Table of Contents

Abstract...........................................................................................................................................i
Acknowledgement..........................................................................................................................iii
Table of Contents............................................................................................................................1
List of Figures......................................................................................................................................5
List of Tables......................................................................................................................................7

Chapter 1: Introduction .................................................................................................................. 8
  1.1. Definition of Thesis and Dissertations..................................................................................... 8
  1.2. Thesis versus Dissertation......................................................................................................... 9
  1.3. What is an E-TD?.................................................................................................................... 9
  1.4. E-TD Repository..................................................................................................................... 9
  1.5. Background Study.................................................................................................................. 10
  1.6. Project Overview.................................................................................................................. 10
  1.7. Problem statement................................................................................................................ 12
  1.8. Project Objectives................................................................................................................ 12
  1.9. Project Scope........................................................................................................................ 13
  1.10. Expected outcome.............................................................................................................. 13
  1.11. Summary............................................................................................................................ 16

Chapter 2: Literature Review........................................................................................................ 17
  2.1. History of Electronic Thesis and Dissertation (E-TD).......................................................... 17
  2.2. Networked Digital Library of Theses and Dissertations....................................................... 18
      2.2.1. Membership.................................................................................................................. 18
      2.2.2. Collection Size.............................................................................................................. 21
      2.2.3. Access Statistics........................................................................................................... 23
      2.2.4. Worldwide Release....................................................................................................... 25
  2.3. The E-TD system of Worcester Polytechnic Institute........................................................... 27
  2.4. The E-TD system of George Washington University......................................................... 28
  2.5. The E-TD System of Brigham Young University................................................................. 30
  2.6. The E-TD system of Pennsylvania State University............................................................ 31
  2.7. Metadata Standard.............................................................................................................. 32
  2.8. Metadata Standard Used by NDLTD.................................................................................. 33
  2.9. File Format.......................................................................................................................... 34
  2.10. Grid Computing and Data Grid........................................................................................ 35
  2.11. Open Grid Service Architecture (OGSA) Framework..................................................... 36
  2.12. Globus Toolkit.................................................................................................................... 36
      2.12.1. Global Digital Library Grid (GDLG)........................................................................... 37
  2.13. Storage Resource Broker.................................................................................................. 38
      2.13.1. SRB Architecture....................................................................................................... 39
Chapter 1: Introduction ................................................................. 4
1.1. Research Background ............................................................... 4
1.2. Objectives of E-TD Repository System .......................................... 5
1.3. Research Questions .................................................................... 5
1.4. Data Source .............................................................................. 6
1.5. System Concept ......................................................................... 7
1.6. Thesis Description ...................................................................... 7
1.7. Thesis Metadata ......................................................................... 8
1.8. Thesis Ontology ........................................................................ 9
1.9. Thesis Browsing ...................................................................... 10
1.10. Thesis Searching ...................................................................... 10
1.11. Thesis Creation ....................................................................... 11
1.12. Thesis Update ...................................................................... 11
1.13. Thesis Indexing ....................................................................... 12
1.15. Thesis Security ....................................................................... 12
1.16. Thesis Access ....................................................................... 12
1.17. Summary .............................................................................. 16

Chapter 2: System Architecture .................................................... 17
2.1. System Architecture ................................................................ 17
2.2. System Implementation ............................................................ 22
2.3. Systems using Metadata Catalog .................................................. 23
2.3.1. NSF National Science Digital Library Program ................................. 23
2.3.2. Metadata Catalog (MCAT) ................................................................. 39
2.3.3. SRB Server ........................................................................... 40
2.3.4. SRB Client ........................................................................... 41
2.3.5. Replicated Catalog .................................................................. 41
2.3.6. Organizations using SRB ............................................................... 42
2.3.7. National Science Foundation (NSF) National Science Digital Library Program ................................................................. 43
2.14. Ontology and Semantic Web ....................................................... 44
2.15. Systems using ontology .............................................................. 45
2.15.1. Semantic Digital Library Services (SDLS) .................................................. 45
2.15.2. JeromeDL - Semantic interoperability in Digital Libraries ..................... 46
2.15.3. Swoogle: A Semantic Web Search and Metadata Engine ...................... 47
2.16. Summary .............................................................................. 49

Chapter 3: Methodology ............................................................... 50
3.1. Research Methodology ............................................................... 50
3.2. Relevant Lifecycle Models ........................................................... 54
3.2.1. V-Model ............................................................................. 54
3.2.2. Rapid Application Development Model ............................................ 55
3.3. Life Cycle Model for the E-TD Repository System ............................. 56
3.4. Summary .............................................................................. 57

Chapter 4: System Analysis ............................................................ 58
4.1. Requirements of E-TD Repository System ........................................ 58
4.2. Functional requirements ............................................................. 59
4.2.1. Access Authentication ............................................................... 59
4.2.2. Membership registration ............................................................ 60
4.2.3. Thesis Submission ................................................................. 60
4.2.3.1. Inserting Metadata ................................................................. 60
4.2.3.2. Access Constraints ................................................................. 61
4.2.4. Viewing E-TDs Online ............................................................ 61
4.2.5. Downloading E-TDs ............................................................... 61
4.2.6. Change password ................................................................. 61
4.2.7. Adding Ontology ................................................................. 62
4.2.7.1. Add New Term ................................................................. 62
4.2.7.2. Add Synonym ................................................................. 62
4.2.7.3. Add Relationship ................................................................. 62
4.2.8. Theses Searching ................................................................. 62
4.2.8.1. Simple Searching ................................................................. 62
4.2.8.2. Ontology-based Searching ...................................................... 63
4.2.9. Thesis Browsing ................................................................. 63
4.3. Non-functional requirements ..................................................... 63
4.3.1. Modularity ................................................................. 63
4.3.2. Usability ................................................................. 64
4.3.3. Robustness ............................................................. 64
4.3.4. Maintainability and Expandability .............................. 64
4.3.5. Security ................................................................. 65
4.3.6. Reliability and Availability .......................................... 65
4.3.7. Fault tolerance ........................................................ 65
4.4. Structured Systems Analysis and Design Methodology (SSADM) .... 66
4.5. Major Tools of SSADM ..................................................... 67
4.5.1. Logical Data Modeling ............................................... 67
4.5.2. Data Flow Modeling ................................................... 68
4.5.3. Entity Event Modeling ................................................... 73
4.6. Dublin Core Metadata Elements of the E-TD System .......... 77
4.7. Extending Dublin Core Metadata Standard for this E-TD System .... 80
4.8. Summary .................................................................. 83

Chapter 5: System Design ...................................................... 84
5.1. Overview of E-TD Architecture ......................................... 84
5.1.1. Presentation Layer ...................................................... 84
5.1.2. Application Layer ...................................................... 85
5.1.3. The Data Grid Layer .................................................. 85
5.1.4. The Storage System .................................................... 86
5.2. Ontology of Software Testing ............................................ 88
5.3. Database Design ........................................................... 91
5.4. Interface Design ........................................................... 95
5.5. User Interface Design Principles ......................................... 98
5.5.1. User Familiarity ......................................................... 98
5.5.2. Consistency .............................................................. 100
5.5.3. Minimal Surprise ...................................................... 102
5.5.4. Recoverability ......................................................... 104
5.5.5. User Guidance ......................................................... 106
5.6. Summary ................................................................ 107

Chapter 6: System Implementation .......................................... 108
6.1. Hardware Development Environment ................................. 108
6.2. Software Development Environment ................................... 109
6.2.1. Operating System .................................................... 109
6.2.2. Web Server ............................................................ 109
6.2.3. Database ............................................................... 110
6.2.4. Graphic Creation ...................................................... 110
6.2.5. Browser ................................................................. 111
6.3. The implementation of Data Grid Environment ..................... 111
List of Figures

Figure 2.1: Student and committee choice for ETD availability from Virginia Tech (2668 ETDs as of July 17, 2000) ......................................................... 26
Figure 2.2: E-TD system of WPI ................................................................. 27
Figure 2.3: E-TD system of George Washington University .................. 29
Figure 2.4: E-TD System of Brigham Young University ....................... 30
Figure 2.5: E-TD Website of PENNSTATE ............................................... 31
Figure 2.6: Globus Toolkit Architecture .................................................. 37
Figure 2.7: SRB Architecture ................................................................. 39
Figure 2.8: Java SRB Admin Tool ........................................................... 40
Figure 2.9: Replicated Catalog ............................................................... 41
Figure 2.10: Swoogle Advanced Semantic Search Interface .................. 47
Figure 2.11: Swoogle Query Result ....................................................... 48
Figure 3.1: Activities taken in carrying out this E-TD System ................... 53
Figure 3.2: The V Model ............................................................... 54
Figure 3.3: The Rapid Application Development Model ....................... 56
Figure 4.1: Entity Relationship Diagram of E-TD System ..................... 67
Figure 4.2: Level 0 Data Flow Diagram for E-TD Repository System .... 69
Figure 4.3: Level 1 Data Flow Diagram for E-TD Repository System .... 71
Figure 4.4: Level 2 Data Flow Diagram for Searching Theses by Ontology Process ... 72
Figure 4.5: Member Entity Life History .............................................. 74
Figure 4.6: Thesis Entity Life History .................................................. 74
Figure 4.7: Metadata Entity Life History ............................................. 75
Figure 4.8: Term Entity Life History .................................................. 75
Figure 4.9: Synonym Entity Life History ............................................ 76
Figure 4.10: Ontology Entity Life History .......................................... 76
Figure 4.11: Browse by Program ........................................................ 81
Figure 4.12: List of E-TDs by Program ................................................. 81
Figure 4.13: Search by Matric Card Number ........................................ 82
Figure 4.14: Result by Matric Card Number ........................................ 83
Figure 4.15: Components of E-TD Architecture .................................. 87
Figure 5.1: Components of E-TD Architecture .................................. 87
Figure 5.2: Ontology of Software Testing Terms ................................. 89
Figure 5.3: Ontology of Software Testing Terms (Test Design Technique) ... 90
Figure 5.4: Example of synonyms in software testing terms ............... 94
Figure 5.5: Some of the terms of ontology in software testing ............. 95
Figure 5.6: Basic Interface Design ..................................................... 96
Figure 5.7: Member Login Page of E-TD Repository System ................ 96
Figure 5.8: Theses Searching Page of E-TD Repository System .......... 97
Figure 5.9: Insert Metadata Page of E-TD Repository System ............ 97
Figure 5.10: Save and Reset ............................................................. 99
Figure 5.11: Browse, Upload and Reset .............................................. 99
Figure 5.12: Consistency of the size of banner and navigation bar ....... 100
Figure 5.13: Consistency of the size and colour of header and text ........ 101
Figure 5.14: Save New Term ............................................................ 102
Figure 5.15: Successful Message (saving term) .................................. 103
Figure 5.16: Saving New Synonym .................................................... 103
Figure 5.17: Successful Message (saving synonym) ............................ 104
Figure 5.18: Confirmation Message to Remove Thesis File ................. 104
Figure 5.19: Upload Thesis Page of E-TD Repository System ............. 105
Figure 5.20: Error Message

Figure 5.21: E-TD Submission Guidelines

Figure 5.22: E-TD Online Help Page of E-TD Repository System

Figure 7.1: Modules of E-TD Repository System

Figure 7.2: Result of Question 1 (Part A)

Figure 7.3: Result of Question 2 (Part A)

Figure 7.4: Result of Question 3 (Part A)

Figure 7.5: Result of Question 4 (Part A)

Figure 7.6: Result of Question 5 (Part A)

Figure 7.7: Result of Question 6 (Part A)

Figure 7.8: Login

Figure 7.9: Upload ETD

Figure 7.10: Browse Theses

Figure 7.11: Search Theses

Figure 7.12: Ontology-based Searching

Figure 7.13: Viewing Metadata and ETDs online

Figure 7.14: Feel Comfortable in Using This E-TD System

Figure 7.15: Graph for the comparison of the result from simple search engine and ontology search engine

Figure 7.16: Simple Search Engine

Figure 7.17: Search Result Returned from Simple Search Engine

Figure 7.18: Ontology Search Engine

Figure 7.19: Search Results Returned from Ontology-based Search Engine

Figure 7.20: Search Results Returned from Ontology-based Search Engine
List of Tables

Table 2.1: NDLTD Membership.................................................................21
Table 2.2: NDLTD collection size..............................................................23
Table 2.3: Access Log Statistics from the VT-ETD collection...................24
Table 2.4: Access by Non-US Sites...........................................................25
Table 2.5: Comparison between SDLS, JeromeDL, Swoogle and E-TD Repository System.................................................................49
Table 4.1: Symbols of Data Flow Diagram...............................................68
Table 4.2: Symbols of Entity Life History..................................................73
Table 5.1: Term Table..............................................................................91
Table 5.2: Sample rows in Term table......................................................92
Table 5.3: Synonym Table......................................................................93
Table 5.4: Sample rows of Synonym table..............................................93
Table 5.5: Ontology Table......................................................................94
Table 5.6: Sample rows of Ontology Table.............................................94
Table 5.7: User Interface design principles.............................................98
Table 7.1: Test Results of Integration Testing........................................122
Table 7.2: Test Results of System Testing...............................................125
Table 7.3: Comparison of the test result from simple search engine and ontology search engine.........................................................136
Chapter 1: Introduction

Nowadays, modern technology is rapidly progressing in the countries all over the world. Therefore, it is not surprising that most of the activities change to online systems. The technology progress has a great impact on the education sector.

As our road goes ahead to e-communications and information age, the availability of scholarly material in e-format becomes increasingly important and the movement of publishing Theses and Dissertations (TDs) in electronic format is truly becoming a global phenomenon.

For being an introduction, the definition of thesis and dissertation, the difference between thesis and dissertation, the definition of Electronic-Theses and Dissertations (E-TDs), background study, project overview, problem statement, project objectives, project scope and what is the expected outcome of this research will be discussed in this chapter.

1.1. Definition of Thesis and Dissertations

A thesis can be defined as a dissertation on a particular subject in which one has done original research, as one presented by a candidate for a diploma or a degree, especially a master's degree whereas a dissertation is a written essay, treatise or thesis, especially one written by a candidate for the degree of Doctor of Philosophy (Stein and Urdang, 1966).
1.2. Thesis versus Dissertation

The “Dissertation” and “Thesis” are both used interchangeably even though they are not the same. The difference between them has become the difference between the degrees themselves. The term “Thesis” is used to designate the paper submitted to the Master’s degree program while the term “Dissertation” is designated to the paper applied for the Ph.D/doctorate degree (Rachel, 2002).

1.3. What is an E-TD?

An Electronic Thesis or Dissertation (E-TD) has the same content as the traditional paper version such as text, figures, tables, footnotes, and references. However, being electronic, an E-TD can be displayed on the World Wide Web and be retrieved and archived electronically (Marilyn, 2006).

1.4. E-TD Repository

A repository is a place where data are stored and maintained. A repository can be (Wikipedia, 2008):

- a place where data is stored
- a place where specifically digital data are stored
- a site where eprints are located
- a place where multiple databases or files are located for distribution over a network,
a computer location that is directly accessible to the user without having to travel across a network;

a place where anything is stored for probable reuse;

a place to store digital data.

In this case, as it is E-TD Repository System, it is a place where online and searchable E-TDs are stored and located for distribution over a network.

1.5. Background Study

An E-TD system was initially launched by Virginia Technology (VT) in 1996. Nowadays, hundreds of E-TD systems are accessible on the Internet. Even though each of the E-TD system has different representation, they all have the same functions such as uploading, searching and browsing theses.

