FACETED SEARCH OF OPEN EDUCATIONAL RESOURCES USING THE DESIRABILITY INDEX

ISHAN SUDEERA ABEYWARDENA

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY UNIVERSITY OF MALAYA KUALA LUMPUR MALAYSIA

2015

Abstract

The open educational resources (OER) movement has gained considerable momentum in the past few years. According to the Paris OER Declaration, OER can be defined as

"*teaching, learning and research materials in any medium, digital or otherwise, that reside in the public domain or have been released under an open license that permits no-cost access, use, adaptation and redistribution by others with no or limited restrictions. Open licensing is built within the existing framework of intellectual property rights as defined by relevant international conventions and respects the authorship of the work*".

With this drive towards making knowledge open and accessible, a large number of OER repositories have been established and made available online throughout the world. However, the limitation of existing search engines such as Google, Yahoo!, and Bing to effectively search for useful OER that are useful or fit for teaching purposes is a major factor contributing to the slow uptake of the movement. As a major step to solve this issue, the researcher has designed, developed and tested *OERScout*, a technology framework based on text mining solutions. Utilizing the concept of faceted search, the system allows academics to search heterogeneous OER repositories for useful resources from a central location. Furthermore, the *desirability* framework has been conceptualized to parametrically measure the usefulness of an OER with respect to openness, accessibility and relevance attributes.

The objectives of the project are: (i) to identify user difficulties in searching OER for academic purposes; (ii) to identify the limitations of existing OER search methodologies with respect to locating fit-for-purpose resources from heterogeneous repositories; (iii) to conceptualize a framework for parametrically measuring the

suitability of OER for academic use; and (iv) to design a technology framework to facilitate the accurate centralized search of OER from heterogeneous repositories.

The major contributions of this research work are twofold:

The first contribution is a conceptual framework which can be used by search engines to parametrically measure the usefulness of an OER, taking into consideration the openness, accessibility and relevance attributes. The advantage of this framework is that, using the well-established four R's and ALMS frameworks, it can restructure search results to prioritize the resources which are the easiest to reuse, redistribute, revise and remix. As a result, academics practicing the Open and Distance Learning (ODL) mode of delivery can locate resources which can be readily used in their teaching and learning.

The second contribution is a search mechanism which uses text mining techniques and a faceted search interface to provide a centralized OER search tool to locate useful resources from the heterogeneous repositories for academic purposes. One of the key advantages of this search mechanism is its ability to autonomously identify and annotate OER with domain specific keywords. As a result, this search mechanism provides a central search tool which can effectively search for OER from any repository regardless of the technology platforms or metadata standards used. Another major advantage is the utilization of the conceptual framework which can parametrically measure the usefulness of an OER in terms of fit-for-purpose. As a result, academics are able to easily locate high quality OER from around the world which best fit their academic needs.

Abstrak

Seja akhir-akhir ini, pergerakan Sumber Pendidikan Terbuka (SPT) telah mula bermomentum. Menurut Deklarasi SPT Paris , SPT boleh ditakrifkan sebagai

"*bahan-bahan pengajaran, pembelajaran dan penyelidikan dalam berbagai jenis medium, digital atau sebaliknya, yang berada di domain awam atau dikeluarkan sebagai lesen terbuka yang membenarkan akses percuma, digunakan, disesuaikan dan diedarkan semula oleh orang lain tanpa sekatan atau sekatan minima. Pelesenan terbuka dibina dalam kerangka hak-hak harta intelek seperti yang ditakrifkan oleh konvensyen antarabangsa yang berkaitan dan menghormati pengarang kerja itu*".

Dengan kewujudan pemanduan baru ini ke arah menjadikan ilmu pengetahuan lebih terbuka dan mudah diakses, banyak repositori SPT telah dibina dan disediakan dalam talian untuk kegunaan seluruh dunia. Walau bagaimanapun, pembatasan enjin carian yang sedia ada seperti Google, Yahoo!, dan Bing dalam carian SPT yang boleh dipakai atau yang menetapi ciri-ciri pengunaan untuk tujuan pengajaran merupakan satu faktor utama yang menyumbangkan kepada kelembaban pergerakan itu pada keseluruhannya. Sebagai langkah utama dalam penyelesaian masalah yang tertera, projek ini mengesyorkan OERScout, satu rangka kerja teknologi berdasarkan pengunaan perlombongan teks. Mengunakan konsep carian pelbagai aspek, sistem ini membolehkan para akademik mencari beraneka repositori SPT untuk sumber-sumber yang berguna dari satu lokasi utama. Tambahan pula, rangka kerja keinginan adalah berkonsep pengukuran secara berparameter kesesuaian SPT berdasarkan sifat-sifat keterbukaan, akses dan sifat-sifat berkaitan.

Objektif projek ini adalah (i) untuk mengenal pasti secara terperinci sebab, dari perspektif pengguna, yang menyumbang kepada ketidakupayaan mencari SPT untuk

tujuan akademik, (ii) untuk mengenal pasti batasan kaedah carian SPT yang sedia ada khususnya bagi carian sumber yang tepat dari pelbagai repositori, (iii) untuk mengkonsepsikan satu rangka kerja bagi mengukur secara berparameter kesesuaian SPT untuk penggunaan akademik; dan (iv) untuk mereka bentuk satu kerangka teknologi yang akan memudah dan memusatkan carian SPT yang berupaya memberi keputusan tepat dari pelbagai repositori.

Sumbangan utama kerja-kerja penyelidikan ini adalah berlipat ganda: Sumbangan pertama adalah rangka konsep yang boleh digunakan oleh enjin carian untuk mengukur secara berparameter kebergunaan SPT, dengan mengambil kira keterbukaan, akses dan sifat-sifat berkaitan. Kelebihan rangka kerja ini adalah dengan penggunaan rangka kerja 4'R' dan ALMS yang termuka, ia boleh menyusun semula hasil carian dengan mengutamakan sumber yang paling mudah diguna, diagih, disemak dan dicampurkan semula. Ini akan membolehkan , ahli akademik yang mengamalkan kaedah pengajian secara terbuka dan jarak jauh (ODL) mengesan sumber yang sedia digunakan dalam pengajaran dan pembelajaran mereka.

Sumbangan kedua adalah mekanisme carian menggunakan teknik perlombongan teks dan berbagai carian antara muka dalam menyediakan satu alat carian terpusat SPT untuk mencari sumber-sumber yang berguna daripada pelbagai repositori untuk tujuan akademik. Satu kelebihan utama mekanisme carian ini adalah keupayaan untuk mengenal pasti secara autonomi identiti SPT melalui anotasi SPT yang menggunakan kata kunci domain tertentu. Hasilnya, mekanisme carian ini menyediakan alat carian terpusat yang mampu mencari dengan berkesan SPT dari mana-mana repositori tanpa mengambil kira platform teknologi atau standard metadata yang digunakan. Satu lagi kelebihan utama ialah penggunaan rangka kerja yang mengukur secara berparameter kebergunaan SPT dari aspek tepat diguna. Hasilnya, ahli akademik terutama mereka

yang berada di Selatan Global akan dapat mencari dengan mudah SPT yang berkualiti tinggi dari seluruh dunia yang berupaya memenuhi keperluan akademik masing-masing.

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **Ishan Sudeera Abeywardena**
Registration/Matric No: **WHA100051**
Name of Degree: **Doctor of Philosophy**
Title of Thesis ("this Work"): **FACETED SEARCH OF OPEN EDUCATIONAL RESOURCES USING THE DESIRABILITY INDEX**
Field of Study: **Computer Science**

I do solemnly and sincerely declare that:

1) I am the sole author/writer of this Work;
2) This Work is original;
3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

_____

Candidate's Signature   Date: 12th March 2014

Subscribed and solemnly declared before,

_____

Witness's Signature   Date: 12th March 2014

Name: N.S. Abeyawardena

Designation: National Programme Coordinator, Ministry of Plantation Industries
          Smallholder Plantation Entrepreneurship Development Programme
          No. 172, 2nd Floor, Elvitigala Mawatha, Colombo 08, Sri Lanka.

Acknowledgements

Table of Contents

List of Figures

List of Tables

List of Abbreviations

List of Appendices

Appendix	Description
A	Abeywardena, I.S., Chan, C.S., & Tham, C.Y. (2013). OERScout Technology Framework: A Novel Approach to Open Educational Resources Search. *International Review of Research in Open and Distance Learning, 14(4),* 214-237.
B	Abeywardena, I.S., Raviraja, R., & Tham, C.Y. (2012). Conceptual Framework for Parametrically Measuring the Desirability of Open Educational Resources using D-index. *International Review of Research in Open and Distance Learning, 13(2),* 104-121.
C	Abeywardena, I.S., Chan, C.S., & Balaji, V. (2013). OERScout: Widening Access to OER through Faceted Search. *Proceedings of the 7th Pan-Commonwealth Forum (PCF7), Abuja, Nigeria.*
D	Abeywardena, I.S., & Chan, C.S. (2013). Review of the Current OER Search Dilemma. *Proceedings of the 57th World Assembly of International Council on Education for Teaching (ICET 2013), Thailand.*
E	Abeywardena, I. S., Tham, C.Y., Chan, C.S., & Balaji. V. (2012). OERScout: Autonomous Clustering of Open Educational Resources using Keyword-Document Matrix. *Proceedings of the 26th Asian Association of Open Universities Conference, Chiba, Japan.*
F	Abeywardena, I. S., Dhanarajan, G., & Chan, C.S. (2012). Searching and Locating OER: Barriers to the Wider Adoption of OER for Teaching in Asia. *Proceedings of the Regional Symposium on Open Educational Resources: An Asian Perspective on Policies and Practice, Penang, Malaysia.*
G	Abeywardena, I. S., Dhanarajan, G., & Lim, C.K. (2013). Open Educational Resources in Malaysia. In G. Dhanarajan & D. Porter (Eds.), *Open Educational Resources: An Asian Perspective. Commonwealth of Learning and OER Asia (ISBN 978-1-894975-61-2),* 119-132.
H	Dhanarajan, G., & Abeywardena, I. S. (2013). Higher Education and Open Educational Resources in Asia: An Overview. In G. Dhanarajan & D. Porter (Eds.), *Open Educational Resources: An Asian Perspective. Commonwealth of Learning and OER Asia (ISBN 978-1-894975-61-2),* 3-18.
I	Survey Instrument: A study of the current state of play in the use of Open Educational Resources (OER) in the Asian Region.
J	User Manual: OERScout
K	User Test Feedback Form and User Feedback Summary: OERScout

CHAPTER 1
INTRODUCTION

Chapter 1 : Introduction

With the new drive towards accessible and open information, Open Educational Resources (OER) have taken center stage after being first adopted at a UNESCO forum in 2002. An early definition of OER is "*web-based materials, offered freely and openly for use and re-use in teaching, learning and research*" (Joyce, 2007, p. 1). The Paris OER Declaration (UNESCO, 2012, p. 1) provides a more comprehensive definition:

"teaching, learning and research materials in any medium, digital or otherwise, that reside in the public domain or have been released under an open license that permits no-cost access, use, adaptation and redistribution by others with no or limited restrictions. Open licensing is built within the existing framework of intellectual property rights as defined by relevant international conventions and respects the authorship of the work".

The global demand for education is currently not met through existing conventional educational institutions, especially in the developing countries or the 'Global South'. This deficiency is further heightened when 'excluded communities' which have limited access to education due to geographic, demographic and sociographic circumstances are considered. OER initiatives such as the Massachusetts Institute of Technology's (MIT) *Open Courseware* (OCW) initiative, the Rice University's *Connexions* initiative and the Commonwealth of Learning's (COL) *WikiEducator* initiative have provided high quality learning material for use and re-use through the *Creative Commons* (CC) licensing scheme. The ability to freely use and modify the content for teaching and learning purposes has boosted the drive towards OER for educating the masses.

Asian countries such as India, China, Japan, Korea and Vietnam have made the move towards the use of OER but are still in the process of making the use of OER 'accepted'

practice among academics, due to various inhibitors. One such inhibitor is the inability to effectively search for OER that are academically useful and of an acceptable academic standard. This limitation is further amplified by the heterogeneity of the large number of repositories available and their constant expansion.

During the past decade, various technologies and platforms such as *Wiki* and *Rhaptos* have emerged to support the OER repositories. Although such technologies provide native search mechanisms, searching for useful OER is still predominantly done using mainstream search engines such as Google, Yahoo! and Bing. This has added to the inability to locate academically useful OER as these search engines are not specifically designed for this purpose. In addition, there is as yet no measure available for search engines to parametrically measure the usefulness of a resource for academic purposes. Furthermore, there exists no search engine capable of allowing academics to easily navigate through the search results to pinpoint OER of an acceptable academic standard. As a solution to these issues, this research project proposes a technology framework which can parametrically measure the usefulness of an OER for academic purposes. In addition, it utilizes text mining techniques to provide a mechanism for easily navigating through the search results to pinpoint OER of an acceptable academic standard.

This chapter is organized into four sections. The first section provides a general overview of OER, introduces the problems to be addressed and the research objectives. Sections two and three discuss the proposed research approach and the significance of the research project respectively. Section four provides an outline of the other chapters.

1.1 Problem Statements and Research Objectives

The World Wide Web (WWW) provides cost effective information, rapid revision and democratized access (Crowley, Leinhardt, & Chang, 2001). Modern day education is very much dependent on technology and the global flow of information which is underpinned by the accessibility of technology. In this new global paradigm, OER play a major role, as academics have to increase their competitiveness both for obtaining funding and for improving knowledge (Rawsthorne, 2007). In recent years, global OER initiatives have been established by many organizations, including UNESCO, COL and the United Nations (UN). Among these initiatives are 'Education for All' from the UN and World Bank (Geith & Vignare, 2008), the Open e-Learning Content Observatory Services (OLCOS) (Baumgartner, et al., 2007), OER Africa (OER Africa, 2009), the African Virtual University (AVU) (Bateman, 2006), China's Open Resources for Education (CORE) (Downes, 2007), Japan's Open Courseware Consortium (JOCW) (Fukuhara, 2008), Teacher Education for Sub-Saharan Africa (TESSA) (Moon & Wolfenden, 2007), the European educational digital library project 'Ariadne' (Duval, et al., 2001), eVrest, which links Francophone minority schools across Canada (Richards, 2007), and Blended Learning Open Source Science or Math Studies (BLOSSOMS) (Larson & Murray, 2008). Over the past decade, these initiatives have accumulated large volumes of OER which are made openly available to the public for use and re-use. However, ironically, the sheer volume of the resources available and the increasing number of repositories have become a major stumbling block in terms of locating *fit-for-purpose* (Calverley & Shephard, 2003) resources for academic purposes.

The most common method of OER search is via generic search engines such as Google, Yahoo! or Bing. Even though this method is the most commonly used, it is not the most effective, as discussed by Pirkkalainen & Pawlowski (2010, p. 24) who argue that

"*…searching this way might be a long and painful process as most of the results are not usable for educational purposes*". Consequently, alternative methods such as Social-Semantic Search (Piedra, Chicaiza, López, Tovar, & Martinez, Finding OERs with Social-Semantic Search, 2011), DiscoverEd (Yergler, 2010) and OCW Finder (Shelton, Duffin, Wang, & Ball, 2010) have been introduced. Semantic web based alternatives such as Agrotags (Balaji, et al., 2010) have also been proposed, which build ontologies of domain specific keywords to be used for classification of OER belonging to a particular body of knowledge. However, the creation of such ontologies for all the domains discussed within the diverse collection of OER would be next to impossible. This research project attempts to provide viable solutions for these problems.

1.1.1 Problem Statements

A majority of the existing OER initiatives are based on established web based technology platforms and have accumulated large volumes of quality resources which are shared openly. However, **one limitation inhibiting the wider adoption of OER is the current inability to effectively search for academically useful OER from a diversity of sources** (Yergler, 2010).

This limitation in locating *fit-for-purpose* resources is further heightened by the heterogeneity of the vast array of OER repositories currently available online. As a result, West & Victor (2011) argue that **there is no single search engine which is able to locate resources from all the OER repositories**.

According to Dichev & Dicheva (2012), **one of the major barriers to the use and re-use of OER is the difficulty in finding quality OER matching a specific context** as it can take as much time as creating one's own materials.

Unwin (2005) argues that **the problem with open content is not the lack of available resources on the Internet but the inability to effectively locate suitable resources for academic use**.

The Paris OER Declaration (UNESCO, 2012, p. 1) states the challenge for more research in this area as a need to *"**encourage the development of user-friendly tools to locate and retrieve OER that are specific and relevant to particular needs**"*.

In sum, **this research project aims to demonstrate how to facilitate the effective centralized search of Open Educational Resources (OER) from heterogeneous repositories for academic purposes**.

1.1.2 Research Objectives

The objectives of this research project are to:

- **Identify user difficulties in searching OER for academic purposes;**

- **Identify the limitations of existing OER search methodologies with respect to locating fit-for-purpose resources from the heterogeneous repositories;**

- **Conceptualize a framework for parametrically measuring the suitability of an OER for academic use;**

- **Design a technology framework to facilitate the accurate centralized search of OER from the heterogeneous repositories.**

1.2 Research Approach

The research approach adopted in this project consists of six distinct phases, as shown in Figure 1.1.

Phase 1 of the project conceptualizes the problem by identifying the key issues which need to be addressed. In addition, a literature review explores existing findings within the problem domain.

Based on the literature review, a survey instrument is developed in Phase 2. This instrument is used to gather information from key stakeholders on their experience in OER search. The key variables are identified through the survey data analysis. Parallel to the survey, desk research is conducted to review past research projects which had attempted to address similar or related issues.

The variables identified from the survey and the insights gained from the desk research are fed into Phase 3, where a conceptual technology framework is developed. This framework addresses the problem of measuring the usefulness of an OER for academic purposes.

Phase 4 concentrates on the implementation of the conceptual framework, using a software based prototype system. The prototype system is tested and evaluated in a controlled environment during Phase 5.

The project lifecycle is documented in Phase 6 in the form of a thesis. The complete research project spans a duration of 36 months. The duration and the deliverables for each phase are shown in Table 1.1.
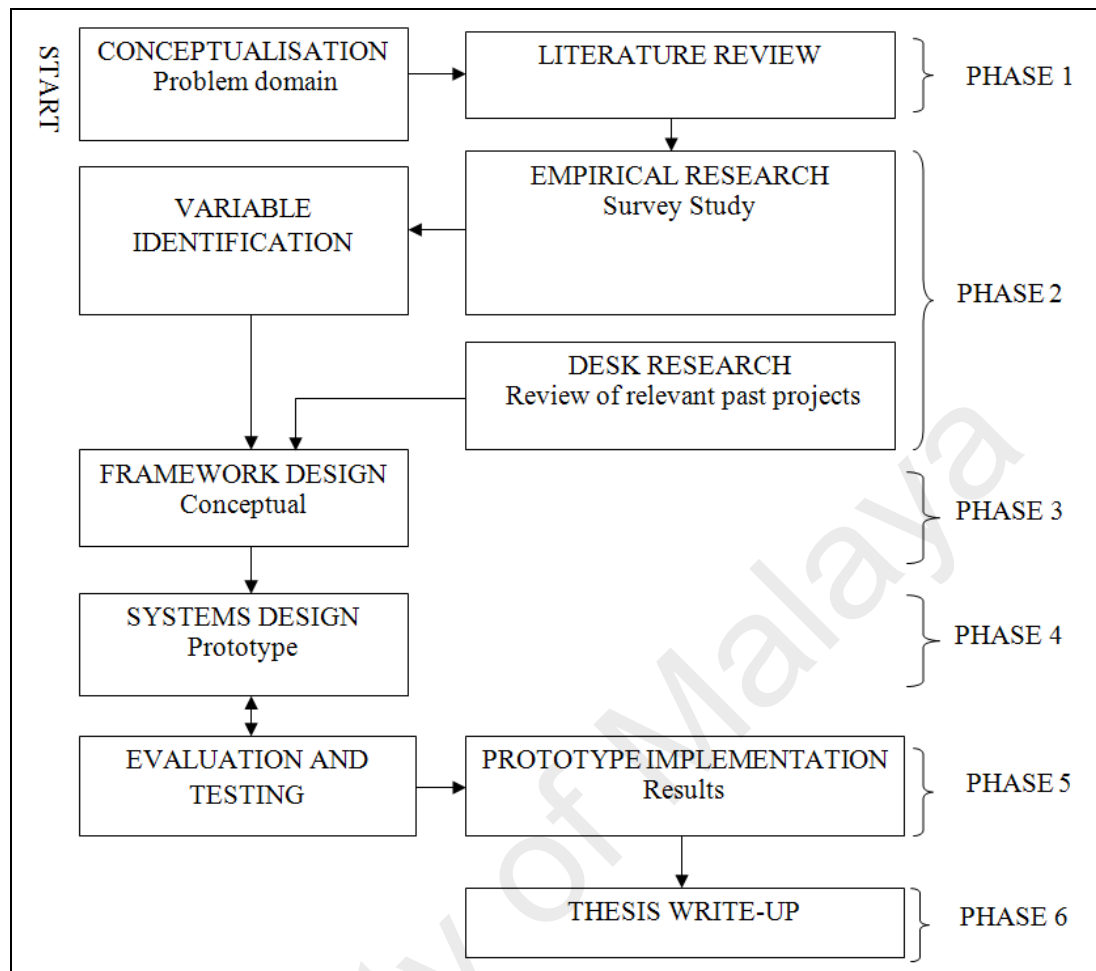
Figure 1.1 Six phases of the research project. This figure documents the flow and the relationships between the phases.

