View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Massey Research Online

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Web-course Search Engine

A thesis presented in partial fulfillment of

the requirements for the degree of

Master of Science in Computer Science

at Massey University, Palmerston North, New Zealand.

Xiaohua Gong

2002

Acknowledgements

Thanks to my supervisor Kinshuk for his patience, ideas, faith and enthusiasm. His support made the year a lot easier to bear.

Abstract

The World Wide Web is an amazing place that people's lives more and more rely on. Especially, for the young generation, they spend a significant amount of their play and study time using the Internet. Many tools have been developed to help the educational users in finding educational resources. These tools include various search engines, Web directories and educational domain gateways. Nevertheless, these systems have many weaknesses that made them unsuitable for the specific search needs of the learners.

The research presented in this thesis describes the development of the *Web-course search engine*, which is a friendly, efficient and accurate helper for the learners to get what they want in the vast Internet ocean. The most attractive feature of this system is that the system uses one universal language, which lets the searchers and the resources "communicate" with each other. Then the learner searchers can find the Web-based educational resources that are most fit to their needs and course providers can provide all necessary information about their courseware. This universal language is one widely acceptable Metadata standard. Following the Metadata standard, the system collects exact information about educational resources, provides adequate search parameters for search and returns evaluative results. By using the *Web-course search engine*, the learners and the other educational users are able to find useful, valuable and related educational resources more effectively and efficiently.

Some improvement suggestions of the search mechanism in the World Wide Web have been brought forward for the future research as a result of this project.

Table of Contents

Chapter 1 Outline.....1

1.1 Problem Analysis.....1

1.2 Solving Problems.....2

1.3 Implementation.....3

1.4 Summary.....3

Chapter 2 Standard and Metadata.....5

2.1 Introduction.....5

2.2 Why standard?.....5
2.2.1 Examples of standards.....6
2.2.2 Benefits of using a standard in a Web search system.....7

2.3 Metadata standards.....8

2.4 Examples of Metadata standards.....11
2.4.1 Learning Object Metadata (LOM).....11
2.4.2 Dublin Core.....15
2.4.3 EdNA Metadata.....17
2.4.4 Gateway to Educational Materials (GEM) Metadata Standard.....20

2.5 Comparison of various metadata standards.....22
2.6 Conclusion.....26

Chapter 3 Survey of various Web-based search systems.....28

3.1 Introduction.....28

3.2 Systems dedicated to educational area.....28
3.2.1 Education World.....28
3.2.2 GEM (The Gateway to Educational Materials).....32
3.2.3 EdNA Online.....38

3.3 Common search engines or Web directories used for both educational and other purposes.....41
3.3.1 YAHOO.....42
3.3.2 Northern Light.....44
3.3.3 HotBot.....45
3.3.4 Lii.org (Librarians' index to the Internet).....47

3.4 Discussion on common search engines.....50

3.5 Conclusion.....50

Chapter 4 Using Metadata Standard to construct "Web-course Search engine".....52

4.1 Introduction.....52

4.2 Functionality of the "Web-course Search engine".....52

4.3 Metadata Standard of the "Web-course Search engine".....53
4.3.1 Summary of the "Metadata Standard".....54
4.3.2 Mapping LOM.....62
4.3.3 Characteristic of the "Metadata Standard".....64

4.4 Information Collection, Search criterion and search result.....67
4.4.1 Information Collection.....67
4.4.2 Search criterion.....73
4.4.3 Search result.....76

4.5 Database structure.....79
4.5.1 Database structure underlying search mechanism.....79
4.5.2 Definition.....81
4.5.3 Referential constraints.....84
4.5.4 Relationships between tables.....85

4.6 Conclusion.....85

Chapter 5 Implementation of The Web-course search engine.....88

5.1 3-tiered Architectures.....88

5.2 Functionality Specification.....89
5.2.1 Scenarios.....89
5.2.2 Functional Specification.....93

5.3 Structure.....94
5.3.1 Model-View-Controller design pattern (MVC).....94
5.3.2 EJB-centric and Web-centric design.....95

