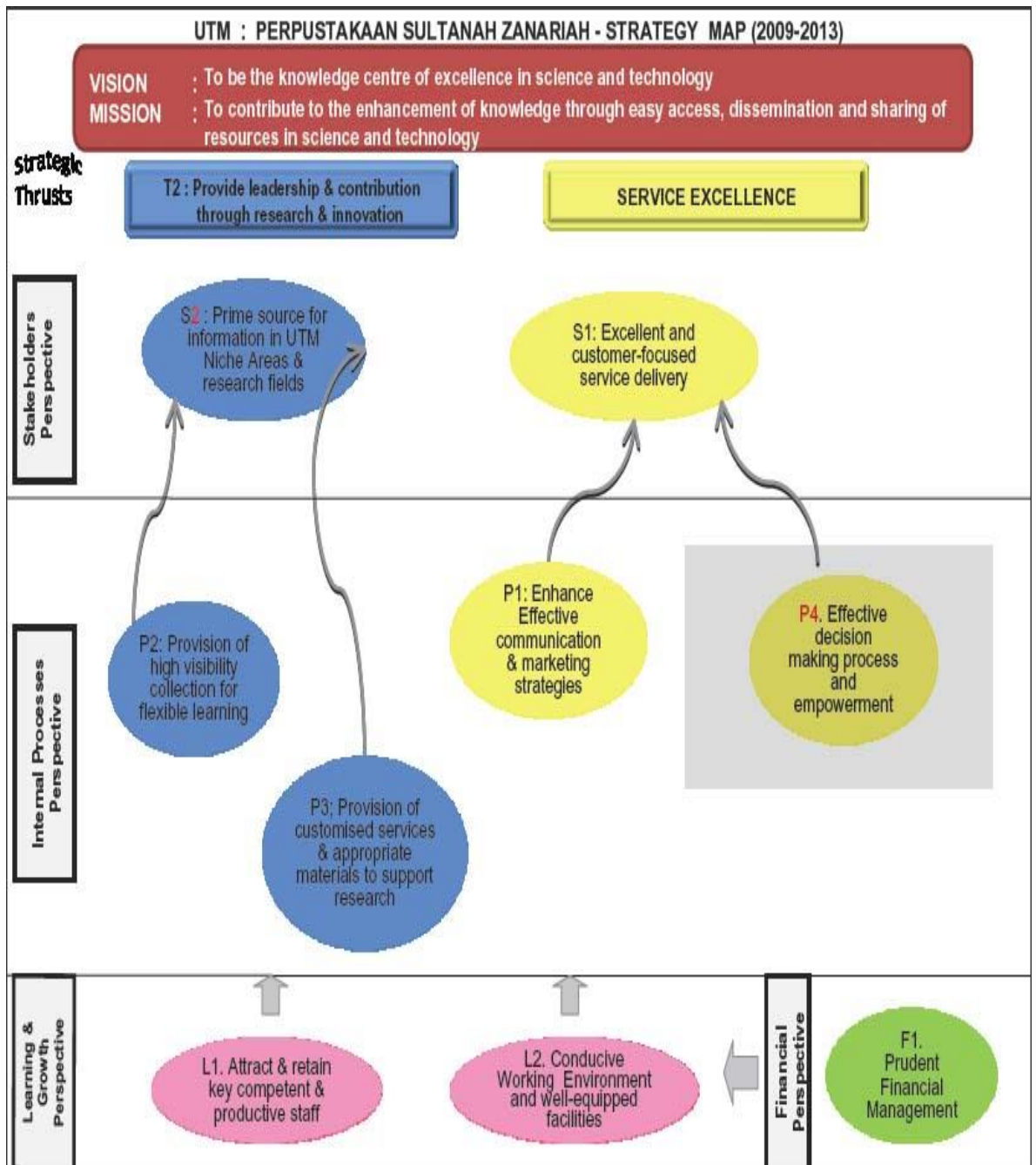


Figure 1.1: UTM Library Strategy Map (2009-2013) (<http://portal.psz.utm.my/quality>)

### **1.3 PROBLEM STATEMENTS**

The alliance of businesses and universities to create a new paradigm of tertiary education, and the emergence of virtual universities, calls into question many of our basic assumptions about the role of the academic libraries, and the sanctuary of its future. Retaining and expanding their customer base, plus focusing more commitment in meeting their customers' expectations are the only ways for academic libraries to survive in this volatile environment (Cullen, 2001). The library organisation nowadays definitely relies on bibliographic data and digital sources outcome to plan for activities and future development. However, few have made use of data mining as an advantage. (Nicholson, 2003).

#### **1.3.1 Unorganised and Hidden Data Scattered Within the Library Bibliographic Data Reservoirs**

Decision-making errors and judgement biases are becoming serious consequences in the evidence-based decision making culture. Noordin (2008) claimed that lot of valuable organisation's knowledge is usually not noticed, stored and utilized. Insights generated by library administrators can create an advantage or a missed opportunity. Therefore, this shifting requirement demands a skilled, knowledgeable library staff (Connor, E., 2008). Since the early 1970s, scholars in the areas of management information systems (MIS) and decision support systems (DSSs) have recognized the important roles of computer-based information systems which play an important role in supporting managers in their semi-structured or unstructured decision making activities (Khademolqorani & Hamadani, 2013)

#### **1.3.2 Misinformation of Decision Making**

Most libraries are performing analysis on the data captured in their database (Nicholson & Stanton, 2003). This analysis is useful because it provides the library with information on what resources are being used in the library or how these resources are being used. These types of analysis are driven by the librarians who perform the

analysis based on their hypothesis or from their experience to get the statistical information. Schachter (2006) suggested when it comes to administrative decisions; the librarians often approach decision-making in a similar unstructured way. Hence, libraries need to be independent so as to broaden their sources in order to establish a concrete strategic plan (Schachter, 2006). Therefore, the purpose of this research is to discover a method to develop communities which are based on issue pattern and search for applications which the library can apply as tools to identify the behaviour and interests of users in their relation with a particular type of books or reference materials. Next, the relationship is visualised through a graph to demonstrate the extracted information more vividly and effectively (Chen, Lin & Wu, 2004).

Productivity is defined as the use of various inputs such as information or processes to generate the required products and services which can be applied as a measure of progress or efficiency. Hence, the challenge faced by the library administrators is to maintain a positive productivity trend. In order to ensure that productivity continues to increase, available information needs to be accessed and utilised effectively (Connor, 2008).

### **1.3.3 Lack of Expertise, Experience and Interest in Data Mining**

In the past, the librarians had to depend upon surveys to gather user information. However, via data mining, similar patterns are able to be discovered without wasting the user's time or the taxpayer's money on surveys (Estabrook, 1996). From a survey done by Mento & Rapple (2003), majority of the librarians (90%) agreed that although data mining is a valuable tool which could facilitate libraries in the future and has actually been integrated into the curriculum at many academic institutions, yet it was still being wasted by the organisation (Mento & Rapple, 2003).