Table 1.1 Duration and deliverables for each of the research phases.

Phase	Activity	Duration	Deliverables
1	Conceptualization	6 months	Problem statement, Literature review
2	Variable definition	3 months	Web based survey instrument, Set of variables
3	Framework design	6 months	Conceptual framework
4	System design	9 months	Prototype system
5	Evaluation and testing	6 months	Test results
6	Thesis write-up	6 months	Thesis

1.3 Research Contributions

Section 1.1.1 discussed the main problems encountered in terms of effective OER search. In this regard, the contributions of this research project are twofold:

1. A major problem in OER search is the difficulty in finding quality OER matching a specific context suitable for academic use. This is due to the lack of a framework which can measure the usefulness of OER in terms of fit-for-purpose, taking into consideration the key attributes of an OER. **The first contribution of this research project is a conceptual framework which can be used by search engines to parametrically measure the usefulness of an OER, taking into consideration the attributes of openness, accessibility and relevance**.

   o The advantage of this framework is that, using the well-established four R's and ALMS frameworks, it can restructure search results to prioritize the resources which are the easiest to reuse, redistribute, revise and remix. As a result, academics practicing Open and Distance Learning (ODL) can locate resources which can be readily used in their teaching and their students' learning.

2. Another major problem encountered is the inability to effectively search for academically useful OER from a diversity of sources. The lack of a single search engine which is able to locate resources from all the heterogeneous OER repositories further adds to the severity of this issue. **The second contribution of this research project is to develop a novel search mechanism which uses text mining techniques and a faceted search interface to provide a centralized OER search tool to locate useful resources from the heterogeneous repositories for academic purposes**.

   o One of the key advantages of this novel search mechanism is the ability to autonomously identify and annotate OER with domain specific keywords.

9

This removes human error with respect to annotation of metadata as it is done in a consistent and uniform manner by the system. As a result, this novel search mechanism provides a central search tool which can effectively search for OER from any repository, regardless of the technology platforms or metadata standards used.

o Another major advantage of this novel search mechanism is the utilization of the conceptual framework which can parametrically measure the usefulness of an OER in terms of fit-for-purpose. This ability allows the search mechanisms to restructure the search results returned from numerous repositories, giving priority to the most open, most accessible and most relevant resources. As a result, academics are able to easily locate high quality OER from around the world which best fit their academic purposes.

1.4 Outline of Chapters

This introductory chapter provides a brief overview of the concept of OER followed by the research problem within broader domain of OER. It then defines the research objectives and discusses the methodology used. The chapter also outlines the significance of the research within the academic community.

Chapter 2 reviews recent literature relevant to the problem domain. The first section explores the concept of OER in detail with respect to definition, copyright, resource formats, creation, curation, delivery, policy, funding, sustainability, impact and future direction. The key focus of the remaining sections are on current issues with OER search, OER curation, existing OER search approaches and approaches to knowledge extraction. The chapter also provides a detailed discussion of existing methodologies and technologies while establishing the need for an improved methodology for OER search.

Chapter 3 is a detailed discussion on the methodology used in this research project. The key areas covered are the initial survey study, the design of the conceptual framework, the design of the technology framework and the prototype development.

Chapter 4 provides the results of the project. Four sets of results are discussed with respect to the survey, conceptual framework, prototype implementation and user tests.

Chapter 5 is the discussion chapter. This chapter critically reviews the complete project with respect to the problem statement, objectives, methodology and results.

Chapter 6 provides conclusions of the project. It also highlights the contributions, advantages and the applications beyond the scope of this project. It also discusses the future direction of the project.

Summary

OER comprise a relatively new phenomenon which is widely regarded as a means of increasing access to education. The free and open nature of OER allows the academic community to legally use, reuse, remix and redistribute material without paying royalties to publishers. This distinct characteristic of OER is of special benefit to developing countries in the 'Global South' which are struggling to meet increasing demand for education. It should also be noted that in common law countries small portions of copyright material can be used, reused, revised and remixed under "fair dealing" or "fair use". However, this rule does not allow full use of the material.

Despite generous funding by governments and non-governmental organizations alike, OER are still yet to become mainstream academic practice, due to a range of inhibitors. One of the major inhibitors contributing to the slow uptake of OER is the inability to search for materials which are suitable for academic purposes. A key aspect of this limitation is the current inability of mainstream search engines such as Google, Yahoo! and Bing to locate OER which are of an acceptable academic standard. As a solution to this issue, this research project proposes a conceptual technology framework which can parametrically measure the usefulness of an OER for academic purposes. It also proposes a technology framework which utilizes text mining techniques to facilitate the effective zeroing in on materials which are of acceptable academic standard.

This chapter has provided a general introduction to the problem domain, a brief overview of OER, the problem statement, research objectives and the significance of the research project. It has also outlined the research approach adopted for the project. The next chapter will review the relevant literature describing the concept of OER, problem domain and the technology research which address related or similar issues.

CHAPTER 2

LITERATURE REVIEW

Chapter 2 : Literature Review

There have been many research initiatives, both academic and commercial, in the recent past with respect to providing a viable solution to the OER search problem. These initiatives range from standardization of metadata to innovative approaches such as the semantic web. However, many of these projects haven't proceeded beyond the prototype stage, indicating of the difficulty of the issue and the volatile nature of the whole concept of OER.

This chapter looks at how the academic community has attempted to provide solutions to the OER search dilemma. The literature review will explore the extent of the current problem, some established standards and a few important technologies which can be directly utilized to provide alternatives to the existing OER search mechanisms.

The rest of the chapter is divided into five sections. Firstly, the constantly evolving concept of OER is discussed with respect to definition, copyright, media formats, creation, curation, delivery, funding models, sustainability, impact and future direction. This provides a backdrop to the current OER search dilemma addressed through this research project. The second section tries to identify the extent of the current OER search problem and the reasons behind the inadequacy of existing mainstream search engines. The third section discusses existing OER metadata standards used for OER curation. The fourth section looks at some of the more promising initiatives in OER search. The last section highlights a few technologies which are utilized in this research project to provide an innovative solution to the problem.

2.1 Open Educational Resources

With the dramatic changes taking place in Higher Education (HE) within the past 10 years, academics have had to adopt new cost effective approaches in order to provide individualized learning to a more diverse student base (Littlejohn, Falconer, & Mcgill, 2008). In this context, OER have the potential to become major sources of freely reusable teaching and learning resources, especially in higher education, due to active advocacy by organizations such as UNESCO, COL, Organization for Economic Co-operation and Development (OECD); and the International Council for Open and Distance Education (ICDE).

2.1.1 Definition

The definition of OER has evolved since its inception in 2002. However, it is generally accepted that OER are web based educational materials which are freely and openly available for use, reuse, remix and redistribution. It is noted that OER can exist in forms other than web based material.

The openness and freedoms of OER are governed by a set of globally accepted conventions. These can be best explained through the four R's model proposed by Hilton, Wiley, Stein & Johnson (2010).

The four R's model:

- *Reuse* – the most basic level of openness. People are allowed to freely use all or part of the unaltered, verbatim work.

- *Redistribute* – people can share copies of the work with others.

- *Revise* – people can adapt, modify, translate, or change the form of the work.

- *Remix* – people can take two or more existing resources and combine them to create a new resource.

The openness of a resource increases with the number of 'R's governing the freedoms, as shown in Figure 2.1.
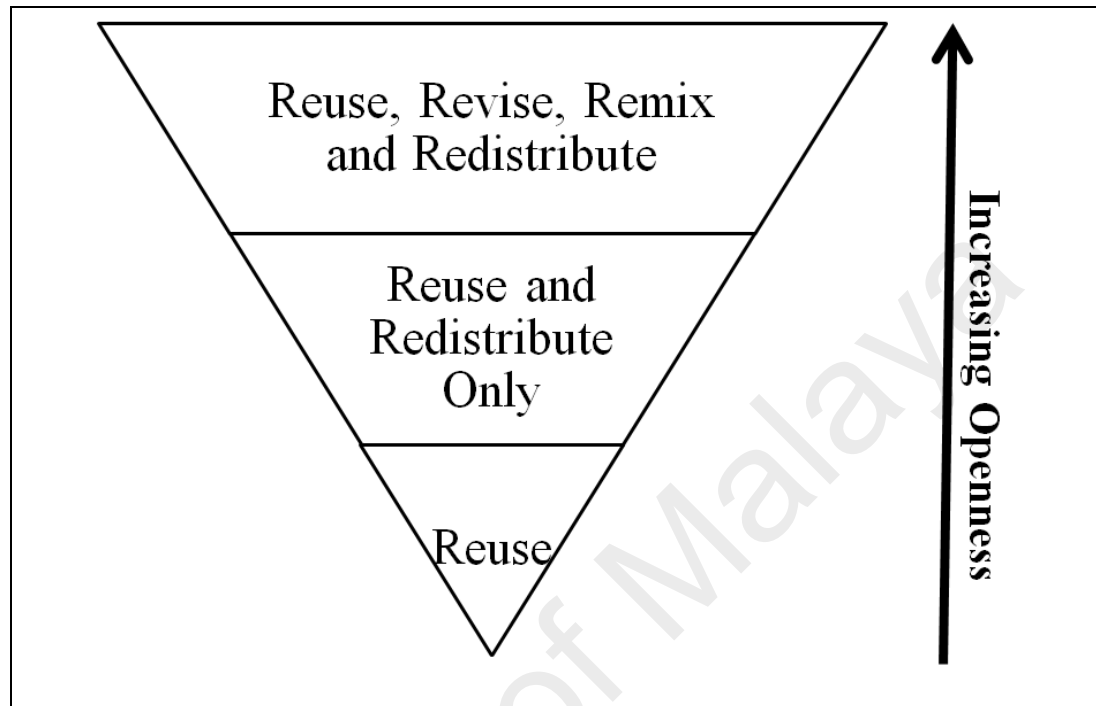
Figure 2.1 Increasing openness of the four R's: adapted from (Hilton et al., 2010).

2.1.2 Copyright

With the opening up of content to a global audience come the challenges of managing copyright and intellectual property rights (IPR). According to Fitzgerald (2006, p. 4)

"…*while the new digital technologies possess an enormous capacity to disseminate knowledge, copyright law will play a key role in determining the legality of any such act*".

Currently there are several open content licensing (OCL) schemes such as the Creative Commons (CC) and *GNU Free Documentation Licensing,* among others (Hylén, 2006). These schemes introduce certainty and clarity in terms of obtaining permission to legally use the work of others. There are also institutional or group specific licenses

such as the *BC Commons* (Stacey, 2006) which limits the usage of published resources to a particular group or institution.

Among the various licensing schemes available, the CC licensing scheme is arguably the most widely used due to its simplicity, legal robustness and the large number of regional chapters. This licensing scheme is currently in its fourth generation as 'Creative Commons 4.0' which was officially launched at the end of 2013. However, 'Creative Commons 3.0', which was its immediate predecessor, is by far the most widely used at present. The Creative Commons 3.0 licensing scheme can be divided into two forms: (i) *unported,* which abides by international copyright law and is not subject to regional jurisdictions; and (ii) *ported,* which is a version customized to suit the copyright laws of a particular region or jurisdiction. The Creative Commons 3.0 *unported* license grants six specific freedoms as shown in Table 2.1.

Table 2.1 Creative Commons 3.0 unported licensing scheme: adapted from **(Creative Commons)**.

	License	Freedoms Granted
1.	Attribution (CC BY)	This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.
2.	Attribution-ShareAlike (CC BY-SA)	This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to "copyleft" free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.

3.	Attribution-NoDerivs (CC BY-ND)	This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.
4.	Attribution-NonCommercial (CC BY-NC)	This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.
5.	Attribution-NonCommercial-ShareAlike (CC BY-NC-SA)	This license lets others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms.
6.	Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)	This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially.

2.1.3 Media Formats

Despite the fact that OER were initially limited to text based material and are still predominantly in text based formats (Wiley, 2006), they are not restricted by the media types or the file types used. Many modern OER are released as images, movie clips, animations, datasets, audio clips, podcasts etc., providing rich multimedia based material for use and reuse. These multimedia resources are made available through large repositories such as YouTube (video), Flickr (images) and iTunesU (podcasts) under the CC licensing scheme. Repositories such as YouTube even provide native software applications such as the YouTube Video Editor which facilitates the easy reuse and remixing of these multimedia based resources in an online setting. However, the accessibility of these resources, with respect to the four R's, needs to be considered before using them for teaching and learning purposes.

The accessibility governing various formats of OER can be best explained using the ALMS analysis proposed by Hilton et al. (2010).

The ALMS analysis:

- *Access to editing tools* – how accessible are the software tools needed to reuse the resource? The accessibility depends on the cost and availability (e.g. can free and open source software (FOSS) be used to edit a resource instead of proprietary software).

- *Level of expertise required to revise or remix* – how easy is it to revise or remix a resource without advanced technical skills or specialist knowhow? For example, text based documents can be easily revised or remixed in contrast to movie clips or animations.

- *Meaningfully editable* – can the resource be reused or remixed with less time and effort than is needed to create it from scratch? For example, scanned documents are difficult to reuse or remix. It is better to create them from scratch in an editable format.

- *Source file access* – does the resource provide access to an editable source file which can be used to reuse or remix? For example, an animation might not provide the editable storyboard needed to reuse or remix.

2.1.4 Creation and Curation

Creation and use of OER are very much dependent on the technologies which enable collaboration and information sharing. In recent times, many projects and initiatives have enabled the development and sharing of OER over the web. Web 2.0 is commonly known as the platform which largely underpinned the rise of OER. Within this context, social software have taken center stage in terms of enabling learners and educators to create and share OER using wiki, blogging and social networking (Piedra, Chicaiza,

Tovar, & Martinez, 2009). Among these technologies, Wiki plays a central role in the present day OER arena. According to Leuf & Cunningham (2001), Wiki is a software tool that promotes and mediates discussion and collaboration between different users. In addition to *WikiEducator*, projects such *Wikibooks*, *Wikimedia Commons* and *Wikiversity* are also among the popular Wiki based OER repositories. The *Wikipedia* OER repository is largely credited as the pioneer user of the Wiki concept. Another widely used platform for OER creation is *Rhaptos,* developed by Rice University. This platform hosts the popular *Connexions* OER repository which allows the easy creation, use and re-use of text based learning objects (LO). The *Rhaptos* platform is currently also being used by other repositories such as Vietnam Open Educational Resources (VOER) under Free and Open Source Software (FOSS) licenses. When considering institutional OER repositories, the popular DSPACE repository system is the most commonly used due to its compatibility with existing library systems and protocols. However, DSPACE only acts as a repository of content and does not provide features which facilitate reuse and remix of resources.

According to McGreal (2010), modern OER repositories can be classified into three categories:

- *Content repositories* – hosts content internally within the repository (e.g. Connextions, WikiEducator);
- *Portal repositories* – provides searchable catalogues of content hosted in external repositories (e.g. OER Commons, DOER); and
- *Content and portal repositories* – hosts content internally in addition to providing catalogues of content hosted externally (e.g. MERLOT, JORUM).

The attribute common to all of these repositories is the use of metadata for resource curation. Among the key metadata are title, description, keywords, license type and

author. These are defined according to established metadata standards such as Dublin Core Metadata Initiative (DCMI) and IEEE Learning Object Metadata (IEEE LOM). However, key concerns regarding OER curation are the standardization of metadata across repositories and ensuring the integrity of the metadata annotated by content creators. The manual cataloguing of OER has also become an issue, due to the constant expansion in volume.

2.1.5 Delivery

With the availability of more and more OER, research has also turned towards the effective delivery of these materials through the Internet. Among the methods of delivery are Digital Talking Books (DTB) which speak out the content (Brasher, 2007); audio, video podcasts and Really Simple Syndication (RSS) (Cann, 2007); naturalistic video conferencing (Tomadaki & Scott, 2007); knowledge mapping (Shum & Okada, 2007); instant messaging (Little, Eisenstadt, & Denham, 2007); accessible SCORM content (Douce, 2007); Open Learning Objects (OpenLO) (Fulantelli, Gentile, Taibi, & Allegra, 2007) and LeMill web community for sharing OER (Toikkanen, 2008). Complementing this research, other studies are being conducted which aim to identify the effectiveness of the delivery methods using techniques such as non-intrusive eye tracking, remote desktop sharing and browsing logs (San Diego, 2007).

2.1.6 Funding and Sustainability

Another key area of debate and constant research is the sustainability of OER. According to Wiley (2006, p. 5), sustainability in an OER project can be defined as "…*the ability of a project to continue its operations*". Sustainability can be divided into financial sustainability and resource sustainability. There are many sustainability models such as the MIT model, Rice model and USU model, among others, all of which have their own benefits and drawbacks (Wiley, 2006). However, the major hurdle any OER initiative needs to overcome is the financial sustainability. Most of the

modern OER initiatives can be categorized into the funding models proposed by (Downes, 2007). Each model has its own merits and demerits. OER initiatives need to pay special attention to the funding model adopted to ensure long term sustainability.

Funding models for OER initiatives:

- *Endowment Model* – the project is sustained on the interest earned on base funding.

- *Membership Model* – organisations interested in the project are invited to contribute a certain sum as seed funding or recurring subscriptions.

- *Donations Model* – the project receives donations from the wider community.

- *Conversion Model* – the users of the free material are converted into paying customers for value added services.

- *Contributor-Pay Model* – the contributors pay for the maintenance of materials which are made freely available.

- *Sponsorship Model* – draws income from sponsorship such as advertising.

- *Institutional Model* – an institution assumes responsibility for the initiative and absorbs the costs.

- *Governmental Model* – national governments or government agencies assume responsibility for the initiative and absorbs the costs.

- *Partnerships and Exchanges* – multiple institutions and organisations communally contribute to the project.

2.1.7 Impact

Farber (2009, p. 28) states that:

"*just as the Linux operating system and other open source software have become a pervasive computer technology around the world, so too might OER materials become the basis for training the global masses*".

This statement clearly outlines the significance of OER as a global movement. Claims have also been made by Caswell et al. (2008) that the move towards OER can contribute to reduce the costs of learning. Initiatives such as OCW, Connexions and WikiEducator help those who reuse these freely available materials in bringing the costs down. As a result, institutions and individuals globally can adapt and reuse material without investing in developing them from scratch. Therefore, OER can contribute to broaden access and provide equity in education. This is especially important for countries in the Global South such as India, which has 411 million potential students, of which only 234 million enter school at all, less than 20% reach high school and less than 10% graduate (Kumar, 2009).

2.1.8 Future Direction

The concept of OER is subject to constant evolution. The latest incarnation of this concept is in the form of Massive Open Online Courses (MOOC). Daniel (2012) argues that the concept of MOOC is also constantly evolving, trying to define itself within the open education movement. In his article he quotes the Wikipedia definition which states that "*a MOOC is a type of online course aimed at large-scale participation and open access via the web*" (Daniel, 2012, p. 3). It is accepted that the concept of MOOC originated in Canada in the form of cMOOC (Billsberry, 2013). According to Baggaley (2013), Stanford University's *Udacity*, which was launched in February 2012, is credited as the first xMOOC. It also doubles as a commercial entity providing services to new MOOC startups. In April 2012 *Coursera* was launched by Stanford followed by Harvard and MIT who launched *edX* in May 2012.

Coursera claimed to have, at the end of 2012, over 1.4 million learners enrolled in more than 200 courses offered by 33 partner institutions (Lewin, 2012) (DeSantis, 2012). By 2014 this number had increased to 6 million "Courserians" (Knox, 2014, p.

165). All of these comprehensive courses are openly and freely made available to global learners, potentially bridging the knowledge divide. However, it should also be noted that xMOOCs are not always made available as OER. With the rapidly expanding volume of MOOC being offered throughout the world, the necessity for purpose built search mechanisms capable of locating useful resources will continue to increase. Within this context, the next section discusses the problem investigated by this research project.

2.2 The OER Search Dilemma

OER are fast gaining attraction within the academic community as a viable means of increasing access and equity in education. The concept of OER is of especial significance to marginalized communities where distance education is prominent due to the inability of conventional brick and mortar institutions to cope with the growing demand (Lane, 2009). However, the wider adoption of OER by academics has been inhibited due to various socio, economic and technological reasons (D'Antoni, 2009). One of the major technological inhibitors is the current inability to search for OER which are academically useful and are of an acceptable academic standard.

2.2.1 Issues Related OER Search

In his study into identifying the inhibiting factors for reuse experienced by content developers in developing countries, Hatakka (2009) points out that the most inhibiting factor is the inability to locate '*relevant*' material for a particular teaching or learning need. Relevance in this context is best explained by William Geoffman (1964) who argues that relevance is a measure of information conveyed by a document relative to a search query. However, Geoffman also states that the relationship between the document and the query is not sufficient to determine the relevance.

The subjects of Hatakka's study attribute the inability to locate relevant material to (i) the inability to locate resources which fit the scope of the course in terms of context and difficulty; (ii) the lack of awareness with respect to how 'best' to search for material on the Internet; and (iii) the inability to choose the most appropriate resources from the large number of resources returned by search engines such as Google.

Affirming the findings of Hatakka's study, Shelton et al. (2010, p. 316) argue that:

"Well-studied and commercialized search engines like Google will often help users to find what they are seeking. However, if those searching do not know exactly what they are looking for, or they do not know the 'proper' words to describe what it is that they want, the searching results returned are often unsatisfactory".

