E-RESOURCE SEARCHER (ERS) SYSTEM

CHEN LIANG

UNIVERSITI UTARA MALAYSIA 2010

KOLEJ SASTERA DAN SAINS
(College of Arts and Sciences)
Universiti Utara Malaysia

PERAKUAN KERJA KERTAS PROJEK
(Certificate of Project Paper)

Saya, yang bertandatangan, memperakukan bahawa
(I, the undersigned, certifies that)

CHEN LIANG
(806158)

calon untuk Ijazah
(candidate for the degree of) **MSc. (Information Technology)**

telah mengemukakan kertas projek yang bertajuk
(has presented his/her project of the following title)

E-RESOURCE SEARCHER (ERS) SYSTEM

seperti yang tercatat di muka surat tajuk dan kulit kertas projek
(as it appears on the title page and front cover of project)

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan.
(that this project is in acceptable form and content, and that a satisfactory knowledge of the field is covered by the project).

Nama Penyelia
(Name of Supervisor) : **ASSOC. PROF. ABDUL NASIR ZULKIFLI**

Tandatangan
(Signature) : _____ Tarikh (Date) : 20/10/10

Nama Penilai
(Name of Evaluator) : **DR. AZMAN YASIN**

Tandatangan
(Signature) : _____ Tarikh (Date) : 20/10/10

PERMISSION TO USE

In presenting this project in partial fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this project in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence by the Dean of Postgraduate Studies and Research. It is understood that any copying or publication or use of this project or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my project.

Requests for permission to copy or to make other use of materials in this project, in whole or in part should be addressed to:

Dean of Postgraduate Studies and Research
College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok
Kedah Darul Aman
Malaysia

ABSTRACT

With the continue increase of the information, information retrieval (IR) has been becoming more important than ever before. Particularly in the digital library, generally it has a large plenty of E-Resource for the user to search and use. Therefore, this paper attempts to design an E-Resource searcher system which helps the student to find E-Resource faster and accurately among the ocean of E-Resource. The paper introduces the IR concepts, search engine working procedure, and the prominent open-source search library - Apache Lucene in details. The ERS search functionality primarily relies on the Lucene as full-text search core and partly depends on the traditional SQL query search for the common search. The research also presents feasible search architecture for the integration of the traditional database operation and search index database, which is appropriate for the most of the web-application embedded search. In addition, the IR (Information Retrieval) test collection is adopted to test the ERS search performances in precision and recall capabilities. The result is satisfactory and meets the research requirement.

ACKNOWLEDGEMENT

My deepest gratitude goes first and foremost to Prof. Abdul Nasir bin Zulkifli, my supervisor, for his constant encouragement, inspirations and guidance. He has walked me through all the stages of the writing of this thesis. Without his consistent and illuminating instruction, this thesis could not have reached its present form.

Second, I feel grateful to all the lecturers in the College of Arts and Sciences, Information Technology department who once offered me valuable courses and advice during my study.

Last my thanks would go to my beloved family for their loving considerations and great confidence in me all through my study period. I also owe my sincere gratitude to my friends and my fellow classmates who gave me their help and time in listening to me and helping me work out my problems during the difficult course of the thesis.

TABLE OF CONTENTS

LIST OF FIGURE

LIST OF TABLE

LIST OF ABBREVIATION

Ajax(Asynchronous JavaScript and XML)
DAO (Data Access Object)
DML (Data Manipulate Language)
ERS (E-Resource Searcher)
GUI (Graphic User Interface)
I/O (Input / Output)
IOC (Inversion of Control)
IR (Information Retrieval)
JDBC (Java Database Connectivity)
LAN (Local Area Network)
MVC (Model – View- Controller)
SQL (Structured Query Language)
UML (Unified Modeling Language)
VSM (Vector Space Model)
WWW (World Wide Web)

CHAPTER ONE

INTRODUCTION

This chapter briefly introduces the primary parts of this research area. It also defines the motivation to conduct this study. It is followed by the objectives, the significance, scope of this research, and the organization of this study.

