ABSTRACT

The World Wide Web (WWW) is gaining popularity by providing an alternative way for individuals to get the information they require easily and quickly through the Internet. This is threatening the existence of libraries, which have been one of the main resources for information. The librarians in return are proactively planning and improving their services to meet the challenges imposed by the World Wide Web. Libraries are using the technologies of the Internet to provide new and better services to their patrons. Although the librarians are being proactive, they require information to make strategic decisions that would improve their services for their patrons. One way of getting this information is to perform data mining on the data captured in the libraries' systems. Data mining or also known as knowledge discovery from databases is a process of getting useful information or knowledge by analysing the data for patterns. The capability of data mining to solve some of the organisation's persistent and pressing issues has encouraged new information technology users to implement this system. The Extraction, Transformation and Loading (ETL) processes are required to extract, transform and load the data into a database for the data mining activities. In this dissertation, firstly, a survey was conducted on the use of data mining in the libraries of Higher Education Institutes and Universities in Malaysia. Secondly, a data extraction and data transformation application was developed for extraction, cleaning and transforming data for data mining. Finally, a prototype data mining system was implemented for the library integrating the developed data extraction and data transformation application with other suitable tools and methodology.

Acknowledgement

Firstly, I would like to thank my supervisor, Dr. Teh Ying Wah, who has given valuable direction and guidance towards completing this dissertation.

I would also like to thank the survey participants, who consist of the Higher Education Institutes' libraries and Universities' libraries in Malaysia, for their kind participation in the survey research.

I am also grateful to the Faculty of Computer Science and Information Technology staff for their support at the various level of my dissertation, especially Ms Papu and Mr. Sim.

I would also like to thank my sister, Vijaya L. Raj for her guidance and support when the going was tough.

Last but not the least; I would also like to dedicate my gratitude to my family, Mrs. R. Jayanthi, my wife, who had always been supportive, encouraging and understanding during the time I spend on this dissertation and my sons Rakesh and Shatesh for just being there.

Rabindranath Raj

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ARL	Association of Research Libraries
BI	Business Intelligence
CRISP-DM	CRoss-Industry Standard Process for Data Mining
DAMIM	Data Mining Methodology
ETL	Extraction, Transformation and Loading
FTP	File Transfer Protocol
GRI	Generalised Rule Induction
ILS	Integrated Library System
KDD	Knowledge Discovery in Databases
OLAP	Online Analytical Processing
SEMMA	Sample, Explore, Modify, Model, Assess
SPSS	Statistical Package for the Social Sciences
WWW	World Wide Web