In an attempt to identify the effectiveness of mainstream search engines such as Google in locating relevant OER, Dichev et al. (2011) of the Winston-Salem State University conducted an experiment by comparing Google side by side with native search mechanisms of OER repositories. To narrow the Google search in terms of OER, the advanced search feature '*free to use, share or modify, even commercially*' was used. Alongside Google, native search mechanisms of 12 OER repositories were used to search for material in the computer science domain. The repositories were: Connexions, MIT OpenCourseWare, CITIDEL, The Open University, OpenLearn, OpenCourseWare Consortium, OER Commons, Merlot, NSDL, Wikibooks, SOFIA, Textbook Revolution and Bookboon. Table 2.2 shows the comparison between Google and native OER search mechanisms in locating relevant material. It is apparent from this comparison that the native search mechanisms are more effective than Google in terms of locating relevant material.

Table 2.2 The comparison between Google and native search mechanisms of OER repositories in terms of locating relevant material.

No	Search Keywords	Found Relevant	
		Google	OER
1	analysis algorithms spanning trees	0	0
2	artificial intelligence reasoning systems	0	0
3	computer architecture pipelining	1	3
4	computer networks congestion control	1	6
5	database management normalization	1	9
6	data structures binary trees	2	3
7	introduction programming java inheritance	1	3
8	operating systems synchronization java	1	1
9	programming languages data types	1	0
10	software engineering agile development	1	4
11	artificial intelligence definite clause grammars	0	0
12	e-commerce web services	0	0
13	hardware media security blue-ray prm	0	0
14	information system security access matrix model	0	0
15	information system security public key infrastructure	0	0
16	web programming php arrays	3	2
17	web programming xml xpath	0	0

Commenting on the inability of mainstream search engines such as Google to effectively locate OER, Pirkkalainen & Pawlowski (2010, p. 24) state that *"... searching this way might be a long and painful process as most of the results are not usable for educational purposes"*. Furthermore, they argue that search mechanisms native to OER repositories are capable of locating resources with an increased relevance. However, a problem is the choice of repositories within the large global pool. Levey (2012, p. 134) relates this to her experience working in the African 'AgShare' project:

"Despite numerous gateways, it is not always easy to identify appropriate resources. How a resource is tagged or labelled is one problem. Poor information retrieval skills is another. Furthermore, academics are busy".

This inadequacy with respect to searching for OER from a diversity of sources gives rise to the need for new alternative methodologies which can assist in locating relevant resources. Ideally these search tools should return materials which are relevant, usable and from a diversity of sources (Yergler, 2010). Yergler further suggests that the reliance on a full text index and link analysis of mainstream search engines impede the process of discovery by including resources not necessarily educational. As such, *"increasing the relevance of the resources returned by a search engine can minimize the time educators need to spend exploring irrelevant resources"* (Yergler, 2010, p. 2).

The Paris OER Declaration (UNESCO, 2012), which is a global non-binding declaration signed by many governments, declares the need for more research into OER search as:

"i. Facilitate finding, retrieving and sharing of OER: Encourage the development of user-friendly tools to locate and retrieve OER that are specific and relevant to particular needs. Adopt appropriate open standards to ensure interoperability and to facilitate the use of OER in diverse media" (UNESCO, 2012, p. 1).

This declaration is the culmination of a global effort towards establishing a roadmap for the future development of the OER movement. The above recommendation made with respect to OER search reaffirms the need for new and more effective OER search methodologies within the context of locating relevant material for particular teaching and learning needs.

2.2.2 Metadata

The majority of existing search methodologies, including mainstream search engines such as Google, work on the concept of metadata for locating educational resources. The use of metadata as opposed to full text search makes the search process faster and more efficient.

According to Anido et al. (2002, p. 359),

"Educational metadata provides information about educational resources…. As the available educational resources grow and grow, the need for metadata becomes apparent. The lack of information about the properties, location or availability of a resource could make it unusable…. Metadata contributes to solve this problem by providing a standard and efficient way to conveniently characterize resource properties".

In terms of characterizing resource properties, the quality of the metadata is an important factor. Barritt et al. (2004) argue that the quality of metadata can be evaluated from two different perspectives, namely (i) its validity in terms of its ability to describe the resource; and (ii) its usefulness for "searchability" and how well it supports retrieval of the resource.

Considering metadata standards, there are many being used to systematically annotate educational resources. However, Devedzic et al. (2007) argue that the term "standard" is used colloquially by the e-learning community to describe:

- *official standard*: a set of definitions, requirements, formats, and design guidelines for e-learning systems or their components that a recognized standards organization has documented and approved.

- *de facto standard:* the same as an official standard, but widely accepted only by the community and industry—that is, lacking formal approval from a recognized standardization body.

- *specification:* a document on the same issues as an official standard, but less evolved; usually developed and promoted by organizations or consortia of partners from academia, industry, and educational institutions. It captures a rough consensus

in the e-learning community and is used as a de facto standard in system and content development.

- *reference model:* an adapted and reduced version of a combination of standards and specifications focusing on architectural aspects of an e-learning system, definitions of parts of the system, and their interactions.

According to the definition of the word "standard", the authors categorize a few of the current standards which are widely used in the academic community as shown in Table 2.3.

Among these various standards, the IEEE Learning Object Metadata (LOM) (IEEE Learning Technology Standards Committee, 2005) is the official standard adopted by many OER repositories. The standard allows resources to be tagged according to nine key categories. These categories are (i) General; (ii) Life Cycle; (iii) Meta-Metadata; (iv) Technical; (v) Educational; (vi) Rights; (vii) Relation; (viii) Annotation; and (ix) Classification. A schematic representation of the LOM standard is shown in Figure 2.2. Ongoing research constantly looks into extending the LOM standard to identify various facets of education. One example is the addition of the "Competence" category proposed by Sampson (2009) to facilitate competence-based learning. This ability to describe various technical and educational information, in addition to general metadata used for search purposes, makes the LOM standard a popular choice for describing OER.

Table 2.3 Selected e-learning standards (Devedzic et al., 2007).

Standard	Developer	Description	Level	Type
Learning Object Metadata (LOM; http://ltsc.ieee.org/wg12)	IEEE	A conceptual structure for a metadata instance based on a hierarchy of nine categories of metadata elements.	Official standard	Metadata standard
IMS Content Packaging (www.ims global.org/content/packaging/ cpv1p4/imscp_infov1p4.html)	IMS Global Learning Consortium	A set of structures that enable exchanging learning content.	Specification	Packaging standard
IMS Learning Design (LD; www.ims global.org/learningdesign/ldv1p0/ imsld_infov1p0.html)	IMS Global Learning Consortium	A standardized framework for integrating learning (instructional) design with content packages.	Specification	Packaging standard
IMS Learner Information Package (LIP; www.imsglobal.org/profiles/ lipinfo01.html)	IMS Global Learning Consortium	A set of structures that enable recording and managing the learner's characteristics as well as exchanging learner information among applications.	Specification	Learner information standard
Public and Private Information (PAPI) for Learners (PAPI Learner) (www.edutool.com/papi/drafts/08/)	IEEE	Semantics and syntax of learner information; elements for recording and viewing descriptive information about learners from different perspectives.	Specification	Learner information standard
Learning Technology Systems Architecture (LTSA; http://ltsc.ieee. org/wg1/files/IEEE_1484_01_D09 _LTSA.pdf)	IEEE	High-level, pedagogically neutral, content-neutral, culturally neutral, and platform/technology-neutral architecture for learning technology systems based on abstract components.	Reference model	Communication standard (reference architecture)
Sharable Courseware Object Reference Model (Scorm; www.adlnet. org/scorm/history/2004/index.cfm)	ADL Initiative	A unified content and communication reference model for consistent implementation of e-learning systems. The latest release (Scorm 2004) is augmented with learning content sequencing capabilities.	Specification and reference model	A combination of multiple other standards
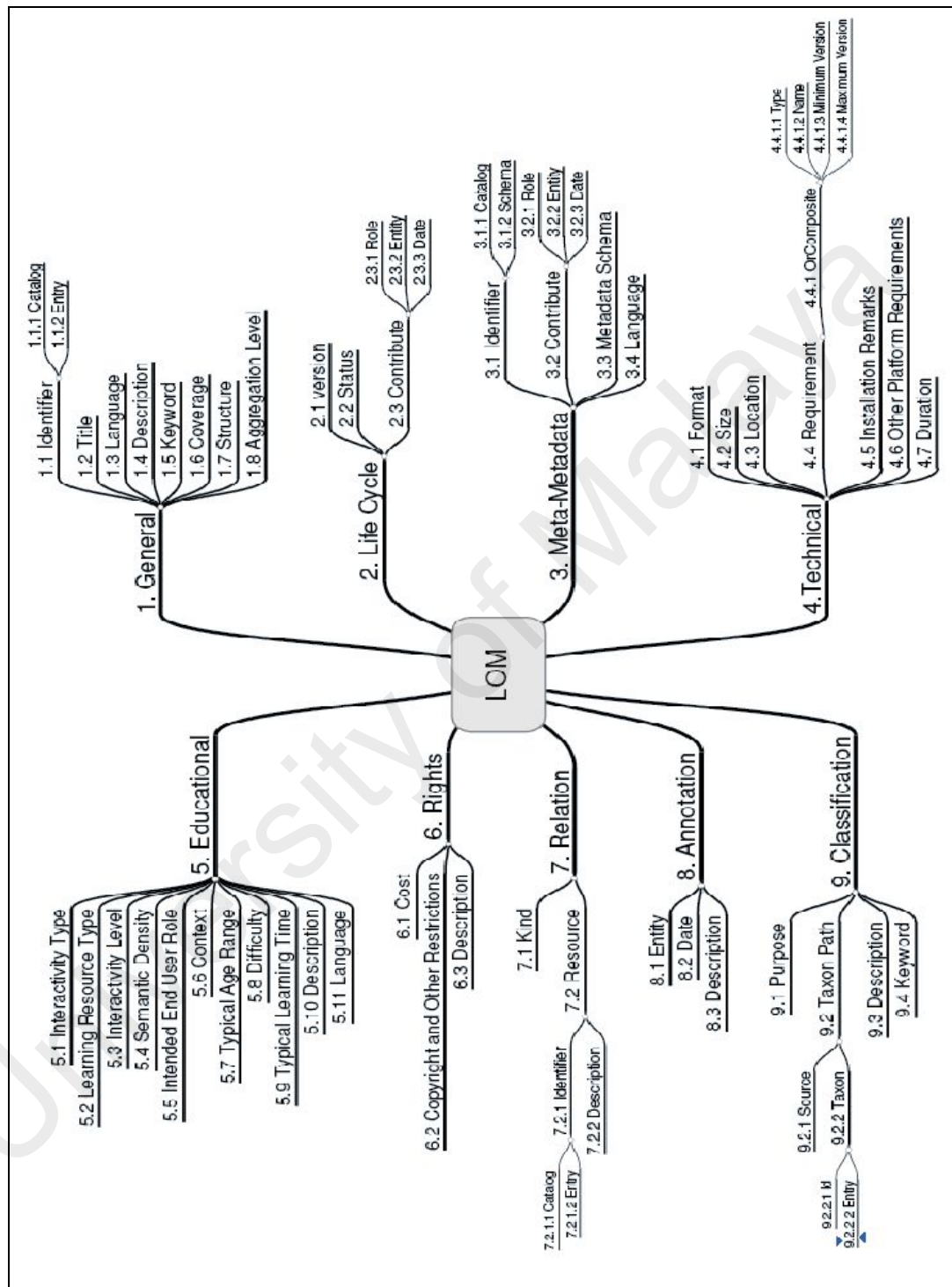
Figure 2.2 A schematic representation of the hierarchy of elements in the LOM data model (Casali et al., 2013).

The LOM standard has been in use since 2002. However, the global academic community has exercised its choice by adopting standards other than LOM such as the Dublin Core Metadata Initiative (DCMI) and IMS Learning Resource Meta-Data (IMS Global Learning Consortium, 2001). Although these standards have reached a high level of interoperability along the years, a truly global standard is needed to facilitate a higher level of accuracy in searching for relevant resources. Potentially answering this call is the Learning Resource Metadata Initiative (LRMI) launched by the Association of Educational Publishers and Creative Commons. This project aims to build a common metadata vocabulary for educational resources. This *common metadata framework* is used for uniform tagging of web based learning resources. According to the official website of the project, the Association believes that:

"*Once a critical mass of educational content has been tagged to a universal framework, it becomes much easier to parse and filter that content, opening up tremendous possibilities for search and delivery*" (Association of Educational Publishers & Creative Commons, p. 1).

The inclusion of LRMI into schema.org, a joint project by Bing, Google and Yahoo! looking at standardizing metadata, is an early indication of the potential global impact of the project.

Regardless of the robustness of existing metadata standards for describing learning resources, these standards still depend solely on the competence of the content creators in terms of metadata annotation. Barton et al. (2003) list some of the key problem areas of metadata as (i) spelling and abbreviations; (ii) author and author contributed fields; (iii) title; (iv) subject; and (v) date. In his study, Tello (2007) classified the errors in metadata as (i) missing - no data were recorded; (ii) syntactic - metadata do not conform to the standards; and (iii) semantic – the metadata values of the elements do

not match the expected information. As such, the human input becomes the weakest link in the whole process. Devedzic et al. (2007, pp. 20-21) describe this issue as:

"... content authors are typically reluctant to provide metadata, so the amount of metadata is usually insufficient.... Thus, a metadata-based query to an LOR for certain LOs might not return the most suitable content for the learner, or learners might have to examine several returned LOs manually to select those that suit their needs. Likewise, it's impossible for authors to predict all possible learning situations when annotating LOs with metadata".

In this context, Brooks & McCalla (2006, p. 52) argue that:

"...metadata formats are typically created with the notion that some human will both be the producer and consumer of the metadata and the learning object content itself. We believe that such heavy reliance on human intervention is costly and mitigates against real-time adaptivity to individual learner needs. Moreover, when annotating learning objects, humans often do not fill in all the fields and even when they do, inter-rater reliability is often quite low.... The lack of reliability between metadata authors appears to be a general trend...".

In sum, the work done by Cechinel et al. (2009), which used graduate students to manually annotate learning resources, suggests that metadata standards such as LOM still have much room for improvement.

2.3. Important OER Search Initiatives

Jones (2007, p. 155) described modern academic repositories as *"next generation"* repositories which have made the shift from independent stand-alone to distributed, federated and highly integrated applications and services. Although there are many repositories of this sort, the number and diversity of resources in these repositories is a major issue when selecting appropriate resources satisfying both teacher and learner requirements (Ouyang & Zhu, 2008). This is especially noteworthy as the popular choices for searching these repositories remain their native search mechanisms and mainstream search engines such as Google. However, there have been several initiatives over the past few years which focus on providing viable solutions to this particular issue from a global perspective. These initiatives can be broadly categorized into federated search and semantic search. Despite showing initial promise, only a handful of these solutions have proceeded beyond the prototype stage. Out of these, the ones which have become global players are mainly commercial ventures or global federations backed by philanthropic funding. The next two sections describe some of the more exciting projects which have emerged to show great potential in both the federated search and semantic search domains.

2.3.1 Federated Search

Among the existing OER search approaches, federated search is considered to be the most viable from a global perspective. This viability arises from the fundamental task of federated search engines which is to search a group of independent collections, and to effectively merge the results they return for queries (Shokouhi & Si, 2011). Pawlowski & Bick (2012, p. 210) state that *"There is currently a strong trend to federate repositories to enable search and re-use for a large number of repositories".* This is achieved either by federated search across different repositories at runtime or by periodically harvesting metadata for offline searching. The authors further speculate

that semantic web technologies will be increasingly used for OER search in the coming years. However, they raise the question of how this can be viably achieved to facilitate ease of use and the retrieval of large amounts of relevant resources.

BRENHET[2] proposed by De la Prieta et al. (2011) is a multi-agent system (MAS) which facilitates federated search between learning object repositories (LOR). It uses an *"Organisational View (OV)"*, as shown in Figure 2.3, to provide federated search facilities in a social environment. In the OV, the search and retrieval of (LO) is segmented into five distinct aspects which are (i) mission; (ii) services; (iii) producers; (iv) product; and (v) consumers. This allows the system to map the relationships among these aspects to efficiently bring better quality LO, produced by the various producers, to the clients who are students, editors and teachers. Based on experimental results, the authors claim that the prototype system is fully satisfactory as the number of results returned has significantly improved without increasing the query time.
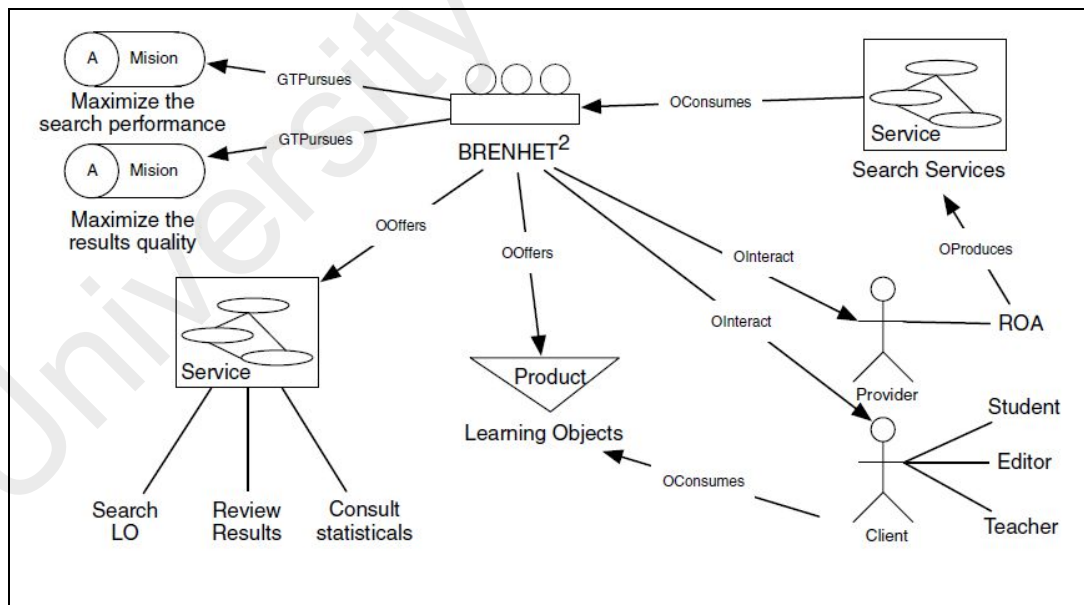
Figure 2.3 Diagram of the organizational model (De la Prieta et al., 2011).

The OpenScout system proposed by Ha et al. (2011) is another example of federated OER search. It copies metadata from existing repositories to create a searchable index of resources accessible from a single location (Figure 2.4). One of the key features of

OpenScout is the adoption of a faceted search approach where users can filter search results according to the properties of the resources. OpenScout currently concentrates on resources in one domain, management education. Another limitation to the expansion of the system is its dependence on external federated metadata.
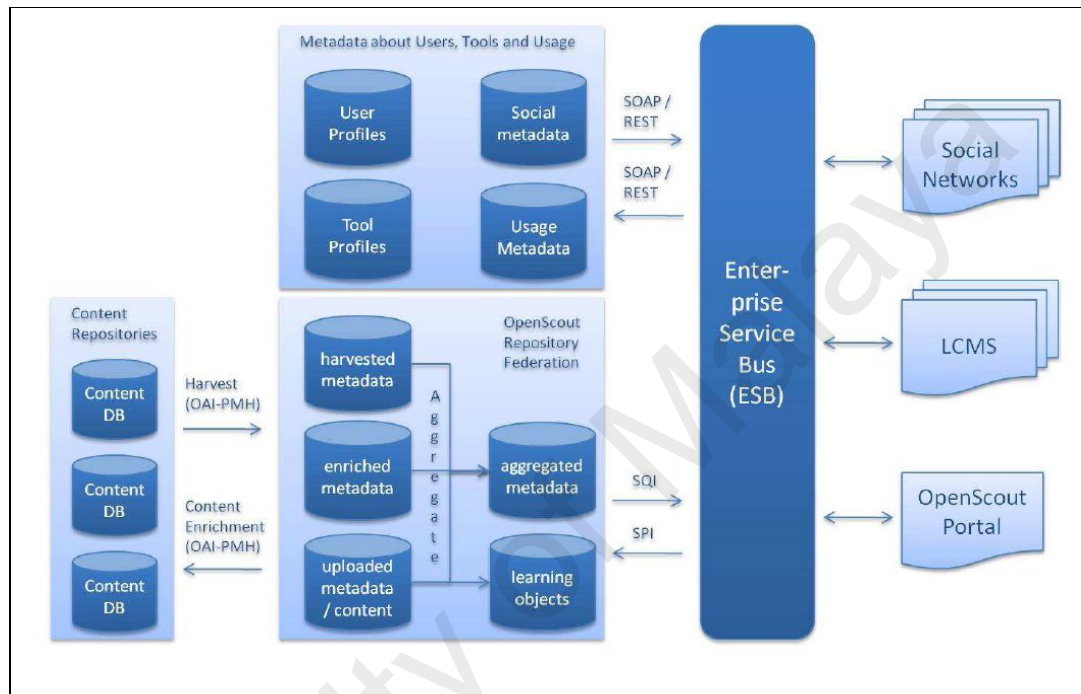
Figure 2.4 The OpenScout architecture (Ha et al., 2011).