1.1 Introduction

With the development of the network, surfing on the web to require information has become an important part of our daily lives. The majority of the web users rely on the search functions to obtain information and information materials. Thus, the requirements of the information search on the web particularly intranet have risen significantly (Zhou et al., 2008). As a result, web-based search has become one of the most essential parts of the web application (Vishal et al, 2010).

Now it is fairly common that many digital libraries in the world offer the huge amount of the E-Resources such as E-book, E-journal, and E-magazine on their website for the library user to view, share and study on line. On one hand, E-Resource can represent live and vivid the content for the students to better understanding the content. On the other hand, E-Resource also can be easily shared with the users and free downloaded. It undoubtedly facilitates the students study, helps them work more efficiently (Anantha Rao K, 2009).

The contents of the thesis is for internal user only

REFERENCES

Aghajanyan, A. (2008). HttpUnit Tutorial. Retrieved 9 10, 2010, from http://httpunit.sourceforge.net/doc/tutorial/index.html

Anantha Rao K, R. (2009). Library Microsite: An E-Resource Search. *7th International CALIBER(Convention on Automation of Libraries in Education and Research)-2009* (pp. 145-151). Pondicherry University, Puducherry: INFLIBNET Centre, Ahmedabad.

ApacheLuceneOverview. (2010). Apache Lucene - Overview. Retrieved 8 10, 2010, from http://lucene.apache.org/java/docs/index.html

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern Information Retrieval.* New York: Addison-Wesley.

Conallen, J. (2000). *Building Web Application With UML.* New York: Addison-Wesley Object Technology Series.

Erik, H., Otis, G., & Michael, M. (2009). *Lucene In Action (Second Version).* Greenwich CT, New York: Mnanning.

Gang, L., Wei, S., & Zhe, Q. (2006). *Conquer Ajax + Lucene for Building Search Engine.* BeiJing: People Post Press.

Ivar, J., Magnus, C., Patrik, J., & Gunnar, O. (1993). *Object-Oriented Software Engineering.* Edinburgh Gate Essex CM20 2JE England: Addison Wesley.

James, R., Ivar, J., & Grady, B. (2006). *The Unified Modeling Language Reference Manual.* New York: Addison Wesley.

Jie, Y., Fangfang, L., & Jie, G. (2010). Discovering Collaborative Users Based On Query Context for Web Information. *International Conference on Futrue Computer and Communication* (pp. V3-1 - V3-5). Wuhan, China : IEEE.

JiuXian. (2009). Lucene Search Principle And Coding Anaysis. BeiJing, China: JavaEye offical website.

L.Ledford, J. (2009). *Search Engine Optimization.* Indianapolis,Indiana: Wiley Publishing, Inc.

Lashkari, A. h., Mahdavi, F., & Ghomi, V. (2009). A Boolean Model In Information Retrieval For Search Engines. *2009 International Conference on Information Management and Engineering* (pp. 385-389). Kuala Lumpur, Malaysia : IEEE.

Manning, C. D., Prabhakar, R., & Hinrich, S. (2008). *Introduction To Information Retrieval.* London England: Cambridge University Press.

Nezhad, A. B., & Perlis, M. D. (2009). Information Retrieval On The World Wide Web And Active. *Information Retrieval on the WWW and Active Logic.*

Norshuhada, S., & Shahizan, H. (2010). *Design Research In Software Development.* Sintok: Universiti Utara Malaysia Press.

Oracle-SunDeveloperNetwork. (2010). Java BluePrints Model-View-Controller. Retrieved 8 10, 2010, from http://java.sun.com/blueprints/patterns/MVC-detailed.html

Otis, G., & ErikHatcher. (2005). *Lucene In Action.* Greenwich CT: Mmanning.

PingBing, L. (2005). *Information Retrieval Based On The Lucene.* XiAn: Electronic secience and technology University.

Qian, L., & Wang, L. (2010). An Evaluation of Lucene For Keywords Search in Large-scale Short Text Storage. *2010 International Conference On Computer Design And Applications(ICCDA 2010)* (pp. V2-206 - V2-209). Qinhuangdao, China : IEEE.