## **1.4 RESEARCH QUESTIONS AND OBJECTIVES**

Libraries have to be more proactive in providing the information and services required by their patrons in order to be one of the main sources of information in today's Information Technology world. Libraries today are required to provide useful and reliable information to their patrons. To ensure their patrons do not go elsewhere to get

their information. The library should be able to proactively define the needs of their patrons, and therefore provide in advance the information or services required by their patrons.

Libraries should venture into data mining where analysis on the data captured can provide predictive information based on the patterns generated. This information can help the libraries study the trend of the resources their patron may want to use in the future or the period of time certain resources are in demand. The knowledge captured by the data mining activity is data-driven and covers areas that may not have been thought of. The reasons why the libraries should use data mining are (Nicholson and Stanton, 2003):

- to improve or increase the services demanded or required by the patrons,
  - to get information on appropriate books to purchase for the use of their patrons,
  - to justify the library purchases of resources to the top management or the funding bodies,
- and
- for allocation of resources required in future which is difficult to predict.

The primary goal of carrying out this study is to explore the type of data warehouse which would allow for exploration of patterns of connections between the authors, works, libraries and users (Nicholson, 2005).

Hence, this study is focusing on two main questions:

### **Research Questions:**

1. What are the anticipated emerging pattern and trend constructed from statistical analysis of the data from the library data reservoirs related to the user/patron and circulation information?
2. How data mining could be applied in a library as a knowledge discovery tool?

One way to understand these issues and harness the large amounts of raw data created through digital library use is through bibliomining, or the application of statistical and pattern recognition tools to the data associated with the library systems (Nicholson, 2003).

### **Research Objectives:**

1. To determine the related transaction data for data warehouse development.
  - a. User/patron information
  - b. Circulation information
2. To explore the patterns and trends among these data reservoirs using bibliomining analysis.
3. To construct a bibliomining framework for library as a knowledge discovery tool

### **1.5 BACKGROUND INFORMATION OF THE ORGANIZATION WHERE THE STUDY WAS CONDUCTED**

Perpustakaan Sultanah Zanariah (PSZ) occupies a central location at the Universiti Teknologi Malaysia (UTM) main campus in Skudai. It has a branch at the UTM City Campus, Kuala Lumpur and also branches at several faculties, learning centers and Centers of Excellence. PSZ was officiated by Her Majesty Sultanah Zanariah, the Chancellor of University Teknologi Malaysia on 3rd February 1991. The library is a four storey building with a seating capacity of 3,422 and a collection of nearly half a million volumes and with a total of 179 staff.

As an integral component of the academic programme, PSZ supports the university's teaching, learning, research, consultancy and publication activities. Its services and collection development activities are geared towards fulfilling the need for library materials and information in the university's core area of Science and Technology. Nevertheless, PSZ also has a good Humanities and Social Science collection to support courses in these areas which are offered by several faculties (PSZ Annual Report 2010).

UTM Library has infinitely developed from traditional toward digital services since the year 2000. The demand from the users, institutional, international organization and the fast development of technology have affected tremendously on the Library either on its budget cost or decision making mode. Libraries definitely rely on bibliographic data and digital sources outcome to plan for activities and foreseeing future development. However, few have used these as advantages. Thus, this research

scope area will focus on the Universiti Teknologi Malaysia Library (UTM Library) as it has nearly 1 million records of collection in different formats.

The process of library automation at PSZ started in 1986 with only an in house library system called BERLIAN. It cater only two library work process module such as Library Catalogue and OPAC. Today most of the library's operations and services are computerized and integrated. All processes including materials acquisition, indexing, circulation and information searching are conducted through the Computerised Library System known as SirsiDynix (Workflows 3.3.1J). Information on all materials available at PSZ can be accessed through **LESTARI** (Library Electronic SysTem And Research Information).

PSZ Digital Library project was established since year 2001 which cost approximately RM2.8 million. The digital library project that illustrated in figure 1.2 consist with three (3) main module that is;

- i.** Advance Integrated Library System
- ii.** Cyberdocs & Fulcrum Knowledge Server (Document Management System)  
and,
- iii.** Library Applications System