Another promising initiative is the Global Learning Object Brokered Exchange (GLOBE) initiative which uses a federated search approach. The GLOBE consortium, which was founded in 2004, has now grown to 14 members representing America, Asia, Australia, Europe and Africa. GLOBE acts as a central repository of IEEE LOM educational metadata harvested from various member institutional repositories. Users are provided with a single sign-on query interface where they can search for resources across repositories, platforms, institutions, languages and regions. As of February 2012 the total number of metadata harvested and available through GLOBE is 817,436 (Yamada, 2013). The consortium is currently expanding its reach to more institutions worldwide. However, the work done by Ochoa et al. (2011) on the GLOBE repository

suggests that although the initiative is promising, there is much room for improvement with respect to the accuracy of the harvested metadata.

One of the more exciting technologies unveiled recently is the *Blue sky* project by the global publishing giant Pearson. This custom search engine specifically concentrates on searching for OER with an academic focus. The platform allows instructors to search for e-book chapters, videos and online exercise software from approximately 25 OER repositories distributed worldwide. However, it gives precedence to e-book material published under Pearson. Irrespective of this possible bias towards its own products, Associate Professor David Wiley of Brigham Young University states that "*the more paths to OER there are in the world, the better*" (Kolowich, 2012, p. 2).

2.3.2 Semantic Search

Semantic search is derived from semantic web technologies, where people are considered as producers or consumers and machines as enablers. The enablers gather, remember and search pools of data making the users' lives easier (Gruber, 1993). According to Gruber's definition, an ontology specifies the conceptualization of a specific domain in terms of concepts, attributes and relationships. When expressed in a formal language, these ontologies can be interpreted and processed by machines.

The OER-CC ontology, introduced by Piedra et al. (2010), is one example of the use of semantic web to better search OER. The ontology was created by combining the LOM2OWL ontology (García, Alonso, & Sicilia, 2008) describing learning resources and the CC ontology created using the METHONTOLOGY (Corcho, Fernández-López, Gómez-Pérez, & López-Cima, 2005) guidelines. The authors claim that the prototype system resulted in the *"short-term improvement"* in information retrieval during their experiments. Piedra et al. (2011, p. 1200) further extended this experiment into the domain of *Social-Semantic Search* through the MIT OCW repository where

they report *"…the semantic search is answering questions reasonably well where data are available".*

Casali et al. (2013) propose another prototype system (Figure 2.5) which builds an ontology based on the IEEE LOM standard. The system combines various existing Application Programming Interfaces (API) in the semantic web domain through an *"Assistant"* built using the Java programming language. The prototype system performs three actions namely restriction, extraction and validation. The authors claim that this *"Assistant"* prototype helps users with respect to loading metadata through automation.
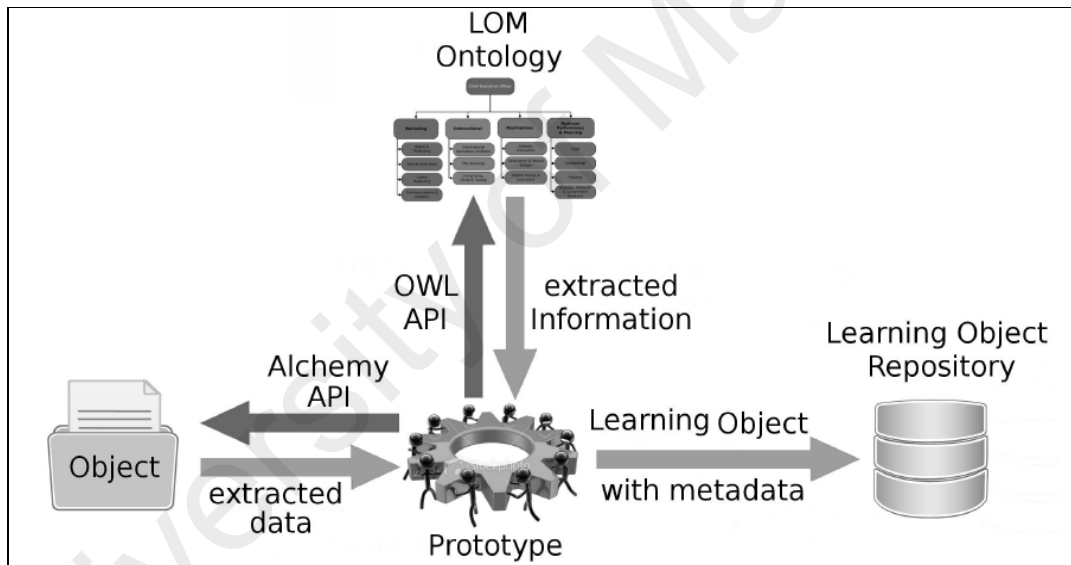
Figure 2.5 The Assistant prototype: Interactions and functionalities (Casali et al., 2013).

Shelton et al. (2010) of the *"Folksemantic"* project propose a hybrid search system for OCW and OER which combines (i) *OCW Finder* - a lightweight interface for searching OCW; and (ii) *OER Recommender* – a content based recommendation system which uses TF-IDF weighting scheme to make recommendations based on metadata such as title, keywords and description. The metadata are harvested from the existing repositories using RSS. RSS provides a "feed" of frequent updates to a particular webpage as full or summarized text coupled with metadata. This allows the syndication

of metadata automatically by the system. The hybrid system uses a semantic web approach to recommend resources based on relevance, attention type, attention details, attention recency and article history. However, the authors state that more research needs to be conducted with respect to recommending resources of higher academic quality in addition to relevance.

A more specific example of the use of semantic web is the *"Agrotags"* project which concentrates on tagging resources in the agriculture domain (Balaji, et al., 2010). The initial ontology was created using an existing base of more than 40,000 words related to agriculture. A module called *"Agrotagger"* was developed to be used as a plug-in module for popular repositories and content management systems (CMS) where resources will be automatically tagged according to the ontology. Agrotagger executes three main tasks, namely (i) identify the agriculture related terms in a document; (ii) create a bag of tags for use based on the identified terms; and (iii) use statistical techniques to calculate the suitability of these terms as keywords. The workflow of Agrotagger is shown in Figure 2.6. One major limitation of this approach is the extensive human input required to create the initial ontology. Therefore, the application of the system to domains other than agriculture remains a challenge.
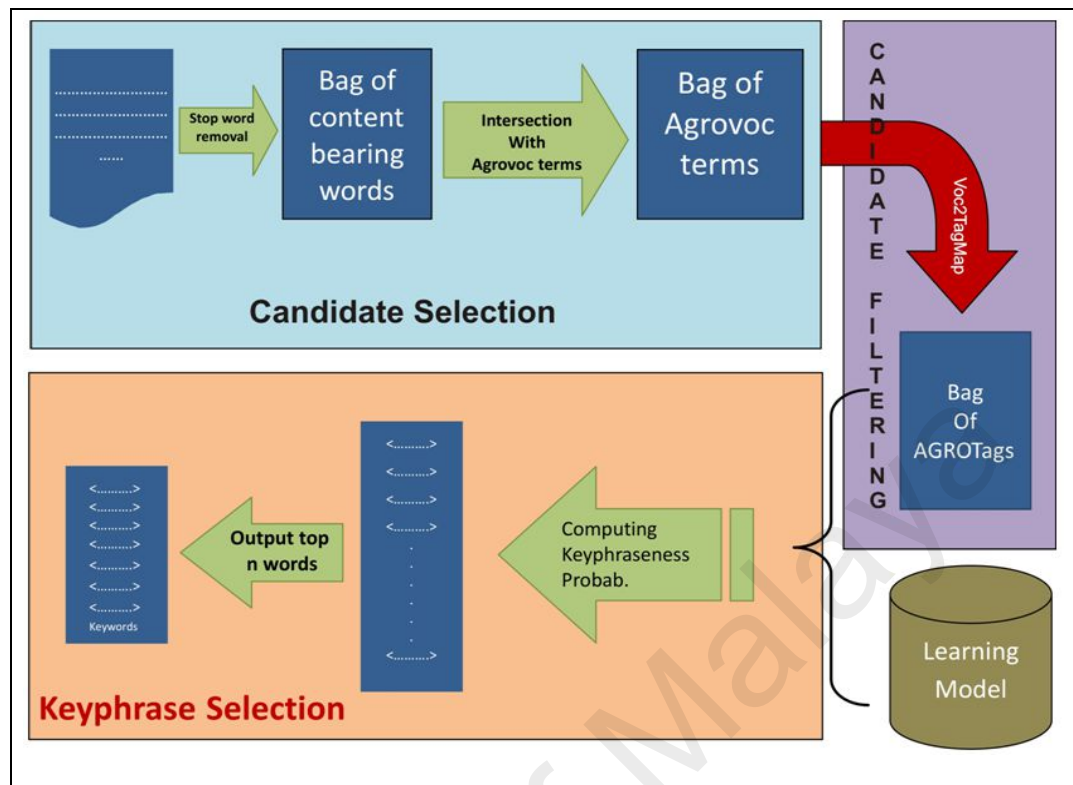
Figure 2.6 Workflow of Agrotagger (Balaji et al., 2010).

With reference to these technology initiatives which aim to provide viable solutions to the OER search dilemma, it can be noted that both the federated search and semantic search methodologies have inherent strengths and drawbacks. However, the common issue faced by both approaches is the high dependence on human annotated metadata. This issue has a snowballing effect, as the accuracy of the entire search methodology becomes a function of the accuracy of the metadata annotated by the content creators. Furthermore, the multiple standards used when annotating metadata pose an additional challenge to the search mechanisms in terms of standardization.

Summary

The literature review chapter has discussed the concept of OER, the extent of the current OER search dilemma, key OER metadata standards widely used and some of the important OER search initiatives which have been initiated in the recent past. It has also reviewed the key technology trends used by these search initiatives to provide viable solutions to the issues at hand.

Following the overview of OER provided in the first section, the second section examined the literature which highlights how the global pool of OER has grown tremendously over the past decade. The studies cited in this section provide a holistic view of the dilemma faced by academics with respect to searching and locating OER which are acceptable for teaching and learning within this large resource pool. It also reveals from an empirical perspective the challenges faced by academics in terms of using generic search engines such as Google in locating these resources.

The third section introduced the key metadata standards which are used by a majority of existing OER repositories. Among these standards, DCMI and IEEE LOM are the most common. It also looked at new attempts to create global standards, such as the LRMI, and how these new standards will influence the way academics search for OER.

The last section introduced recent prominent OER search initiatives. It classified these initiatives into federated search and semantic search to identify the merits and demerits of each in terms of large scale implementation.

Chapter 3 will discuss the methodology used in this project with a view of achieving the research objectives.

CHAPTER 3
METHODOLOGY

Chapter 3 : Methodology

As mentioned in Section 1.2 of the Introduction chapter, the methodology adopted in this research project consists of six distinct phases. An extensive literature review is conducted in Phase 1 to conceptualize the problem domain. Phase 2 consists of a survey study which is used to identify the extent of the problem with respect to OER search. Desk research is conducted simultaneously with the survey to probe the case studies to identify the limitations of existing OER search technologies. The findings of the desk research and literature review were presented in the previous chapter. Phase 3 involves the design of a framework, taking into consideration the variables identified from the survey and the desk research. The main objective of this phase is to design a conceptual framework for parametrically measuring the suitability of OER for academic purposes. Phase 4 is used to design a technology framework which encapsulates the conceptual framework designed in Phase 3. This framework utilizes text mining techniques to facilitate precise searching of suitable OER for academic purposes. The implementation of the framework is achieved through a prototype software system which is evaluated and tested in Phase 5.

This chapter is organized into four sections. The first section looks at the empirical research conducted in the form of the survey study. It details the design, data collection and analysis of the survey. The second section discusses the conceptual framework design highlighting the rationale behind the framework, definitions, scales used, calculations and concept verification. Section 3 details the technology framework. This includes the algorithm, the Keyword-Document Matrix (KDM) and the desirability calculation. The prototype system is discussed in section 4, which presents the system architecture and interface design.

3.1 Empirical Research

Recently, an Asian regional group of researchers (collaborators) from China, Hong Kong SAR, India, Indonesia, Japan, South Korea, Malaysia, Philippines and Vietnam, who are currently active in the OER arena, jointly conducted a study to elicit an understanding of the OER landscape in the Asian region. This study aimed to gather information regarding (i) the use of digital resources; (ii) the use of OER; and (iii) the understanding of copyright from both an individual's as well as an institution's perspective. Approximately 580 responses were gathered from academics who have had some exposure to the concept of OER.

3.1.1 Overview

The survey study was part of a sub-project (sub-project 7) funded by the International Development Research Centre (IDRC) of Canada (grant code # 102791) through an umbrella study on Openness and Quality in Asian Distance Education. The study commenced in March 2010 and was conducted over a duration of 27 months. The main objective of this study was to establish, qualitatively and quantitatively, the extent of OER use by institutions and/or individuals in the developing parts of Asia. The specific objectives of the study aimed to (i) determine the demand for and use of digital resources including OERs; (ii) establish regional capacities to develop and or use OERs; (iii) determine, list and describe the range of OER activities in the region; (iv) list and describe the methods adopted for the creation of OERs; (v) identify policy, legal and technological issues relating to the use of OERs; (vi) identify/determine requirements of quality and their relevance in the OER environment; and (vii) undertake an economic analysis of the OER development and use. The project concluded in December 2012.

The target population of the survey was the academic community of Higher Education Institutions (HEI). The reasons which determined this selection included (i) the availability of digital infrastructure and resources in HEIs; and (ii) the familiarity of HEI's academic staff with the availability and use of digital resources. No contrived sampling method was used in the identification of the target population. Respondents were self-selected to respond to the survey at the country level. Collaborators (Table 3.1) representing the various regions and HEIs/organizations in Asia participated in the project. The complete project consisted of four aspects which are (i) survey study; (ii) discussion groups; (iii) focus groups; and (iv) case studies. Only the results from the survey study are considered for the purposes of this Thesis.

Table 3.1 Collaborators of the project representing the various regions and HEIs in Asia.

	Collaborator	Region	Institution/Organization
1	Prof Dr G. Dhanarajan (principle investigator) Mr I.S. Abeywardena (co-investigator)	Malaysia	Wawasan Open University (WOU)
2	Prof Dr Li Yawan Dr Li Ying	China	Open University of China (OUC)
3	Dr K.S. Yuen Mr A. Wong	Hong Kong SAR	Open University of Hong Kong (OUHK)
4	Dr V. Balaji Assoc Prof Dr. B. Harishankar	India	Commonwealth of Learning (COL) University of Madras
5	Dr Daryono	Indonesia	Universitas Terbuka Indonesia (UTI)
6	Prof Dr T. Yamada	Japan	Open University of Japan (OUJ)
7	Assoc Prof Dr P. Arinto	Philippines	University of the Philippines Open University (UPOU)
8	Prof Dr Y. Kim	South Korea	Korea National Open University (KNOU)
9	Dr M. Do	Vietnam	Vietnam OER (VOER) Foundation

3.1.2 Survey Instrument

The survey instrument was collaboratively created using an iterative method to ensure that it addressed all the objectives of the study. The items of the instrument were predominantly adapted under the CC license from existing validated survey instruments. A draft instrument was created and tested with collaborators before the final form was adopted. The survey instrument was split into two parts targeting (i) individuals who have experience in OER; and (ii) competent authorities of institutions who can comment holistically on the institution's practice of OER. Only the responses from the first cohort were considered in this research work to understand the OER search habits of individuals. Each part covered four major areas, namely (i) personal and/or institutional profile; (ii) information relating to digital infrastructure and resources; (iii) information relating to practice and policy on OER; and (iv) information relating to copyright. The 84 independent items in the instrument covered multiple domains including teaching background, types and sources of digital resources used, personal digital collections, how digital resources are used in teaching, motivations for using digital resources, motivations for not using digital resources, barriers and frustrations, and support and assistance. The complete instrument is available in Appendix I.

3.1.3 Data Collection and Analysis

The data collection was done using hard copies of the survey instrument as well as an online version. The online version was delivered using the *Survey Monkey* (surveymonkey.com) platform. The hardcopy versions were made available in English, Mandarin, Vietnamese and Korean languages. The respondents had the option to select English, Japanese or Bahasa Indonesia as the preferred language for the online version. The online version was also set up to enable respondents to choose the sections they wanted to respond to and skip the rest. This allowed the capturing of partially complete

responses. The data collection was conducted over a period of three months. During the data processing stage, all the surveys completed in hardcopy format were manually entered onto the online system by a research assistant (RA). The qualitative feedback received in languages other than English was loosely translated into English and entered into the system.

Upon completion of the data collection, the online survey was closed and the complete dataset was extracted as tab delimited data. This dataset was processed using the FOSS statistical analysis software package PSPP (https://www.gnu.org/software/pspp/). The incomplete responses such as the ones missing the names and contact information were removed from the dataset. Responses received from countries outside the scope of the study were also removed. The final dataset consisted of 420 valid responses from individuals and 98 valid responses from institutional representatives. Subsequently, the dataset was made available to the collaborators for their own data analysis purposes. Crosstab and frequency analyses were conducted on various dimensions of the datasets. The findings discussed in Section 4.1 of the Results chapter concentrates solely on the individuals' perspective on OER search.

3.2 The Conceptual Framework

As discussed by Hilton et al. (2010) the use and reuse of OER depends on two factors which are (i) the permission; and (ii) the technologies needed. However, at present, all the three types of OER repositories (content repositories; portal repositories; and content and portal repositories explained in Section 2.1.4) consider only the relevance of a resource to the search query when locating internal and/or external resources. This is due to the dependence of the search on keywords or metadata which do not necessarily provide information on the various attributes of OER (Atenas & Havemann, 2013). Thus, the rank of the search result is not a direct indicator of the suitability of a resource, as it does not take into consideration the permission nor the technologies needed for successful use and reuse. This challenge is further heightened by the common use of OER formats such as Portable Document Format (PDF) which renders resources useless with respect to reuse (Baraniuk, 2007). The inability of average users to use the available technological tools to re-mix the resources (Petrides, Nguyen, Jimes, & Karaglani, 2008) adds to this dilemma. Furthermore, as resources are constantly added to these repositories (Dholakia, King, & Baraniuk, 2006), a static method of defining the suitability for use and reuse within the metadata becomes an impossible task.

3.2.1 Rationale

In the academic community, the perceived quality of an academic publication or a resource is largely governed by peer review. However, with the present day influx of research publications being made available online, the peer review mechanism becomes inefficient as not all experts can review all publications. As such, an alternative method of measuring the quality of a publication or a resource is needed. According to Buela-Casal & Zych (2010, p. 271),

"*If an article receives a citation it means it has been used by the authors who cite it and as a result, the higher the number of the citations the more utilized the article. It seems to be an evidence of the recognition and the acceptance of the work by other investigators who use it as a support for their own work*".

Therefore, at present, the number of citations received is widely accepted as an indication of perceived quality of an academic publication or resource.

As the styles of citation for academic publications are well established, search mechanisms such as Google Scholar (scholar.google.com) have a usable parametric measure for providing an indication of the usefulness of a publication for academic research. Although there are similar established styles of citation and attribution for OER, these styles are still not widely practiced when using, reusing, remixing and redistributing. As such, it is extremely difficult for a search mechanism to autonomously identify the number of citations or the number of attributions received by a particular OER. Providing potential solutions to this issue are systems such as AnnotatEd (Farzan & Brusilovsky, 2006) which uses web based annotations; use of brand reputation of a repository as an indication of quality; allowing users to review resources using set scales (Hylén, 2006); and the "Popularity" in the Connexions repository, which is measured as percentile rank of page views/day over all time. Despite these very specific methodologies, there is still no generic methodology available to enable search mechanisms to autonomously gauge the usefulness of a particular OER for teaching and learning purposes.

3.2.2 Definitions

OER are available in multiple media formats including text, images, audio, video, animations and games. However, only texts are considered in this research work. As such, the usefulness of a text based OER for a particular teaching or learning need can

50

only be accurately assessed through reading the content by users. The user makes the final decision on the suitability of a resource in his/her context (Nash, 2005). As this is quite a subjective exercise due to ones needs differing from another's, it is extremely difficult for a software based search mechanism to provide any indication of usefulness with respect to *fit-for-purpose*. However, when considering the use and re-use of an OER, there are other aspects of a resource which are fundamental to the usefulness of that particular resource and can be parametrically measured by a software based mechanism. The first of such aspects is the relevance of a resource to a user's needs. This can be assessed from the search rank of a resource against a search query. The second aspect is the openness of a resource with respect to the four R's (Section 2.1.1). The third aspect is the accessibility of the resource with respect to the ALMS analysis (Section 2.1.3). Therefore, the usefulness of an OER with respect to (i) the level of openness; (ii) the level of access; and (iii) the relevance; can be defined as the **_desirability_** of an OER, indicating how desirable it is for use and reuse for one's needs. Within the requirement of being able to use and reuse a particular OER, these three parameters can be defined as:

(i)   *level of openness*: the permission to use, reuse, remix and redistribute the resource;

(ii)   *level of access*: the technical keys (following the ALMS analysis) required to unlock the resource; and

(iii)   *relevance*: the level of match between the resource and the needs of the user.

