CITATION ENDORSEMENT FRAMEWORK FOR DIGITAL REPOSITORY

ABDIRIZAK SA'ID GHEDI

Universiti Utara Malaysia

MASTER OF SCIENCE IN INFORMATION TECHNOLOGY
UNIVERSITI UTARA MALAYSIA
2016

Permission to Use

In presenting this thesis in fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for review. I further agree that permission for the copying of this thesis in any way, in full or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis 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 thesis. Requests for permission to copy or to make other use of materials in this thesis, in whole or in part should be addressed to:

Universiti Utara Malaysia

Dean of Awang Had Salleh Graduate School of Arts and Sciences

UUM College of Arts and Sciences

Universiti Utara Malaysia

06010 UUM Sintok

Abstrak

Repositori digital adalah salah satu platform yang berguna untuk menguruskan bahan-bahan digital terutama persidangan prosiding. Setakat ini, prosiding persidangan telah diterbitkan sebagai versi cetak dan digital, versi digital telah disimpan dalam repositori digital dan maklumat nukilan setiap artikel dikumpul. Had model yang sedia ada ialah petikan secara automatik diambil dari pangkalan data yang tertentu dan yang lain dari pengarang artikel yang yang mempunyai maklumat mengenai pemetikan tidak boleh mengemaskini maklumat yang masuk ke tabung. Oleh itu kajian ini mencadangkan rujukan rangka kerja sokongan untuk repositori digital. Permohonan pembangunan Rapid (RAD) telah digunakan untuk pembangunan sistem dan temu bual telah dilakukan untuk mendapatkan kajian pakar sistem. Analisis tema kajian pakar menunjukkan bahawa peserta bersetuju bahawa repositori digital yang baru boleh meningkatkan perkongsian maklumat aktif.

Universiti Utara Malaysia

Kata kunci: perkongsian maklumat, mencari maklumat, repositori digital petikan pengendorsan

ABSTRACT

Digital repository is one of the useful platform to manage digital materials especially conference proceedings. To date, conference proceedings have been published as printed and digital versions, the digital version has been stored in digital repository and the citation information of each article is collected. The limitation of the existing model is that the citation is automatically retrieved from certain databases and the other authors of that article who have information about the citation cannot update the information in to the repository. Therefore this study proposed citation endorsement framework for digital repository. Rapid application development (RAD) was used for the system development and interview were done to get the expert review of the system. The thematic analysis of the expert review shows that the participants agreed that the new digital repository can improve the citation indexing by allowing the author to update the citation information.

Universiti Utara Malaysia

Keywords: information sharing, information retrieval, digital repository, citation endorsement

Acknowledgement

In the Name of Allah, the Most Gracious and Most Merciful

First and foremost, all praise to Allah for providing me with the strength, perseverance, and wisdom to have this work done.

I would like to express my deep and sincere gratitude to my supervisors, Mr. Wan Hussain Wan Ishak and Dr. Norliza Katuk for the patience, humble supervision and the advice I received from them in the course of this dissertation. The supervision and support that they gave me truly helped the progression and smoothness of this research. Furthermore, I would like to thank my lovely parents, Sa'id Ghedi and Anab Einte. They were my role models and the best example for me to follow in my life. They never stop supporting me through my whole life and without them I would have never been what I am today. I would also like to thank my brother Mohamed Sa'id and my sister Muna Sa'id for supporting me financially and morally.

Last but not least, I want to thank my cousins, Jamiilo Ahmed, Abdul Kadir Shekh ali, Maryan Shekh Ali, and Fadumo Shekh Ali for their constant demonstrations of love and nonstop moral and financial support throughout my years of study.