Some of the search engines of these E-TD systems do not give effective responses in searching E-TDs. Theses systems require the users to spend a lot of time and effort to search for theses and dissertations which are relevant to them. Besides, the metadata captured in some E-TD systems is not very helpful for searching the E-TDs from the database. For instance, it is found that only one keyword is allowed to be entered for one E-TD submission in the E-TD system of George Washington University.

1.6. Project Overview

With the development of Internet and the rapid advancement of information and network technology, E-TD system has become a hot method in recent years. Access to global theses, dissertations and literatures are continually increasing in large quantities. To cope with the ever-changing environment resulting from the knowledge explosion,
more and more institutions have tried to change their manual theses management system into digitized format.

The E-TD system is getting more and more popular in the universities as it could save shelf space, be more easily searched and retrieved online. Of particular importance are the students who write the works and often wish to be more expressive than black ink on white paper allows them to be. Through the Internet, students can freely retrieve TDs without the restriction of time and place. If all TDs are captured electronically and are freely shared, it can help scholars and researchers to share knowledge on worldwide scale.

In this research, the E-TD system generally includes both digital collections of TDs and services that facilitate archiving and retrieving of E-TDs. Besides, the submission of thesis is done electronically through this online system. When submitting a thesis or dissertation to this system, it is required to input some metadata about the theses or dissertations like the author, title, abstract, subject code, program, supervisor, year, etc. Those metadata could be used later for precise search. One of the most promising metadata schemas used for digital object is Dublin Core which is used for this E-TD system.

In the Internet era, Semantic Web came as a model of semantic retrieval in the web environment. The traditional means and techniques for information storage and retrieval are required to be modified to suit the changed needs. In this research, the traditional searching service of TDs has been changed into ontology-based searching service to represent domain knowledge in machine processable form. As faster and meaningful information retrieval is important in E-TD systems, the searching service of this system is improved with the ontology-based searching technique.
1.7. Problem statement

Although a lot of E-TD systems have been implemented by the Universities all over the world, almost all of these E-TD systems are just simple ones. Without data grid, different database schemas cannot be linked together to access the data from one interface to all different database systems. The data grid technology allows users to access location-independent data: users do not need to know where files are located and how they are stored.

The effective search and browse services are one of the most important goals of E-TD systems. Normally, a lot of time and effort are wasted in searching for all of the available information about theses from the database. The ontology-based searching mechanism provides the ability to describe the concepts of the domain and the relationships between them.

1.8. Project Objectives

The objectives of this study are to implement an E-TD system using data grid and enhance searching of E-TDs by incorporating semantic web technique.

In this research, the implementation of data grid technology is studied in the development of E-TD system. The aim of using data grid technology is to allow users to access location-independent data: users do not need to know where files are located and how they are stored.

Besides, semantic web technology is also incorporated. The semantic web technology provides the ability to characterize relationships between named entities of the domain. To save time and effort of retrieving information about theses from the database, the ontology-based searching technology is applied in this project. For
example, if you were searching theses related to *Unit Testing*, you might also want to find the theses related to all the terms that are related to *Unit Testing*. This can be fulfilled by using ontology which can formalize to describe the semantics. The software testing ontology is implemented and searching service is improved with ontology-based searching mechanism in this system.

1.9. Project Scope

The scope of this project is to archive Computer Science theses and dissertations in electronic format in the repository of E-TD System. Moreover, the ontology-based searching technique is implemented in this study. Even though there are a wide range of terms used in Computer Science, the scope for ontology is narrowed down to software testing terminologies. The uploaded theses to this E-TD system can be in one of these three formats: *.pdf or *.doc or *.zip.

1.10. Expected outcome

This research is to archive computer science theses and integrate data grid technology to this E-TD system as to make the best use of the power of data grid in this project. Moreover, it can save cost and time of the students for their commutation to and fro the library.

The SRB is a federated server system (Rajasekar *et al.*, 2003). In this research, two SRB servers are configured to form a federation to service client requests. The federated SRB implementation provides unique advantages: improve performance, data reliability and availability and fault tolerance.
Replication is core functionality in SRB (Rajasekar). The MCAT replication is implemented in this research. Even though there are two MCATs operating in two different zones, the overall system behaves as though it is a single zone with two replicated MCATs. The MCATs synchronize metadata between them, so that each contains the same information as the MCAT of the other zone. This model provides a completely replicated system which has a high degree of fault-tolerance for MCAT failures. The user will not miss any access to E-TDs in this project even if one MCAT becomes non-functional.

With this E-TD system in place, the students from our faculty will be able to find the full texts of related works easily, to read literature reviews prepared by the seniors. Since this is an online system, it is available for 24 hours for anyone from anywhere. Therefore, this system can benefit the students who are working and having the problem to visit the library during daytime.

The students are likely to benefit financially from this system. Publishing electronically should save them the costs of preparing at least some of the paper copies now required. There may be lower fees for the faculty as no shelf space required to store the hard copies of theses. It is anticipated that the system will provide solutions to solve the problems of manually searching and getting theses from library and submitting theses. Creating theses and dissertations in e-format and publishing it online has several benefits compared to writing a traditional paper thesis and dissertation. This ETD system would have numerous benefits for researchers, librarians and students with expected outcomes below:

- Better access to *Theses and Dissertations (TD)*
  - TD is available 24 hours
− TD is accessible worldwide provided if the user has a valid account.

Less expense to the students and the library

− no paper costs
− no shelf space

Greater freedom for students to demonstrate creatively the result of their research

− more flexible

User friendly and easy to use

Development of faculty staffs and students as electronic scholars

Reduce costs to produce the final product for students

Improve access to the information in graduate research by making it immediately available to huge audiences through the Internet.

Advantages from the library's point of view:

More timely public access to current research

− Freely available worldwide.

Serve more users with reduced staff

No shelf space required for storage

Reduce the problem of storing theses

More efficient interaction with faculty
All in all, the incorporation of data grid and semantic web technology helps to produce an E-TD system which is more scalable and efficient in retrieving and searching of theses.

1.11. Summary

The overview of the whole project and the problems of the existing E-TD systems are discussed, the objectives for this research were set, the scope of this study was defined and the expected outcome of this research was conferred in this chapter.
Chapter 2: Literature Review

The literature review presents the background study about knowledge and information to do this research. It is the information gathering process which is part of an important research phase in this project.

This chapter starts with a discussion on history of ETD followed by the review of current E-TD systems, review of semantic web and ontology, review of grid computing and data grid technology and review of methodologies. The enough information for the implementation of this research is obtained at the end of this chapter.

2.1. History of Electronic Thesis and Dissertation (E-TD)

The first workshop for Electronic Thesis and Dissertation (E-TD) project was held in 1987 in Ann Arbor arranged by University Microfilms (UMI) and attended by representatives of Virginia Tech, the University of Michigan, SoftQuad, and ArborText (Erickson, 1997 and Edward et al 1999). Yur Rubinski from SoftQuad developed the initial Document Type Definition (DTD) for theses and dissertations and Virginia Tech (VT) funded the development of DTD and the project continued at VT.

With funding in 1993 from Southeastern Universities Research Association (SURA), VT was able to develop tools for submitting E-TDs in Standard Generalized Markup Language (SGML) and in Adobe’s Portable Document Format (PDF). In 1996, Virginia Tech launched a pilot project to implement a distributed digital library of E-TDs. VT provides software and technical support to member institutions of the Networked Digital Library of Theses and Dissertations (NDLTD). Virginia Tech began requiring all theses to be submitted electronically since 1997. Initially, VT accepted both Adobe’s PDF format and Standard Generalized Markup Language (SGML) format.
In aiming of serving theses and dissertations in electronic format available worldwide, UMI also announced to accept the full text of all newly submitted theses of electronic version in late 1997 (Matthew, 1996).

2.2. Networked Digital Library of Theses and Dissertations

The Networked Digital Library of Theses and Dissertations (NDLTD) is an international organization dedicated to promoting the adoption, creation, use, dissemination and preservation of electronic analogues to the traditional paper-based theses and dissertations (Ann, 2008).

The NDLTD has progressed in it’s membership, collection size, access and worldwide availability (Marcos et al., 2001).

2.2.1. Membership

NDLTD memberships shown in Table 1 are the list analyzed in August 2001. These various partners represent 23 countries: Australia, Brazil, Canada, China, Colombia, Germany, Greece, Hong Kong, India, Italy, Mexico, Netherlands, Norway, Russia, Singapore, South Africa, South Korea, Spain, Sudan, Sweden, Taiwan, the USA, and the United Kingdom. In addition, 11 of the registered NDLTD members (marked with an asterisk in the table below) have started accepting the submission of theses and dissertations in electronic format by early 2002 (Marcos et al., 2001).

<table>
<thead>
<tr>
<th>USA Universities (524)</th>
<th>International Universities (512)</th>
<th>Institutions (156)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air University (Alabama)</td>
<td>Alicante University (Spain)</td>
<td>Cinemedia</td>
</tr>
<tr>
<td>Alicante University</td>
<td>Australian National University</td>
<td>Coalition for Networked</td>
</tr>
<tr>
<td>University</td>
<td>Country</td>
<td>Organization</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>--------------------------------</td>
<td>---------------------------------------------------</td>
</tr>
<tr>
<td>Baylor University</td>
<td></td>
<td>Information</td>
</tr>
<tr>
<td>Brigham Young University</td>
<td></td>
<td>Committee on Institutional Cooperation</td>
</tr>
<tr>
<td>California Institute of Technology</td>
<td></td>
<td>Consorci de Biblioteques</td>
</tr>
<tr>
<td>Clemson University</td>
<td></td>
<td>Univers. Catalunya</td>
</tr>
<tr>
<td>College of William and Mary (Illinois)</td>
<td></td>
<td>Diplomica.com</td>
</tr>
<tr>
<td>Concordia University</td>
<td></td>
<td>Dissertationene Online</td>
</tr>
<tr>
<td>East Carolina University</td>
<td>(Australia)</td>
<td>Dissertation.com</td>
</tr>
<tr>
<td>East Tennessee State University*</td>
<td></td>
<td>ETDweb</td>
</tr>
<tr>
<td>Florida Institute of Technology</td>
<td></td>
<td>Ibero-American Sci. &amp; Tech. Ed. Cons. (ISTEC)</td>
</tr>
<tr>
<td>Florida International University</td>
<td></td>
<td>National Documentation Centre (NDC, Greece)</td>
</tr>
<tr>
<td>George Washington University</td>
<td></td>
<td>National Library of Portugal</td>
</tr>
<tr>
<td>Louisiana State University*</td>
<td></td>
<td>OhioLINK</td>
</tr>
<tr>
<td>Marshall University</td>
<td></td>
<td>OCLC</td>
</tr>
<tr>
<td>Massachusetts Institute of Technology</td>
<td></td>
<td>Organization of American States (OAS)</td>
</tr>
<tr>
<td>Miami University of Ohio</td>
<td></td>
<td>SOLINET</td>
</tr>
<tr>
<td>Michigan Tech</td>
<td></td>
<td>Sudanese National</td>
</tr>
<tr>
<td>Mississippi State University</td>
<td></td>
<td>Electronic Library (Sudan)</td>
</tr>
<tr>
<td>Montana State University</td>
<td></td>
<td>Solinet UNESCO</td>
</tr>
<tr>
<td>Naval Postgraduate School</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Jersey Institute of Technology</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University</td>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>---------------------------</td>
<td></td>
</tr>
<tr>
<td>New Mexico Tech</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Carolina State University*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Western University</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pennsylvania State University</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regis University</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rochester Institute of Technology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texas A&amp;M University</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Colorado</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Florida</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Georgia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Hawaii at Manoa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Iowa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Kentucky</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Maine*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of North Texas*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Oklahoma</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Pittsburgh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Rochester</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of South Florida</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Tennessee, Knoxville</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Tennessee, Memphis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Texas at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rand Afrikaans University (South Africa)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhodes University (South Africa)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shanghai Jiao Tong University</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(China)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>St. Petersburg State Technical U.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Russia)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>State University of Campinas</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Brazil)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sudanese National Electronic Library (Sudan)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universidad de las Américas Puebla (México)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universitat Autonoma de Barcelona (Spain)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universitat d'Alacant (Spain)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universitat de Barcelona (Spain)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universitat de Girona (Spain)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universitat de Lleida (Spain)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universitat Oberta de Catalunya</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Spain)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universitat Politecnica de Catalunya</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universitat Politecnica de Valencia (Spain)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universitat Pompeu Fabra (Spain)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universitat Rovira i Virgili (Spain)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University</td>
<td>Institution</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------------</td>
<td>---------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Austin*</td>
<td>Université Laval (Québec, Canada)</td>
<td></td>
</tr>
<tr>
<td>University of Utrecht</td>
<td>University of Bergen (Norway)</td>
<td></td>
</tr>
<tr>
<td>University of Virginia</td>
<td>University of Antioquia (Medellin, Colombia)</td>
<td></td>
</tr>
<tr>
<td>University of West Florida</td>
<td>University of British Columbia</td>
<td></td>
</tr>
<tr>
<td>University of Wisconsin, Madison</td>
<td>(Canada)</td>
<td></td>
</tr>
<tr>
<td>Vanderbilt University</td>
<td>University of Guelph (Ontario, Canada)</td>
<td></td>
</tr>
<tr>
<td>Virginia Commonwealth University</td>
<td>University of Hong Kong*</td>
<td></td>
</tr>
<tr>
<td>Virginia Tech*</td>
<td>University of Melbourne (Australia)</td>
<td></td>
</tr>
<tr>
<td>West Virginia University*</td>
<td>University of Mysore (India)</td>
<td></td>
</tr>
<tr>
<td>Western Michigan University</td>
<td>University of New South Wales</td>
<td></td>
</tr>
<tr>
<td>Worcester Polytechnic Institute</td>
<td>University of Pisa (Italy)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>University of Queensland (Australia)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>University of Sao Paulo (Brazil)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>University of Sydney (Australia)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>University of Utrecht (Netherlands)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>University of Waterloo (Canada)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Uppsala University (Sweden)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wilfrid Laurier University (Canada)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1: NDLTD Membership (Marcos et al., 2001)

### 2.2.2. Collection Size

Table 2.2 shows the result of an online survey conducted by Gail McMillan in July 2001 and represents only 27 institutions that responded to the survey. The collection size grows from a few dozen at VT in 1996, to 4,328 ETDs in March 2000, to
7,268 ETDs in July 2001 (Marcos et al., 2001).

<table>
<thead>
<tr>
<th>University/Institution</th>
<th>ETD Collection size</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADT: Australian Digital Thesis Program (Australia)</td>
<td>238</td>
</tr>
<tr>
<td>University of Bergen (Norway)</td>
<td>45</td>
</tr>
<tr>
<td>California Institute of Technology</td>
<td>2</td>
</tr>
<tr>
<td>Consorci de Biblioteques Universitaries de Catalunya (Spain)</td>
<td>151</td>
</tr>
<tr>
<td>East Tennessee State University</td>
<td>106</td>
</tr>
<tr>
<td>Humboldt-University (Germany)</td>
<td>430</td>
</tr>
<tr>
<td>Louisiana State University</td>
<td>3</td>
</tr>
<tr>
<td>Mississippi State University</td>
<td>33</td>
</tr>
<tr>
<td>MIT</td>
<td>62</td>
</tr>
<tr>
<td>North Carolina State University</td>
<td>301</td>
</tr>
<tr>
<td>Pennsylvania State University</td>
<td>83</td>
</tr>
<tr>
<td>Pontifical Catholic University (PUC) (Brazil)</td>
<td>90</td>
</tr>
<tr>
<td>Gerhard Mercator Universitat Duisburg (Germany)</td>
<td>126</td>
</tr>
<tr>
<td>Universitat Politecnica de Valencia (Spain)</td>
<td>189</td>
</tr>
<tr>
<td>University of Florida</td>
<td>174</td>
</tr>
</tbody>
</table>
2.2.3. Access Statistics

The collection size of theses and the number of access to NDLTD has continued to increase each year. The access logs of E-TDs collections at Virginia Tech has been periodically analyzed and the result from 1997 to 2000 are shown in the table below (Marcos et al., 2001).