5.4 The Model of Model-View-Controller.....96
5.4.1 State in the Web-course search engine.....97
5.4.2 Persistent Data.....99

5.5 The View of Model-View-Controller.....100
5.5.1 Interaction Interface.....100
5.5.2 Screen's structure.....103
5.5.3 Model-View Synchronization.....104

5.6 The Controller of Model-View-Controller.....106
5.6.1 EJB-centric controller.....106
5.6.2 Web-centric controller.....110

5.7 Deployment.....112
5.7.1 Deployment Descriptors.....112

5.7.2 Packaging the Application.....112

5.8 Transactions & Security.....114
5.8.1 Transactions.....114
5.8.2 Security.....115

5.9 Summary.....120

Chapter 6 Future of The Web-course search engine.....121

6.1 Searcher Membership Module.....121

6.2 Metadata Editor.....121

6.3 Cooperation With Different Search Systems.....122

6.4 Summary.....124

References.....125

Appendix A: Screenshots of the system.....130

Appendix B: Outline of LOM Metadata Elements.....135

Appendix C: Dublin Core Metadata Elements.....138

List of Figures

Chapter 1 Outline

Figure 1.1 Architecture of the system.....4

Chapter 2 Standard and Metadata

Figure 2.1 Nutrition information.....6
Figure 2.2 Hierarchy/tree structure of LOM.....13
Figure 2.3 Relationships of EdNA/DC and GEM/DC.....22
Figure 2.4 Inheritance Relationship (e.g. the LOM inherits from Dublin Core).....24
Figure 2.5 Relationships of Dublin Core, GEM, LOM and EdNA.....26

Chapter 3 Survey of various Web-based search systems

Figure 3.1 "ADD A SITE" page.....29
Figure 3.2 "Lesson Plan Submit" page.....30
Figure 3.3 Search results.....32
Figure 3.4 GemCat.....35
Figure 3.5 GEM Search.....35
Figure 3.6 GEM search result.....36-37
Figure 3.7 EdNA search results.....40
Figure 3.8 YAHOO Search Options.....42
Figure 3.9 YAHOO search results.....43
Figure 3.10 Northernlight.com Power Search.....44
Figure 3.11 Northern Light search results.....45
Figure 3.12 Hotbot advanced search.....46
Figure 3.13 Hotbot search results.....46-47
Figure 3.14 lii.org-advanced searches.....49
Figure 3.15 lii.org search results.....49-50

Chapter 4 Using Metadata Standard to construct "Web-course Search engine"

Figure 4.1 Relationships of Dublin Core, LOM, and "New Metadata Standard".....67
Figure 4.2 Information-sheet.....67-72
Figure 4.3 Search request.....74-76
Figure 4.4 Search result.....77-79
Figure 4.5 Relationships between tables.....87

Chapter 5 Implementation of The Web-course search engine

Figure 5.1 A 3-Tier Architecture.....89
Figure 5.2 Mixed EJB-centric and Web-centric design.....96
Figure 5.3 Fine-grained Objects.....99
Figure 5.4 A Template.....104
Figure 5.5 Model-View Synchronization.....105
Figure 5.6 CatalogWebImpl accesses Data.....106
Figure 5.7 EJB-centric controlling process.....110

Figure 5.8 Web-centric controlling process.....111
Figure 5.9 The enterprise beans protected by Web components.....118

Chapter 6 Future of The Web-course search engine

Figure 6.1 Cooperation between systems.....122
Figure 6.2 Insides a search system.....123

List of Tables

Chapter 2 Standard and Metadata

Table 2.1 EdNA Metadata Elements.....18-19
Table 2.2 GEM Metadata Elements.....20-21
Table 2.3 Dublin Core and LOM relation.....22-23
Table 2.4 GEM mapping LOM.....24-25
Table 2.5 EdNA mapping LOM.....25