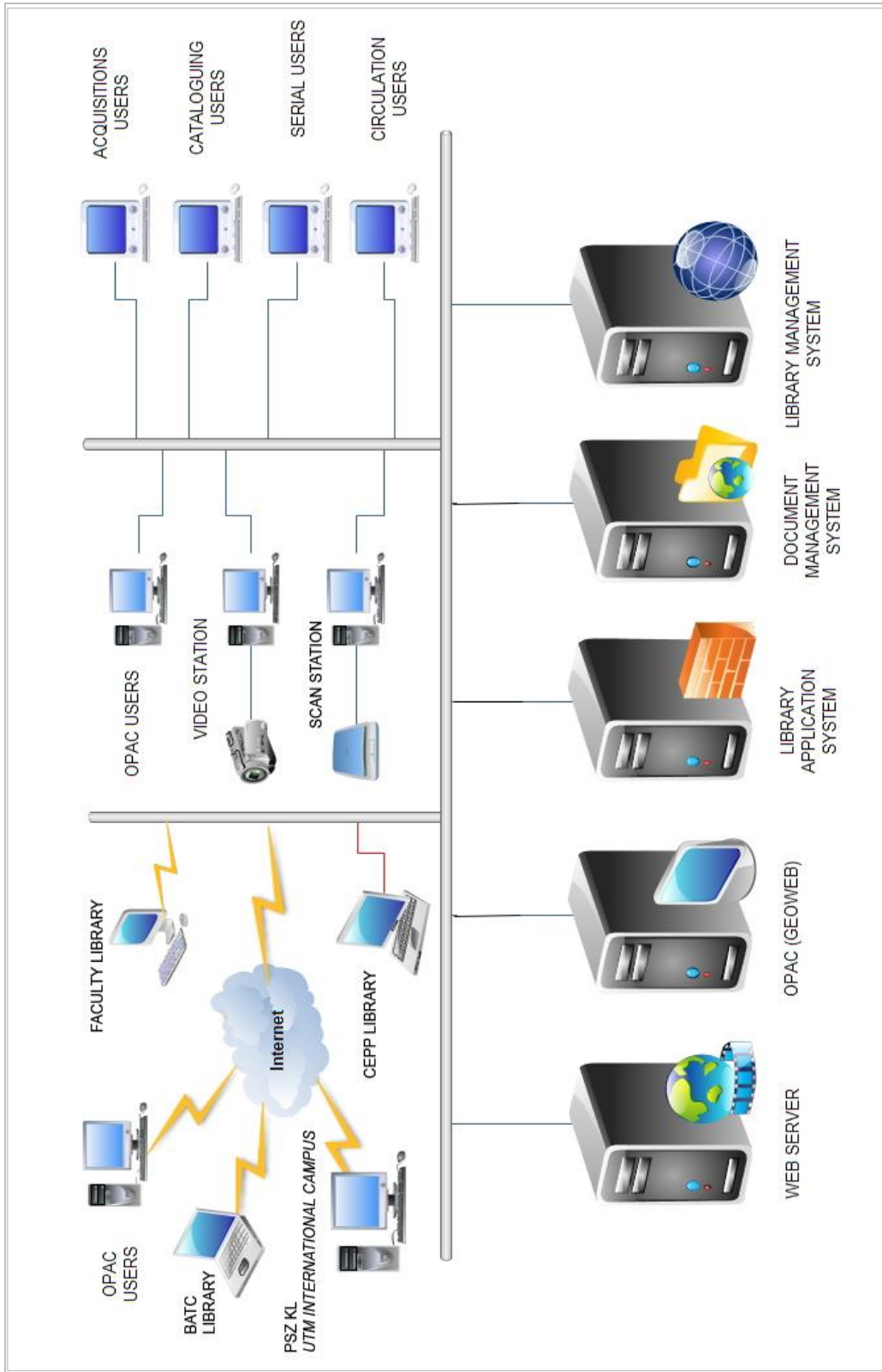


Figure 1.2: UTM Digital Library Systems

Integration between library systems that illustrate from figure 1.3 describes the relations with the user (staff) and University student information system (patron) that was also established by this project. The main task in doing this digital library project is the data migration regarding 300,000 library bibliographic records, 600,000 of works (copy), 2,000 journal titles and 20,000 patron records. All of the library workflow was re-design and plan with a time phase project. The library module was recognised and the library applications and the digitize document was separated to ease in doing integration and data bibliographic record (Automation Development Department 2010).



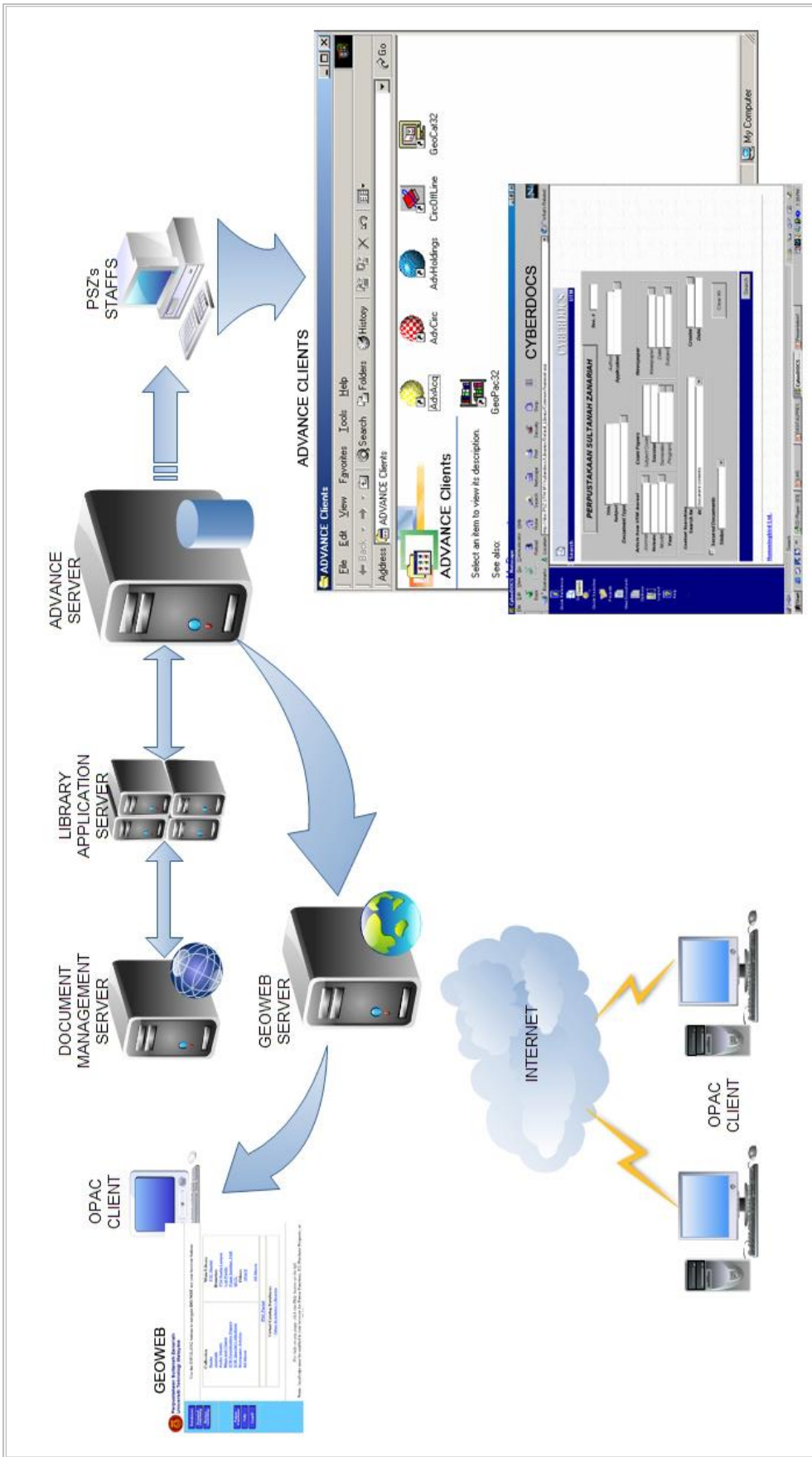
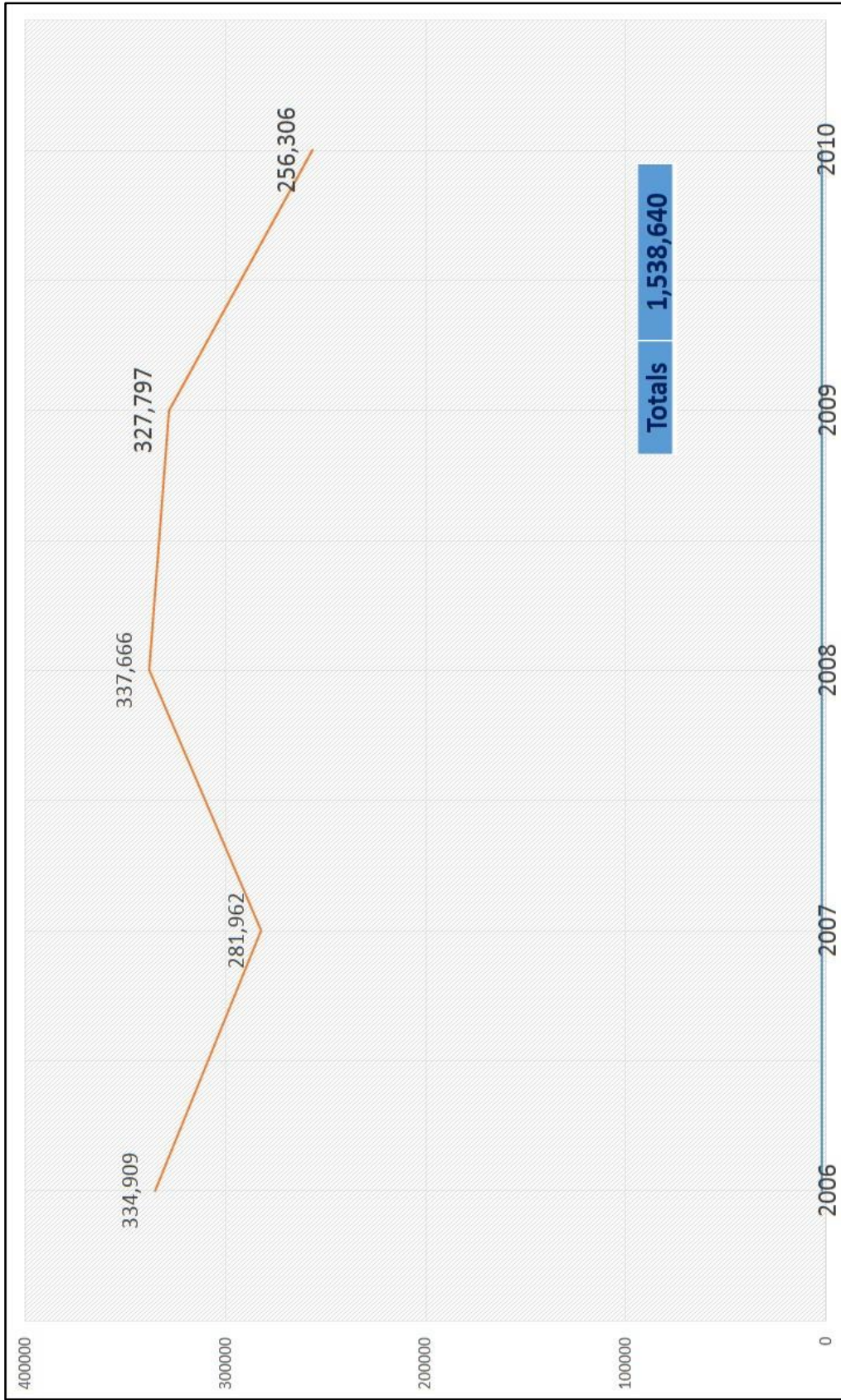