<table>
<thead>
<tr>
<th>Institution</th>
<th>Accesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Georgia</td>
<td>121</td>
</tr>
<tr>
<td>University of Iowa</td>
<td>6</td>
</tr>
<tr>
<td>University of Kentucky</td>
<td>19</td>
</tr>
<tr>
<td>University of Maine</td>
<td>27</td>
</tr>
<tr>
<td>University of North Texas</td>
<td>337</td>
</tr>
<tr>
<td>University of South Florida</td>
<td>25</td>
</tr>
<tr>
<td>University of Tennessee</td>
<td>12</td>
</tr>
<tr>
<td>University of Tennessee, Knoxville</td>
<td>28</td>
</tr>
<tr>
<td>Uppsala University (Sweden)</td>
<td>178</td>
</tr>
<tr>
<td>Virginia Tech</td>
<td>3393</td>
</tr>
<tr>
<td>West Virginia University</td>
<td>1006</td>
</tr>
<tr>
<td>Worcester Polytechnic Institute</td>
<td>83</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>7268</strong></td>
</tr>
</tbody>
</table>

*Table 2.2: NDLTD collection size (Marcos et al., 2001)*
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Requests for PDF files (mostly full ETDs)</td>
<td>221,679</td>
<td>481,038</td>
<td>117.0%</td>
<td>578,152</td>
<td>20.2%</td>
</tr>
<tr>
<td>Requests for HTML files (mostly tables of contents and abstracts)</td>
<td>165,710</td>
<td>215,539</td>
<td>30.1%</td>
<td>260,699</td>
<td>21.0%</td>
</tr>
<tr>
<td>Requests for multimedia</td>
<td>1,714</td>
<td>4,468</td>
<td>160.7%</td>
<td>12,633</td>
<td>182.7%</td>
</tr>
<tr>
<td>Distinct files requested</td>
<td>6,419</td>
<td>21,451</td>
<td>234.2%</td>
<td>16,409</td>
<td>-23.5%</td>
</tr>
<tr>
<td>Distinct hosts served</td>
<td>29,816</td>
<td>57,901</td>
<td>94.2%</td>
<td>87,804</td>
<td>51.6%</td>
</tr>
<tr>
<td>Average data transferred daily</td>
<td>156,089</td>
<td>219,132</td>
<td>40.4% MbMB</td>
<td>382</td>
<td>74.4% MbMB</td>
</tr>
<tr>
<td>Data transferred</td>
<td>55,637</td>
<td>78,107</td>
<td>40.4%</td>
<td>137 GbGB</td>
<td>75.6%</td>
</tr>
</tbody>
</table>

Table 2.3: Access Log Statistics from the VT-ETD collection (Marcos et al., 2001)

The seven countries which have an increasing number of accesses each year except Germany in the 97/98 – 98/99 are shown in Table 2.4. Most of the accessing countries are European countries.
### Table 2.4: Access by Non-US Sites (Marcos et al., 2001)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>United Kingdom</td>
<td>6,735</td>
<td>1</td>
<td>11,347</td>
<td>1</td>
<td>68.5%</td>
<td>25,583</td>
<td>1</td>
<td>125.5%</td>
</tr>
<tr>
<td>Malaysia</td>
<td>876</td>
<td>16</td>
<td>4,190</td>
<td>6</td>
<td>378.3%</td>
<td>16,147</td>
<td>2</td>
<td>285.4%</td>
</tr>
<tr>
<td>France</td>
<td>2,138</td>
<td>7</td>
<td>4,797</td>
<td>5</td>
<td>124.4%</td>
<td>14,960</td>
<td>3</td>
<td>211.9%</td>
</tr>
<tr>
<td>Germany</td>
<td>6,727</td>
<td>2</td>
<td>3,374</td>
<td>9</td>
<td>-49.8%</td>
<td>14,384</td>
<td>4</td>
<td>326.3%</td>
</tr>
<tr>
<td>Canada</td>
<td>3,413</td>
<td>4</td>
<td>9,632</td>
<td>3</td>
<td>182.2%</td>
<td>13,543</td>
<td>5</td>
<td>40.6%</td>
</tr>
<tr>
<td>Spain</td>
<td>590</td>
<td>18</td>
<td>3,647</td>
<td>8</td>
<td>518.1%</td>
<td>9,918</td>
<td>6</td>
<td>171.9%</td>
</tr>
<tr>
<td>Italy</td>
<td>1,430</td>
<td>12</td>
<td>3,095</td>
<td>10</td>
<td>116.4%</td>
<td>9,300</td>
<td>7</td>
<td>200.5%</td>
</tr>
</tbody>
</table>

**2.2.4. Worldwide Release**

According to the copyright issue of NDLTD, some electronic documents are allowed to be viewed worldwide, some are not allowed to be viewed at all and some are mixed which means that some portions (e.g., particular chapters) have restricted access. Most of the VT students allow their documents to be viewed worldwide (see Figure 2.1) (Marcos et al., 2001).
Figure 2.1: Student and committee choice for ETD availability from Virginia Tech

(2668 ETDs as of July 17, 2000) (Marcos et al., 2001)
2.3. The E-TD system of Worcester Polytechnic Institute

Since July 01, 2002, the policy of Worcester Polytechnic Institute (WPI) states that all the Masters Theses and Doctoral Dissertations are to be submitted electronically (Worcester Polytechnic Institute, 2007). This means that the university no longer accepts paper copies of theses and dissertations. The WPI is also a member of NDLTD. Information of E-TD at WPI is available on the web at http://www.wpi.edu/+etd.

![Figure 2.2: E-TD system of WPI](image)

The E-TDs at WPI are made accessible through the World Wide Web. Abstract of all E-TDs are available worldwide. Most of the body of the E-TD is openly available for worldwide access. However, some E-TDs are made readable only from the computers within the WPI campus (Worcester Polytechnic Institute, 2007).

The E-TDs can be browsed by author or department and search by title, author,
abstract or department. It can also be searched by advisor name and keyword. When the users browse the theses, all the E-TDs are not able to be displayed on one page since there are too many E-TDs in the collection. Users can view the result by selecting certain alphabet that makes the arrangement of theses look more systematic and clear. WPI also provide a table which display the statistic about distribution of the theses and dissertations for each department.

The E-TD is captured in PDF format at WPI. This file format retains all formatting information about a document including multimedia information to be viewed on any computer. The WPI also provide a quick link to Adobe Acrobat Reader® to allow the users who need to install it to view the E-TDs.

The WPI provides a guideline of how to submit E-TDs. This will help users on how to convert the document to PDF format, how to modify E-TDs online and display some important announcement to students. The interface of WPI is simple and clear. Index and hyperlink of the website are also systematic and clear for users.

2.4. The E-TD system of George Washington University

The students from George Washington University (GWU) can now publish their theses/dissertations in electronic format (George Washington University, 2007). The ETD can include high resolution graphics, charts, pictures, and photographs. All the theses and dissertations are made available worldwide. The worldwide users can access these E-TDs via http://www.gwu.edu/~etds/.
The weaknesses of the ETD system from GWU are as follows:

מזווג

The collection of ETD is limited. It just has 17 ETDs in the collection which makes users often fail to search the ETDs they need.

מזווג

The search engine does not give an excellent response. May be it is because of few ETDs in the collection at the GWU site.

מזווג

Since only one keyword for one E-TD is captured by the system, users need to know the exact keyword to search that E-TD. Thus, the probability of getting the correct E-TD from system is very low.

מזווג

No guideline available for users on how to get started to search ETDs. Users have to spend their time and energy to explore by themselves.
2.5. The E-TD System of Brigham Young University

The E-TDs of Brigham Young University (BYU) can be accessed at http://etd.byu.edu/. The content of the E-TDs at BYU are accessible worldwide (Brigham Young University). The users can browse all E-TDs or can search E-TDs by author, title, subject, abstract and department. However, password is required for the E-TD submission to this system.

The E-TD is captured in PDF format at BYU. As only the PDF file format is allowed for the thesis submission to this E-TD system, the guideline on how to convert to PDF is provided. The interface of BYU E-TD is simple and clear. It has 1320 of E-TDs in the collection.
2.6. The E-TD system of Pennsylvania State University

The Pennsylvania State University (PSU), Information Technology Services, Digital Library Technologies, and University Libraries started an initiative to allow students to submit theses and dissertations to be submitted and archived electronically since the fall of 1998 (Pennsylvania State University, 1998).

The **User Id** and **Password** is required for thesis submission. The body of all the thesis at Penn State is publicly available to all the users. Users can browse theses by author or program. Moreover, the E-TD at PSU can be searched by author, title, abstract, program, committee member name and keyword. The E-TD system at Penn State allows the students to integrate video, sound and colour images to their work. The Penn State University is also the member of NDLTD.

![Figure 2.5: E-TD Website of PENNSTATE](image-url)
2.7. Metadata Standard

The most well-known metadata initiative for digital libraries is the Dublin Core Metadata Element Set (Michael, 1997). As the E-TD system is also a part of digital library, most of the E-TD systems use Dublin Core as a metadata standard for information retrieving. It has the specific aim of supporting resource discovery in a network environment.

The Internet has many virtues, but it is not designed specifically for information retrieval. However, it is possible that if Dublin Core metadata is attached to or embedded in theses submitted to E-TD system, a new generation of E-TD system could collect this metadata and make it available through improved searching services.

*Difference between data and metadata* (Suleman, 2002):

- Data refers to digital objects or digital representations of objects
- Metadata is information about the objects (e.g. title, author, etc.)

Metadata forms the core of the E-TD systems. It is the major component of the data input stage; it is used for generating various views for display on the Web and it serves as the basis for search and retrieval. Metadata usually contains useful links to the source digital objects.

To use and benefit from metadata, a common format is needed for expressing it which should be designed to be machine-readable. The Dublin Core metadata set consists of 15 metadata elements as follows (Dublin Core Metadata Initiative, 2001):

1. Title
2. Creator
3. Subject
2.8. Metadata Standard Used by NDLTD

The ETD-Metadata Standard (ETD-MS) is a standard set for metadata elements used by NDLTD to describe an electronic thesis or dissertation. All Dublin Core elements, except the element Relation, are represented by ETD-MS and one additional metadata element (thesis.degree) is included in the ETD-MS format.

The metadata used by NDLTD are as follows (Atkins, 2007):

1. dc.title
2. dc.creator
3. dc.subject
4. dc.description
2.9. File Format

The two most common file formats used by universities with active E-TD programs are PDF and XML. Choice of file format is based on several factors including ease of creation of the document, searchability, and long-term archiving (migration) issues, etc. Currently, PDF is the most common file format accepted by institutions with E-TD system.

With the widespread use of word processors, it is highly likely that a student will have written his/her thesis using any word processor. Therefore, a conversion from any type of file format to PDF file will be needed before submission. The conversion from of any format to PDF is very simple and easy to produce a basic E-TD for online preservation and access.

Adobe Acrobat’s Portable Document Format (.pdf) is recommended in most of the E-TD systems since it retains all format codes and graphic images, appearing as the original paper document and also because it is easily portable. In addition, .pdf files can
be indexed and searched by keywords. Apart from being an open standard, it maintains the integrity of the document. It can be converted to PostScript format, which can be used for electronic delivery and printed directly. In UNIX, the software for conversion of PostScript to PDF is also available. This produces essentially an electronic version of a paged document. The conversion can easily be done by using converter software (such as PDF995, Adobe Acrobat Distiller).

2.10. Grid Computing and Data Grid

The term Grid computing originated in the early 1990s (Wikipedia, 2007). Grid Computing is the creation of a "virtual supercomputer" by using a network of geographically dispersed computers. A data grid is a grid computing system that deals with data — the controlled sharing and management of large amounts of distributed data (Wikipedia, 2006).

Many scientific and engineering applications need to access large amounts of distributed data (Wikipedia, 2006). The number and size of the collection of these data has been growing rapidly in recent years and will continue to grow.

Current large-scale data grid projects, which are using SDSC Storage Resource Broker (SRB) as the underlying data grid technology, include the Biomedical Informatics Research Network (BIRN), the Southern California Earthquake Center (SCEC), and the Real-time Observatories, Applications, and Data management Network (ROADNet) (Wikipedia, 2007).
2.11. Open Grid Service Architecture (OGSA) Framework

The OGSA framework is intended to facilitate the seamless use and management of distributed, heterogeneous resources (Foster et al., 2002). It supports the creation, maintenance, and application of ensembles of services maintained by Virtual Organizations.

The development of OGSA represents an evolution of the Globus Toolkit and Web Services (Foster et al., 2002). The OGSA extends support for transient, stateful service instances with existing Web services technologies. Although Globus toolkit exists the key concepts such as factory, registry, reliable and secure invocation, etc., it is less general and flexible form than OGSA, and without the benefits of a uniform interface definition language.

2.12. Globus Toolkit

The Globus Toolkit, developed by the Globus Alliance and many others all over the world, is an open source letting people share computing power, databases, and other tools securely online across corporate, institutional, and geographic boundaries without sacrificing local autonomy (University of Chicago, 1998). The toolkit includes software services and libraries for resource monitoring, discovery, and management, plus security and file management.

The parts of Globus that are impacted most by the OGSA are (Berman et al., 2003):

- The Grid resource allocation and management (GRAM) protocol.
- The information infrastructure, metadirectory service (MDS-2), used for information discovery, registration, data modeling and a local registry.
- The Grid security infrastructure (GSI), which supports single sign-on, restricted delegation and credential mapping.

A growing number of projects and companies are using the Globus Toolkit to unlock the potential of grids for their cause.

2.12.1. Global Digital Library Grid (GDLG)

The data grid technology is used in GDLG to establish an integrated electronic library to link libraries, museums and archives to form an enormous virtual museum by the power of data grid to share the resources among different digital institutions and allow end users to have instant access to any and all integrated resources (Chao-Tung, 2005).

The GDLG project use Globus Data Grid which comes with two layers namely data grid core services and high level components. The lower layer are Data Grid Core Services, and the upper layer are High Level Components. See Figure 2.6.

![Figure 2.6: Globus Toolkit Architecture](image-url)
The strengths of this system are as follows:

- The dispersed resources are linked together by data grid.
- By using one single interface, the end users are able to search for files which are dispersed between different digital libraries.
- Once more and more digital libraries and museums becomes members of this grid and the grid grows, there will be more resources integrated and shared among themselves.
- More information are organized in one enormous virtual environment to meet reader needs.

This GDLG project is similar to the E-TD Repository System as both of these systems make use of data grid technology. But, the GDLG project uses Globus toolkit whereas SRB data grid middleware is used in the E-TD Repository System. Both systems have the same function like allowing end-users to search for resources dispersed in a heterogeneous environment by using one single interface.

2.13. Storage Resource Broker

The Storage Resource Broker (SRB) is a client-server middleware developed by San Diego Supercomputer Centre (SDSC) (San Diego Supercomputer Centre, 2006a). The SRB has been used in production since 1997 (NIEeS, 2007). It is finding applications in a number of grid-computing and data management projects across the range of physical and biological/medical sciences.
2.13.1. SRB Architecture

The architecture consists of three components: the metadata catalog (MCAT) service, SRB servers and SRB clients connected to each other via network.