Chapter 3 Survey of various Web-based search systems

Table 3.1 Metadata elements compare.....30

Chapter 4 Using Metadata Standard to construct "Web-course Search engine"

Table 4.1 Definition of the "Metadata Standard" 55-61
Table 4.2 LOM Mapping.....62-64
Table 4.3 Some Elements of "Metadata Standard".....65-66
Table 4.4 Definitions of tables.....81-84

Chapter 1 Outline

"The World Wide Web facilitates co-operative teaching and learning, offering an exciting potential for sharing both the learning resources as well as the tasks of creating new learning resources through *international collaboration.*" (Kinshuk., Patel, A., & Russell, D. (1999)). The methods by which learners find information on the Web are tools like search engines, Web directories and metasearch engines. (Loren, Will & Brian. (1999)). Information banks (Perkins. (1991)), search engines/systems are the indispensable component of this *international collaboration.* The learners and the other educational users use these search systems to index, search, and find teaching and learning resources. This thesis describes the *Web-course search engine*, developed in this project, which is particularly focused on indexing, searching and finding educational resources. The system attempts to eliminate various deficiencies of commonly available search engines.

1.1 Problem Analysis

There are many search tools/systems for resources available on the World Wide Web. However, currently, these tools/systems cannot help users to find Web-Based educational resources effectively and efficiently. Eric, Steve, William & C.Lee. (1999) pointed out that metasearch engines, which rely on other search tools, can considerably increase coverage, but they are still limited by the engines they use with respect to the number and quality of results.

If a student wants to find some educational material for assignment or research on the Web, for example, for studying English, he/she will use some sort of keywords like "study English" in a search engine, such as Yahoo, Excite, or Google. The search engine will then display a number of the Web addresses about this topic. Not all of these links will be of his/her interest because the student may have certain requirements such as a high level course, some kind of multimedia inter-action course and Mac Platform only, cost free, using the Cambridge English, and so on. How can the student submit such a search request using currently available searching engines?

Literature suggests that it is hard to get accurate results from these search engines. There are various reasons for this phenomenon.

One reason is that these search engines just search the information inside the HTML document and HTML is not flexible enough to express and contain information about the content. It only has few meta tags for this purpose.

Another important reason is that there is no established standard or rule to describe the special material such as study English course. Because there is no established standard, courses providers are not able to provide uniform information to search engines, and search engines could not provide a standard request sheet for users to submit queries. In other words, there is not good communication between query user, search engine and courses' providers. Metadata is information about an object. As the number of educational objects grows exponentially and the needs for learning expand equally noticeably, the lack of information or metadata about educational objects places a serious and fundamental constraint on ability to discover, manage and use objects.

The chapter 2 will introduce the Metadata concept, explain the benefits of using Metadata on a Web search system, and then compare several Metadata Standards that are commonly accepted. The chapters 3 will analyze in-depth the problems existed in some commonly available search system in the WWW by discussing a survey of various Web-based search systems, and grouping them to systems that are dedicated to educational area, and common search engines or Web directories used for both educational and other purposes.

1.2 Solving Problems

For constructing an effective and efficient search system, the *Web-course search engine*, some problems have to be solved. First is to find a Metadata standard, which is suitable for describing the Web-Based educational course material. Second is to collect the information about Web-based courses following this Metadata standard. Final is to provide the request sheet for the searcher to submit query and search

results/response sheet, which can clearly and exactly express the information about Web-based course.

The chapter 4 will describe the functionality of the *Web-course Search engine*, the *Metadata standard* used in this system, and the details of the course information collection, search criterion, search results and database structure.