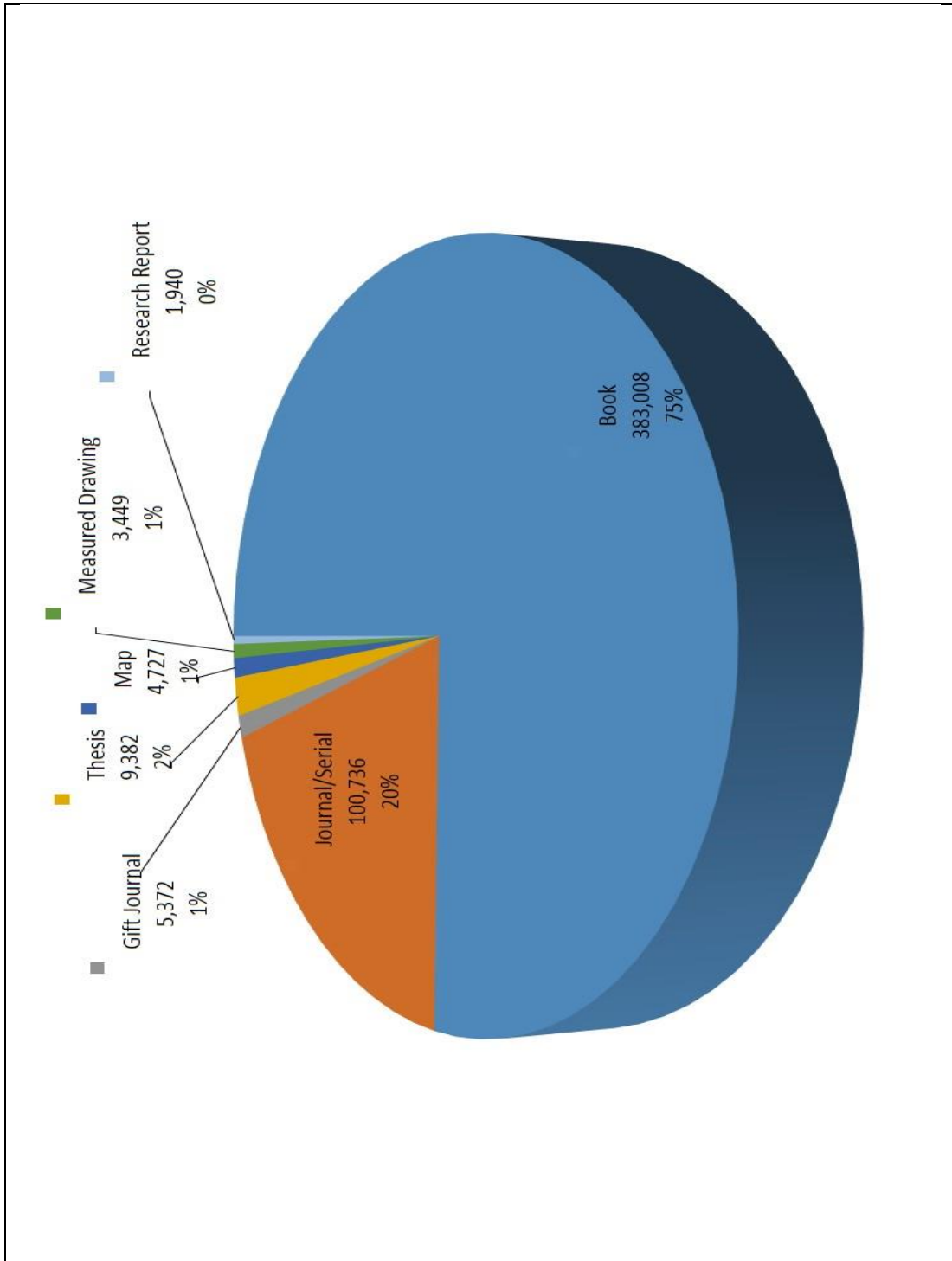
Figure 1.3: UTM Digital Library System: Integration Overview

The huge Library bibliographic data could be statistically patterned and clustered to indicate relationship among them. Bibliomining which is equipped with Circulation history of UTM library relating to the patrons and works (authors) has achieved more than 1.5 million transactions from 2006 till 2010 (Graph 1.1). The Library database stores all the information about users and the books loaned transactions, all transactions between the library user and their works and the different material types of books in the Library.