![SRB Architecture Diagram]

2.13.2. Metadata Catalog (MCAT)

The **MCAT** stores metadata associated with data sets, users and resources managed by the SRB. The MCAT server handles requests from the SRB servers. These requests include information queries as well as instructions for metadata creation and update (San Diego Supercomputer Center, 2006).

MCAT consists of two components:

a) MCAT Database

   i) Is a metadata repository.

   ii) It stores information used by the SRB system which includes:

      - **Internal system data** required for running the system
      - **Application (user) metadata** regarding data sets being brokered by SRB.
iii) It also maps between a file’s SRB address, and its storage location.

b) MCAT SRB Server (MES-MCAT Enabled Server)

i) At least one SRB Server of a zone must be installed on a node that can access the MCAT database.

ii) Work directly against the MCAT database to provide SRB Services.

iii) All other SRB Servers interact through MES as it has additional software and database scripts over other SRB servers.

2.13.3. SRB Server

The **SRB server** is responsible for carrying out tasks to satisfy the client requests. It handles complex tasks that include interacting with various types of storage system and OS/hardware architecture, interacting with the MCAT service and performing I/O on behalf of the clients (San Diego Supercomputer Center, 2006).

User Info, Zone Info, Location and Resource (logical resource or physical resource) can be managed by using SRB Admin Tool. See Figure 2.8.

![SRB Admin Tool](image)

*Figure 2.8: Java SRB Admin Tool (SAT)*
2.13.4. SRB Client

The **SRB Client** uses a set of APIs for sending requests and receiving response to/from the SRB server to access every storage systems managed by the SRB.

2.13.5. Replicated Catalog

![Replicated Catalog Diagram]

**Figure 2.9: Replicated Catalog**

The whole system behaves as a single zone with replicated MCATs even though there are multiple MCATs operating in distinct zones (Wan et al, 2003). The metadata about users, resources, collections, containers and data objects are all synchronize between all MCATs, so that each contains the same information as any of its sister MCATs. An object created in one zone is registered as an object in all other sister zones and any associated metadata is also replicated. Hence, the view from every zone is the same. This provides a completely replicated system which has a high degree of fault-tolerance for MCAT failures (Wan et al, 2003). The user will not miss any access to data even if their local MCAT becomes non-functional.
2.13.6. Organizations using SRB

The SRB is being used by the following organizations (Kenneth):

- **Digital Libraries**
  - UCB, Umich, UCSB, Stanford, CDL
  - National Science Foundation National Science Digital Library
    (NSF NSDL) - UCAR / DLESE

- **NASA Information Power Grid**

- **Astronomy**
  - National Virtual Observatory
  - 2MASS Project (2 Micron All Sky Survey)

- **Particle Physics**
  - Particle Physics Data Grid (DOE)
  - GriPhyN

- **Medicine**
  - Digital Embryo (NLM)

- **Earth Systems Sciences**
  - ESIPS
  - LTER

- **Persistent Archives**
  - NARA
  - LOC
As the E-TD system is part of Digital Library (DL), NSF NSDL is reviewed among the organizations which are using SRB data grid.

2.13.7. National Science Foundation (NSF) National Science Digital Library Program

The National Science Digital Library (NSDL) was built since 1994 and the NSF began the NSDL pilot program in 2000 and the full program was launched in 2001 with projects in core integration, collections, services, and targeted research (Laura et al., 2002). These efforts have led to a description of the digital library as a learning environments and resources network for science, mathematics, engineering, and technology education (Zia, 2001).

The aim of building NSDL project is to construct the premier portal to preserve the current and future high-quality educational and research content in digital format for the learners and researchers. Moreover, it is managed actively for the users to access quality collections and services reliably from anywhere at anytime. The NSDL is available both within and without the network of NSF (Zia, 2001).
2.14. Ontology and Semantic Web

Web was originally invented by Tim Berners-Lee (Sean). The Semantic Web is not a separate Web but an extension of the current one, in which information is given precise meaning and enabling computers and people to work in collaboration (Berners-Lee et al., 2001). The Semantic Web will be as decentralized as possible. The properly designed Semantic Web can assist the evolution of human knowledge as a whole.

The Semantic Web relies heavily on the formal ontologies for the purpose of comprehensive and transportable machine understanding. Therefore, the success of the Semantic Web depends strongly on the increase of ontologies. Ontologies are metadata schemas, providing controlled vocabulary of concepts, each with an explicitly defined and machine processable semantics (Maedche and Staab, 2001).

Ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary (Uschold and Gruninger, 1996). Thus, ontology can be used in the communications between people and information systems.

An ontology is a “shared specification of a conceptualisation” (Gruber, 1993). Ontologies are designed for enabling knowledge sharing and reuse. Ontology provides a structured way of describing knowledge. Ontologies can be seen as special kinds of graphs describing the entities that exist in a domain, their properties and the relations between them (Agissilaos, 2005).

The most important types of relations in ontology are is-a and part-of relationships. Most ontologies include is-a relationship (i.e. “motorbike is a vehicle”) and part-of relationship (i.e., “a wheel is-part of a motorbike”). Although there are other kinds of relationships available, these two are mostly used in ontologies as it can
be used to create hierarchies of concepts in a domain. Ontologies can enhance the functioning of the Web in many ways. They can be used in a simple fashion to improve the accuracy of Web searches, the search program can look for only those pages that refer to a precise concept instead of all the ones using ambiguous keywords.

The benefit of using ontology in a system is that ontology provides a standard specification of concepts in the specific domain. Although ontology has been an active research area in the past decade, there is no ontology reported in the literature for software testing domain.

2.15. Systems using ontology

This section describes several systems that make use of ontology.

2.15.1. Semantic Digital Library Services (SDLS)

The ontology used in SDLS is discussed in this section. This semantic digital library service aims to bring semantic retrieval in digital libraries using semantic web technologies with the key tasks such as semantic discovery, ontology-based annotation, semantic mediation and context matching (Guha et al., Year). The domain of agriculture is chosen for implementing these services among various digital repositories of agriculture domain. Various agriculture metadata schemas and thesaurus has been studies and are used by this system.
2.15.2. JeromeDL - Semantic interoperability in Digital Libraries

JeromeDL project is implemented to deploy SemanticWeb technology for user management and search (Ding et al.). JeromeDL project has made an effort to bring semantic interoperability among the digital repositories using different bibliographic metadata standards like Dublin Core, BibTeX, MARC21, etc. JeromeDL has been designed with Semantic Web technologies like Resource Description Framework (RDF), Friend-of-a-Friend (FOAF), and ontologies. The features of JeromeDLs middleware are viewing resources, searching and browsing, users’ profile management and resources management. JeromeDL is implemented in Java and available as open-source licence and it is continually enhanced with semantic features. The database of JeromeDL prototype has been filled with 100 resources provided MARC21, BibTeX and semantic descriptions. The JeromeDL also provides searching service with and without query expansion with semantics.

In JeromeDL project, the DublinCore Metadata is modified to include additional definitions of keywords and catalog classifications of resource. The search algorithm of JeromeDL consists of three major steps (Ding et al., Year):

Step 1: Fulltext index search on the resources’ contents and users’ annotations on resources.

Step 2: Bibliographic description search consisting of MARC21 and BibTeX formats.

Step 3: User-oriented search with semantics, based on the semantic description of the resources and information about most interested categories (regarding the user that issued the query).
2.15.3. Swoogle: A Semantic Web Search and Metadata Engine

Swoogle is a crawler-based indexing and retrieval system developed using the
JENA for Semantic Web Documents (SWDs) written in RDF or Ontology Web
Language (OWL) (Sebastian et al., 2005). JENA is a Java framework for building
Semantic Web applications. It provides a programmatic environment for RDF and
OWL and includes a rule-based inference engine. Swoogle is an ongoing project. The
metadata of Swoogle is stored in MySQL database and indexed about 11,000 SWDs.
Please refer to *Figure 2.10* and *Figure 2.11*.

![Swoogle Advanced Semantic Search Interface](image)

*Figure 2.10: Swoogle Advanced Semantic Search Interface*
Figure 2.11: Swoogle Query Result
The three systems reviewed in section 2.15.1, section 2.15.2 and section 2.15.3 and this E-TD repository system are compared in the following table. Please refer to *Table 2.5.*

**Table 2.5: Comparison between SDLS, JeromeDL, Swoogle and E-TD Repository System**

<table>
<thead>
<tr>
<th>No.</th>
<th>Compared By</th>
<th>SDLS</th>
<th>JeromeDL</th>
<th>Swoogle</th>
<th>ETD Repository System</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Domain of Ontology</td>
<td>Agriculture</td>
<td>-</td>
<td>-</td>
<td>Software Testing</td>
</tr>
<tr>
<td>2.</td>
<td>Language used in implementing ontology</td>
<td>-</td>
<td>Java</td>
<td>Jena</td>
<td>C language</td>
</tr>
<tr>
<td>3.</td>
<td>Metadata Standard</td>
<td>-</td>
<td>Dublin Core, BibTeX, MARC21, etc</td>
<td>-</td>
<td>Dublin Core Metadata Standard</td>
</tr>
</tbody>
</table>

**2.16. Summary**

As a conclusion, the existing E-TD systems, the metadata standards, file format used by E-TD systems, the Globus and SRB Data Grid, Ontology and Semantic Web and the different types of life cycle models are reviewed in this chapter.
Chapter 3: Methodology

This chapter discusses the steps involved in carrying out this research. The lifecycle model used for the development of the E-TD system is also discussed in detail in this chapter.

3.1. Research Methodology

The methodology defines what the activity of the study is, how to proceed, how to measure progress and what constitutes success (Patton, 1990). A series of procedures and the techniques used in this study is described in sequential steps. Please refer to Figure 3.1.

1. Gather Information: The related articles regarding data grid technology, Dublin Core metadata standard, semantic web and ontology, existing E-TD systems and other similar research papers are reviewed to understand the background knowledge of E-TD systems, advantage of using data grid technology and using semantic web techniques. Other existing E-TD systems are also reviewed to identify the limitation of E-TD systems without data grid and the weaknesses of the conventional searching services used in existing E-TD systems.

By reviewing a couple of existing E-TD systems, it is found that there is no E-TD system implemented using data grid technology which allows to link different database schemas and allows users to access location-independent data. Without data grid, it is not possible to access data stored in the different database schema by using one user interface.
In the conventional searching technique of other existing E-TD systems, it is time consuming in searching all semantically related E-TDs. This is because there is no ontology searching technique.

2. **Review and choose a suitable lifecycle model for development:** The V-Model and Rapid Application Development (RAD) Model are reviewed and RAD model was chosen for this system as it allows iteration and self-correction.

3. **Define functional and non-functional requirements of the system:** System analysis is concerned with discovering what the new system is required to do. Requirements are set of functionalities and constraints that the end-user expects from the system. All possible requirements to be developed for this E-TD system are captured in this step.

4. **Develop Logical System:** The Structured System Analysis and Design Methodology (SSADM) is the common structured analysis methodology which is used to develop the logical system of this project. The logical design of the E-TD system such as Data Flow Diagram, Entity Relationship Diagram and Entity Life History are produced in this stage.

5. **Define ontology of software testing terminologies:** The software testing terms are first categorized based on the definitions written in the paper named “the Standard glossary of terms used in Software Testing”. Then, the ontology of software testing terminology is defined in this stage.

6. **Design E-TD system:** It is important to define what are going to be created and what the E-TD system should look like. Before the implementation is started, the architecture of the E-TD system, database design for ontology and interface representation of the E-TD system are defined.
7. **Setup and configure SRB data grid:** The *Oracle Database 10g* and *Postgresql database* are installed prior to SRB configuration. After creating the empty database, the SRB server (version 3.4.2) is downloaded from SRB website and installed and configured on Linux platform by using the SRB installation notes.

8. **Implement E-TD system:** The work is divided into small modules/units. Each of these modules, functions and the user interface are implemented using HTML and C language. The C Client APIs are used as intermediary to communicate with SRB data grid. The ontology database is created using Postgresql RDBMS and ontology search engine is developed using C language.

9. **Test E-TD System:** Initially, each and every unit is tested individually to verify that each unit meets its specification. Then, the individual program units or modules are integrated and tested to ensure that all the modules/units coordinate between each other. Finally, the system is tested to ensure that it meets its requirements as a whole.

10. **E-TD system is produced:** Finally, the E-TD system is produced with all the functional and non-functional requirements defined in stage 4. The integration of data grid technology to this E-TD system has a high degree of fault-tolerance, the availability of resources to the user is higher and the system is more reliable as the metadata in the MCAT is replicated. Even when one MCAT or SRB server is unavailable, the other MCAT will take over to fulfill the client request. The incorporation of ontology-based searching technique improves the search results of the E-TD system.
1. Gather Information
(Review data grid technologies, ontology, and existing E-TD systems)

2. Review and choose a suitable life cycle model for development
(Review the related life cycle models and RAD model has been chosen)

3. Define functional and non-functional requirements of the system

4. Develop Logical System
(Produce DFD, ERD and ELH of the E-TD Repository System)

5. Define ontology of software testing terminologies
(Categorize software testing terms and define ontology of software testing)

6. Design E-TD system
(Produce E-TD architecture, design database to capture the ontology of software testing terms, and define the E-TD interface design)

7. Set up and configure SRB data grid
(Download SRB server software form SRB website and install and configure SRB data grid)

8. Implement E-TD system
(Implement user interface using HTML and C language and create ontology database using Postgresql RDBMS and implement ontology search engine using C language)

9. Test E-TD system
(Each unit is tested; all the units are integrated and tested again, test the whole system using test cases and the system is tested by end-users for acceptance testing)

10. E-TD system is produced

Figure 3.1: Activities taken in carrying out this E-TD System
3.2. Relevant Lifecycle Models

The software development life cycle (SDLC) is a framework for understanding and developing information systems and software successfully (Bentley, 2001). The V-Model and Rapid Application Development (RAD) Model are suitable for the development of this system and therefore reviewed here together with their own strengths and weaknesses.

3.2.1. V-Model

The V model rectifies some, but not all, of the problems that are noted with the waterfall model. The immediately apparent difference is that products are identified at the end of each stage. This is a significant improvement. It shows how there is a relationship between the products at the early stages of the life cycle and the later stages. The V model lines up these products on either side of the V making the relationships clear (Bentley, 2001).

![Figure 3.2: The V Model (Bentley, 2001)]
Advantages:

- Simple and easy to use.
- Explicitly illustrates the output from each phase.
- Higher chance of success over the waterfall model due to the development of test plans early on during the life cycle.

Disadvantages:

- Does not work well for big projects.
- No early prototypes are available as the software is developed during the implementation phase.

3.2.2. Rapid Application Development Model

The Rapid Application Development (RAD) model is a more radical approach that has gained a lot of strength over the past ten years. It started from a realization that development does not need to proceed in the ‘big bang’ waterfall style. It is better to produce some of the system as early (as rapidly) as possible (Bentley, 2001).

Along with this idea of iterating through the design, code and test cycle many times is the view that prototyping is essential to obtain valid systems that will be effective and that will be owned by business users.
Advantages:

- Working model of the product (prototype) can be seen quickly;
- The product is being evaluated with end-user before the next iteration starts;
- Early involvement of the end-user.

Disadvantages:

- Require large development team members for large scalable project.

3.3. Life Cycle Model for the E-TD Repository System

The RAD Model has been chosen as a life cycle model to be used throughout the development of this system. Iteration allows for effectiveness and self-correction. Normally, human beings almost never perform a complex task correctly at the first time.
However, people are extremely good at making an adequate beginning and then making many small refinements and improvements (Steve, 1996).