Universiti Utara Malaysia

Table of Contents

Permission to Use ……..…….... ii

Abstrak…… iii

Abstract……..... iv

Acknowledgment.. v

Table of Content ……. vi

List of Tables…….. ix

List of Figures……. x

List of Abbreviations..... xi

1.0 CHAPTER ONE INTRODUCTION 1

1.1 General Overview ……. 1

1.2 Background of Study…… 3

1.3 Problem Statement ……. 5

1.4 Research Questions …….. 6

1.5 Objectives ……. 6

1.6 Scope ……..... 6

1.7 Significance of Study …… 7

1.8 Organization of the Dissertation ……... 7

2.0 CHAPTER TWO LITERATURE REVIEW ……... 8

2.1 General Overview ..…….. 8

2.2 Information Storage and Retrieval System ……..... 9

2.2.1 Types of Information Retrieval ……. 12

2.2.2 Information Retrieval Model..... 14

2.2.3 Search Engine..... 16

2.3 Electronic Libraries ……... 17

2.3.1 Digital Library..... 18

2.3.2 Digital Repository 20

2.4 Institutional Digital Repository …… 21

Universiti Utara Malaysia

2.4.1 Digital Content and Archive 22
2.4.2 Types of Digital Repository 23
2..4.3 Features of Digital Repository 25
2.5 Citation Management 26
2.5.1 Citation Indexing Services 26
2.5.2 Citation Extraction Method 33
2.6 Collaborative Information Sharing 34
2.6.1 Collaborative Digital Archive 34
2.6.1.1 Scholarly Publishing 35
2.6.1.2 Open Access 36
2.6.2 Active Information Sharing 37
2.6.3 H-index and the 10-index 41
2.7 Summary 43

3.0 CHAPTER THREE METHODOLOGY 44

3.1 General Overview 44
3.2 Developing the Digital Repository Model 45
3.3 Integrating the Citation Endorsement Model into the Repository 46
3.3.1 Interface Design 46
3.3.1.1 Flow Chart/Transition Diagram 47
3.4 Validating the Proposed Citation Endorsement Framework 48
3.4.1 Procedure 49
3.4.2 Materials 49
3.4.3 Participants 49
3.5 Thematic Analysis Method 50
3.6 Summary 50

4.0 CHAPTER FOUR CITATION ENDORSEMENT FRAMEWORK 51

4.1 General Overview 51
4.2 Requirement Analysis 51
4.2.1 Identification of the Digital Repository Features 52
4.2.2 Identification of the Actors and Roles 52
4.2.3 Identification of the Functionality 52
4.3 Interface Design 54

Universiti Utara Malaysia

4.4 Modeling of the Citation Endorsement Process ……56
4.5 Integrating Digital Repository with Citation Endorsement Framework ……. 58
4.6 Summary ……... 67

5.0 CHAPTER FIVE EXPERT EVALUATION 68

5.1 General Overview …… 68
5.2 Result of the Thematic Analysis ….. ……. 68
5.3 Discusion …….. 79
5.4 Summary ……... 80

6.0 CHAPTER SIX CONTRIBUTION AND FUTURE WORK…… 81

6.1 General Overview …… 81
6.2 Contribution of the Study …….. 82
6.3 Revisit of Research Objectives ……. 82
6.4 Conclusion …… 83
6.5 Limitation …….. 84
6.6 Future Work ……... 84

REFERENCES 85

APPENDIX A 93

APPENDIX B 103

APPENDIX C 104

APPENDIX D 105

APPENDIX E 106

APPENDIX F 107

APPENDIX G 108

APPENDIX H 109

APPENDIX I 110

Universiti Utara Malaysia

List of Tables

Table 2.1: Comparison of five citation indexing systems ……. 30

Table 3.1: The main stages of the study …… 44

Table 4.1: The entities, functions, and requirements for the model ….....53

Table 5.1: The years that the experts have been on the duty ……..……69

Table 5.2: The familiarity of the experts towards digital repositories ……... 69

Table 5.3: The experience of the experts in using indexing and citation systems.. 70

Table 5.4: The answer of the experts on their first impression of this repository system ……..……..71

Table 5.5: The opinion of the experts on the systems functions and capabilities ... 72

Table 5.6: The answer of the experts towards the citation indexing feature …… 73

Table 5.7: The opinion of the experts towards the idea of allowing the author to add their own citation .....…….. 74