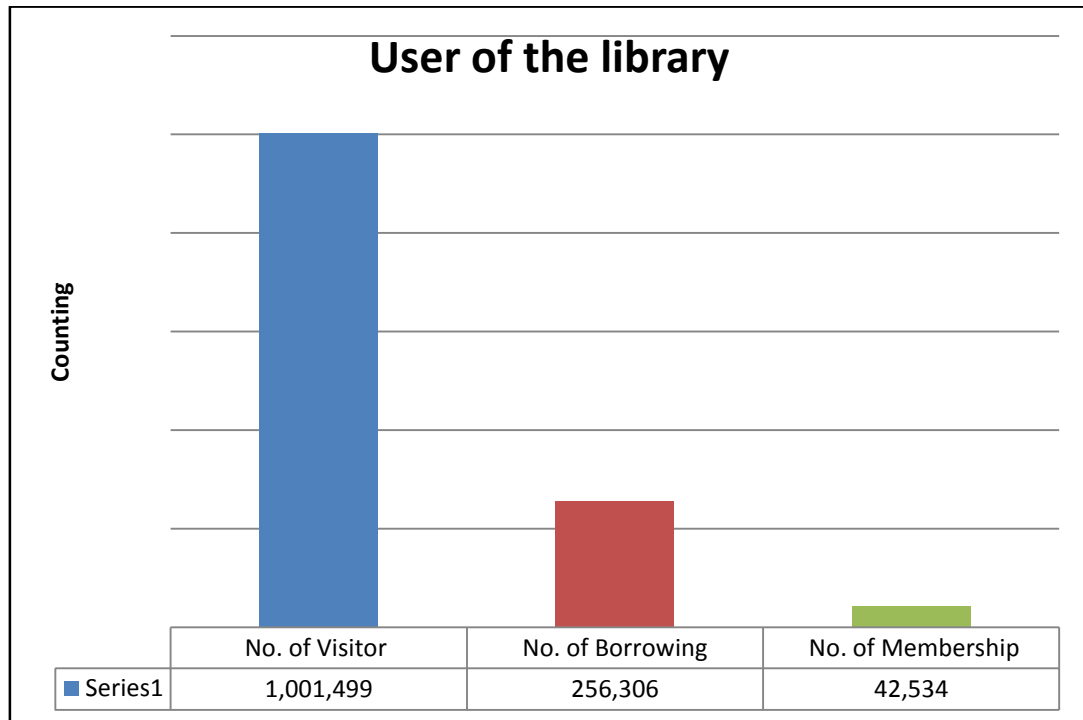
Graph 1.1: Number of Borrowing Transactions from 2006 – 2010

Determined from graph 1.2, books are the highest materials being loan by the library user and followed by journals and theses. These transaction data log history is range from 2006 till 2010 and it overall circulation transaction was more than 500,000 of works (Library Automation Department Report, 2010). This transaction is due to all borrowing process history by the user but this statistic does not reflect the total of UTM library collections as a whole.



Graph 1.2: Number of Collection of Printed Titles (Cumulative)

From graph 1.3, it is clear that more than a million users has visited the library and only a quarter of visitors borrow material from the library. It seems that the library is a conducive place to study and to do research within database via online.



Graph 1.3: Library Usage (2010)

The aim of this study is to discover useful information or rules from numerous data and utilise them in decision making matters (Linoff & Berry, 2011). In addition, the technique can also be applied on library resource service in order to understand users' borrowing behaviours based on the library borrowing database so as to assist library managers in their decision making process (Chen, Lin and Wu, 2004).

## **1.6 THEORETICAL FRAMEWORK**

Figure 1.4 shows bibliographic records which are collected from the Library data source. The selected data was carefully singled out in order to predict on how these variables could effect the decision making process of the Organization, which was considered by reviewing related research by Scott Nicholson on library decision making. The bibliographic record subsequently analysed to produce statistical data that depicts the patterns and trends of the library data.