### 3.4. Summary

The research methodology in conducting this E-TD system is defined in this chapter. The steps involved in this study were explained. This chapter also reveals the strengths and weaknesses of each V model and RAD model, which is important in making a justification in choosing suitable software process model for this system. As a result, the RAD Model has been chosen as a guideline throughout the development of this E-TD repository system.
Chapter 4: System Analysis

System analysis refers to the process of investigation and analysis of the system to be built. A detailed analysis of the existing system must form a basis for the requirements of the new system. Functional and non-functional requirements are the most important in the definition of the requirements of a new system. System analysis is the vital part of the whole development life cycle process (Benyon-Davies, 2002). The logical system of the product is developed in this phase.

4.1. Requirements of E-TD Repository System

Generally, all the existing E-TD systems have some common functions such as uploading E-TDs, inserting metadata, viewing and downloading E-TDs, theses searching and browsing. All these common functions are built in this E-TD system as well. Moreover, the data grid technology is being used in this E-TD system and metadata in one MCAT is replicated in another MCAT of different SRB server. Therefore, this E-TD system has a high degree of fault-tolerance for MCAT failures. With the replicated MCAT, the reliability of the system and availability of resources (E-TDs) is also improved. Compared to the existing E-TD systems, the additional modules or functions in this project are ontology-based searching and adding ontology.
4.2. Functional requirements

These are statements of services the system should provide, how the system should react to particular inputs and how the system should behave in particular situations. Functional requirements describes an interaction between the system and it’s environment. In some cases, the functional requirements may also explicitly state what the system should not do (Sommerville, 1996 & Pressman 2001).

The followings are the functional requirements of the ETD system developed in this study:

4.2.1. Access Authentication

Access authentication serves as a security measure in this ETD system. The access authentication module provides separate access types for members of the system and guests. A login authentication is required to make sure unauthorized users cannot access the restricted areas of the system.

The login module is for user authentication to prevent any unauthorized access to the system. The restricted areas of the system such as uploading thesis, viewing the body of theses, downloading theses, inserting metadata, modifying metadata, changing access constraint of theses, changing password, registering new user, adding ontology to the system, viewing the member information and viewing relationships in ontology are allowed to be accessed by the staffs, students and admin. The adding ontology module is available only to the E-TD system admin and uploading thesis and inserting metadata for the thesis are allowed only to the clerical staff. The changing access constraint module is available to the admin, clerical staff and supervisors.

Only authorized users such as students, clerical staff, supervisors, moderators and admin are required to enter their username and password to login to the system.
Other users can access the system by using “guest” account. The guest users can browse and search E-TDs in the collection. They are allowed to view the metadata and abstract of E-TDs in the collection. However, the guest users are neither allowed to view the body of E-TDs in the collection nor download E-TDs from the system.

4.2.2. Membership registration

Not all the modules in this E-TD system are available to all the users. The students and staffs must request the new account from the E-TD system admin to access a certain part of this E-TD system such as uploading thesis, viewing the body of theses, downloading theses, inserting metadata, modifying metadata, changing access constraint of theses, changing password, registering new user, adding ontology to the system, viewing the member information and viewing relationships in ontology.

4.2.3. Thesis Submission

This module is to upload E-TDs to this system. The E-TDs can be submitted in either Microsoft Word (.doc) file format or Adobe Portable Document Format (.pdf) or compressed file (.zip). After uploading a thesis to this system, it is required to key in the metadata of the thesis to this system and change the access constraint of the thesis. All these tasks are done by the clerical staff.

4.2.3.1. Inserting Metadata

This module is to capture the metadata of E-TDs. The 14 elements of metadata captured by this system are: Title, Author, Matric No., Program, Subject Code, First Supervisor, Second Supervisor, Year, Language, Format, Publisher, Rights, Abstract and Keyword.
4.2.3.2. Access Constraints

When the E-TDs are first uploaded to the system, the *all* access is given only to the person (i.e., clerical staff) who uploaded the E-TDs. He/She needs to change the access constraint of E-TDs to *all* access for supervisor and admin. Once the E-TD is viewed and approved to make it public, the supervisor or admin will change the access constraint of the E-TD to *read* access for all the students and other members to let them to view and/or download the body of the E-TD.

In access constraint module, the *all* access means all the operations such as view or download the body of the thesis file, modify the metadata of the thesis file, remove the thesis file from the system and change the access constraint of the thesis file is allowed. The *read* access means only viewing and downloading of the thesis is allowed. The users who have *read* access cannot modify the metadata of E-TDs and remove the E-TDs from the system.

4.2.4. Viewing E-TDs Online

The E-TDs can be viewed online except the .zip file if the user has a valid account.

4.2.5. Downloading E-TDs

In this module, the authorized users can download E-TDs from this system.

4.2.6. Change password

The authorized users can change their password at any time by using the change password module.
4.2.7. Adding Ontology

This module is an additional module which is not presented in other E-TD systems. The E-TD administrator can add new ontologies to the system by using this module. There are 3 sub-modules under this module:

4.2.7.1. Add New Term

At first, the administrator needs to add all new terms and their definitions to the system.

4.2.7.2. Add Synonym

After adding new terms, the administrator must enter the synonym(s) of the term(s).

4.2.7.3. Add Relationship

The last task is to add the is-a/type-of relationship to the system.

4.2.8. Theses Searching

The Search Theses allows the users to formulate the search using different fields either by including or excluding terms. There are two types of searching mechanisms applied in this system. One is simple searching and the other one is ontology-based searching.

4.2.8.1. Simple Searching

In this module, users can search E-TDs by Title, Author, Program, Year, Language, Abstract and Keyword. The wildcard searching (‘*’) is also provided in this module.
4.2.8.2. Ontology-based Searching

Additionally, the searching service is also enhanced with ontology. The conventional searching technique of existing E-TD systems just provide direct keyword matching and/or keyword matching using wildcards ("*").

In this module, users still search E-TDs by keyword. However since the semantic searching technique is applied in this system, the ontology of the searched keyword is retrieved from the Ontology Database before searching E-TDs from the repository is started. Consequently, the result of the ontology-based searching of E-TDs is more comprehensive (more documents are returned) than the result returned from a simple search engine of other existing E-TD systems.

4.2.9. Thesis Browsing

Anyone can browse all the theses stored in the E-TD repository or browse theses by program or subject code or language or year.

4.3. Non-functional requirements

There are constraints on the services or functions offered by the system. Non-functional requirements are often harder to deal with than functional ones, because their impact is generally not localized to any particular module of the system, but cuts across the whole system (Sommerville, 1996 & Pressman 2001).

The followings are the non-functional requirements of this E-TD system:

4.3.1. Modularity

Modularity means breaking down a large or complex system into smaller or more manageable program modules so that distinct functions of objects could be
isolated from one another. This characteristic makes the testing and maintenance much easier. In this E-TD system, modularity is applied from the beginning, as this will lead to easy modification (the functional independence of program components) and enhancement in the future.

4.3.2. Usability

Usability is the ease of use and user friendliness of a software product. This E-TD system is developed with a well-formed graphical user interface to make the users easy to interact with the system.

4.3.3. Robustness

Robustness refers to the quality of the system that is able to handle or at least avoid disaster when faced with unexpected circumstances such as when giving the improper or invalid data. An error message will be prompted when an error is detected by this E-TD system. It behaves “reasonably” even in unforeseen circumstances (e.g., incorrect input).

4.3.4. Maintainability and Expandability

A system is maintainable if the program modules can be easily modified and tested in the case of updating the modules to meet new requirements. It also can be understood, adapted, corrected, enhanced or move to a different computer system.

Expandability is the degree to which architectures, data or procedural design can be extended. The E-TD system should be able to be extended to accommodate more functionality in the future. This will allow the progression and advancement of technology to take part in the future of the system.
4.3.5. Security

Security is an important feature of any system to resist unauthorized usage of the critical information inside the database. Every system should have security feature in order to prevent unauthorized access to data and to diagnose accidental corruption of program and/or data. Therefore, some restricted areas of this E-TD system can only be accessed by authorized users.

4.3.6. Reliability and Availability

With the use of data grid technology in this research, the MCAT of one SRB server is replicated in different storage systems (Postgresql RDBMS) under the control of different SRB server to provide load balancing. Therefore, this E-TD system is more reliable and the probability of the availability of E-TDs to the user is higher compared to other E-TD systems.

4.3.7. Fault tolerance

When the first MCAT server is unavailable, the system will be automatically redirecting access to a replica MCAT on different SRB server. Hence, users will not miss any access to E-TDs in this system. This function is not provided in other E-TD systems.
4.4. Structured Systems Analysis and Design Methodology (SSADM)

The SSADM is the most highly developed conventional method for system analysis and design. It is used in the development of systems, but it does not cover the entire system life cycle. But it has advantages over the other methods available. One of the main advantages is that SSADM builds up several different views of the system which are used to cross-check one another. Another advantage of SSADM over a number of methods is that it combines techniques for the analyst; it gives guidance on how and when to use them (John, 1996 & Benyon-Davies, 2002).

SSADM’s objectives are to (John, 1996 & Benyon-Davies, 2002):

- Improve project management and control;
- Make more effective use of experienced and inexperienced development staff;
- Develop better quality systems;
- Make projects resilient to the loss of staff;
- Enable projects to be supported by computer-based tools such as computer-aided software engineering systems.

It uses a range of techniques to specify system requirements and to design an appropriate IT solution. It is a well-proven methodology that is freely available and openly documented.
4.5. Major Tools of SSADM

The diagrammatic tools of SSADM are (Bentley, 1996):

- Logical Data Modeling
- Data Flow Modeling
- Entity Event Modeling

4.5.1. Logical Data Modeling

Logical Data Modeling is at the heart of SSADM. This modeling technique is used to show what information should be held by the system, the entities or things that are of importance in a system and how these entities related to each other. This model describes the complete picture of the data used and stored by the system. (John, 1996 & Benyon-Davies, 2002)

A student has 1 thesis or more and a thesis is owned by 1 student. 1 thesis has more than 1 metadata and the metadata is related to only one thesis. One synonym includes many terms and one term is included in many synonyms. One ontology includes many terms and one term is included in many ontologies. This is shown in Figure 4.1.

![Entity Relationship Diagram of E-TD System](image)

Figure 4.1: Entity Relationship Diagram of E-TD System
4.5.2. Data Flow Modeling

The data flow model is used for representing the information flows of a system. Data flow diagram is a fundamental building block of a lot of analysis and design methodologies. It describes how the data flows between elements in a system, between the system and external entities and where data is stored. The DFD starts from a high level diagram (context diagram) decomposing to many lower-level diagrams. (John, 1996 & Benyon-Davies, 2002)

There are different sets of symbols available for data flow diagrams, depending on accepted conventions. The DFD symbols used in this study is from Business System Analysis book of National Computing Centre (NCC) Education.

<table>
<thead>
<tr>
<th>Symbols</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="User.png" alt="User" /></td>
<td>External Entity</td>
</tr>
<tr>
<td><img src="Registration.png" alt="Registration" /></td>
<td>Process</td>
</tr>
<tr>
<td>![D1 User](D1 User.png)</td>
<td>Data Store</td>
</tr>
<tr>
<td>![User Info](User Info.png)</td>
<td>Data Flow</td>
</tr>
</tbody>
</table>

*Table 4.1: Symbols of Data Flow Diagram (Benyon-Davies, 2002)*

The Level 0 DFD for E-TD Repository System is shown in Figure 4.2. The E-TD Admin provides User Details to the system for new user registration. The system will display Success Info to the admin once the registration is done.

The Clerical Staff upload Thesis File to the system and key in the Metadata
Info. The system gives the Success Info to the Clerical Staff. He/She needs to change the access right of the thesis to “all” for Supervisor and ETD Admin. The Success Info of changing access right is given by the system. The Clerical Staff can give Browse Query/Search Query to the system and the system gives the Result back to the Clerical Staff. The Clerical Staff can View Thesis File or Download Thesis File from the system. The system gives the Metadata Info of the Thesis and the Thesis File back to the Clerical Staff.

Other entities can also be interpreted in the same way as the Clerical Staff entity.

Figure 4.2: Level 0 Data Flow Diagram for E-TD Repository System
The **Member** entity includes ETD Admin, Supervisors, Moderators, Clerical Staffs and Students. In the **Browsing Theses** process, when the process receives **Browse Query** from Member or Guest, it gets the **Thesis Info** from Thesis data store and **Metadata Info** from Metadata data store and returns this information back to the **Member** or **Guest**. When the entity with the same name is repeated in the diagram, the entity with a line at the top-left corner is used in all the repeated entities. Please refer to **Figure 4.3**.
Figure 4.3: Level 1 Data Flow Diagram for E-TD Repository System
Among the nine processes in the Level 1 DFD, process number 8 is decomposed into Level 2 DFD. Please refer to Figure 4.4. The Member entity includes anyone registered in this E-TD Repository System. There are four sub-processes under this Process.

A Member/Guest gives Search Query to the system. The Searching Term sub-process receives the query and retrieves Term Id of the searched keyword from the Term table and passes it to the Searching Synonym sub-process and Searching Ontology sub-process. Both of these sub-processes retrieves Synonym Info and Ontology Info from their related data stores and passes the Info to the Finding Theses sub-process which retrieves Metadata Info and Thesis Info from the data stores and returns the Ontology-based Searched Result to the Member and Guest user.

Figure 4.4: Level 2 Data Flow Diagram for Searching Theses by Ontology Process
4.5.3. Entity Event Modeling

The entity event modeling mainly based on an entity rather than the whole system. This model shows how entities change over their life. The sequence, selection and iteration of events affecting an entity are shown using a graphical notation (John, 1996 & Benyon-Davies, 2002). The ELH symbols used in this study is from Business System Analysis book of National Computing Centre (NCC) Education.

<table>
<thead>
<tr>
<th>Symbols</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Thesis is uploaded" /></td>
<td>Sequence: One action is followed by another in sequence</td>
</tr>
<tr>
<td><img src="image" alt="Thesis is browsed" /></td>
<td>Selection: The symbol ‘o’ in the top right-hand corner of a box denotes selection.</td>
</tr>
<tr>
<td><img src="image" alt="Thesis is browsed" /></td>
<td>Iteration: The symbol ‘*’ in the top right-hand corner of a box denotes iteration.</td>
</tr>
</tbody>
</table>

*Table 4.2: Symbols of Entity Life History (Benyon-Davies, 2002)*
**Member is registered** is the first event to have an effect on **Member**. Before the **Member is deleted** from the system, he/she can change password many times. Please refer to Figure 4.5.

![Figure 4.5: Member Entity Life History](image)

**Thesis is uploaded** is the first event to have an effect on **Thesis**. Before the **Thesis is removed** from the system, users can select any actions to be done on the **Thesis** entity. Thesis is browsed or searched or viewed or downloaded or change the access constraint of a thesis. These actions can be done many times. Thus, the iteration symbol ‘*’ is used at the corner of the box. Please refer to Figure 4.6.

![Figure 4.6: Thesis Entity Life History](image)
*Metadata is inserted* is the first event to have an effect on *Metadata*. Before the *Metadata is deleted* from the system, metadata is viewed or searched or updated many times. Please refer to *Figure 4.7*.

![Figure 4.7: Metadata Entity Life History](image)

*Term is inserted* is the first event to have an effect on *Term*. Before the *Term is deleted* from the system, *Term is retrieved* many times. Please refer to *Figure 4.8*.