As each of these mutually exclusive parameters are directly proportional to the desirability of an OER, the desirability can be expressed as a three dimensional measure (Figure 3.1).
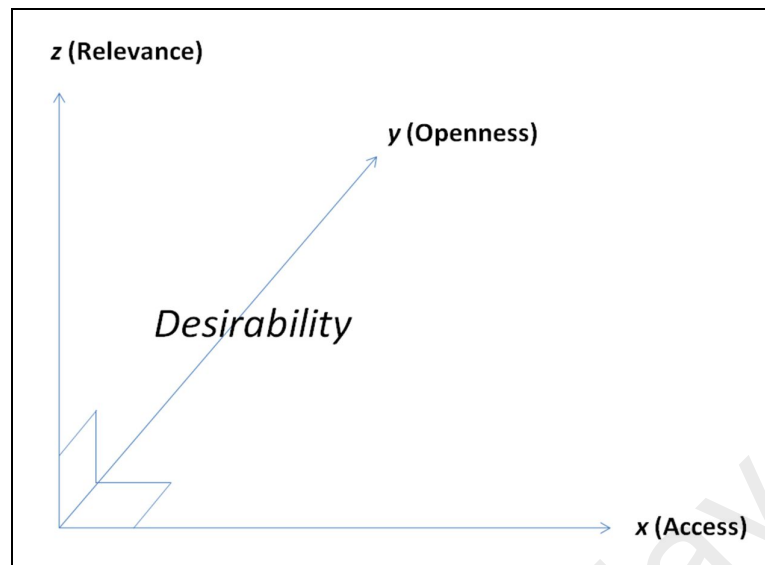
Figure 3.1 The three attributes used in the calculation of the desirability of an OER.

3.2.3 The Scales

In order to parametrically calculate the *desirability* of an OER, each of the parameters discussed in section 3.2.2 needs to be given a numeric value based on a set scale. These scales are defined as follows:

(i) The level of openness is defined using the four R's of openness as shown in Table 3.2. The values 1 to 4 are assigned to the four R's where 1 corresponds to the lowest level of openness and 4 corresponds to the highest level.

Table 3.2 The level of openness based on the four R's of openness.

Permission[a]	Value[b]
Reuse	1
Redistribute	2
Revise	3
Remix	4

[a]Permission granted by the copyright holder of a material. [b]The value assigned to each permission according to importance during the calculation of the desirability.

(ii) The level of access is defined on a scale of 1 to 16 using the ALMS analysis. As shown in Table 3.3, the value 1 corresponds to the lowest accessibility and value 16 to the highest accessibility.

Table 3.3 The level of access based on the ALMS analysis.

Access[a]				Value[b]
A	**L**	**M**	**S**	
Low	High	No	No	1
Low	High	No	Yes	2
Low	High	Yes	No	3
Low	High	Yes	Yes	4
Low	Low	No	No	5
Low	Low	No	Yes	6
Low	Low	Yes	No	7
Low	Low	Yes	Yes	8
High	High	No	No	9
High	High	No	Yes	10
High	High	Yes	No	11
High	High	Yes	Yes	12
High	Low	No	No	13
High	Low	No	Yes	14
High	Low	Yes	No	15
High	Low	Yes	Yes	16

[a]Accessibility of a resource with respect to the ALMS analysis. [b]The value assigned to each level of access according to the ease of access during the calculation of the desirability. **A**: access to editing tools; **L**: level of expertise required to revise or remix; **M**: meaningfully editable; **S**: source-file access.

(iii) The relevance of a resource to a particular search query is measured using search rank. The relationship between relevance and search rank as argued by Saracevic (1975, p. 148) is stated as

"It has been accepted explicitly or implicitly that the main objective of an IR system is to retrieve information relevant to a user queries. The logic of search and retrieval is based on the algebra of sets, Boolean algebra, which is well formulated and thus easily applicable to computer manipulations. Inherent in the application of this logic is the fundamental assumption: those documents (answers, facts, data) retrieved are also those relevant to the query; those not retrieved are not relevant. In some systems documents can be ordered (evaluated, associated) as to their relevance and retrieved when some specified threshold is reached, and presented in some ordered form; but even here the assumption that retrieved/not retrieved corresponds to relevant/not relevant still holds true".

According to Vaughan (2004), users will only consider the top ten ranked results for a particular search as the most relevant. Vaughan further suggests that users will ignore the results below the top 30 ranks. Based on this premise, the scale for the relevance is defined as shown in Table 3.4, where the value 1 is the least relevant and value 4 is the most relevant.

Table 3.4 The level of relevance based on search rank.

Search rank[a]	Value[b]
Below the top 30 ranks of the search results	1
Within the top 21-30 ranks of the search results	2
Within the top 11-20 ranks of the search results	3
Within the top 10 ranks of the search results	4

[a]Ranking of the search results returned for a particular search query. [b]The value assigned to each set of search results returned according to the search rank during the calculation of the desirability.

3.2.4 Calculations

Based on the scales discussed in section 3.2.3, the *desirability* of an OER is defined as the volume of the cuboid (Figure 3.2) calculated using Equation 1.

desirability* = *level of access* x *level of openness* x *relevance (1)

As a result, the *desirability* becomes directly proportional to the volume of the cuboid.
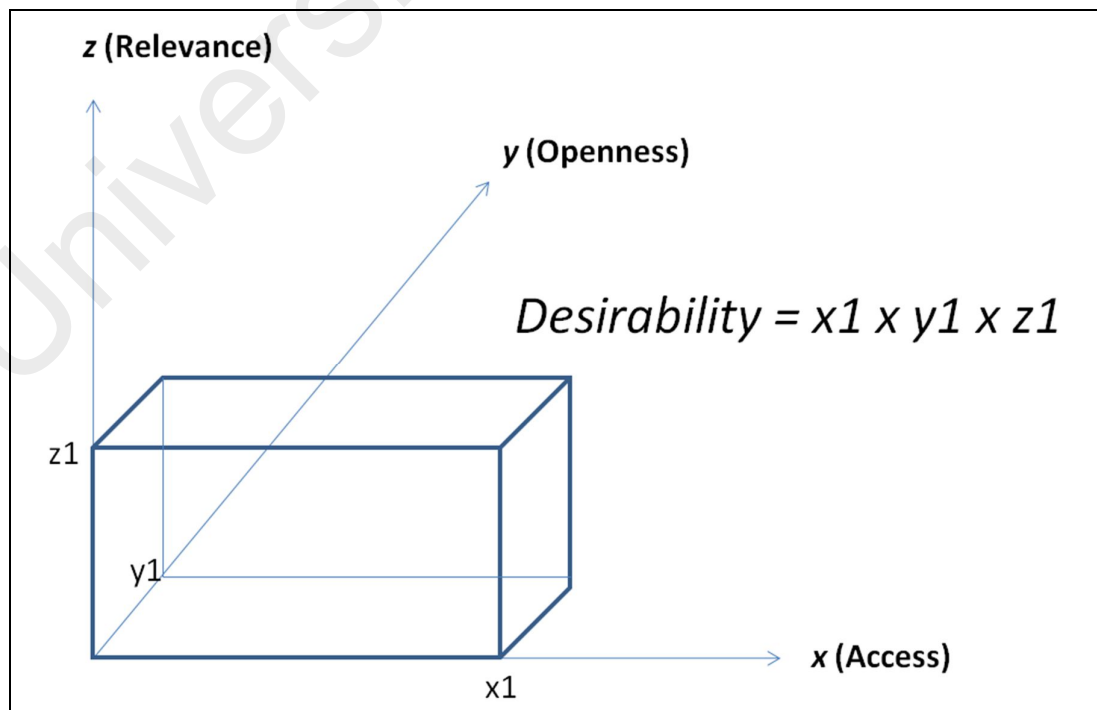
Figure 3.2 Calculation of desirability as a function of access, openness and relevance.

By normalizing the values indicated in Table 3.2, Table 3.3 and Table 3.4 to make the scales uniform, the **_D-index_** of an OER can be calculated using Equation 2. (The value 256 is used to normalize the access, openness, and relevance parameters. It is the product of the values 16, 4, and 4, respectively, which correspond to the highest value assigned to each parameter.)

D-index = (*level of access* x *level of openness* x *relevance) / 256*　　　　(2)

Based on the above calculation, a resource becomes more desirable as the *D-index* increases on a scale of 0 to 1 where 0 is the least desirable and 1 is the most desirable.

3.2.5 Verification of Concept

The most commonly used methods for locating OER are generic search mechanisms and repository specific search mechanisms. However, both of these types only consider the relevance of the resource, either by matching the title and description or the keywords to the search query provided by the user. Therefore, the top search results are not always the most desirable as they might be less open or less accessible. The *D-index* is specifically designed to overcome this limitation by taking into consideration the openness and accessibility of an OER in addition to the relevance.

When applying the *D-index* to an OER repository, the level of access, as discussed in Table 3.3, needs to be implemented using the file types of the OER, where their features are mapped against the ALMS. The level of openness (Table 3.2) needs to be measured using the copyright licensing scheme under which the resource was released. The de facto scheme used in most repositories is the CC licensing scheme (Section 2.1.2). However, other specific licensing schemes such as the *GNU Free Documentation License* can also be used for this purpose as long as they can be categorized into the four levels of openness constituting the desirability. Table 3.5 maps

the six CC licenses to the four R's of openness. It should be noted that the level of openness of the CC licenses starts at the redistribute level. Based on the four R's, it can be interpreted that the most restrictive licenses are CC BY-ND and CC BY-NC-ND as they prohibit derivations. CC BY and CC BY-SA are the most open licenses. Despite the fact that CC BY-NC and CC BY-NC-SA restrict commercial use, they still embody the all the freedoms of four R's. As such, they are given a higher value.

Table 3.5 Openness based on the CC license.

Permission[a]	Creative Commons (CC) license[b]	Value[c]
Reuse	None	1
Redistribute	Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) Attribution-NoDerivs (CC BY-ND)	2
Revise	Attribution-NonCommercial-ShareAlike (CC BY-NC-SA) Attribution-ShareAlike (CC BY-SA)	3
Remix	Attribution-NonCommercial (CC BY-NC) Attribution (CC BY)	4

[a]Permission granted by the copyright holder of a material. [b]Creative Commons (CC) license which corresponds to the permission granted by the copyright holder. [c]The value assigned to each CC license, according to importance, during the calculation of the desirability.

To verify the proposed *D-index* concept, an experiment was carried out in the widely used OER Commons (oercommons.org) repository. This repository was specifically selected for the experiment due to (i) the repository providing users with a native search mechanism to locate OER; and (ii) the variety of OER available in different levels of openness and access. The repository was searched using the term "calculus" to locate OER on the topic of calculus in mathematics. The term "calculus" was intentionally selected for the experiment due to the large number of OER written and made available on the topic. Only the top 40 search results, returned based on relevance, were considered in the experiment as users tend to ignore results below the rank of 30 (Vaughan, 2004). Out of the 165 resources returned as results, three resources at the post-secondary level of different search rank were chosen for comparison (Table 3.6) to demonstrate the application of the *D-index*.

Table 3.6 Selected search results at post-secondary level returned by the OER Commons search mechanism for the search term "calculus".

Resource	Title	Search rank	License	File type
A	Calculus I	2	Creative Commons Attribution-Noncommercial-Share Alike 3.0 (CC BY-NC-SA)	PDF
B	Topics in calculus	8	Creative Commons Attribution-Noncommercial 3.0 (CC BY-NC)	webpage (HTML)
C	Calculus I (MATH 151)	23	Creative Commons Attribution 3.0 Unported (CC BY)	MS Word

The file type, search rank and license of each resource in Table 3.7 was then compared with Table 3.3, Table 3.4 and Table 3.5 respectively to identify the parameters required to calculate the *D-index* (Table 3.7).

Table 3.7 Parameters required for calculating the D-index.

Resource	Relevance	Openness (four R's)	Access (ALMS)				
			A	L	M	S	Value
A	4	3	Low	High	No	No	1
B	4	4	High	Low	Yes	Yes	16
C	2	4	Low	Low	Yes	Yes	8

Referring to Table 3.7 we can see that the search mechanism ordered the results according to the relevance, where *Resource A* is the most relevant. However, *Resource A* is less open and less accessible when compared with *Resource B*. Table 3.8 shows how the results would be re-organized when the *D-index* is applied to the same search results.

Table 3.8 After applying the D-index to the same search results shown in Table 3.7.

Resource	Relevance	Openness	Access	*D-index*
B	4	4	16	1.00
C	2	4	8	0.25
A	4	3	1	0.05

From the results in Table 3.8, it can be seen that *Resource B* is the most desirable OER for use and reuse due to its level of openness and access, even though *Resource A* was the most relevant.

More experiments were conducted on two other widely used OER repositories to verify the validity of the *desirability* conceptual framework. The results of these experiments are discussed in Section 4.2 of the Results chapter.

3.3 OERScout Technology Framework

As discussed in Section 2.2, the most common OER search method is generic search engines such as Google, Yahoo! or Bing. However, this method is not the most effective. Though these generic search engines provide advanced facilities to define various filter criteria, they are not tailored to effectively locate OER which are the most desirable for a particular academic purpose. As such, OER consumers need to resort to frequenting various OER repositories to search for relevant and useful materials. However, this too has become a cumbersome and time consuming task as the number of repositories and the volume of each repository keeps on expanding. In addition, users are spending an extended amount of time on these repositories conducting multiple searches using repository specific search mechanisms (Figure 3.3); and by so doing limit the scope and the variety of OER available to them (Abeywardena, 2013). Ultimately, the user is stuck in a scenario where the use of these materials is not a choice but a lack of options.
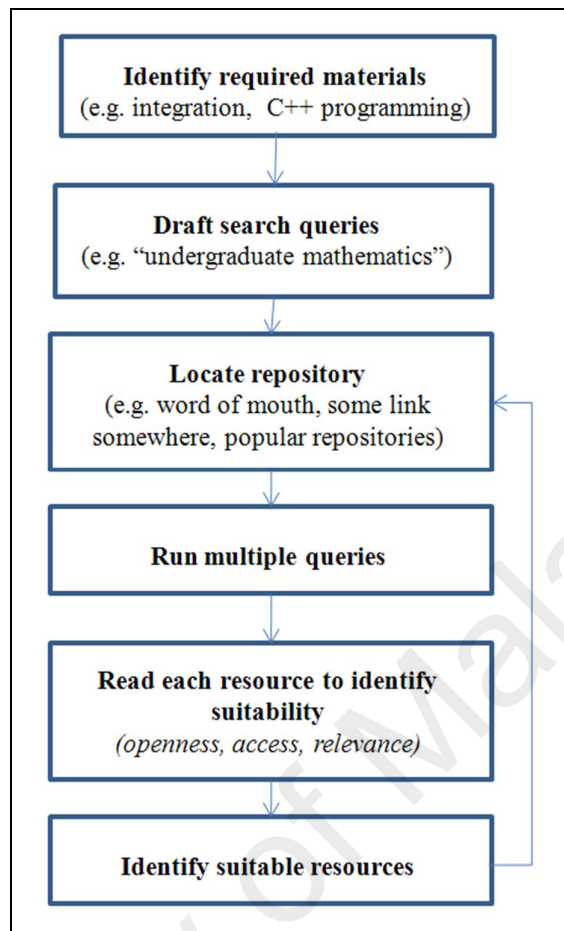
Figure 3.3 The flow of activities in searching for suitable OER on heterogeneous repositories based on personal experience (Abeywardena, 2013). These activities will need to be repeated on multiple repositories until the required resources are located.

Another factor inhibiting effective OER search is the heterogeneity of OER repositories. Within the context of parametric web based search, this disparity can be broadly attributed to (i) the lack of a single metadata standard; (ii) the lack of a centralized search mechanism; and (iii) the inability to indicate the usefulness of an OER returned as a search result.

Metadata provides a standard and efficient way to conveniently characterize educational resource properties (Anido, et al., 2002). The majority of existing search methodologies; including mainstream search engines such as Google, work on the concept of metadata for locating educational resources. However, it can be argued that the annotation of resources with metadata cannot be made 100% accurate or uniform if done by the creator(s) of the resource (Barton, Currier, & Hey, 2003; Tello, 2007;

Devedzic, Jovanovic, & Gasevic, 2007; Brooks & McCalla, 2006; Cechinel, Sánchez-Alonso, & Sicilia, 2009). Therefore the use of human annotated metadata in performing objective searches becomes subjective and inaccurate. A possible way to overcome this inaccuracy and to ensure uniformity of metadata is to utilize a computer based methodology which considers the content, domain, and locality of the resources, among others, for autonomously annotating metadata. As a solution to these issues, this phase of the project proposes the ***OERScout*** technology framework to accurately cluster text based OER by building a searchable matrix of autonomously mined domain specific keywords.

3.3.1 The Algorithm

As discussed in Section 2.4 of the Literature Review, mainstream search engines, federated search, and semantic search are the current key OER search methodologies. However, all of these methodologies depend on human annotated metadata for approximating the usefulness of a resource for a particular need. Given the limitations of human annotated metadata with respect to accurately and uniformly describing resources, the effectiveness of search becomes a function of the content creators' ability to accurately annotate resources. Therefore, the OERScout system uses text mining techniques to annotate resources using autonomously mined keywords.

The OERScout text mining algorithm is designed to "read" text based OER documents and "learn" which academic domain(s) and sub-domain(s) they belong to. To achieve this, a *bag-of-words* approach is used due to its effectiveness with unstructured data (Feldman & Sanger, 2006). The algorithm extracts all the individual words from a particular document by removing noise such as formatting and punctuation to form the *corpus*. The corpus is then *tokenized* into the *list of terms* using the *stop words* found in the Onix Text Retrieval Toolkit (Lextek), as shown in Figure 3.4.
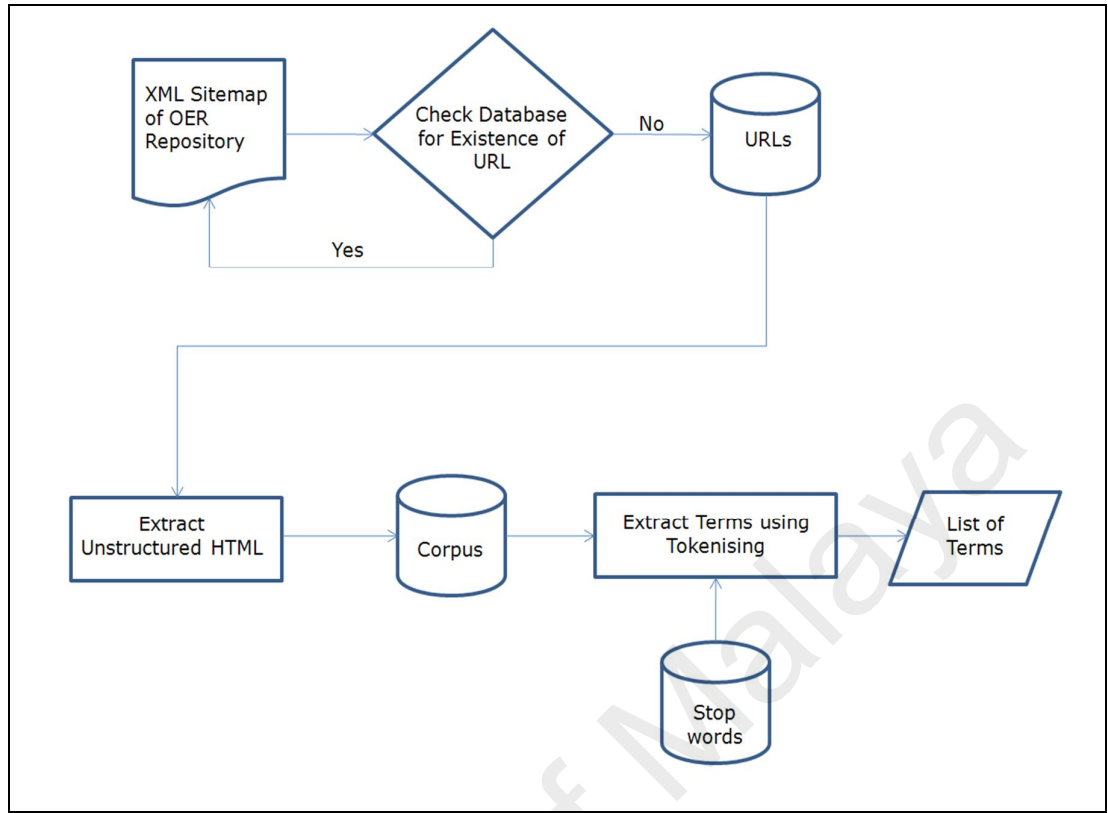
Figure 3.4 The List of Terms is created by Tokenising the Corpus using the stop words found in the Onix Text Retrieval Toolkit.

The content describing terms are extracted from the list of terms for the formation of the term document matrix (TDM) by applying the term frequency–inverse document frequency (TF-IDF) weighting scheme. The weight of each term (TF-IDF) is calculated using Equation 3 (Feldman & Sanger, 2006):

$$(TF\text{-}IDF)_t = TF_t \ x \ IDF_t \tag{3}$$

TF_t denotes the frequency of a term t in a single document. IDF_t denotes the frequency of a term t in all the documents in the collection [$IDF_t = Log \ (N/DF_t)$] where N is the total number of documents in the collection and DF_t is the total number of documents containing the term t in the collection. The probability of a term t being able to accurately describe the content of a particular document as a keyword decreases with the number of times it occurs in other related and non-related documents. For example the term "introduction" would be found in many OER documents which discuss a

variety of subject matter. As such the TF-IDF of the term "introduction" would be low compared to terms such as "operating systems" or "statistical methods" which are more likely to be keywords.