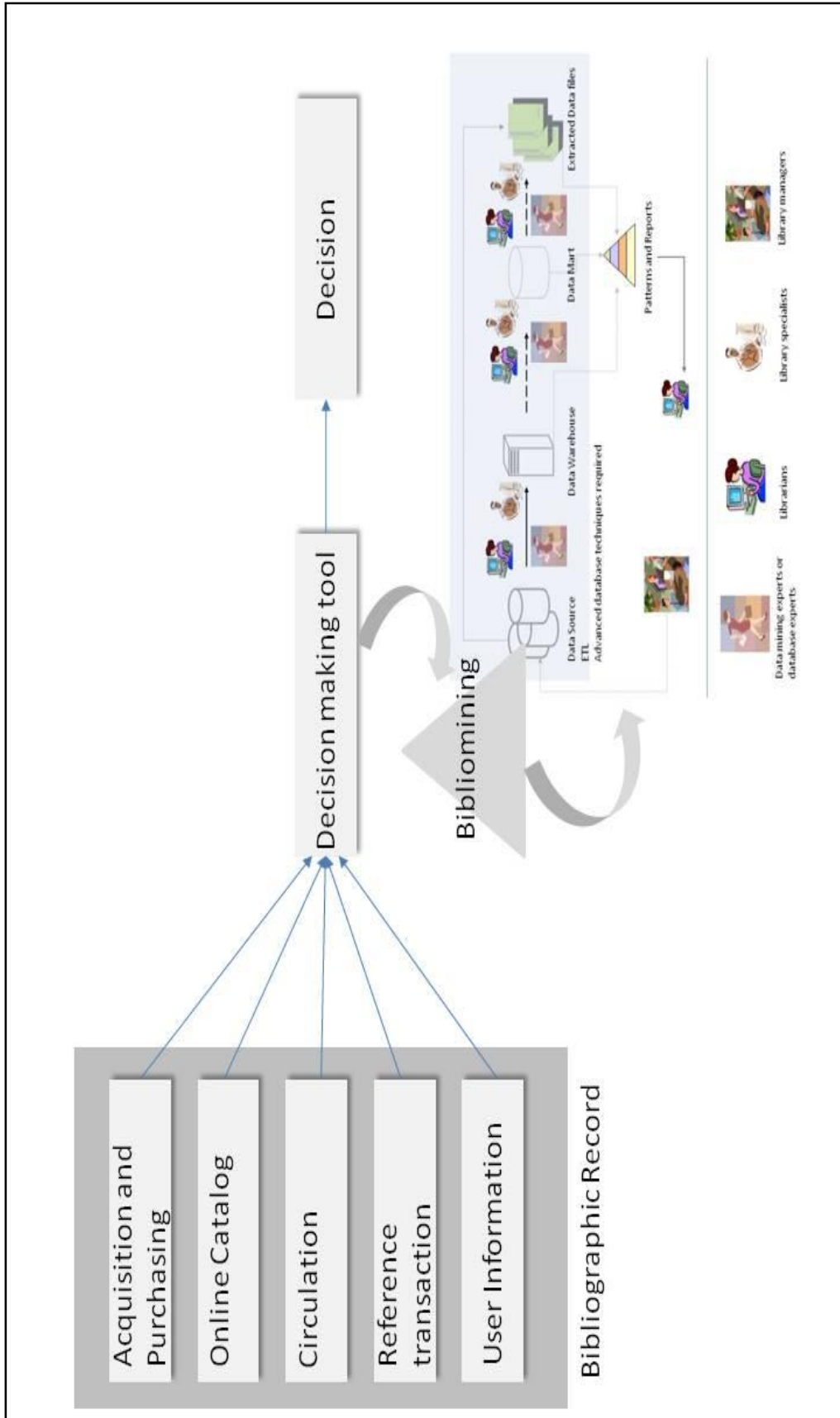


Figure 1.4 : Theoretical Framework (Han & Kamber, 2006)



Many researches indicated that demographic variables play the most important role in the circulation allocation operation in performing bibliographic bibliomining. These include (1) the size of faculty, (2) the size of library users (3) works or author (4) the adequacy of the library collection in an academic discipline (5) the size of type of courses (6) the amount of online catalogue (7) the past record in use of allocated funds, and (8) the circulation statistics. Many researchers also have employed demographic variables in their research hence to the enhancement of library management (Kao, Chang, Lin, 2003; Ichise, Takeda, and Ueyama, 2005; Lavoie, Dempsey & Lynn, 2006; Goswami, Verma & Krishnan, 2010; Nicholson, 2005).

In addition, the framework was derived from a book by Han & Kamber, 2006, which stated that data mining could easily be done by extracting data from a single data warehouse which could be constructed via a process of data cleaning, data integration, data transformation, data loading, and periodic data refreshing. Bibliographic records and other sources contain a wealth of information which is able to support the creation and maintenance of such views (Lavoie, Dempsey & Lynn, 2006).

## **1.7 SIGNIFICANCE OF THE STUDY**

For decades, the corporate sectors have exploited technological advances to better market and deliver products and services to customers via the techniques of data mining and yet the technique was not widely used in libraries. However, with the current emphasis on evidence-based decision making, libraries are beginning to utilize their system and the user generated data. Data mining usually involves a significant endeavor to extract embedded and potentially useful information from large undiscovered data sets (Connaway & Dickey, 2008).

Most of academic libraries have made huge investments in creating and maintaining rich, structured information which describe the resources in their collections by effectively changing to digital integrated library systems. Library bibliographic bibliomining fields are tools which are capable of visualise how libraries manage their cost, staff activity, customer service, user needs, marketing, popular books, circulation, reference transaction, quality of collection, educational programmes etc. It also represents a potential value in terms of knowing more about the characteristics of library collections; generating interesting and innovative data displays; providing intelligence to support a range of library decision-making needs,

including collection development, digitization and preservation activities (Lavoie, Dempsey& Lynn, 2006).

Designing new data analysis model using data mining process for UTM Library is seen as easing continuing standard in order to extract data in the future plan. Selected fields and attributes will be collected and placed in a data warehouse for pre-processed and could be analysed for other needs for future bibliomining process.

## **1.8 CONCLUSION**

Library bibliomining is about identifying topics, creating a data warehouse, refining and exploring data, plus evaluating results. Nevertheless, its main purpose is to operate as a catalyst for library managers in making precise and supportable decisions (Wu, 2003; Shieh, 2009). Digital library has stimulate users and organizations to perceive in a way that digital library is not just an ordinary physical library but it is equivalent to the World Wide Web (www) with the needs to improve areas such as performance, organization, functionality, and usability (Xie, 2008).

Therefore, organizations which embraced digital libraries must seized the opportunity in implementing bibliographic analysis using bibliomining. Moreover, bibliomining assists academic libraries in matters pertaining to cost efficiency, increasing user satisfaction and gaining more leverage in the decision making process.

As this chapter highlighted the nature of this research study, chapter two (2) discusses the theoretical and conceptual framework of the study.

## REFERENCES

- Abdul Hakim Mohammed, Maimunah Sapri & Maizan Baba. (2006). *Pengurusan Fasilitas*. Skudai: Penerbit Universiti Teknologi Malaysia.
- Agrawal, R., Imieli, T. & Swami, A. (1993). Mining association rules between sets of items in large databases. *SIGMOD Rec.*, 22(2), 207-216.
- Azam, I., Sohrawardi, S.J., Das, H.S., Alam, M.S., Alvy, M.S. & Rahman, R.M. (2013). Bibliomining on North South University library data. *Digital Information management (ICDIM)*, pp.235-240.
- Baker, T. L. (1999). *Doing Social Research*. McGraw-Hill.
- Banejee, K. (1998). Is data mining right for your library? *Computers in Libraries*, vol. 18, p.28-31.
- Bayley, L. (2009). Practicing what we preach: A case study on the application of evidence based practice to inform decision making for public services staffing in an academic health sciences library. *The New Review of Academic Librarianship*, vol. 15(2), p.235-252.
- Berry, M. J. A. & Linoff, G. (2000). *Mastering data mining: the art and science of customer relationship management*. New York: Wiley Computer Pub.
- Booth, A. & Brice, A. (2004). *Evidence-based Practice for Information Professionals: A Handbook*. London: Facet Publishing.
- Booth, A. & Brice, A. (2003), "Clear-cut? Facilitating health librarians to use information research in practice", *Health Libraries Review*, vol. 20 no.1, pp.45-52.
- Booth, A. (2006). Counting what counts: performance measurement and evidence-based Practice. *Performance Measurement and Metrics*, vol. 7, iss. 2, p.63 – 74.
- Broadus, R. N. (1987). Toward a definition of "bibliometrics". *Scientometrics*, vol. 12, no 5-6, p. 373-379.
- Buckland, M. (2003). Five grand challenges for library research, *Library Trends* 51(4).
- Casey, A.M. (2011). *Strategic Priorities and Change in Academic Libraries*. Simmons College Graduate School of Library and Information Science (Doctoral dissertation). Central Michigan University, Michigan.