![Figure 4.8: Term Entity Life History](image)
*Synonym is inserted* is the first event to have an effect on *Synonym*. Before the *Synonym is deleted* from the system, *Synonym is retrieved* many times. Please refer to Figure 4.9.

![Figure 4.9: Synonym Entity Life History](image)

*Ontology is inserted* is the first event to have an effect on *Ontology*. Before the *Ontology is deleted* from the system, *Ontology is retrieved* many times. Please refer to Figure 4.10.

![Figure 4.10: Ontology Entity Life History](image)
### 4.6. Dublin Core Metadata Elements of the E-TD System

The selected DC metadata elements with its own extended ones are used in this research. The followings are descriptions of the common Dublin Core metadata elements and the descriptions on how do these metadata elements relate to the proposed E-TD system.

**dc.title**

<table>
<thead>
<tr>
<th>Element</th>
<th>DC Description</th>
<th>E-TD Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc.title</td>
<td>A name given to the resource.</td>
<td>For this E-TD system, this is the title of the work as it appears on the title page or equivalent.</td>
</tr>
</tbody>
</table>

**dc.creator**

<table>
<thead>
<tr>
<th>Element</th>
<th>DC Description</th>
<th>E-TD Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc.creator</td>
<td>An entity primarily responsible for making the content of the resource.</td>
<td>For this E-TD system, this field is appropriate for the author(s) of the work.</td>
</tr>
</tbody>
</table>

**dc.subject**

<table>
<thead>
<tr>
<th>Element</th>
<th>DC Description</th>
<th>E-TD Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc.subject</td>
<td>The topic of the content of the resource.</td>
<td>For this E-TD system, this field is the subject code of the thesis.</td>
</tr>
</tbody>
</table>
### dc.description

<table>
<thead>
<tr>
<th>Element</th>
<th>DC Description</th>
<th>E-TD Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc.description</td>
<td>An account of the content of the resource.</td>
<td>For this E-TD system, this is the full text of the abstract.</td>
</tr>
</tbody>
</table>

### dc.publisher

<table>
<thead>
<tr>
<th>Element</th>
<th>DC Description</th>
<th>E-TD Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc.publisher</td>
<td>The institution archiving the work.</td>
<td>In this case, FCSIT.</td>
</tr>
</tbody>
</table>

### dc.contributor

<table>
<thead>
<tr>
<th>Element</th>
<th>DC Description</th>
<th>E-TD Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc.contributor</td>
<td>An entity responsible for making contributions to the content of the resource.</td>
<td>For this E-TD system, this is the supervisor (or) advisor of the work.</td>
</tr>
</tbody>
</table>

### dc.date

<table>
<thead>
<tr>
<th>Element</th>
<th>DC Description</th>
<th>E-TD Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc.date</td>
<td>A date associated with an event in the life cycle of the resource.</td>
<td>For this E-TD system, this should be the date that appears on the title page</td>
</tr>
</tbody>
</table>

### dc.format

<table>
<thead>
<tr>
<th>Element</th>
<th>DC Description</th>
<th>E-TD Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc.format</td>
<td>The physical or digital</td>
<td>In this case, the .doc, .pdf</td>
</tr>
</tbody>
</table>
manifestation of the resource. In the case of an electronic thesis or dissertation, this should contain a list of the electronic format(s) in which the work is stored and/or delivered.

<table>
<thead>
<tr>
<th>dc.language</th>
<th>DC Description</th>
<th>E-TD Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc.language</td>
<td>A language of the intellectual content of the resource. This should be the primary language in which the work is recorded.</td>
<td>In this case, English and Malay.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dc.rights</th>
<th>DC Description</th>
<th>E-TD Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc.rights</td>
<td>Information about rights held in and over the resource. Typically, this describes the conditions under which the work may be distributed, reproduced, etc., how these conditions may change over time, and whom to contact regarding the copyright of the work.</td>
<td>In this case, University Malaya.</td>
</tr>
</tbody>
</table>
4.7. Extending Dublin Core Metadata Standard for this E-TD System

The Dublin Core metadata standard already contains relevant elements like Title, Creator, Description, etc. However, these elements give basic details about the context of a thesis or dissertation. Two extra metadata elements are added to the DC metadata elements for this E-TD system to enable searching for the E-TDs by student’s matriculation number and by the name of the degree program. The followings are the extended metadata elements and the screenshots of using the two added metadata elements in this E-TD system. Please refer to Figure 4.11, Figure 4.12, Figure 4.13 and Figure 4.14.

thesis.program

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>thesis.program</td>
<td>Name of the degree associated with the work as it appears within the work. For this study, Masters of Computer Science or Bachelor of Computer Science.</td>
</tr>
</tbody>
</table>
Figure 4.11: Browse by Program

Figure 4.12: List of E-TDs by Program
thesis.maticno

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>thesis.maticno</td>
<td>The registration number of a student.</td>
</tr>
</tbody>
</table>

Figure 4.13: Search by Matric Card Number
4.8. Summary

The Entity Relationship Diagram (ERD), Data Flow Diagram (DFD), and Entity Life History (ELH) have been used for the analysis of this ETD system. The analysis diagrams used for this project are described in this chapter.
Chapter 5: System Design

System Design is a process through which requirements are translated into a modular representation of the software that can be assessed for quality before coding started (Bentley, 2001). System design is the only way to translate users’ requirement into a finished system accordingly.

The discussion of this chapter is mainly on the architecture design, database design and User Interface (UI) design of E-TD system.

5.1. Overview of E-TD Architecture

The architecture of this E-TD system is derived according to the functional and non-functional requirements of the system. This E-TD system is based on client-server architecture and allows users to access location-independent data: the user does not need to know where files are located and how they are stored. This system consists of Presentation Layer, Application Layer, Data Grid Layer and Storage System. The architecture is shown in Figure 5.1.

5.1.1. Presentation Layer

This layer consists of user interface of the E-TD system represented to the users who can interact with the system through this. It includes web browser which is the intermediary which accepts the input from the user and send it to the SRB via C Client APIs and get the result back from the system to the user.
5.1.2. Application Layer

This layer consists of the client application of the E-TD system which uses a set of C Client APIs (C Client Application Programming Interfaces) to communicate with the SRB Master. The client library sends requests using pre-defined request stubs to the SRB agent, and receives and parses replies from the SRB agent to the Web Browser.

5.1.3. The Data Grid Layer

There are two Metadata Enabled Servers in this layer. One server is named as fsktm.geranium2.um.edu.my and the other one is fsktm.geranium3.um.edu.my. The two MCATs of the two hosts synchronize metadata between each other. The synchronization process is done using the Szonesync.pl script in each machine. The script should be periodically run by the administrator to keep the metadata in the MCAT of local zone up-to-date with the MCAT in another zone in the Federation. Please refer to Appendix A to see the federation of the two SRB servers and the synchronization process.

Each of these host consist of two separate components, SRBMaster and SRBServer (i.e., SRB agent). The process of SRBMaster is to listen on a well-known port for connection request from client. Once the connection from the client is established, the SRBMaster launches a new SRBServer process on another port to service the connection. The SRB server is responsible for carrying out tasks to satisfy the client requests.

The Metadata Catalog (MCAT), the heart of SRB, is a database system to store metadata associated with data sets, users and resources brokered by the SRB. By querying MCAT, clients can easily find distributed data objects, replicate, transfer or synchronize data, perform sophisticated queries, and many other functions.
5.1.4. The Storage System

The two types of RDBMS, the Oracle and Postgresql, are installed in the SRB server named fsktm.geranium2.um.edu.my. The Oracle Database is to keep the metadata of the MCAT of fsktm.geranium2.um.edu.my and the Postgresql Database is to store the ontology of software testing terminologies.

However, only Postgresql Database is installed in the host named fsktm.geranium3.um.edu.my. This database keeps the metadata of the MCAT of this machine.
Figure 5.1: Components of E-TD Architecture
5.2. Ontology of Software Testing

The software testing terms used in this study are from Standard glossary of terms used in Software Testing (Erik, 2006). The terms are categorized based on the definitions in the glossary.

The term informal review is a synonym of ad hoc review or vice versa. The informal review, formal review and inspection has is-a relationship to review. Please refer to Figure 5.2 and Figure 5.3. Other terms in the figures can be interpreted in the same way. All the is-a relationships between terms are to be interpreted in top-down direction.
Figure 5.2: Ontology of Software Testing Terms
Figure 5.3: Ontology of Software Testing Terms

(Test Design Technique)
5.3. Database Design

The database design refers to the tables, columns, relationships, keys and indexes of which a database is comprised (Benyon-Davies, 2002a).

The following is the database design for the Ontology searching of this E-TD system. Please see Appendix B to view other tables of the MCAT which is used in this system.

<table>
<thead>
<tr>
<th>Table Name: Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field Name</td>
</tr>
<tr>
<td>term_id</td>
</tr>
<tr>
<td>term</td>
</tr>
<tr>
<td>def</td>
</tr>
</tbody>
</table>

*Table 5.1: Term Table*

The Term table (*Table 5.1*) stores all the software testing terms and definition of each term. The term_id is the primary key of this table. Some sample rows in Term table are shown in *Table 5.2*.

<table>
<thead>
<tr>
<th>Term_id</th>
<th>Term</th>
<th>Def</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>component testing</td>
<td>The testing of individual software components.</td>
</tr>
<tr>
<td>2</td>
<td>program testing</td>
<td>See component testing.</td>
</tr>
<tr>
<td>3</td>
<td>unit testing</td>
<td>See component testing.</td>
</tr>
<tr>
<td>4</td>
<td>black-box testing</td>
<td>Testing, either functional or non-</td>
</tr>
</tbody>
</table>
functional, without reference to the internal structure of the component or system.

<table>
<thead>
<tr>
<th>5</th>
<th>integration technique</th>
<th>Testing performed to expose defects in the interfaces and in the interactions between integrated components or systems. See also component integration testing, system integration testing.</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>system testing</td>
<td>The process of testing an integrated system to verify that it meets specified requirements.</td>
</tr>
<tr>
<td>7</td>
<td>testing</td>
<td>The process consisting of all life cycle activities, both static and dynamic, concerned with planning, preparation and evaluation of software products and related work products to determine that they satisfy specified requirements, to demonstrate that they are fit for purpose and to detect defects.</td>
</tr>
<tr>
<td>Field Name</td>
<td>Type</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>syn_id</td>
<td>serial</td>
<td>Primary Key</td>
</tr>
<tr>
<td>term_id</td>
<td>varchar (50)</td>
<td>Foreign Key</td>
</tr>
<tr>
<td>synonym</td>
<td>varchar (50)</td>
<td>Not Null</td>
</tr>
</tbody>
</table>

Table 5.3: Synonym Table

The Syn table stores all the synonyms of each term in the Term table. The syn_id is the primary key and term_id is the foreign key of this table. Some of the rows stored in this database are shown in Table 5.4.

<table>
<thead>
<tr>
<th>syn_id</th>
<th>term_id</th>
<th>synonym</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>component testing</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>program testing</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>component testing</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>unit testing</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>program testing</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>unit testing</td>
</tr>
</tbody>
</table>

Table 5.4: Sample rows of Synonym table

The component testing and program testing are synonyms of term_id: 3 (i.e., unit testing). The component testing and unit testing are synonyms of term_id: 2 (i.e., program testing). The program testing and unit testing are synonyms of term_id: 1 (i.e., component testing). Please refer to Figure 5.4.
Table Name: Ontology

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ontology_id</td>
<td>Serial</td>
<td>Primary Key</td>
</tr>
<tr>
<td>term_id</td>
<td>varchar (50)</td>
<td>Foreign Key</td>
</tr>
<tr>
<td>ontology</td>
<td>varchar (50)</td>
<td>Not Null</td>
</tr>
</tbody>
</table>

Table 5.5: Ontology Table

The Ontology table stores all the ontologies of each term in the Term table. The ontology_id is the primary key and term_id is the foreign key of this table. Some of the rows stored in this database are shown in Table 5.6.

<table>
<thead>
<tr>
<th>ontology_id</th>
<th>term_id</th>
<th>ontology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>unit testing</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>system testing</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Testing</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>Testing</td>
</tr>
</tbody>
</table>

Table 5.6: Sample rows of Ontology Table
The term_id:4 (i.e., black-box testing) is-a unit testing.

The term_id:4 (i.e., black-box testing) is-a system testing.

Again, term_id:3 (i.e., unit testing) is-a testing.

The term_id:6 (i.e., system testing) is-a testing.

Please refer to Figure 5.5.

![Diagram of testing types]

*Figure 5.5: Some of the terms of ontology in software testing*

### 5.4. Interface Design

The user interface design creates an effective communication medium between the end-users and the system. An interface which is difficult to use will result in a high level of user errors. Therefore, usability is vital role in designing the user interface. Usable websites are logical, consistent and clear to people who use them.

The interface design is the representation of the system. The table format is used to make the interface design of this E-TD system to be convenient and consistent. The interface design of this system is divided into 3 parts, which is the banner part, menu part and content part. Figure 5.4. is the example of the user interface screen design.
Figure 5.6: Basic Interface Design

The following screen design is the **Member Login Page** of this system which has the main banner at the top, navigation bar on the left and content of the page right below the main banner.

Figure 5.7: Member Login Page of E-TD Repository System
The Figure 5.8 and Figure 5.9 are Insert Metadata Page and Searching Theses Page of the system. After login to this system, the navigation bar is not necessary to display to the user so that these pages contain only the main banner and the content.

**Figure 5.8: Theses Searching Page of E-TD Repository System**

**Figure 5.9: Insert Metadata Page of E-TD Repository System**
5.5. User Interface Design Principles

There are some general UI design principles which are applicable to all user interface design (Sommerville, 1996). These principles are also considered when designing the interface of this E-TD system.

<table>
<thead>
<tr>
<th>Principle</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>User familiarity</td>
<td>The interface should use terms and concepts which are drawn from the experience of the anticipated class of user.</td>
</tr>
<tr>
<td>Consistency</td>
<td>The interface should be consistent in that comparable operations should be activated in the same way.</td>
</tr>
<tr>
<td>Minimal surprise</td>
<td>Users should never be surprised by the behaviour of a system.</td>
</tr>
<tr>
<td>Recoverability</td>
<td>The interface should include mechanisms to allow users to recover from their errors</td>
</tr>
<tr>
<td>User guidance</td>
<td>The interface should incorporate some form of context-sensitive user guidance and assistance.</td>
</tr>
</tbody>
</table>

Table 5.7: User Interface design principles (Sommerville, 1996)

5.5.1. User Familiarity

This is the important principle in designing user interface. The terms used for this E-TD system are considered to be user familiar. As this system is designed for use by Computer Science students and staffs, the terms used for the operations in this system are *Save, Reset, Browse, Upload, Download, View* and so on. Please refer to *Figure 5.10* and *Figure 5.11*. 
Figure 5.10: Save and Reset

Figure 5.11: Browse, Upload and Reset
5.5.2. Consistency

Interface consistency means that commands and menus of the system should be in the same format (Sommerville, 1996). Consistent interface reduce user learning time. The size of the banner and navigation bar, the size and colour for header and text are all same and consistent in every page of the entire system. Please refer to Figure 5.12 to see the consistency of the size of banner and navigation bar between the two screenshots.

Figure 5.12: Consistency of the size of banner and navigation bar
Throughout the whole system, *brown* is used as the colour for all the headers and *black* is used for the text. The font size for the header is 16pt and for text is 12pt. Please see *Figure 5.13* as examples.

*Figure 5.13: Consistency of the size and colour of header and text*
5.5.3. Minimal Surprise

Users become particularly irritated and confused when a computer system behaves in an unexpected (Sommerville, 1996). It is vital for the interface designer to ensure that comparable actions have comparable effects. If something completely different happens, the user is both surprised and confused.