Due to the document lengths of OER, the TF value of certain words will be quite high. As a result, there will be a considerable amount of noise being picked up while identifying the keywords. However, the large number of documents available in OER repositories will also increase the DF value of words. This reduces the IDF value which results in a lower TF-IDF value and the reduction of noise picked up as keywords. As such, the TF-IDF weighting scheme allows the system to refine its set of identified keywords at each iteration. Therefore, the TF-IDF weighting scheme is found to be suitable for extracting keywords from the OER documents.

3.3.2 Keyword-Document Matrix

The **_Keyword-Document Matrix (KDM)_**, a subset of the TDM, is created for the OERScout system by matching the autonomously identified keywords against the documents as shown in Figure 3.5.

	$Keyword_1$	$Keyword_2$	…………	$Keyword_n$
$Document_1$	√			√
$Document_2$	√		√	
…………..		√		
$Document_n$		√		√

Figure 3.5 The KDM, a subset of the TDM, is created for the OERScout system by matching the autonomously identified keywords against the documents.

The formation of the KDM (Figure 3.6) is achieved by (i) normalizing the TF-IDF values for the terms in the TDM; and (ii) applying the Pareto principle (80:20) empirically (Milojević, 2010) for feature selection, where the top 20% of the TF-IDF values are considered to be keywords describing 80% of the document.
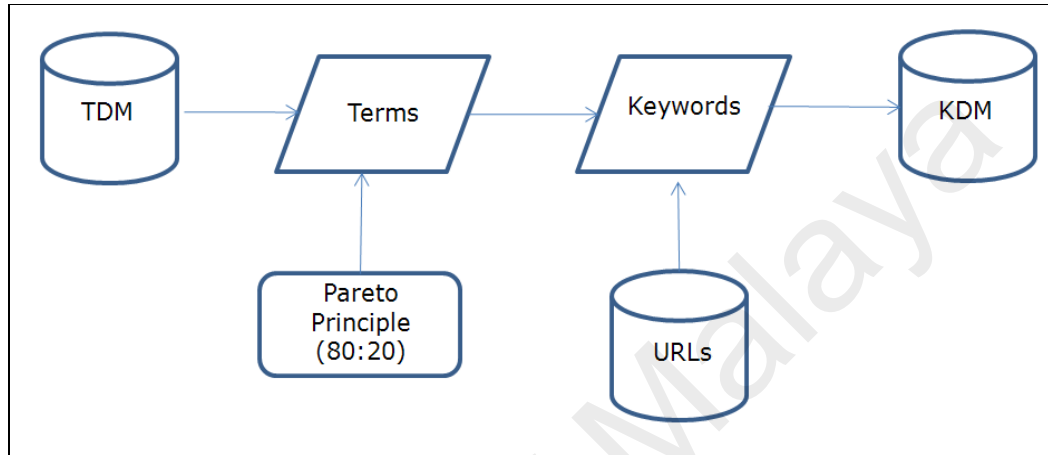
Figure 3.6 Formation of the KDM by normalizing the TF-IDF values of the terms in the TDM and applying the Pareto principle empirically for feature selection.

3.3.3 Calculation of the Desirability

The desirability of each document in the KDM is calculated using Equation 2. The openness of the document is calculated using the CC license of the document (Table 3.5). The accessibility is calculated by extracting the file type of each document, as shown in Table 3.3. This version of OERScout is built to index documents of type PDF (.pdf), webpage (static and dynamic web pages which include .htm, .html, .jsp, .asp, .aspx, .php etc.), TEXT (.txt) and MS Word (.doc, .docx) as these file types were found to be the most commonly used for text based OER (Wiley, 2006). The value for each file type was calculated with reference to Table 3.9 based on access to editing tools (A); level of expertise required to revise or remix (L); whether meaningfully editable (M); and access to source file (S). The relevance of a document (Table 3.4) to a particular search query is calculated using the TF-IDF values of the keywords which are stored as additional parameters of the KDM.

Table 3.9 Accessibility based on the file type.

File Type[a]	Access (ALMS)[b]				
	A	**L**	**M**	**S**	**Value**
PDF	Low	High	No	No	1
MS Word	Low	Low	Yes	Yes	8
webpage[c]	High	Low	Yes	Yes	16
TEXT	High	Low	Yes	Yes	16

[a]File type of the document. [b]Level of access calculated according to Table 4. [c]Static and dynamic web pages which include .htm, .html, .jsp, .asp, .aspx, .php etc.

One of the key observations made during the calculation of the desirability is that some OER repositories do not use the CC licensing scheme as the standard for defining copyright. However, these repositories explicitly or implicitly mention that the resources are freely and openly available for use and reuse. One example is the National Programme on Technology Enhanced Learning (NPTEL) repository of India which had its own open license prior to adopting CC licenses in 2012. Furthermore, a resource is copyright by default if there is no indication of a license. Due to the inability of the current OERScout system to determine the level of openness of these resources, a value of zero was assigned to any resource which did not implement the CC licensing scheme. As such, the desirability of these resources was reduced to zero due to the ambiguity in the license definition. This feature spares the user from legal complications attached to the use and re-use of resources which are copyright or do not clearly indicate the permissions granted.

3.4 Prototype Development

The OERScout technology framework has been implemented using a prototype system which consists of (i) a set of server tools for building the KDM; and (ii) a client user interface for querying the KDM. The two components operate independently of each other, connected only through the KDM. This prototype system has been used to conduct the verification and testing of the technology framework.

3.4.1 System Architecture

The server tools were developed using the Microsoft Visual Basic.NET (VB.NET 2010) programming language. The corpus, *List of Terms*, TDM and KDM are implemented using the MySQL Community Server relational database platform. A stored procedure (SP) architecture is used for database transactions. The OER resources are fed into the system using sitemaps based on extensible markup language (XML) which contain the uniform resource locators (URL) of the resources. The KDM is accessed online using the client user interface via a web service as shown in Figure 3.7.
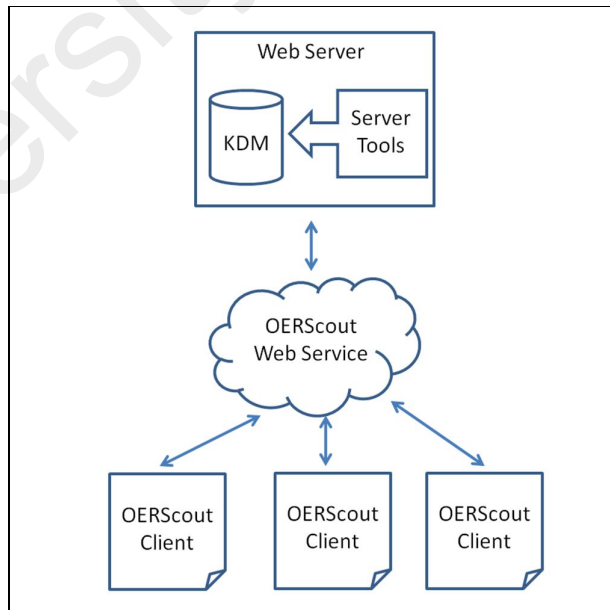
Figure 3.7 OERScout deployment architecture which has a web server hosting the KDM, a web service for accessing the KDM, and a Microsoft Windows based client interface.

3.4.2 User Interface

The client user interface (Figure 3.8) is designed to be user friendly, simple and intuitive. VB.NET was the language of choice for the interface design due to its rich design components and ease of use. The search terms are input by the user through a one dimensional search box. Multiple search terms can be input separated by commas. The "Scout" button is used to commit the search. Once the search is committed the system utilizes the same methodology used for creating the List of Terms (Figure 3.4) to remove noise and identify potential keywords for search. This reduces duplicate results being returned. For example, a search conducted for "operating system" and "operating systems" will return the same results.

Using the same algorithm discussed in section 3.3.1, the search terms are *Tokenized* into a list of words which consist of meaningful search keywords. This *Tokenizing* process removes noise such as *stop words* and punctuation, resulting in a more accurate search query.
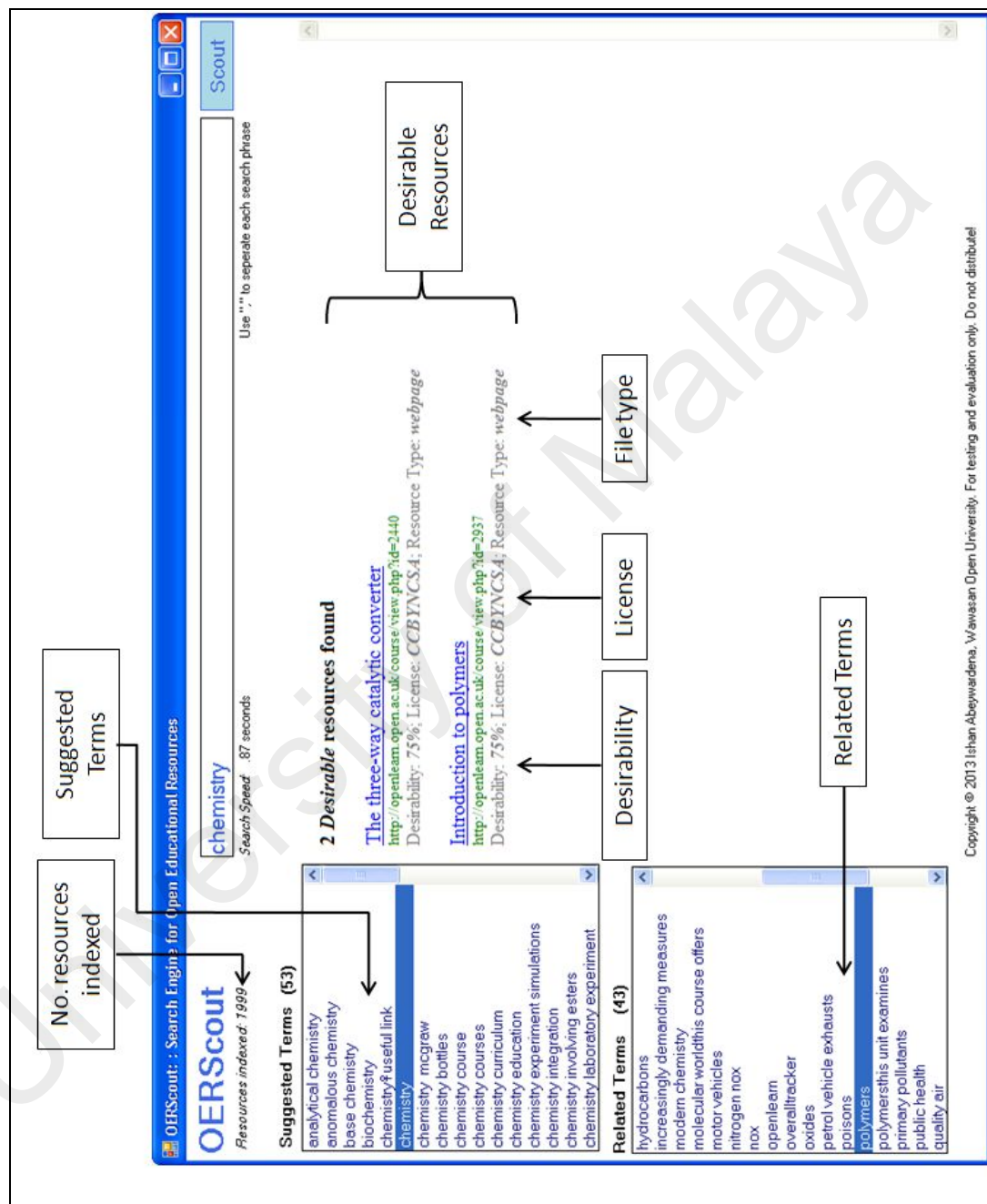
Figure 3.8 OERScout client interface used for testing the system.

3.4.3 Faceted Search Approach

Faceted search is a hybrid search approach which combines parametric search and faceted navigation (Tunkelang, 2009). According to Dash et al. (2008, p. 3), *"First, it smoothly integrates free text search with structured querying. Second, the counts on selected facets serve as context for further navigation"*.

Search engines have undergone rapid evolution in the past decade due to global technological giants such as Google providing innovative approaches to free-text search. In his book *Faceted Search,* Daniel Tunkelang (2009) of Google explains how previous search technologies morphed into the faceted search approach. According to Tunkelang, the earliest search engines used the Boolean retrieval model, which limited the flexibility and increased the complexity of the search query. Abandoning this method, information retrieval (IR) researches adopted a free-text query approach (Hobson, et al., 1997) which provided increased flexibility in creating search queries. This method cast a wide net to return results based on rank. Although not as accurate as Boolean retrieval, many search engines still follow the free-text query approach incorporating the ranked retrieval framework. Another approach used in searching for information, especially on the World Wide Web, is the directory approach. The advantage of this approach is the organization of content based on set taxonomies. This allowed users to navigate categories and sub categories to ultimately arrive at the information they seek. However, Tunkelang observes that the creators of the taxonomies themselves and the users frequently disagree on the categorization of the content as this is a subjective exercise. For example, a resource on mobile learning could be categorized under technology and education. Where it will be categorized to avoid duplication is a subjective decision made by the creators of the taxonomy whereas the user might have a different opinion. Therefore, users will have to learn to

think like the creators to find the relevant information. Figure 3.9 shows the Open Directory Project, which is among the earliest directories.

Figure 3.9 The Open Directory Project (captured June 7, 2013 from http://www.dmoz.org/).

When considering faceted search, Marti Hearst of UC Berkley, who was the lead researcher in the popular *Flexible information Access using Metadata in Novel COmbonations* (Flamenco) faceted search project, argues that *"A key component to successful faceted search interfaces (which unfortunately is rarely implemented properly) is the implementation of keyword search"* (Hearst, 2006, p. 4).

In simpler terms, modern faceted search combines free-text querying to generate a list of results based on keywords which can then be refined further by the user using a

Boolean, structured or directory approach. To achieve this functionality, faceted metadata need to be extracted from documents using text mining techniques. A few general strategies are (i) exploit latent metadata such as document source, type, length; (ii) use rule based or statistical techniques to categorize documents into predetermined categories; and (iii) use an unsupervised approach such as terminology extraction to obtain a list of terms from the document (Tunkelang, 2009).

Typical interaction between a faceted search interface and the user is explained by Ben-Yitzhak et al., (2008) as (i) type or refine a search query; or (ii) navigate through multiple, independent facet hierarchies that describe the data by drill-down (refinement) or roll-up (generalization) operations. Koren et al., (2008, p. 1) further explain this interaction as:

"The interfaces present a number of facets along with a selection of their associated values, any previous search results, and the current query. By choosing from suggested values of these facets, a user can interactively refine the query."

Ultimately, faceted search allows users to quickly drill down into a more focused set of search results using the initial results set.

There have been many academic research projects on faceted search. Among the earliest are: query previews which reduces search steps by eliminating zero-hit queries, University of Maryland (Donn, Plaisant, & Shneiderman, 1996); view-based search which allows users to interact directly with the database using views, University of Huddersfield (Pollitt, Smith, Treglown, & Braekevelt, 1996); Flamenco which defines task-oriented search interfaces across a wide variety of domains, UC Berkeley (Hearst, 2006); The Relation Browser which allows users to quickly explore a document space using dynamic queries, University of North Carolina (Capra & Marchionini, 2008); University of Southampton's mSpace: a new interaction design for user-determined

content which support preview queues, dimensional sorting spatial context (Karam & Zhao, 2003); and MIT's Parallax: which offers "set-based browsing" which extends faceted search shift views between related sets of entities (Anderson, 2007).

The most noted applications of faceted search come from the e-commerce industry. Among these are: ENDECA, which provided faceted search branded as "Guided Navigation" to e-commerce sites such as Wal-Mart and Home Depot; eBay Express which acts as a typical shopping site rather than the eBay online auction (Ebay's Express to take on Amazon, 2006); and Amazon's "Project Ruby" an experimental faceted search site for its multi-store apparel department (Cox, 2002). In the open source front CNET's Solr project (Hostetter, 2006) and the popular CMS *Drupal* provide faceted search features.

In contrast to the static list of search results produced by the generic search engines, OERScout employs a faceted search approach by providing a dynamic list of *Suggested Terms* which are related to the search term(s), as shown in Figure 3.10. The user is then able to click on any of the *Suggested Terms* to access the most *desirable* OER from the repositories indexed in the KDM. Furthermore, based on the selection by the user, the system will provide a list of *Related Terms* which enable the user to drill down further, zeroing in on the most suitable OER for their teaching needs. The results of the search are shown in descending order of the desirability. The license type and resource type are also indicated, along with the desirability, allowing the user to make a quick assessment on which resources best suit their needs.
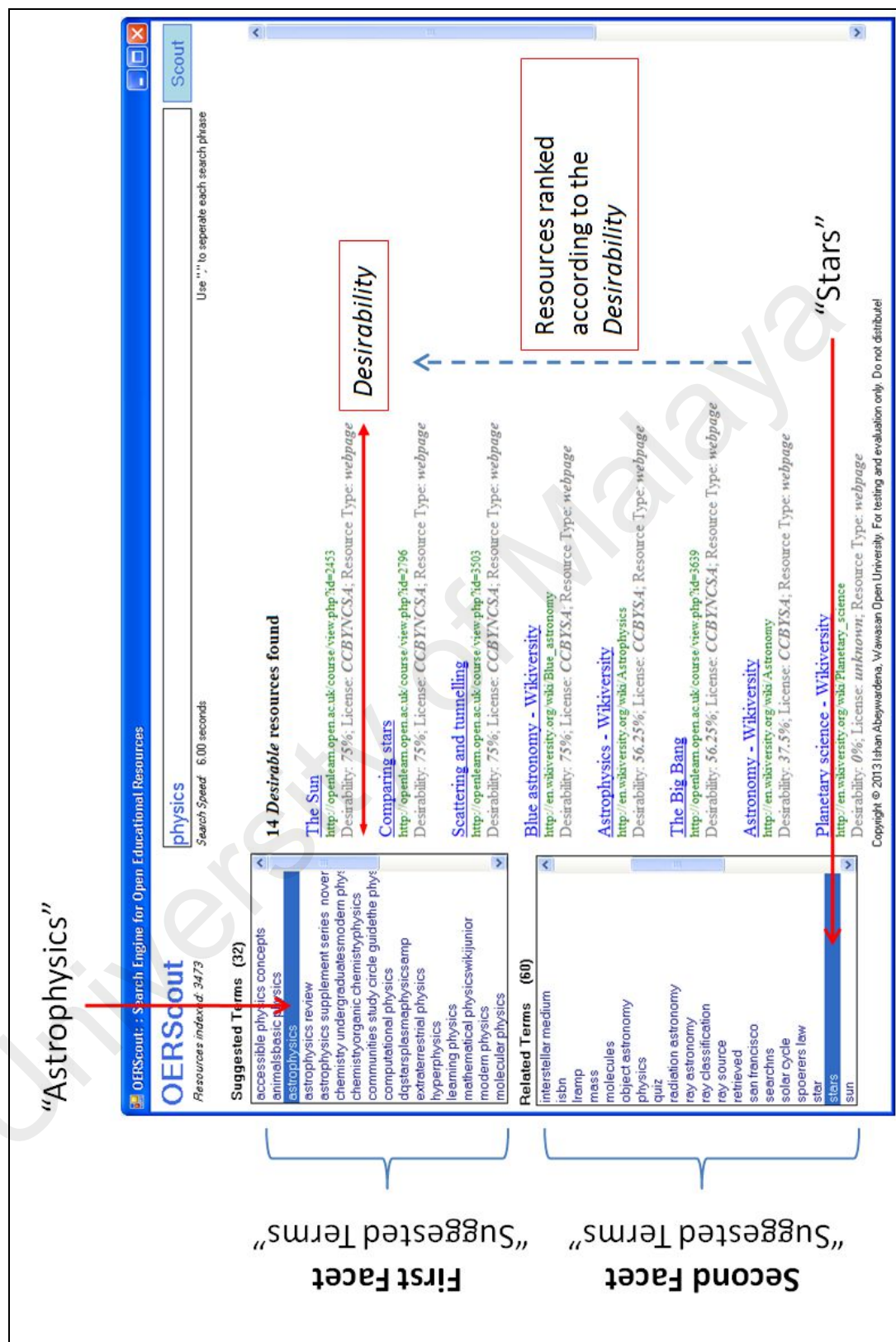
Figure 3.10 OERScout faceted search user interface. The figure shows a search conducted for Physics: Astrophysics: Stars.

Summary

This chapter has detailed the methodology used in the research project. The complete research project is divided into six phases. The first phase looks at the conceptualization of the problem domain. This is achieved through extensive literature review.

The second phase concentrates on empirical research where a survey instrument is designed and used to elicit an understanding of the current situation with respect to OER search. The variables identified during the survey study are used to design the conceptual desirability framework, which takes into consideration the openness, access and relevance attributes of OER to parametrically measure their usefulness for academic purposes.

During Phase 4, the desirability framework is implemented using the OERScout technology framework, which uses text mining techniques to create a KDM of autonomously identified domain specific keywords. The complete *OERScout* technology framework is then developed into a prototype system which consists of server tools and client interface which access the KDM using a web service architecture. The OERScout user interface adopts a faceted search approach which allows users to quickly zero-in on the resources they are after.

The next chapter discusses the results of the survey study, desirability framework implementation, OERScout prototype implementation and user tests.