Table 5.8: The opinion of the experts towards the citation verification process …. 75

Table 5.9: The expert's recommendation of other features ……..... 75

Table 5.10: The opinion of the experts on what they would change about this system ……...…. 76

Table 5.11: The aspect of the repository that the experts don't like or disagree ….. 77

Table 5.12: The opinion of the experts on the missing things in the repository system ……...…….. 77

Table 5.13: The answer of the experts answer on what particular they like about the system ……...…… 78

Universiti Utara Malaysia

List of Figures

Figure 2.1: Snapshot of Google search engine …….. 17

Figure 2.2: The active information sharing model for Vrepo …… 39

Figure 3.1: Add citation feature in the feature in the digital repository ……... 47

Figure 3.2: The flow of updating citation information ...……..... 48

Figure 4.1: The interface of the digital repository ....……. 55

Figure 4.2: The flow of the digital repository interface ….……... 56

Figure 4.3: Digital repository with citation endorsement framework …….. 57

Figure 4.4: The data flow diagram of the system …... ...……..... 59

Figure 4.5: The system flow of the admin ……...……..... 60

Figure 4.6: The system flow of the author …….. ……..... 61

Figure 4.7: The system flow of the author of the cited paper ……..... 62

Figure 4.8: The login interface of the system …….. ……..... 63

Figure 4.9: The author profile page of the system …… ……..... 64

Figure 4.10: The form that is used to upload articles into the repository 65

Figure 4.11: The form that is used to add citation into the repository …..... 66

Universiti Utara Malaysia

List of abbreviations

API	Application Program Interface
ARL	Association of Research Libraries
ATU	Attitude Towards Usage
BI	Behavioral Intention
CSS	Cascading Style Sheets
DFD	Data Flow Diagram
DOIS	Digital Object Identifier
EDI	Electronic Data Interchange
HTML	HyperText Markup Language
IP	Internet Protocol
JSPS	Java Server Pages
LDAP	LightWeight Directory Access Protocol
PBC	Perceived Behavioral Control
PEU	Perceived Ease of Usage
PU	Perceived Usefulness
RAD	Rapid Application Development
SOC	School of Computing
TAM	Technology Acceptance Model
UML	Unified Modeling Language
UTAUT	Unified Theory of Acceptance and Use of Technology

Universiti Utara Malaysia

UUM	Universiti Utara Malaysia
WDL	World Digital Repository
CORA	Coriolis Ocean database ReAnalysis

Universiti Utara Malaysia

CHAPTER ONE

INTRODUCTION

1.1 General Overview

A digital repository is a mechanism for managing and storing digital content. Repositories can be subject or institutional in their focus. Storing content into an institutional repository allows staff and institutions to manage and preserve that record and so they expand full value from it. A repository can support research, learning, and administrative processes. Repositories use open access to certify that the content they contain is available and that it can be searched and retrieved for future use. The use of these established open accesses allows mechanisms to be set up which import, export, identify, store and retrieve the digital content within the repository (Marshall, 1997).

Digital repositories might contain an extensive variety of content for a different purposes and users. The kind of record goes into a repository is presently less a matter of technological or software ability, and more a policy choice made by each institution or administrator. Usually records can include research results such as journal articles or research data, e-theses, e-learning materials and teaching objects, and administrative data. Some repositories only take in particular materials such as theses or journal papers, while other repositories look for to gather any credible scholarly work produced by the institution.

The contents of the thesis is for internal user only

References

Agrawal, R., Evfimievski, A., & Srikant, R. (2003, June). Information sharing across private databases. In *Proceedings of the 2003 ACM SIGMOD international conference on Management of data* (pp. 86-97). ACM.

Anzaroot, S., & McCallum, A. (2013). A new dataset for fine-grained citation field extraction. In ICML Workshop on Peer Reviewing and Publishing Models.

Armbruster, C., & Romary, L. (2009). Comparing repository types: challenges and barriers for subject-based repositories, research repositories, national repository systems and institutional repositories in serving scholarly communication. *Research Repositories, National Repository Systems and Institutional Repositories in Serving Scholarly Communication (November 23, 2009)*.