Therefore, the user interface of this E-TD system is designed that all the same actions like clicking the save button in any page of the system will save all the data entered in the text box to the database. Likewise, all the comparable actions are also designed to have the same effect in this system.

For example, by clicking the “Save” button on the Add New Term page and Add Synonym page will save the records entered in the text box and give the successful message to the user. See Figure 5.14, Figure 5.15, Figure 5.16 and Figure 5.17.

![Figure 5.14: Save New Term](image-url)
**Figure 5.15: Successful Message (saving term)**

**Figure 5.16: Saving New Synonym**
5.5.4. Recoverability

Users inevitably make mistakes when using a system (Sommerville, 1996). These mistakes should be recovered by the system. In this E-TD system, for instance, when a user mistakenly clicks the **Remove Thesis File** function, the system will confirm with the user before removing the selected thesis file. So that, the user can recover from his/her mistake. Please see the **Figure 5.18**.

![Confirmation Message to Remove Thesis File](Image)
If the user click the upload button without entering the Thesis Title and choose the file to upload, the system will give the error message “The Thesis Title and/or Upload File cannot be empty. Please check and resubmit”. Please refer to Figure 5.19 and Figure 5.20.

Figure 5.19: Upload Thesis Page of E-TD Repository System

Figure 5.20: Error Message
5.5.5. User Guidance

In this E-TD system, the E-TD submission guidelines are provided for the students to follow in preparing their theses to submit to this system. Please see Figure 5.21.

![Figure 5.21: E-TD Submission Guidelines](image)

Besides, online help page is also available for the user to help them in using the system. See this page in Figure 5.22.
5.6. Summary

The software design requires experience as well as systematic planning because only with the proper software design, the project can be carried out faster and easier. After the design of this system is completed, the project would be continued with the system implementation phase.
Chapter 6: System Implementation

The system implementation and development is the process of transforming the design model of the system into reality or workable product.

A lot of tools and techniques are needed to implement this system such as tools used in designing the interface of the system, hardware and software requirements needed to develop this system and the data grid middleware used to integrate heterogeneous resources.

6.1. Hardware Development Environment

As this E-TD system is built on top of the data grid, two PCs are required to be used for this system. The following is the hardware specification used to develop this E-TD system:

- 1 GHz Pentium Processor
- 512 MB RAM
- 80 GB hard disk
- Other standard desktop PC compliances such as monitor, keyboard, mouse, network card, standard modem for connection, etc.
6.2. Software Development Environment

A variety of software tools were used during the development of this E-TD system. The software tools used for this system are briefly described in this section.

6.2.1. Operating System

The CentOS release 4.2, an Enterprise-class Linux Distribution, is used as the operating system. The software included in this release are:

- Apache HTTP Server-2.0;
- kernel-2.6.9-22.EL;
- JDK1.5.0_05;
- Firefox-1.5;
- MySQL-4.8.

6.2.2. Web Server

The Apache HTTP server, an open-source HTTP server, is used as a web server in the development of this project. Since April 1996, Apache has been the most popular HTTP server on the World Wide Web (The Apache Software Foundation, 1996). It provides a secure, efficient and extensible server that provides HTTP services in sync with the current HTTP standards. The Apache HTTP Server version 2.0 is included with the Red Hat Linux 4.2.1.6.

The features of Apache web server are as follows (Red Hat, Inc):

- **New Apache API:** Modules utilize a new, more powerful set of Application Programming Interfaces (APIs).
- **IPv6 Support:** The next generation IP addressing format is supported.
- **Simplified Directives**: A number of confusing directives have been removed while others have been simplified.

- **Multilingual Error Responses**: When using Server Side Include (SSI) documents, customizable error response pages can be delivered in multiple languages.

- **Multi-protocol Support**: Multiple protocols are supported.

### 6.2.3. Database

There are Oracle and PostgreSQL are the two RDBMS used for this project.

**Oracle**: The *Oracle Database 10g Release 2 (10.2)*, the first database designed for enterprise grid computing, is used as a database management system to store and retrieve related information of E-TD system. The Oracle Database is the most flexible and cost effective way to manage information and applications. Please refer to *Appendix A* for the Oracle Database Installation.

**Postgres**: The *Postgresql (Version 7.4.18)* is another RDBMS used for this project. It is a powerful enterprise class database which has sophisticated features such as Multi-Version Concurrency Control (MVCC), point in time recovery, tablespaces, asynchronous replication, nested transactions (savepoints), online/hot backups, a sophisticated query planner/optimizer, and write ahead logging for fault tolerance (PostgreSQL Global Development Group, 1996).

### 6.2.4. Graphic Creation

The Adobe Photoshop 7.0 is used as a graphic creation tool throughout the creation of graphic used in this system.
6.2.5. Browser

Red Hat Linux comes with the open source browser Mozilla which is used as a web browser to access this E-TD system. Mozilla functions like any other Web browser. It has the standard navigation toolbars, buttons, and menus.

However, other browsers such as Internet Explorer, Netscape Navigator can also be used to access this system.

6.3. The implementation of Data Grid Environment

The Storage Resource Broker (Version 3.4.2) is used as a data grid middleware for this E-TD system. Please refer to Appendix A for the detailed configuration of the SRB data grid, SRB Federation and MCAT replication between two zones.

6.4. C Client APIs

Several sets of C client APIs are used to make the SRB client to perform a number of different types of functions, which are discussed briefly in the following subsections:

6.4.1. Query/update of metadata

A set of APIs are provided for querying and updating the information in the MCAT catalog. For example, the srbGetDataItemInfo API can be used to query the metadata attributes of data items. This set of APIs allows applications to manage metadata associated with data collections, data items, user groups, and storage resources.
As mentioned above, the high-level request handler in the SRB middleware is itself a client to the MCAT catalog and uses these APIs (San Diego Supercomputer Centre, 2006).

**APIs that interact with MCAT**

This is the APIs for querying and updating information in the MCAT catalog.

- **srbGetDataDirInfo** - Get metadata information by querying the MCAT catalog.

- **srbGetDataDirInfoWithZone** - Get arbitrary meta data information on users, data sets or resources by querying the MCAT catalog with an additional zone input.

- **srbRegisterDataset** - Register a SRB data object.

- **srbModifyDataset** - Modify a SRB dataset

- **srbModifyUser** - Modify a user info.

- **srbModifyExtMetaData** - Modify an extensible metadata table in MCAT.

- **srbGetMoreRows** - Get more rows of result from a srbGetDatasetInfo, srbGetDataDirInfo, srbListCollect or srbGetPrivUsers call.

- **clearSqlResult** - Clear the memory associated with a SQL result struct, but the SQL result struct is not freed.

- **srbGetMcatZone** - Get the mcatName of this user.
6.4.2. Connecting to the server

These APIs include the usual connect and disconnect (clFinish) calls. The server performs user authentication when the client tries to connect to the server (San Diego Supercomputer Centre, 2006).

Client/server connection APIs

This set of APIs handles opening and closing of a connection to a SRB server (San Diego Supercomputer Centre, 2006).

- **srbConnect** – Initiate a connection to a SRB server. This call is replacing clConnect.
- **clConnect** – Initiate a connection to a SRB Master.
- **clFinish** – Close an existing SRB connection.
- **clErrorMessage** – Return server error message, if any, associated with the most recent client call.

6.4.3. Creation of data items

When a client issues a request to create an object, the SRB first obtains access control information form MCAT to verify user permissions. If the user has the privilege to create a data item under the specified collection in the requested storage resource, then the system generates a unique physical path name (e.g., UNIX or High Performance Storage System (HPSS) path name, or database LOB id) in that storage resource. Next, a create request is issued to the low-level request handler. If the creation is successful, then the corresponding metadata is registered in MCAT,
including information about the data item name, the associated collection, the corresponding PSR, and the user name. Finally, a data item handle is returned to the client which is used in subsequent read/write calls. If the call to register the data is unsuccessful, a low-level unlink call is issued to remove the data item and an error is returned to the client (San Diego Supercomputer Centre, 2006).

6.4.4. Open/read/write/delete of data items

These APIs are used to open, close, read, write, and unlink of data items as well as APIs to perform replication and grant tickets. The open operation requires querying MCAT to obtain the necessary metadata for the data item. The read/write operations are direct storage operations which need not involve the MCAT service (unless activity logging is turned on). Deletion of a data item requires updates to MCAT to “deregister” the data item (San Diego Supercomputer Centre, 2006).

APIs associated with data collections/items

The following set of APIs are used for performing standard storage operations.

\textit{srbObjOpen} - Open a SRB data object.

\textit{srbObjCreate} - Create a SRB data object.

\textit{srbObjClose} - Close an opened object

\textit{srbObjUnlink} - Unlink an SRB object

\textit{srbObjRead} - Read a block of data from a SRB object into buffer.

\textit{srbObjWrite} - Write content of buffer to a SRB object.
6.5. Implementing Ontology of Software Testing Terms

A known approach for storing large collections of data is to use a relational database management system (RDBMS) (Porto, 2005). Thus, the same approach is used for storing ontology in this research. The Postgresql Database System is used for both storing and retrieving data for ontology. Please refer to Appendix B to view how
the ontology of software testing is stored in the database which has term table, synonym table and ontology table.

6.6. Summary

The system implementation is a step to turn the designed system process flow and data flow to a reality system. The hardware and software development environment are identified in this chapter. After the system is implemented, the project would be continued with the system testing phase.
Chapter 7: Testing and Discussion of Results

This chapter is to present testing done on this E-TD repository system and the discussion of results. Therefore, the discussion will be more on testing to investigate whether the system do what it is supposed to do. Testing must ensure that the system matches its specification and work progressively towards customer confidence in the system. There are different types of testing used for this E-TD system.

7.1. Unit Testing

Unit testing is concerned with testing a module to ensure it satisfies its specification, typically a minispec (Bentley, 1999). A module is the smallest unit of code; but what a module is and what a specification is depend very much on the programming language being used and on the standards for program specification.

There are two complementary approaches (Bentley, 1999):

- **Black Box Testing**: The testers need not have any knowledge of the internal structure and behaviour of the system, being concerned only with what goes in and what comes out of the system black box. In other words, on the functional behaviour of the system. The test designer is only interested in finding circumstances in which the system does not comply with its specification.

- **White Box Testing**: White box testing assumes complete knowledge of the internal structure of a system. It is possible to have a program that is internally correct in some sense but still the wrong program.
As many modules are implemented in this E-TD system, each of the modules is tested independently to ensure that each module is working correctly. In the development of this system, unit testing was conducted after development of each of the components.

All the units/modules are tested and all the errors encountered were rectified. The following is the list of unit testing done on this E-TD system:

- Test whether the user can successfully logged into E-TD System.
- Test whether new ontology(ies) can be added to the database.
- Test whether the ontology stored in the database can be viewed.
- Test whether new members can be registered.
- Test whether password can be changed.
- Test whether metadata can be entered and modified.
- Test whether the browsing result(s) being displayed is correct and matches the given criteria.
- Test whether the search result(s) being displayed is correct and matches the search criteria.
- Test whether access constraints can be changed for different types of users.
- Test whether the thesis file is uploaded to the system.
- Test whether the thesis file can be downloaded from the system.
- Test whether the thesis file can be displayed online if it is .pdf or .doc format.
- Check related database to determine whether records are inserted into the database correctly.
- Test whether member information can be displayed correctly.
- Test whether the user can successfully logout from E-TD System.
7.2. Integration Testing

Integration testing is required when systems are developed in an incremental fashion (Bentley, 1999). This form of testing is used when a module or package is completed, and is intended to uncover errors in the interfaces between this module and all the others.

Integration testing frequently uncovers serious errors in parameter passing between different modules. These errors are typically a result of poor definition of minispecs.
Figure 7.1: Modules of E-TD Repository System
After each module has been tested independently, the interaction between all these modules must be tested again to ensure that the modules can be integrated. In this testing, the functions and procedures in each module are checked and tested carefully. Besides, all links in the web pages are tested. It is to ensure that every hyperlink can lead to the correct destination page. Furthermore, this type of testing was carried out to ensure that the record was being inserted and saved successfully in the database.

<table>
<thead>
<tr>
<th>No.</th>
<th>Test Data</th>
<th>Test Description</th>
<th>Expected Output</th>
<th>Actual Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Click on the hyperlinks such as Home, Member Login, Guest Login, New Account Request, Submission Guidelines, Online Help, About and so on</td>
<td>To check whether every hyperlinks lead to the correct destination pages.</td>
<td>The destination pages were linked correctly and displayed successfully.</td>
<td>As expected</td>
</tr>
<tr>
<td>2.</td>
<td>Enter data in the text boxes and click “Save” button.</td>
<td>To check whether the record is saved into the database.</td>
<td>The record saved successfully.</td>
<td>As expected</td>
</tr>
<tr>
<td>3.</td>
<td>Enter the user name and password and click “Login” button.</td>
<td>To check whether the system login the user and direct to the correct user name.</td>
<td>Login user successfully and display the login user name.</td>
<td>As expected</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>View E-TD online</td>
<td>To check whether the uploaded E-TDs can be displayed correctly</td>
<td>E-TD file displayed successfully.</td>
<td>As expected</td>
</tr>
<tr>
<td>5.</td>
<td>View ontology</td>
<td>To check whether the ontology stored in the database can be displayed.</td>
<td>Ontology of software testing displayed successfully.</td>
<td>As expected</td>
</tr>
<tr>
<td>6.</td>
<td>Change password and use a new password</td>
<td>To check whether a new password entered can be used.</td>
<td>User can login using a new password.</td>
<td>As expected</td>
</tr>
<tr>
<td>7.</td>
<td>Click “Logout” button.</td>
<td>To check whether exit from the system.</td>
<td>Exit and display message to inform the user that he/she is already logout from the system.</td>
<td>As expected</td>
</tr>
</tbody>
</table>

Table 7.1: Test Results of Integration Testing
7.3. System Testing

System testing is used to discover mismatches between the system specification and the system actually produced (Bentley, 1999). It is a form of black box testing because it is not based on any knowledge of the system design. In some cases, system testing is carried out with a similar set of test cases to those that are expected to be used for system acceptance.