CHAPTER 4
RESULTS

Chapter 4 : Results

The previous chapter detailed the methodology used in the project with respect to the six distinct phases. This chapter discusses the corresponding results of the survey study, conceptual framework design, prototype implementation and user tests. The results of the survey study are based on qualitative and quantitative responses provided by 420 academics across Asia. These results are presented in crosstab and frequency analyses formats. The conceptual framework was implemented empirically using three widely used OER repositories which are MERLOT, JORUM and OER Commons. The prototype implementation results are based on a popular content repository (Connexions) and a comprehensive portal repository (DOER). A substantial number of resources such as short articles, full course materials, tutorials varying in domain, locality, length, license and file type were used to autonomously create the KDM. A selected group of OER advocates and practitioners who have at least 3-5 years of experience in the field were invited to provide feedback on the prototype system. The feedback was gathered using an online feedback form following a set test period.

The rest of this chapter is organized into four sections. The first section looks at the responses gathered from the survey study with respect to the need for better OER search technologies. The second section provides evidence of the effect the *D-index* has on relevance based search results. The next section looks at the results of the prototype implementation in a real world scenario. The last section highlights qualitative feedback provided by the expert users with respect to the strengths of the OERScout technology framework and the weaknesses of the prototype.

4.1 Survey Results

For the purposes of this thesis, the data analysis concentrates on 420 responses (N=420)

from nine countries which represent the various Asian regions (Table 4.1).

Table 4.1 Participation rates of academics in the regional study conducted to elicit an understanding of the OER landscape in the Asian region.

	Country / Region	Valid Responses[a] (N)	Percentage[b]
1.	China	75	18%
2.	Hong Kong SAR	40	9%
3.	India	67	16%
4.	Indonesia	42	10%
5.	Japan	12	3%
6.	Malaysia	37	9%
7.	Philippines	36	9%
8.	South Korea	64	15%
9.	Vietnam	35	8%
10.	Other[c]	12	3%
	Total	**420**	**100%**

[a]Number of responses received which were complete with name and contact details of respondent. [b]Percentage of responses received from each country/region with respect to the total number. [c]Bangladesh, Pakistan, Afghanistan and Sri Lanka.

The cohort comprises junior to senior academics from 312 (74.30%) public, 63 (15%)

private not-for-profit and 45 (10.7%) private for-profit institutions, as shown in Table

4.2. The extent of the use of OER by the participants in their teaching is shown in Table

4.3. Their attitudes towards using OER in their teaching are highlighted in Table 4.4.

Table 4.2 Academic and institutional profile of the survey respondents.

Participant Title	Institution Status			Valid Responses (N)
	Public	Private not-for-profit	Private for-profit	
Prof.	80% (20)	8% (2)	12% (3)	100% (25)
Dr.	75.5% (77)	14.7% (15)	9.8% (10)	100% (102)
Mr.	75.7% (168)	14.4% (32)	9.9% (22)	100% (222)
Ms.	66.2% (47)	19.7% (14)	14.1% (10)	100% (71)
Valid Responses (N)	**74.3% (312)**	**15% (63)**	**10.7% (45)**	**100% (420)**

Table 4.3 The extent of use of OER by the survey participants.

Criteria	Yes	No	Unsure	Valid Responses (N)
I have used OER in my teaching in the past	65% (209)	23% (73)	12% (40)	100% (322)
I will use OER in my teaching in the future	80% (253)	5% (16)	15% (46)	100% (315)

Table 4.4 Attitudes towards using OER in teaching.

Attitudes	Agree	Disagree	Neutral	Valid Responses (N)
Reusing OER is a useful way of developing new courses	77% (240)	3.5% (11)	19.5% (61)	100% (312)
Exploring the available OER worldwide will enhance my teaching and raise standards across the University	79.8% (249)	1.9% (6)	18.3% (57)	100% (312)

The OER downloading habits of the participants are shown in Figure 4.1. Table 4.5 shows the extent of use of available search methodologies for locating OER by the respondents who had used OER in the past. These respondents also mentioned that they locate OER through other means such as by word of mouth from colleagues, through Wikipedia and through face-to-face networking. 64% of them further suggested that the lack of awareness of the university OER repository and other OER repositories was a major barrier. 56.6% of the same cohort mentioned that the irrelevance of the available OER to their teaching is also one of the major concerns.
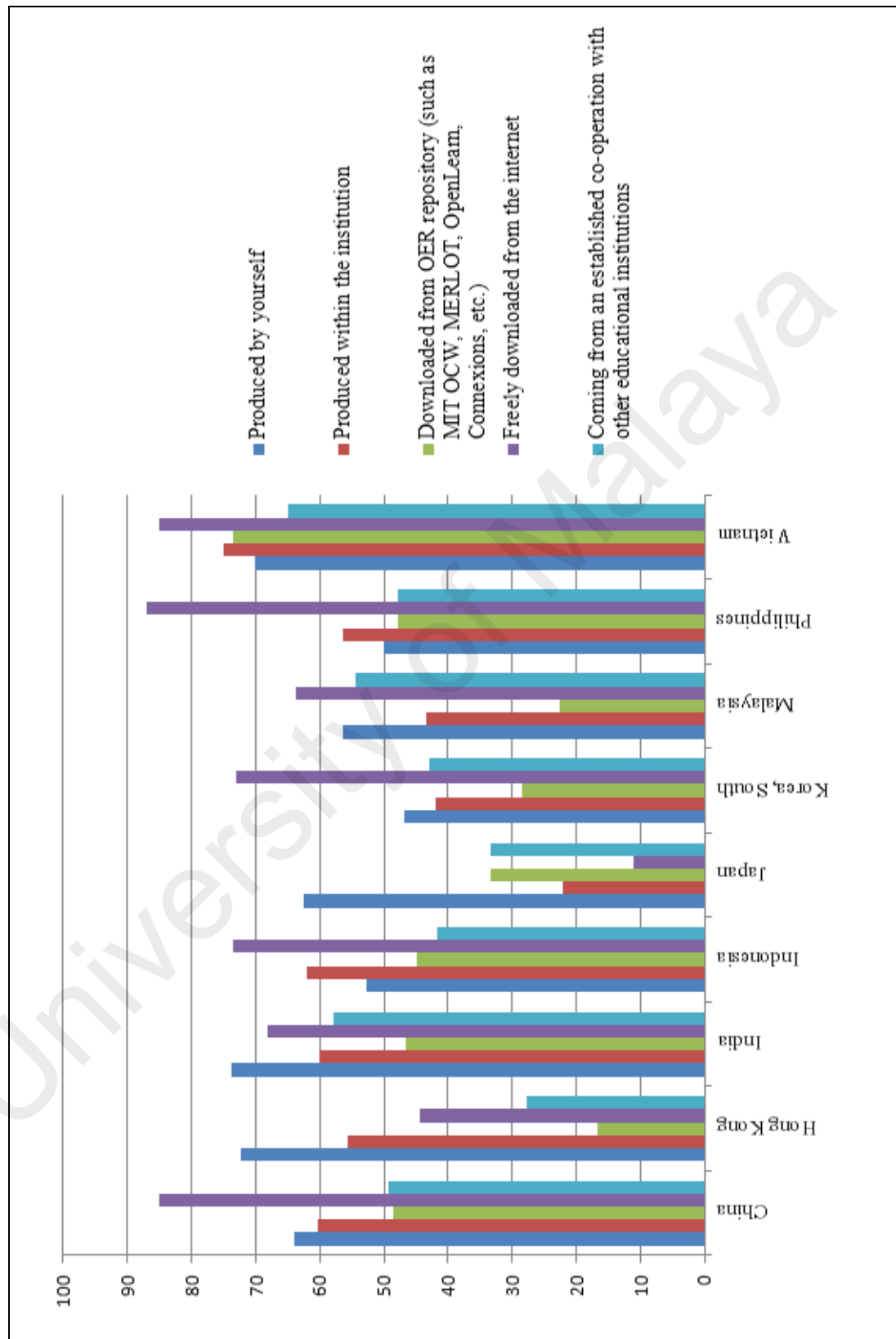
Figure 4.1 OER downloading habits of the participants.

Table 4.5 Comparison between the search methods used by academics for locating OER.

Search Method	Mostly Use	Valid Responses (N)
Generic search engines such as Google[a], Yahoo![b], Bing[c] etc.	96.9% (189)	100% (195)
Specific search engines such as Google Scholar[d]	68.9% (133)	100% (193)
Wikieducator[e] Search facilities	48.2% (92)	100% (191)
Specific search facilities of OER repositories such as OCW[f], Connexions[g] etc.	43.2% (82)	100% (190)
Any other methods for locating OER[h]	33.3% (25)	100% (75)

[a]google.com. [b]yahoo.com. [c]bing.com. [d]scholar.google.com. [e]wikieducator.org. [f]ocw.mit.edu. [g]cnx.org. [h]Word of mouth from colleagues, through Wikipedia and through face-to-face networking.

Table 4.6 shows the respondents' views with respect to the lack of ability to locate specific, relevant and quality OER for teaching. In this context,

- *specific* denotes the suitability of an OER for a particular teaching need. For example, an OER on physics from the final year syllabus of a physics degree would not be suitable for a high school physics class;

- *relevant* denotes the match between the content of the OER and the content needed for a particular teaching need. For example, physical chemistry is not relevant for a teaching need in organic chemistry; and

- *quality* denotes perceived academic standard of an OER for a particular teaching need.

Table 4.6 The importance of locating specific, relevant and quality OER for teaching.

Criteria	Unimportant	Important	Neutral	Valid Responses (N)
Lack of ability to locate specific and relevant OER for my teaching	20.5% (63)	57.4% (176)	22.1% (68)	100% (307)
Lack of ability to locate quality OER for my teaching	13.8% (42)	67.6% (207)	18.6% (57)	100% (306)

4.2 Desirability Framework Results

To verify the accuracy of the proposed *D-index*, experiments were carried out in three widely used OER repositories, OER Commons, JORUM and MERLOT. These repositories were purposely selected due to: (i) the availability of native search mechanisms; and (ii) the variety of OER available in different levels of openness and access. Each repository was searched using the term "calculus" to locate OER on the topic of calculus in mathematics. Only the top 40 search results from each repository, returned based on relevance, were considered in the experiment.

Table 4.7, Table 4.9 and Table 4.11 show the top ten results returned by the repository specific search mechanisms of MERLOT, JORUM and OER Commons respectively for the search term "calculus". Table 4.8, Table 4.10 and Table 4.12 show the top 10 results when the *D-index* is applied to the search results returned by MERLOT, JORUM and OER Commons respectively.

Table 4.7 Top 10 search results returned by MERLOT for the keyword "calculus".

Search Rank	Title	CC License	File Type
1	18.01 Single Variable Calculus	CC BY-NC-SA	PDF
2	Calculus for Beginners and Artists	CC BY-NC-SA	webpage
3	18.01 Single Variable Calculus	CC BY-NC-SA	PDF
4	18.013A Calculus with Applications	CC BY-NC-SA	webpage
5	18.02 Multivariable Calculus	CC BY-NC-SA	PDF
6	Single Variable Calculus	CC BY-NC-SA	PDF
7	Calculus Online Textbook	CC BY-NC-SA	PDF
8	Calculus for Beginners and Artists	CC BY-NC-SA	webpage
9	18.075 Advanced Calculus for Engineers	CC BY-NC-SA	PDF
10	MATH 140 - Calculus I, Summer 2007	CC BY-NC-SA	Protected

Table 4.8 Top 10 results when D-index is applied to the results returned by MERLOT.

Rank After Applying D-index	Original Search Rank	Title	CC License	File Type	*D-index*
1	2	Calculus for Beginners and Artists	CC BY-NC-SA	webpage	0.75
2	4	18.013A Calculus with Applications	CC BY-NC-SA	webpage	0.75
3	8	Calculus for Beginners and Artists	CC BY-NC-SA	webpage	0.75
4	14	Multivariable Calculus	CC BY	webpage	0.75
5	19	MATH 10250 - Elements of Calculus I, Fall 2008	CC BY-NC-SA	webpage	0.56
6	20	18.022 Calculus	CC BY-NC-SA	PDF	0.56
7	22	Single-Variable Calculus I	CC BY	webpage	0.50
8	25	Single-Variable Calculus II	CC BY	webpage	0.50
9	15	Highlights of Calculus	CC BY-NC-SA	Video	0.42
10	21	Calculus I	CC BY	webpage	0.38

Table 4.9 Top 10 search results returned by JORUM for the keyword "calculus".

Search Rank	Title	CC License	File Type
1	Introduction to Calculus	CC BY-NC	Video
2	Introduction to Artificial Intelligence - Neural Networks	CC BY-NC-SA	MS Word
3	Calculus (integration) : mathematics 1 level 4	CC BY	Slides
4	Calculus - Income Growth, Consumption and Savings	CC BY-NC	Video
5	Introduction to Econometrics: EC220	CC BY-NC	PDF
6	Further Mathematical Methods	CC BY-NC-SA	webpage
7	Transient responses : laplace transforms : electrical and electronic principles : presentation transcript	CC BY	Slides
8	Calculus - Determining Marginal Revenue	CC BY-NC	Video
9	Film Series Four - Conclusion	CC BY-NC	Video
10	Finding the optimal number of floors in hotel construction - part one	CC BY-NC	Video

Table 4.10 Top 10 results when D-index is applied to the results returned by JORUM.

Rank After Applying D-index	Original Search Rank	Title	CC License	File Type	*D-index*
1	1	Introduction to Calculus	CC BY-NC	Video	0.75
2	4	Calculus - Income Growth, Consumption and Savings	CC BY-NC	Video	0.75
3	6	Further Mathematical Methods	CC BY-NC-SA	webpage	0.75
4	8	Calculus - Determining Marginal Revenue	CC BY-NC	Video	0.75
5	9	Film Series Four – Conclusion	CC BY-NC	Video	0.75
6	10	Finding the optimal number of floors in hotel construction - part one	CC BY-NC	Video	0.75
7	13	Maths Solutions	CC BY	webpage	0.75
8	11	Finding the optimal number of floors in hotel construction - part two	CC BY-NC	Video	0.56
9	12	Finding the optimal number of floors in hotel construction – Conclusion	CC BY-NC	Video	0.56
10	14	Mathematical analysis	CC BY-NC-SA	webpage	0.56

Table 4.11 Top 10 search results returned by OER Commons for the keyword "calculus".

Search Rank	Title	CC License	File Type
1	Whitman Calculus	CC BY-NC-SA	webpage
2	Calculus I	CC BY-NC-SA	PDF
3	AP Calculus	CC BY-NC-SA	webpage
4	Applied Calculus	Proprietary	webpage
5	A Summary of Calculus	Proprietary	PDF
6	Advanced Calculus	CC BY-NC-SA	PDF
7	Multivariable Calculus	Proprietary	PDF
8	Topics in Calculus	CC BY-NC	PDF
9	Highlights of Calculus	CC BY-NC-SA	Video
10	Vector calculus	Proprietary	webpage

Table 4.12 Top 10 results when D-index is applied to the results returned by OER Commons.

Rank After Applying D-index	Original Search Rank	Title	CC License	File Type	*D-index*
1	1	Whitman Calculus	CC BY-NC-SA	webpage	0.75
2	3	AP Calculus	CC BY-NC-SA	webpage	0.75
3	11	Vector Calculus	GNU Free Documentation License	webpage	0.75
4	9	Highlights of Calculus	CC BY-NC-SA	Video	0.56
5	16	Calculus (Student's Edition)	CC BY-NC-SA	webpage	0.56
6	22	Calculus II (MATH 152)	CC BY	webpage	0.50
7	23	Calculus I (MATH 151)	CC BY	webpage	0.50
8	24	Calculus III (MATH 153)	CC BY	webpage	0.50
9	15	Calculus Revisited, Fall 2010	CC BY-NC-SA	Video	0.42
10	21	Calculus (Teacher's Edition)	CC BY-NC-SA	webpage	0.38

The analysis of Tables 4.7 to 4.12 is provided in Section 5.2 of the Discussion.

4.3 Prototype Implementation Results

The application of the system in a real world scenario was undertaken using the Directory of Open Educational Resources (DOER) of the COL. DOER is a fledgling portal repository which provides an easily navigable central catalogue of OER scattered across the globe. At present, the OER available through DOER are manually classified into 20 main categories and 1158 sub-categories. Despite covering most of the major subject categories, this particular ontology still needs to be expanded by a large degree due to the variety of OER available in an array of subject areas. This expansion, in turn, is a tedious and laborious task which needs to be accomplished manually on an ongoing basis. As a possible solution to this issue, a mechanism was needed for autonomously identifying the subject area(s) covered in a particular OER, in the form of keywords, in order for it to be accurately catalogued.

Given this requirement, the DOER was used as the training dataset for OERScout. In addition to the resources categorized in DOER, 1536 resources from the Rice University's Connextions repository were also included in the training dataset due to: (i) the large number of OER materials available; and (ii) the relatively high popularity and usage rates. An XML sitemap that contains a total of 1999 URLs belonging to the domains of arts, business, humanities, mathematics and statistics, science and technology, and social sciences was created as the initial input. The system was run with the initial input and was allowed to autonomously create the KDM. This training process ensured that the algorithm had an initial set of academic domains and sub-domains which it could use to accurately cluster the resources.

On average, each document required 15-90 minutes to be downloaded, read and learnt by the system, depending on the size and file type. The system took approximately five days to process all the documents in the training dataset. Although the training process required a considerable amount of time due to the lack of optimization and enterprise

scale infrastructure, this process takes place as a background operation at the server. Therefore, once the KDM is created, the end user does not experience any delays during the search process. When implemented, new repositories will be identified for crawling based on referrals by end users. The sitemaps created by the crawlers will be input into the system to be processed. The server tools will continuously run at the server processing new documents and re-visiting processed documents to ensure accuracy.

After completing the run, the system had processed documents of various size, file types and licenses from 11 repositories, representing many regions of the world (Table 4.13). There was a certain amount of noise in the keywords identified due to the limited number of resources indexed in a given domain. However, with more documents being indexed, the expansion of the *List of Terms* will result in larger IDF values which will decrease the TF-IDF value for noise words. This will result in the algorithm rejecting these noise words as keywords, which is the reduction of noise.

Table 4.13 Resources indexed in the KDM based on the initial input.

	Repository	Host Institution	Region	License	File Type	No. Resources Indexed
1.	Connexions	Rice University	USA	CC BY	Webpage	1536
2.	OCW Athabasca	Athabasca University	Canada	CC BY	Webpage	07
3.	OCW Capilano	Capilano University		CC BY-NC-SA	Webpage	19
4.	OCW USQ	University of Southern Queensland	Australia	CC BY-NC-SA	Webpage	10
5.	UCT Open Content	University of Cape Town	South Africa	CC BY-NC-SA	Webpage	63
6.	OpenLearn	The Open University	UK	CC BY-NC-SA	Webpage	242
7.	WikiEducator	COL & Ottago Polytechnic	New Zealand	CC BY-SA	Webpage	38
8.	Unow	University of Nottingham	UK	CC BY-NC-SA	Webpage	27
9.	TESSA	Multiple African Universities	Africa	CC BY-SA	PDF	15
10.	OER AVU	African Virtual University	Africa	CC BY-SA	DOC DOCX PDF	40
11.	WOU OER	Wawasan Open University	Malaysia	Various	PDF	02
	Total					**1999**

4.4 User Test Results

In order to test the functionality of the system from a real-world user's perspective, 27 academics with at least 3-5 years of experience in OER advocacy, creation, use, and reuse were invited to test the system. Out of the 27 experts invited, 19, including six professors, five associate professors, three PhD holders, and four mid-career academics, agreed to test the system and provide feedback. This group of users represented Australia, Brazil, Cambodia, Canada, China, Hong Kong SAR, Indonesia, Malaysia, Pakistan, and Vietnam. They comprised varied backgrounds such as engineering, computer science, electronics, instructional design, distance education, agriculture, biology, law, and library science. The KDM was made available to this group through the OERScout client interface shown in Figure 3.8.

A comprehensive user manual (Appendix J) was provided to the users which outlined how OERScout searched for the most *desirable* resources. The testing was conducted for seven days. The users tested the system by searching for academic material, in the form of OER, to be used in their day-to-day teaching and learning activities. At the end of the test period, the users provided qualitative feedback through a web based feedback form on various aspects of the OERScout framework. The consolidated feedback is shown in Table 4.14. The feedback form and some significant comments are provided in Appendix K.

Table 4.14 Consolidated feedback gathered from the OERScout test users.

	Criteria	Advantages of OERScout	Weaknesses of the prototype
1.	User interface	The user interface is quite simple, friendly, intuitive, un-cluttered and easy to operate. It avoids the hassle of switching between search modes.	Add advanced search features such as year, language, author and type of resources are not available.
2.	"Faceted search" approach which allows users to dynamically generate search results based on suggested and related terms	The ability to drill down using "faceted search" is very useful. It helps to locate resources faster.	As the number of resources grows the list of suggested and related terms will be quite long. Some noise terms are generated along with the keywords.
3.	Ease of use	It is a powerful tool which allows users to easily locate relevant resources.	The number of resources indexed is quite small.
4.	Relevance of the suggested terms generated according to the search query	The suggested terms are quite relevant and cover the scope of the search adequately.	Some unfamiliar noise words were generated as suggested terms.
5.	Use of related terms to effectively zero in on the resources being searched for	The feature is very useful and performs well. The functionality is similar to a thesaurus used by librarians for cataloging.	Many different terms point to the same resource due to the small dataset. Some terms are not related to the domain. Too many terms are generated.
6.	Usefulness of the resources returned with respect to Openness (the ability to use, reuse, revise and remix)	The use of the CC license to locate the most open resources is a useful feature. The value of this feature will increase along with the increase of quantity and quality of OER available.	The licensing scheme needs to be indicated in a more user-friendly manner.
7.	Usefulness of the resources returned with respect to Access (the ease of reuse and remix of resource type)	The resources returned met the criteria of access with respect to use and reuse. Based on the resource type, users can immediately identify how they can use the resource.	This might not be important as the licensing type defines the reuse and remix capabilities.
8.	Usefulness of the resources returned with respect to Relevance (the match between the results and your query)	Currently quite accurate and very useful.	The small size of the dataset limits the relevance.