- Chai, C. L. (2006). *Finding kernel function for stock market prediction with support Vector Regression* (Master's thesis). Universiti Teknologi Malaysia, Johor.
- Chang, C.C. & Chen, R.S. (2006). Using data mining technology to solve classification problems: a case study of campus digital library. *The Electronic Library*, vol. 24, no. 3, pp. 307-321
- Chen, C.C. & Chen, A.P. (2007). Using data mining technology to provide a recommendation service in the digital library. *The Electronic Library*, vol. 25, no. 6, p. 711-24.
- Chen, M. S., Han, J. W. & Yu, P. S. (1996). Data mining: An overview from a database perspective. *Ieee Transactions on Knowledge and Data Engineering*, 8(6), 866-883.
- Chen, T. S., Lin, M. H. & Wu, C. H., (2004). Enhancing library resources usage efficiency by data mining. *IEEE International Conference on Networking, Sensing & Control Taipei*, p.21-23.
- Chen, S.Y., Macredie, R.D. & Liu, X. (2010). Editorial: data mining for understanding user needs. *ACM Transactions on Computer-Human Interaction*, vol. 17, no. 1, article 1.
- Clyde, L.A. (2006). The basis for evidence-based practice: Evaluating the research evidence. *New Library World*, vol.107 (1224/1225), p. 180-192.
- Cotter, L. & Spencer, A. (2007). Evidence-based librarianship – a measure approach to library success. *Elsevier Library Connect Seminar*, Bangkok and Kuala Lumpur, 1-2 August 2007.
- Courtney, H., Kirkland, J. & Viguier, P. (1997).Strategy under uncertainty. *Harvard Business Review*, 75(6), 66-79.
- Connaway, L. S. & Dickey, T. J. (2008) Transforming data into services: delivering the next generation of user-oriented collections and services. Proceedings of the American Society for Information Science and Technology (1550-8390), vol. 45, iss.1; p.1-5.
- Connor, E. (2008). An Introduction to Staff Development in Academic Libraries (Eds.), *Using Data Mining and Analysis Techniques to Understand Academic Librarian Training Requirements* (pp. 187-200). New York, NY: Rout ledge.
- Cook, C. (2002).The maturation of assessment in academic libraries: the role of LibQUAL+™. *Performance Measurement and Metrics*, vol. 3, iss. 2.

- Cullen, R. (2001). Perspectives on user satisfaction surveys. *Library Trends*, vol. 49, p.662- 686.
- Cunningham, S. J., Frank, E., Gedeon, T., Wong, P., Halgamuge, S., Kasabov, N., et al. (1999). *Market basket analysis of library circulation data*. Paper presented at the ICONIP'99. ANZIIS'99 & ANNES'99 & ACNN'99. 6th International Conference on Neural Information Processing. Proceedings (Cat. No.99EX378), Place of Publication: Piscataway, NJ, USA; Perth, WA, Australia. Country of Publication: USA.
- Diaz-Valenzuela, I., Martin-Bautista, M.J., Maria-Amparo, V. & Campaña, J.R. (2013). An automatic system for identifying authorities in digital libraries. *Expert Systems with Applications*, v. 40, iss. 10, p. 3994-4002.
- Dillon, A. (2008). *Accelerating learning and discovery: Refining the role of academic librarians*. Paper presented at the No Brief Candle: Reconceiving Research Libraries for the 21st Century meeting of the Council on Library and Information Resources, Washington, DC.
- Dunham, M. H. (2003). *Data mining: introductory and advanced topics*. Upper Saddle River, N.J.; London: Prentice Hall/Pearson Education.
- Eldredge, J. (2006). Evidence-based librarianship: the EBL process. *Library Hi Tech*, vol. 24, iss: 3, p.341-354.
- Eldredge, J. D. (2004).Inventory of Research Methods for Librarianship and Informatics. *Journal of the Medical Library Association*, vol.92, p. 83–90.
- Estabrook, L. (1996). Sacred trust or competitive opportunity: Using patron records. *Library Journal*, vol. 121(2), p. 48-49.
- Decker R. & Hoppner M. (2006) Strategic planning and customer intelligence in academic libraries. *Library Hi Tech*, vol. 24, no. 4, p. 504-514.
- Fayyad, U. et al. (1996). From Data Mining to Knowledge Discovery in Database. *AI Magazine*, vol. 17, no 3, p. 37-54.
- Feng, S., Zhou, S. & Liu, Y. (2011). Research on data mining in university admissions decision-making. *International Journal of Advancements in Computing Technology*, vol. 3, no. 6.
- Garfield, E., Malin, M. V. & Small, H. (1983). Toward a Metric of Science: The Advent of Science Indicators: citation Data as Science Indicators, in: Elkana, Y., et al. John Wiley & Sons, NY, 1978 (Eds), p. 180. *Essays of an Information Scientist*, vol. 6, p. 580.

- Gold, A. (2007). Libraries and the data challenge: roles and actions for libraries. Cyber infrastructure, Data, and Libraries, Part 2. *D-Lib Magazine*, vol. 13, no. 9-10.
- Goswami, S., Verma S. & Krishnan N. R. (2010). Relationship visualization between books and users based on mining library circulation data. *Intelligent Interactive Technologies and Multimedia '10*.
- Guenther, K. (2000). Applying data mining principles to library data collection. *Computers in Libraries*, vol. 20, no. 4, p. 60-63.
- Guy D.T. & Wouter V.A. (2012). Spaces of information modelling, action, and decision making. *Library Trends*, 61.2, p.304-324.
- Han, J. & Kamber, M. (2006). *Data mining: concepts and techniques*. 2<sup>nd</sup> ed. California: Elsevier pub.
- Han, J., Kamber, M. & Pei, J. (2012). *Data mining: concepts and techniques* (3rd ed.). Amsterdam; Boston: Elsevier/Morgan Kaufmann
- Hajek, P. E. T. R. & Stejskal, J. (2012). Analysis of user behavior in a public library using bibliomining. *Advances in Environment, Computational Chemistry and Bioscience*. p. 339-344.
- Hernon, P. & McClure, C. R. (1990). *Evaluation and library decision making*. New Jersey: Ablex Pub. Co.
- Huang, H., Xiao, J., Yang, Q., Qing Wang, Q. & Wu, H. (2013). Creating process-agents incrementally by mining process asset library. *Information Sciences*, v. 233, p.183-199.
- Ichise, R., Takeda, H. & Ueyama K. (2005). *Community mining tool using bibliography data*. In Proceedings of the 9th International Conference on Information Visualization, p. 953-958.
- Inmon W.H. (2005). *Building the data warehouse*. 4<sup>th</sup> ed. Indianapolis, IN: Wiley Publishing Inc.
- Kamariah Nor Mohd Desa & Norliya Ahmad Kassim. (2011). *Leadership skill in relation to library leader's key performance index*. 1st Malaysia PERPUN International Conference and Workshop on key performance indicators for Libraries, 2011.
- Kao, S.C., Chang H.C. & Lin, C.H. (2003). Decision support for the academic library acquisition budget allocation via circulation database mining. *Information Processing and Management*, vol. 39, p. 133-147.