S.M.Shafi, & A.Rather, R. (2005). Precision and Recall of Five Search Engines For Retrieval of Scholarly Information in the Field of Biotechnology. Retrieved 9 15, 2010, from http://www.webology.ir/2005/v2n2/a12.html

Shengdong, L., Sueqiang, L., Feng, L., & Shuicai, S. (2009). Study on Efficiency of Full-Text Retrieval Based on Lucene. 4. BeiJing, HeBei, China: IEEE.

SunWeiQin. (2006). *Java OOP Programming.* BeiJing: Electric Industry Press.

Tao, H., & Hongyang, C. (2009). The Implement of Searching Engine For Educational Resources Using Text Clustering. *2009 IEEE. International Conference on Granular Computing* (pp. 260 - 263). Nanchang : IEEE.

Thomas, C., & Begg, C. (2004). *Database Solutions.* Edinburgh Gate Harlow Essex CM20 2JE England: Pearson Education Limited.

Tumer, D., Shah, M. A., & Bitirim, Y. (2009). An Empirical Evaluation on Semantic Search Performance of Keyword-Based and Semanitic Search Engines: Google, Yahoo,Msn and Hakia. *2009 fourth International COnference on Internet Monitoring and Protection* (pp. 51 - 55). Venice/Mestre : IEEE.

Vaishnavi, V., & Kuechler, W. (2004). Design Research In Information Systems. Retrieved 8 2, 2009, from http://desrist.org/design-research-in-information-systems/

Vishal, S., Sowmya, v., Vasundhara, T., & Ritesh, K. J. (2010). Share-ken: A Way to Improve Web Search. *2010 International Conference on Recent Trends in Information, Telecommunication and Computing* (pp. 327 - 329). Kochi, Kerala : IEEE.

Vogel, L. (2010). Junit4 Tutorial. Retrieved 9 10, 2010, from www.vogella.de/articles/JUnit/article.html

W.B.Frakes. (1992). *Introduction to information storage and retrieval systems.* UpperSaddle River,NJ USA: Prentice-Hall, Inc.

Wang, N., Li, L., Wang, Y., Wang, Y.-b., & Wang, J. (2008). Research on the Web Information System Developement Platform Based on MVC Design Pattern. *2008 International Conference on Web Intelligence and Intelligent Agent Technology* (pp. 203 - 206). Sydney, NSW : IEEE/WIC/ACM.

Wang, X. H., & Jin, D. (2007). The Comparision between Lucene and Rational Database. *Database and Information Management* , 615-616.

WikiPedia-ExtendedBooleanModel. (2010). WikiPedia Extended Boolean Model. Retrieved 8 10, 2010, from http://en.wikipedia.org/wiki/Extended Boolean Model

WikiPedia-FulltextSearch. (2010). WikiPedia Full-text Search. Retrieved 8 10, 2010, from http://en.wikipedia.org/wiki/full-text search

Wikipedia-InformationRetrieval. (2010). Information Retrieval. Retrieved 8 28, 2010, from http://en.wikipedia.org/wiki/Information_retrieval

WikiPedia-Lucene. (2010). WikiPedia Lucene. Retrieved 8 10, 2010, from http://en.wikipedia.org/wiki/Lucene

WikiPedia-MVC. (2010). MVC Overview. Retrieved 8 10, 2010, from http://en.wikipedia.org/wiki/Model-view-controller

YanHong. (2002). *Java and Pattern.* BeiJing: Electric Industry Press.

Yinan, J., Chunwang, Z., & Xueping, W. (2009). An Empirical Study on Performance Comparison of Lucene and Relational Database. *International Conference on COmmunication Software and Networks* (pp. 336 - 340). Macau : IEEE.

Yong, Z., & Jian-lin, L. (2009). Research and Improvement of Search Engine Based on Lucene. *2009 International Conference on Intelligent Human- Machine Systems and Cybernetics* (pp. 270 - 273). Hangzhou, Zhejiang : IEEE.

Yuri, K., & Jochen, R. (2004). Introducing a Conceptual Information Retrieval (IR) Framework. *Journal of Medical Systems* , 89 - 101.

Zhou, C., Li, Z., & Feng, B. (2008). Research and Implementation of the Small-scale Search Engine Based on Lucene. *2008 International Conference on Computer Sceenece and Software Engineering* (pp. 377 - 380). Wuhan, Hubei : IEEE.