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval* (Vol. 463). New York: ACM press.

Bonito, J. A. (2007). A local model of information sharing in small groups. *Communication Theory*, *17*(3), 252-280.

Universiti Utara Malaysia

Caragea, C., Wu, J., Ciobanu, A., Williams, K., Fernández-Ramírez, J., Chen, H. H., ... & Giles, L. (2014). CiteSeer x: A Scholarly Big Dataset. In *Advances in Information Retrieval* (pp. 311-322). Springer International Publishing.

Cullen, R., & Chawner, B. (2011). Institutional repositories, open access, and scholarly communication: a study of conflicting paradigms. *The Journal of Academic Librarianship*, *37*(6), 460-470.

Dawes, S. S. (1996). Interagency information sharing: Expected benefits, manageable risks. *Journal of Policy Analysis and Management*, *15*(3), 377-394.

Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS biology*, *4*(5), 692.

Universiti Utara Malaysia

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *The FASEB journal*, *22*(2), 338-342.

Fereday, J., & Muir-Cochrane, E. (2008). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. International journal of qualitative methods, 5(1), 80-92.

Giri, R., & Das, A. K. (2011). Indian Citation Index: a new web platform for measuring performance of Indian research periodicals. Library Hi Tech News, 28(3), 33-35.

Giles, C. L., Bollacker, K. D., & Lawrence, S. (1998, May). CiteSeer: An automatic citation indexing system. In Proceedings of the third ACM conference on Digital libraries (pp. 89-98). ACM.

Hensley, M. K. (2011). Citation Management Software: features and futures. Reference & User Services Quarterly, 204-208.

Hogan, A., Harth, A., Umbrich, J., Kinsella, S., Polleres, A., & Decker, S. (2011). Searching and browsing linked data with swse: The semantic web search engine. *Web semantics: science, services and agents on the world wide web*, *9*(4), 365-401.

Jamaludin, Z., & Ishak, W. H. W. (2011). A virtual repository approach to departmental information sharing. *American Journal of Economics and Business Administration*, *3*(1), 18-23.

Jamaludin, Z & Ishak, W.H.W (2007) Repositori Maya: Kit Maklumat Fakulti. Prosiding Seminar Kebangsaan e-Komuniti 2007 "Merapatkan Jurang Digital: Inisiatif Malaysia", Pusat Kajian E-Komuniti, Universiti Kebangsaan Malaysia, 10-11 Dis. 2007

Universiti Utara Malaysia

Johnson, K., & Magusin, E. (2009). *Exploring the digital library: A guide for online teaching and learning* (Vol. 24). John Wiley & Sons.

Kichuk, D. (2015). Loose, Falling Characters and Sentences: The Persistence of the OCR Problem in Digital Repository E-Books. *portal: Libraries and the Academy*, *15*(1), 59-91.

Klingner, J. K., Scanlon, D., & Pressley, M. (2005). How to publish in scholarly journals. *Educational researcher*, *34*(8), 14-20.

Koh, S. C. L., Gunasekaran, A., & Rajkumar, D. (2008). ERP II: The involvement, benefits and impediments of collaborative information sharing. *International Journal of Production Economics*, *113*(1), 245-268.

Lynch, C. A. (2003). Institutional repositories: essential infrastructure for scholarship in the digital age. *portal: Libraries and the Academy*, *3*(2), 327-336.

Minasny, B., Hartemink, A. E., McBratney, A., & Jang, H. J. (2013). Citations and the h index of soil researchers and journals in the Web of Science, Scopus, and Google Scholar. *PeerJ*, *1*, e183.

Universiti Utara Malaysia

Marshall, C. C. (1997, July). Annotation: from paper books to the digital library. In *Proceedings of the second ACM international conference on Digital libraries* (pp. 131-140). ACM.