- Kelly, S. (2006). *Customer intelligence - from data to dialogue*, Wiley, Chichester.
- Khademolqorani, S. & Hamadani, A. Z. (2013). An adjusted decision support system through data mining and multiple criteria decision making. *Procedia - Social and behavioral Sciences*, vol. 73, p. 388-395,
- Kimball, R. & Caserta, J. (2004). *The Data Warehouse ETL Toolkit – Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*, Wiley Publishing, Inc. USA.
- Kimball, R. (1996). *The data warehouse toolkit*. Second edition. Canada: John Wiley.
- Koufogiannakis, D. & Crumley E. (2002). Evidence based librarianship. *Feliciter*, vol. 48(3), p. 112-114.
- Lavoie, B., Dempsey, L.C. & Lynn, S. (2006). Making data work harder. *Library Journal*, vol. 131, iss. 1.
- Liao, S.H, Chu, P.H. & Hsiao, P.Y. (2012). Data mining techniques and applications – a decade review from 2000 to 2011. *Expert Systems with Applications*, v. 39, iss. 12, p. 11303-11311.
- Library of Congress (2003). Understanding marc bibliographic: machine readable cataloguing. 7<sup>th</sup> ed. Website: <http://www.loc.gov/marc/umb/>
- Linoff, G. S. & Berry, M. J.A. (2011). *Data mining techniques: for marketing, sales and customer relationship management*. Indiana, Indianapolis: Wiley Publishing Inc.
- Mary, M. S. & Navjit, B. (2009). A user-centered and evidence-based approach for digital library projects. *Electronic Library*, The, vol. 27, Iss: 3, p.409 – 425.
- Mento, B. & Rapple, B. (2003). *Data mining and data warehousing*. Washington, D.C. Association of Research Libraries, Office of Leadership and Management Services.
- Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks: Sage.
- Miller, B. W. (1990). The conceptual framework as reformation and Counter Reformation. *Accounting Horizons*, Vol. 4, No. 2: 23-32.
- Mishra, A. (2011). *A spatio-temporal simulation framework for group behavior analysis for mobile phone users* (Master Thesis). Kharagpur: Indian Institute of Technology.

- Molz, Redmond K. & Dain, P. (1999). *Civic space/cyberspace: the American public library in the information age*. Cambridge, Mass.: MIT Press.
- Norashikinbinti Ahmad (2003). *Pengkelasan Data Kajicuaca Di Dalam Peramalan Hujan Dari Pelbagai sumber Menggunakan Rangkaian Neural* (Master Thesis). Malaysia: Universiti Teknologi.
- Nicholson, S. & Stanton, J. (2003). Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries. In Nemati, H. & Barko, C. (Eds.). *Organizational data mining: Leveraging enterprise data resources for optimal performance*. Hershey, PA: Idea Group Publishing, p. 247-262.
- Nicholson, S. (2003). The bibliomining process: data warehousing and data mining for library decision making. *Information Technology and Libraries*, vol. 22, p. 4.
- Nicholson, S. (2005). The basis for bibliomining: Frameworks for bringing together usage- based data mining and bibliometrics through data warehousing in digital library services. *Information Processing and Management*, vol. 42, p. 785–804.
- Nicholson, S. (2006). Approaching librarianship from the data: using bibliomining for evidence-based librarianship. *Library Hi-Tech* 24(3): p. 1-2.
- Niyati, A., Amit, K., Harsh, K. & Vaishali, A. (2012) Analysis the effect of data mining techniques on database. *Advances in Engineering Software*. v. 47, iss. 1, p. 164-169.
- Noordin, S. A. (2008). *Knowledge management Competencies: implications for training*. (Thesis, DBA). University of Northumbria, UK.
- Ohio Catalog Library Center (1993). OCLC-MARC Records: 1993 November-Present *Files*. Retrieved October 03, 2011, from <http://www.oclc.org/support/documentation/worldcat/records/subscription/default.htm>
- Pentland, A. (2008). *Honest signals: how they shape our world*. MIT Press, Cambridge, MA.
- Ponniah, R. (2001). *Date warehousing fundamentals, a comprehensive guide from IT professionals*. John Wiley & Sons, Inc., New York.
- Prakash K., Chand P. & Gohel, U. (2004). Application of data mining in library and information services. INFLIBNET, 2nd Convention PLANNER - 2004, Manipur Uni., Imphal.



- Rakesh, A., Johann, C. F. & Raghu, R. (2004). Data mining: the next generation. *Dagstuhl Workshop*.
- Roiger, R. & Geatz, M. (2003). *Data mining: a tutorial-based primer*. Boston, MA: London: Addison Wesley.
- Schachter, D. (2006). The importance of good decision making. *Information Outlook*, vol.10 (4), p. 12-13.
- Schelsky, H. (1967). Grundzüge einer neuen Universität. Eine Denkschrift“, in Mikat, P. and Schelsky, H. (Eds.), *Grundzüge einer neuen Universität. Zur Planung einer Hochschulgründung in Ostwestfalen*, Bertelsmann, Gütersloh, p. 35-70.
- Shie, B.E., Yu, P.S. & Tseng, V.E. (2012). Efficient algorithms for mining maximal high utility item sets from data streams with different models. *Expert Systems with Applications*, v. 39, iss. 17, p. 12947-12960.
- Shieh, J.C. (2009). The integration system for librarians' bibliomining. *The Electronic Library*, vol.28, no. 5, pp. 709-721.
- Tan, P.N., Steinbach, M. & Kumar, V. (2005). *Introduction to data mining, (First Edition)*: Addison-Wesley Longman Publishing Co., Inc.
- Tsai, H. (2013). Knowledge management vs. data mining: Research trend, forecast and citation approach. *Expert Systems with Applications*, v. 40, iss. 8, p. 3160-3173.
- Tsumoto S. & Hong, T. (2011). Special issue on data mining for decision making and risk management. *J Intell. Inf. Syst.*, vol. 36, p. 249–251.
- Young, H. (Ed.) (1983). *The ALA Glossary of Library and Information Science*, Chicago, American Library Association, 1983, p. 22.
- Vercellis, C. (2009). *Business intelligence: data mining and optimization for decision making*. West Sussex, U.K.: John Wiley & Sons Ltd.
- Williams, B.J. & Cole, B. (2013). Mining monitored data for decision-making with a bayesian network model. *Ecological Modelling*, v. 249, p. 26-36.
- Wu, C.H. (2003). Data mining applied to material acquisition budget allocation for libraries: design and development. *Expert Systems with Applications*, vol. 25, p. 401–411.
- Xie, H. I. (2008). Users' evaluation of digital libraries (DLs): Their uses, their criteria, and their assessment. *Information Processing and Management*, vol. 44, p. 1346–1373.

Yin, R. K. (1989). *Case Study Research: Design and Methods*: rev. ed., SAGE Publications, Newbury Park, CA.