The system testing of this E-TD system is to ensure that the entire application works accordingly. The system testing of this E-TD system is done by using a defined test plan and test cases and the results from the testing are listed in Table 7.1 below and see Appendix C for the graphical representation of this test results:

<table>
<thead>
<tr>
<th>No.</th>
<th>Test Data</th>
<th>Test Description</th>
<th>Expected Output</th>
<th>Actual Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Enter invalid password</td>
<td>Whether the user can login or not</td>
<td>Invalid Message</td>
<td>As expected</td>
</tr>
<tr>
<td>2</td>
<td>Without entering user name and password</td>
<td>Whether the user can login or not</td>
<td>Invalid Message</td>
<td>As expected</td>
</tr>
<tr>
<td>3</td>
<td>Enter valid password</td>
<td>Whether the user can login or not</td>
<td>User can login to the system</td>
<td>As expected</td>
</tr>
<tr>
<td>4</td>
<td>Click save button without entering data</td>
<td>Whether the record can be saved or not</td>
<td>Error Message</td>
<td>As expected</td>
</tr>
<tr>
<td>5</td>
<td>Save the record which is already in the database</td>
<td>Whether the record can be saved or not</td>
<td>Error Message</td>
<td>As expected</td>
</tr>
<tr>
<td>6</td>
<td>Save the record</td>
<td>Whether the record can be saved to the database correctly</td>
<td>Record saved and successful message</td>
<td>As expected</td>
</tr>
<tr>
<td>7</td>
<td>Click upload button without entering Thesis Title or choosing thesis file to upload</td>
<td>Whether the thesis file is uploaded or not</td>
<td>Error Message</td>
<td>As expected</td>
</tr>
<tr>
<td>8</td>
<td>Upload the thesis file</td>
<td>Whether the thesis file is uploaded or not</td>
<td>Thesis file uploaded and successful message</td>
<td>As expected</td>
</tr>
<tr>
<td>9</td>
<td>Browse E-TDs without choosing any browse criteria</td>
<td>Whether the result is displayed or not</td>
<td>As no browse criteria is given, all the E-TDs in the system are displayed</td>
<td>As expected</td>
</tr>
<tr>
<td>10</td>
<td>Give browse criteria to browse E-TDs</td>
<td>Whether the result being displayed is correct and matches the given criteria</td>
<td>Result displayed which reflect the browse criteria</td>
<td>As expected</td>
</tr>
<tr>
<td>11</td>
<td>Search E-TDs which is not in the system</td>
<td>Whether the E-TDs can be displayed or not</td>
<td>No record found message</td>
<td>As expected</td>
</tr>
<tr>
<td>12</td>
<td>Search E-TDs by</td>
<td>Whether E-TDs can be</td>
<td>Search result</td>
<td>As expected</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>using simple searching</td>
<td>searched and result being displayed is correct and matches the search criteria</td>
<td>returned which reflect the search criteria</td>
</tr>
<tr>
<td>13</td>
<td>Search E-TDs by using ontology searching</td>
<td>Whether E-TDs can be searched and result being displayed is correct and matches the search criteria</td>
<td>Ontology-based search result returned which reflect the search criteria</td>
<td>As expected</td>
</tr>
<tr>
<td>14</td>
<td>Register new user</td>
<td>Whether the new user can be added to the system or not</td>
<td>New user added and successful message</td>
<td>As expected</td>
</tr>
<tr>
<td>15</td>
<td>Change password</td>
<td>Whether the password can be changed</td>
<td>Password changed</td>
<td>As expected</td>
</tr>
<tr>
<td>16</td>
<td>Show Metadata</td>
<td>Whether the metadata of the selected thesis file can be displayed or not</td>
<td>Metadata displayed</td>
<td>As expected</td>
</tr>
<tr>
<td>17</td>
<td>Viewing the thesis file online</td>
<td>Whether the thesis file can be displayed or not</td>
<td>Thesis File Displayed</td>
<td>As expected</td>
</tr>
<tr>
<td>18</td>
<td>View the thesis file online using “guest” account</td>
<td>Whether the thesis file can be displayed or not</td>
<td>Not Authorized to View</td>
<td>As expected</td>
</tr>
<tr>
<td>19</td>
<td>Downloading thesis file</td>
<td>Whether the thesis file can be downloaded or not</td>
<td>Dialog box pop up to</td>
<td>As expected</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>Condition</td>
<td>Result</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Downloading thesis file using “guest” account</td>
<td>Whether the thesis file can be downloaded or not</td>
<td>Not Authorized to Download Thesis Message</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Remove the thesis file</td>
<td>Whether the thesis file can be removed or not</td>
<td>Thesis file removed</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Change access constraint</td>
<td>Whether the access constraint of the selected thesis file can be changed correctly</td>
<td>Access constraint changed</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Viewing member information</td>
<td>Whether the member info can be displayed or not</td>
<td>Member Information Displayed</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Viewing ontology in the database</td>
<td>Whether the ontology can be displayed or not</td>
<td>Ontology Displayed</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Click the logout button</td>
<td>Whether exit from the system or not</td>
<td>Exit</td>
<td></td>
</tr>
</tbody>
</table>

*Table 7.2: Test Results of System Testing*
7.4. Acceptance Testing

Acceptance testing is a form of black box testing and may include all the same type of tests as system testing (Bentley, 1999). The critical difference is that it is normally carried out by the end-user of the system. The systems analyst may take a lead role in defining these test cases, which should be based directly on the specification for the required system. Acceptance testing is often more correctly considered to be a form of system demonstration rather than a true test.

In this test, it is to confer with end-uses to make sure that the system works according to its expectation. A developer and potential end-users are involved to perform this acceptance testing. A few Master students are selected to use the system based on a set of basic instructions on what this system is about and how they are to use it. A survey form is then distributed to these students to get their opinion regarding the system. Please refer to Appendix D to see the questionnaires used in this acceptance testing. Different people have different views/perspectives of the functionality of the system. The survey question is divided into Part A and Part B. Part A is general question and Part B is regarding the functions and quality of the system. The feedbacks from the end-users regarding this E-TD system are summarized as follows:
Summarized Results for Questionnaires (Part A)

1. Have you visited a web page dedicated for online E-TDs?:
   - 67% of end-users have visited online E-TD systems
   - 33% of end-users have not visited online E-TD systems

![Figure 7.2: Result of Question 1 (Part A)](image)

2. If yes, what is your main concern when using online E-TD systems?
   - 78% of end-users choose “Easy and Convenient”
   - 11% of end-users choose “Fast Response Time”
   - 11% of end-users choose “Others”

![Figure 7.3: Result of Question 2 (Part A)](image)
3. Do you think that online E-TD systems save a lot of time and money?
   - 89% of end-users agree that online E-TD systems save a lot of time and money
   - 11% of end-users do not think online E-TD systems save a lot of time and money

   ![Figure 7.4: Result of Question 3 (Part A)](image)

4. Does the current manual theses management system at our faculty effective?
   - 89% of end-users choose “No”
   - 11% of end-users choose “Yes”

   ![Figure 7.5: Result of Question 4 (Part A)](image)
5. If no, what is the main problem encountered?

- 50% of end-users choose “Library is not open at all time”
- 40% of end-users choose “Hard to find theses from book shelf”
- 10% of end-users choose “Can’t borrow”

![Figure 7.6: Result of Question 5 (Part A)](image)

6. What method of theses management you prefer?

- 89% of end-users prefer electronic method
- 11% of end-users prefer manual method

![Figure 7.7: Result of Question 6 (Part A)](image)
Summarized Results for Questionnaires (Part B)

1. Login with ease
   - 89% of end-users agree that they are able to login to the system with ease.
   - 11% of end-users choose “Undecided”

   ![Figure 7.8: Login](image)

2. Upload E-TD with ease
   - 78% of end-users agree that they are able to upload E-TD to the system with ease.
   - 11% of end-users choose “Undecided”
   - 11% of end-users choose “Disagree”

   ![Figure 7.9: Upload ETD](image)
3. Perform *Browse Theses* function with ease

- 67% of end-users choose “Agree”
- 11% of end-users choose “Undecided”
- 22% of end-users choose “Disagree”

![Figure 7.10: Browse Theses](image)

4. Perform *Search Theses* function with ease.

- 67% of end-users choose “Agree”
- 11% of end-users choose “Undecided”
- 22% of end-users choose “Disagree”

![Figure 7.11: Search Theses](image)
5. Ontology-based search result is more useful compared to conventional searching

- 78% of end-users agree that ontology-based search result is more useful compared to conventional searching.
- 11% of end-users choose “Disagree”
- 11% of end-users choose “Undecided”

![Figure 7.12: Ontology-based Searching](image)

6. Viewing metadata and ETDs online with ease

- 78% of end-users agree that they are able to view metadata and E-TDs online with ease
- 22% of end-users disagree with this question.

![Figure 7.13: Viewing Metadata and ETDs online](image)
7. Comfortable in using this E-TD system

- 67% of end-users choose “Agree”
- 22% of end-users choose “Disagree”
- 11% of end-users choose “Undecided”

As the result, most of the end-users prefer online E-TD systems than manual systems. Most of them also think that online E-TD systems save a lot of time and money. The end-users found most of the functions of this system can be used with ease. Most of them agree that the search results returned from ontology search engine is more useful than the conventional search engine. As a whole, 67% of the end-users feel comfortable in using this E-TD system.

7.5. Discussion Results

Even though hundreds of E-TD systems are available on the Internet, there is no E-TD system integrated with data grid technology. Compared to these existing E-TD systems, this system is built on top of data grid technology and the MCAT is replicated in another SRB server. Hence, if one MCAT is not functioning, the system will be automatically redirecting the client request to a replica MCAT on different SRB server. This is the advantage of this E-TD system compared to other conventional E-TD systems.
The searching techniques used in other existing E-TD systems are just direct keyword matching and keyword matching using wildcard (‘*’). In this system, the simple web searching technique is enhanced with ontology or semantic web technique. Not only the ontology-based searching mechanism is implemented but the simple/conventional searching technique is also provided in this E-TD system.

As there is no other existing E-TD system using ontology-based searching technique, it is not possible to compare the ontology search engine with the searching service of other E-TD systems. Therefore, the search result(s) returned from simple search engine and ontology-based search engine of this E-TD system are compared by using the same search criteria in both search engines built in this study. For example, the keyword used is “black-box testing”. The simple search engine returns only 1 record (E-TD) which directly matches the keyword “black-box testing”. However, the ontology search engine returns 10 records (E-TDs) which not only include the keyword “black-box testing” but also the ontology of “black-box testing”. The ontology search engine first retrieves the ontology of “black-box testing” from the ontology database.

In the first round, the ontology search engine retrieve “unit testing” and “system testing” from the ontology database as “black-box testing” is a type of unit testing and system testing. Then, the search engine retrieve “testing” from the ontology database as the “unit testing” and “system testing” are types of testing.

![Ontology of black-box testing](image)

*Figure 7.15: Ontology of black-box testing*
Likewise, a few other terms are also used in testing the ontology search engine and the results are shown in Table 7.3 and also shown in graph in Figure 7.16. The screenshots are also shown in Figure 7.17, Figure 7.18, Figure 7.19 and Figure 7.20 when testing the search engines using the keyword “black-box testing”.

<table>
<thead>
<tr>
<th>No.</th>
<th>Search Keyword</th>
<th>Number of Record(s) Found (Simple Search Engine)</th>
<th>Number of Record(s) Found (Ontology Search Engine)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>black-box testing</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>2.</td>
<td>bottom-up testing</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>3.</td>
<td>path testing</td>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td>4.</td>
<td>system integration testing</td>
<td>2</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 7.3: Comparison of the test result from simple search engine and ontology search engine

Figure 7.16: Graph for the comparison of the result from simple search engine and ontology search engine
Figure 7.17: Simple Search Engine

Figure 7.18: Search Result Returned from Simple Search Engine
As a whole, this system integrates all the module and sub modules to fulfill the functional and non-functional requirements defined in Chapter 4.
7.6. Summary

In this chapter, different testing methods are used to make sure the system to be bug free. The individual modules of the system are tested using unit testing, each of these modules are integrated and tested again using the integration testing and finally the system is tested as a whole to ensure that all the requirements are met and system does not deviate from any unforeseen circumstances. Additionally, acceptance testing is also carried out together with the potential end-users and the feedbacks are summarized and discussed in this chapter.
Chapter 8: Conclusion

A lot of effort has put in implementing this project. Finally, the effective system can be produced with all the functions proposed in the proposal. The ultimate goal of developing E-TD system using data grid technology and enhancing the conventional searching service by ontology are met.

The valuable knowledge, especially in setting up data grid, was gained throughout the implementation of this project. It was a golden opportunity to get the experience and knowledge in configuring and managing the SRB Data Grid Middleware, Oracle Database Management System, Postgresql Database System, C Programming and C Client APIs to communicate with SRB Server.

Basically, this chapter is to discuss the strengths and weaknesses of the system, problems encountered during the implementation of the system, the future enhancement and to summarize the overall results of the whole system.

8.1. Fulfillment of Project Objectives

This E-TD system was successfully implemented with all the functional and non-functional requirements of the system. The two following objectives are also fulfilled:

- integrating data grid technology to the E-TD system;
- improving the E-TD searching service by incorporating ontology-based searching mechanism.
8.2. System Strengths and Weaknesses

Every system has its own strengths and weaknesses. Although this system covers all the functions proposed in the proposal, it is not a perfect system which still has limitations and weaknesses.

**Strengths**

- users can access the system from anywhere at anytime using any web browser;
- students can share their theses and dissertations worldwide;
- ease and convenient to submit and retrieve E-TDs;
- instant access to any and all integrated resources (theses and dissertations);
- ontology-based searching of ETDs is provided.

**Weaknesses**

- limited to submit *Computer Science* theses and dissertations;
- ontology is limited to software testing.
- currently, ontology database is installed in the host named *fsktm.geranium2.um.edu.my*. If this server is down, the ontology-based searching part will not be working as the ontology of software testing terms cannot be retrieved at all.

Compared to other existing E-TD systems, this system has the following main advantages:

- As this system is built on top of the data grid and MCATs are replicated, users will not miss any access to E-TDs in this system even if one MCAT server is not functioning.
- Searching service of the system is improved with ontology-based searching
mechanism so that it is more searchable and efficient in retrieving and searching E-TDs.

Theses functions are not available in other E-TD systems reviewed in Chapter 2.

8.3. Problems Encountered

As this E-TD system is running on data grid technology, the task of setting up the data grid is required in the implementation of this system. Being lack of knowledge and experience in data grid technology, a lot of unforeseen problems has been encountered during the process of data grid configuration. In order to solve these problems, research and study on related materials have been carried out throughout the development of the system.

8.4. Future Enhancement

In order to improve the quality of the system, some of the features of the system can be enhanced in the future. Even though many other courses/programs are offered by our faculty, this E-TD system allows Computer Science students to upload their theses to this system. The aspect of future work for this system is allowing the students from other majors rather than Computer Science to submit their theses and dissertations to this system. Additionally, analyzing of most popular ETDs in the collection will be included in the future work.

The ontology can also be enhanced to incorporate terms from other domains like software architecture, software requirements, software design, software construction, software configuration management, software engineering management, software quality and so on. Currently, the ontology database is installed in one of the SRB
servers. In the future, the ontology database will be installed in a separate machine so that if the SRB server (i.e., fsktm.gernium2.um.edu.my) is not functioning, it will not affect the retrieving of ontology from the database.

8.5. Summary

Eventually, this E-TD Repository System has been successfully implemented with the objectives stated in the proposal. This system achieved it’s objectives of integrating data grid technology to the system and enhancing the search capability by ontology-based searching technique. Throughout the implementation of this system, precious and useful knowledge have been gained on setting up and configuring data grid, setting up web server, installing Oracle and Postgresql databases, using advanced database C programming, using C Client APIs to connect to SRB servers and constructing data structure for ontology.

Even though the system achieves all functional and non-functional requirements, there are some limitations in current version of the system which needs enhancements in the future.
References


Guha, N., Semantic Digital Library Services, 


Michael, D., *Extending metadata for digital preservation*, (updated May 18, 1997),

NIEeS Info, *Storage Resource Broker*, (updated 15 February 2007),


*Patton MQ* (1990). *Qualitative evaluation and research methods* (2nd ed.). London:
Sage Publications.

Porto, F. (2005), *Ontology Manager*, (created 03 June 2005),

(accessed 10 February 2007).


Rachel (2002), *Thesis vs Dissertation*, (created 01 August 2002),


(accessed 09 September 2007).

San Diego Supercomputer Centre, (updated 15 March 2006), *The SRB client API*,
(accessed 14 March 2007).

San Diego Supercomputer Center, *SRB architecture*, (updated 21 November 2006),


Wikipedia (2008), Repository, (updated 29 March 2008),

Wikipedia (2008), Methodology, (updated 23 June 2008),


Zia, Lee L., Growing a National Learning Environments and Resources Network for Science, Mathematics, Engineering, and Technology Education (created March 2001),