Ohno-Machado, L., Boxwala, A. A., Ehresman, J., Smith, D. N., & Greenes, R. A. (1997). A virtual repository approach to clinical and utilization studies: application in mammography as alternative to a national database. In *Proceedings of the AMIA Annual Fall Symposium* (p. 369). American Medical Informatics Association.

Ostroukh, A. V., & Pomazanov, A. V. (2014). Development of information storage and retrieval system on nanomaterials and nanotechnology. *International Journal of Advanced Studies*, *3*(3), 24-30.

Onwuchekwa, E. O., & Jegede, O. R. (2011). Information Retrieval Methods in Libraries and Information Centers. *African Research Review*, *5*(6), 108-120.

Purcell, K., Brenner, J., & Rainie, L. (2012). Search engine use 2012.Rayman, D., Bertram, C., & Prom, C. (2014). American Library Association Institutional Repository (ALAIR) Report.

Ramalho Correia, A. M., & Carlos Teixeira, J. (2005). Reforming scholarly publishing and knowledge communication: From the advent of the scholarly journal to the challenges of open access. *Online information review*, *29*(4), 349-364.

Universiti Utara Malaysia

Rossi, P. L. (2011). Electronic libraries in partnership: BEEP for Africa. *African Research and Documentation*, *115*, 69-75.

Ross, S. (2012). Digital preservation, archival science and methodological foundations for digital libraries. *New Review of Information Networking*, *17*(1), 43-68.

Schilit, B. N., Price, M. N., & Golovchinsky, G. (1998, May). Digital library information appliances. In *Proceedings of the third ACM conference on Digital libraries* (pp. 217-226). ACM.

Steeleworthy, M., & Dewan, P. T. (2013). Web-based Citation Management Systems: Which One Is Best?. *Partnership: the Canadian Journal of Library and Information Practice and Research*, *8*(1).

Storey, M. A., Best, C., Michaud, J., Rayside, D., Litoiu, M., & Musen, M. (2002, April). SHriMP views: an interactive environment for information visualization and navigation. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems* (pp. 520-521). ACM.

Stuckenschmidt, H., & Van Harmelen, F. (2005). *Information sharing on the semantic web*. Springer Science & Business Media.

Universiti Utara Malaysia

Talja, S. (2002). Information sharing in academic communities: Types and levels of collaboration in information seeking and use. *New Review of Information Behavior Research, 3*(1), 143-159.

Tansley, R., Bass, M., Stuve, D., Branschofsky, M., Chudnov, D., McClellan, G., & Smith, M. (2003, May). The DSpace institutional digital repository system: current functionality. In *Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital libraries* (pp. 87-97). IEEE Computer Society.

Tan, Y. F., Kan, M. Y., & Lee, D. (2006, June). Search engine driven author disambiguation. In *Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries* (pp. 314-315). ACM.

Teregowda, P. B., Urgaonkar, B., & Giles, C. L. (2010, June). Citeseerx: A cloud perspective. In Proceedings of the Second USENIX Workshop on Hot Topics in Cloud Computing.

Tullis, T. S., & Stetson, J. N. (2004, June). A comparison of questionnaires for assessing website usability. In *Usability Professional Association Conference* (pp. 1-12).

Van Leeuwen, T., Moed, H., Tijssen, R., Visser, M., & Van Raan, A. (2001). Language biases in the coverage of the Science Citation Index and its consequencesfor

Universiti Utara Malaysia

international comparisons of national research performance. Scientometrics, 51(1), 335-346.

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, *46*(2), 186-204.

Williams, K., Wu, J., Chen, H. H., Khabsa, M., Caragea, C., Ororbia, A., ... & Giles, C. L. (2014, June). Citeseerx: Ai in a digital library search engine. In *The Twenty-Sixth Annual Conference on Innovative Applications of Artificial Intelligence, IAAI* (Vol. 14).

Zhao, W., & White, G. (2012, November). A collaborative information sharing framework for Community Cyber Security. In *Homeland Security (HST), 2012 IEEE Conference on Technologies for* (pp. 457-462). IEEE.

Universiti Utara Malaysia