1.3 Implementation

The architecture of the *Web-course Search engine* is based on a Web-based three-tier model (Figure 1.1): web browsers as tier 1 (presentation tier), servers (for example, the HTTP server) as tier 2 (business logic tier) running on the Java 2 Platform Enterprise Edition (J2EE), and database system as tier 3. The web browser side includes the course information input sheet for the educational course providers to provide the information about the course according to the *Metadata standard* that is mentioned in chapter 4, request sheet for searcher to submit query/request and search result sheet. On the tier 2, the J2EE provides a suitable containment environment for the business logic components. The J2EE manages these components efficiently and provides a number of services to the components. It uses JDBC to connect the tier 3, database. The data tier uses Cloudscape database Server to control and manage databases access.

The chapter 5 will describe the architecture of the application, which follows the Model-View-Controller design pattern, and how the application uses the deployment, transaction, JNDI service, and security capabilities of the Java 2 Platform Enterprise Edition (J2EE) to simplify component development and provide richer functionality.

1.4 Summary

The *Web-course search engine* eliminates various deficiencies of other search engines as mentioned in chapter 3. The system collects suitable information about educational resources, provides adequate search parameters for searchers to find resources and returns evaluative search results. By using the *Web-course search engine*, the learners and the other educational users are able to index, search and find useful, valuable and

related Web-course resources more effectively and efficiently. This is what educational Web course search mechanisms try to do.

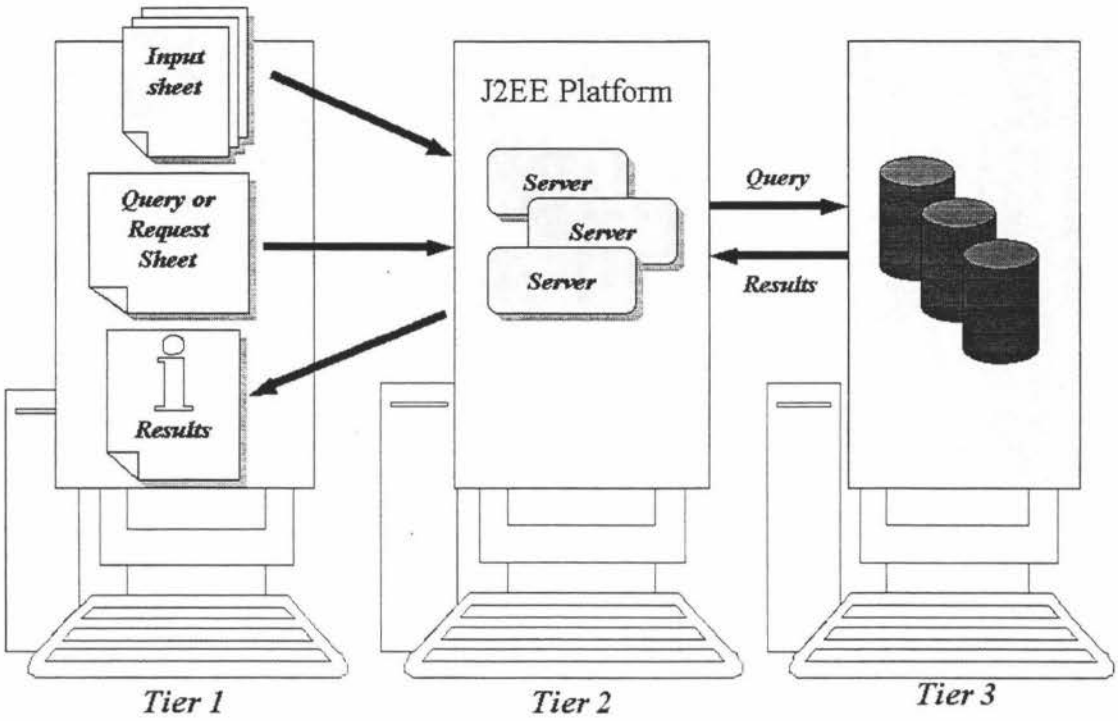

Figure 1.1 *Architecture of the system*

The chapter 6 provides some recommendations for future search engines/systems that are dedicated for finding educational teaching and learning resources.