9.	Effectiveness with respect to identifying the academic domain(s) of a resource	The autonomous identification of academic domains increases the focus of the search and the quality of the resources returned.	The technology shows promise but the number of domains identified is limited due to the size of the dataset.
10.	Use of the desirability for filtering the most useful resources for ones needs	The desirability framework is an interesting idea which will help in identifying resources appropriate for specific needs.	The concept of desirability needs to be explained to the user through the interface.
11.	Effectiveness with respect to locating desirable resources in comparison to mainstream search engines or native search engines of OER repositories[a]	A comparison between the OERScout and conventional search engines cannot be made as they serve different purposes. OERScout is much more focused and addresses some key issues in OER search.	Search engines such as Google have large databases of indexed resources. In this sense they cannot be compared to OERScout.
12.	Innovativeness of the technology framework	The technology framework is quite innovative and can bridge the gap between different metadata standards. The simplicity of the user interface complements the scale of innovation.	The scope of the framework needs to be refined. The system needs to be made available as an online service.
13.	How the wider OER community will be benefited	The technology will benefit the wider OER community as a tool for thought provoking discussion on adopting and adapting resources. It will be very beneficial for the novice user with respect to ease of use and affordability.	At the moment it is only a prototype. More resources need to be indexed before it can benefit the community.

[a] The comments presented under items 11 and 12 of Appendix K have been combined in this section.

Summary

This chapter has presented the results from the various phases discussed in the methodology. The first section discusses the results from the survey study where 420 academics expressed their views on the current OER search situation in their respective regions. The results suggest that there is definitely a need for better search methodologies in these regions. The second section provides empirical evidence of the effect the *D-index* has on relevance based search results. The results are derived from three popular repositories which host comprehensive catalogues of OER. The results from the prototype implementation are discussed in the next section. 1999 resources from 11 repositories were used in the creation of the KDM which hosted resources of varying domains, localities, lengths, licenses and file types. The resulting KDM was used in the expert user tests discussed in the final section. Nineteen experts, including professors, associate professors, PhD holders, and career academics, provided feedback on the system.

The next chapter provides a comprehensive discussion of the results. It will concentrate particularly on the issues of finding useful resources, centralised search mechanisms, users' perspectives, advantages of the system and contributions of the project.

CHAPTER 5

DISCUSSION

Chapter 5 : Discussion

The results chapter presented key information gathered during the project. The first set of results was from the survey study which aimed to gain an understanding of the current issues related to OER search, especially in the Asian region. Being a region which has been heavily investing in the OER movement and benefits from freely available educational material, the dilemma in the Asian region is representative of the wider global problem. The second set of results provided evidence of the use of the desirability framework in parametrically measuring the usefulness of an OER. It demonstrated how search results from the native search mechanisms of popular repositories can be rearranged using the *D-index* to provide users with optimal results. The OERScout prototype implementation results provide an understanding of how the KDM is created for a large dataset of OER which consist of many data types, file formats and subject domains. The last result provides feedback from expert users, who have had at least 3-5 years of experience in OER, on the usefulness and innovation of OERScout in a real world setting.

This chapter takes a critical look at the results gathered during the various phases of the research project. It also discusses how the objectives of the project have been met through the methodology and results. The contributions of the project and the advantages of the research are also highlighted with respect to the objectives.

The rest of the chapter is organized into five key sections. The first section looks at the survey study which identifies the issues at hand. The second section looks at the desirability framework in terms of providing a parametric measure for assessing the usefulness of a resource for a particular teaching or learning need. The third section discusses the advantages of the OERScout prototype system over existing OER search methodologies. The user feedback section concentrates on establishing the usefulness

and practicality of OERScout in a real world setting. The last section reports on the objectives of the project and highlights its contributions.

5.1 The Issues

Section 3.1 of the Methodology chapter provided a detailed overview of the empirical research conducted to identify the extent of the current OER search dilemma in the Asian region. Nine countries representative of sub-regions in Asia were involved in this study, as shown in Table 5.1.

Table 5.1 Representation of Asian sub-regions in the survey responses.

	Country	Region
1.	China	East Asia
2.	Japan	
3.	Hong Kong	
4.	South Korea	
5.	Malaysia	South East Asia
6.	Philippines	
7.	Indonesia	
8.	Vietnam	
9.	India	South Asia

The results of the survey study are presented in section 4.1 of the Results chapter. Of the academics who participated in the survey, 65% had used OER from other academics in their teaching and 80% mentioned that they will use OER in their teaching in the future (Table 4.3). This shows that the use of OER is gaining popularity and wider acceptance in the Asian region. Additionally, referring to Table 4.4, the attitude towards the use of OER is also taking a positive turn as 77% of the participants found OER to be a useful way of developing courses while 79.8% agreed that OER will improve the standard of their teaching. However, even though the use of OER and the attitudes towards it are improving, 57.4% of the academics found that the lack of ability to locate specific and relevant resources was an important inhibitor towards the use of OER (Table 4.6). Furthermore, 67.6% of the academics felt that the lack of ability to locate quality OER was another issue worth consideration.

In order to identify the reasons behind academics not being able to locate relevant OER for their teaching, the method of searching for OER needs to be scrutinized. From Figure 4.1, it is apparent that the majority of academics search for OER which are freely available on the internet, as opposed to using specific OER repositories. Furthermore, Table 4.5 shows that generic search engines such as Google, Yahoo! and Bing are used predominantly for OER search in comparison to the native search mechanisms of OER repositories. From this comparison, it can be seen that many academics depend on generic search mechanisms to locate relevant OER for their teaching purposes. However, the inability of these generic mechanisms to locate relevant OER is highlighted in Section 2.2 of the Literature Review chapter. Only 43.2% of the academics used native search mechanisms of OER repositories which fare better at locating relevant OER. 64% of the academics felt that the lack of awareness of the existence of such repositories was the key contributor to this situation.

The above analysis clearly highlights the issues in the Asian region with respect to OER search. As a result, the following underlying questions are raised:

(i) how can search engines assist in finding resources which are more useful for a particular teaching and learning need?; and

(ii) how does one search for useful OER distributed in heterogeneous repositories using a central mechanism?

5.2 Finding Useful Resources

Section 3.2 of the Methodology chapter introduced the desirability conceptual framework which takes into consideration (i) the level of openness; (ii) the level of access; and (iii) the relevance attributes of an OER to provide a parametric measure of its usefulness. Section 4.2 of the Results chapter detailed the results obtained by applying the *D-index* to three widely used repositories, namely MERLOT, JORUM and OER Commons. This section discusses the significance of these results.

A comparison of Table 4.7 and Table 4.8 on MERLOT show that the original top ten search results (Table 4.7) contain only resources which are released under the CC BY-NC-SA license. This license significantly restricts the user's freedom with respect to the four R's discussed in Section 2.1.1. Also, 6 of the 10 resources returned are in PDF format, which make them difficult to reuse and remix. It is also noted that the resource ranked as number ten is a protected resource, which requires a username and password to access. Looking at Table 4.8 where the results are re-ranked according to the *D-index*, it can be seen that 8 of 10 resources are in HTML/Text formats, which are the most accessible in terms of reuse. Four of the 10 resources are available under the CC BY license (Table 2.1) which makes them the most open resources in the list. Similarly, by comparing Table 4.9 and Table 4.10 we can see that the use of the *D-index* has re-ranked the top ten results so that the most accessible resources are ranked at the top instead of resources which use proprietary software applications. In addition to the textual resources, the video resources returned were given an accessibility value of 12 according to the ALMS (Table 3.3), where: access to editing tools = high; level of expertise required to revise or remix = high; meaningfully editable = yes; and source-file access = yes.

Analyzing Table 4.11 it can be seen that 4 of 10 results returned by the OER Commons search mechanism are copyright protected. As such, these cannot be considered as OER and are the least useful for a user who is searching for open material. A value of 0 for openness was assigned to these resources during the *D-index* calculation. Furthermore, 5 of the top ten results returned by the OER Commons search mechanism were in PDF format. Looking at Table 4.12, it can be seen that the application of the *D-index* has re-ranked the resources to provide 8 out of 10 HTML/Text resources. Also the proprietary contents have been replaced with more open content released under the CC BY and CC BY-NC-SA licenses. The third ranked resource, which is released under the GNU Free Documentation License, was assigned a value of 4 for openness (Table 3.2) during the calculation of the *D-index*.

By referring to the above results it can be concluded that application of the *D-index* greatly improves the effectiveness of the search with respect to locating more suitable resources for use and reuse. This in turn gives academics more flexibility when incorporating the material into their teaching and learning activities.

5.2.1 Application and Limitations

The *D-index* can be incorporated into any search mechanism of an OER repository provided that the resources in the repository are appropriately tagged with the necessary metadata such as title, description, keywords, copyright license and file type. Many OER repositories now require authors to define these basic metadata as standard. As such the use of these parameters to gauge the values for relevance and openness becomes an easier task. However, gauging the access parameter which uses the file type of the OER is a much more challenging task, as some resources consist of multiple files in multiple formats. This can be rectified by breaking a collection of OER into individual LO which allows software applications to determine the file type of the individual OER.

A couple of practical limitations can also be identified with respect to the implementation of the *D-index* in OER repositories. One of these limitations is the desirability being expressed in one dimension, due to fixed copyright licenses and file formats in repositories such as Connexions or Wikieducator. As a result, the *D-index* becomes a function only of the relevance parameter, which would not add much value to the existing search mechanism. Therefore, the *D-index* is best suited for use in repositories such as the OER Commons, MERLOT and JORUM which provide a central location for searching OER distributed in heterogeneous repositories.

The other practical limitation is the subjectivity of the search algorithms used by the various native search mechanisms. That is, search ranks may vary from repository to repository depending on the search algorithm used. In turn, this disparity results in the relevance parameter becoming a function of the search algorithm. Therefore, the need for a single algorithm which measures the relevance of a resource in a uniform manner is brought to attention.

5.3 Centralized Search Mechanism

Section 2.3 of the Literature Review chapter discussed promising recent OER search initiatives. These initiatives can be broadly categorized into federated search and semantic search. The federated search approach harvests metadata from repositories to build a centralized searchable index. The semantic approach depends on the creation of ontologies specific to a particular subject domain, using the metadata. In both cases ensuring the accuracy and uniformity of the user annotated metadata remains a major challenge.

Further to Section 2.3, the accuracy of user annotated metadata is questionable, even though there are a number of established metadata standards being used for OER. Therefore, the accuracy of the search for relevant resources becomes a function of how accurately the content creator can define the metadata. This is of special concern when it comes to the definition of keywords as it is challenging to accurately describe the complete content with a limited number of terms. Furthermore, the enforcement of global metadata standards for large volumes of expanding resources poses a major challenge to existing OER search methodologies. The OERScout technology framework introduced in Section 3.3 could provide a viable solution to these issues.

As discussed in Section 2.2, generic search methodologies such as Google are currently inept at locating relevant OER for a particular teaching need. To illustrate this point, advanced searches were conducted on Google (google.com.my) for the terms "chemistry" and "calculus" respectively. The advanced search parameters were set to search for resources which are *free to use, share or modify, even commercially.* Confirming the statements made in literature, the top five search results returned (Figure 5.1 and Figure 5.2) were from Wikipedia (wikipedia.org) which is an encyclopedia of user created learning objects rather than a repository of credible educational material (Kubiszewski, Noordewier, & Costanza, 2011). This accounts for

50% of the relevant results returned as users will consider only the top ten ranked results for a particular search (Vaughan, 2004). Additionally, two similar searches were conducted on Yahoo! (yahoo.com.my) and on Bing (bing.com). Since both Yahoo! and Bing do not provide mechanisms for conducting searches specifically for OER, the phrase "Creative Commons" was added to the search query to find CC licensed material on these two search engines. The results were similar to the case of Google.

Figure 5.1 Google Advanced Search results for resources on "chemistry" which are free to use, share or modify, even commercially (27th November 2012).
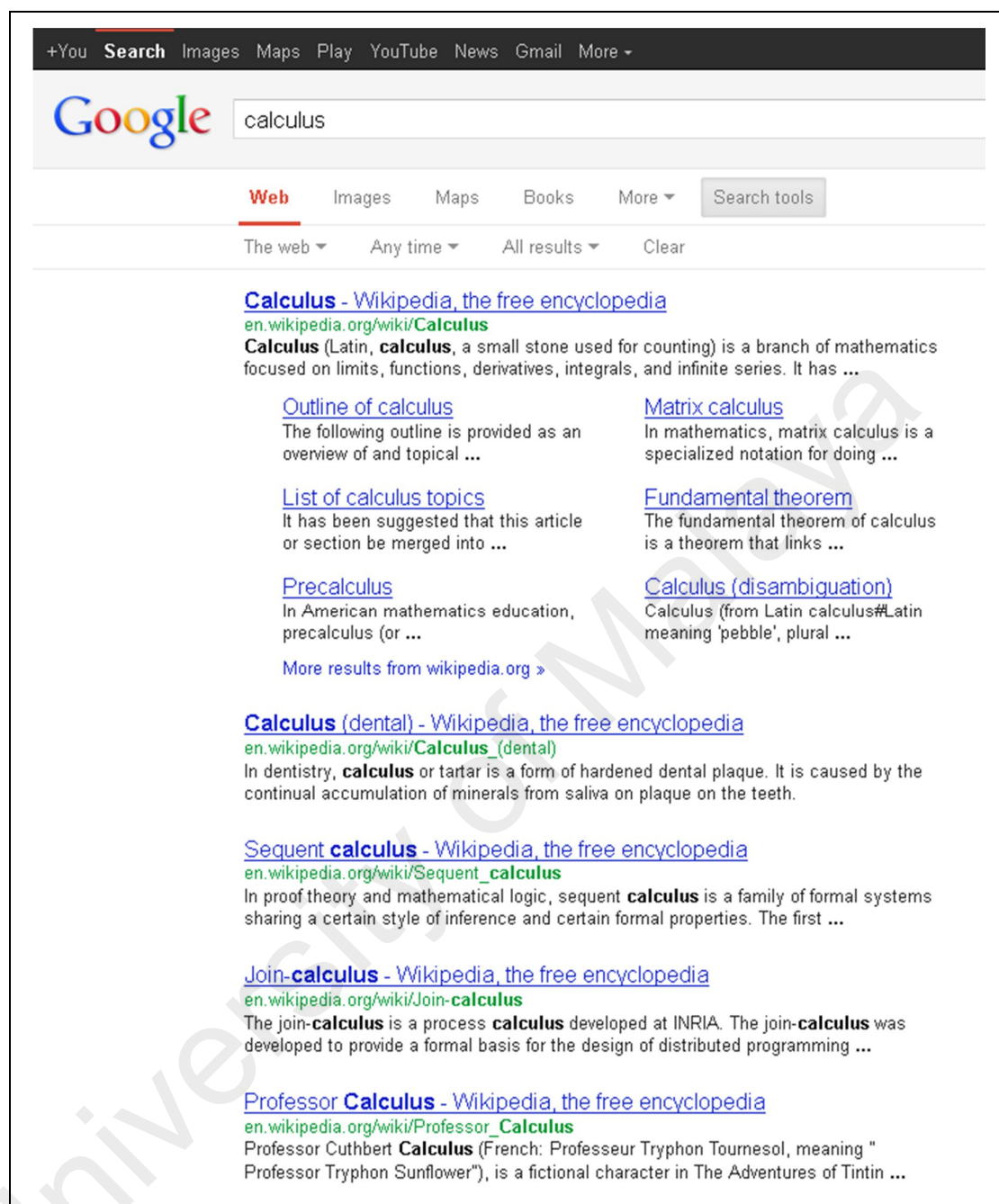
Figure 5.2 Google Advanced Search results for resources on "calculus" which are free to use, share or modify, even commercially (27th November 2012).
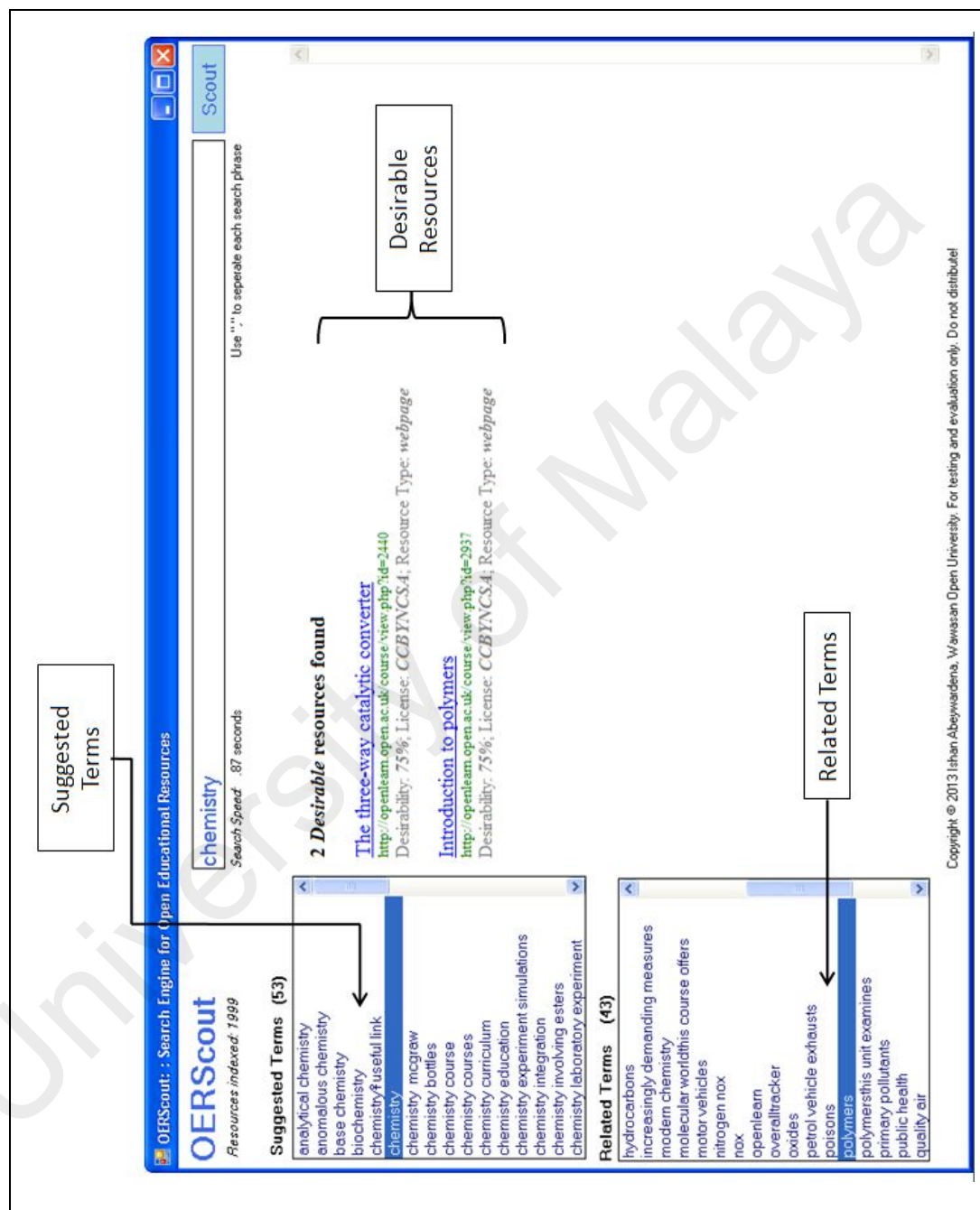
Figure 5.3 A search result for resources on "chemistry: polymers" conducted on OERScout.

Figure 5.3 shows a search conducted for the term "chemistry" on OERScout, based on the KDM explained in Section 4.3 of the Results chapter. In contrast to the static list of search results produced by generic search engines such as Google (Figure 5.1 and Figure 5.2), OERScout employs a *faceted search* approach by providing a dynamic list of *Suggested Terms* which are related to "chemistry". The user is then able to click on any of the *Suggested Terms* to access the most *desirable* OER from all the repositories indexed by OERScout. Furthermore, based on the selection by the user, the system will provide a list of *Related Terms* which enables the user to drill down further to zero in on the most suitable OER for their teaching needs. In this particular example (Figure 5.3), the user has selected "polymers" as the related term to locate two *desirable* resources from the OpenLearn repository of The Open University which is known to host OER of high academic standard. In addition, Figure 5.4 shows the search results returned by OERScout for the search query on "calculus".

Figure 5.4 Search results generated by OERScout for the term "calculus". The desirable resources returned are from Capilano University, The Open University and African Virtual University.

The *desirable* resources returned are from the OCW Capilano of Canada, OpenLearn of UK and OER AVU of Africa. As such, it can be seen that OERScout is a central and dynamic mechanism for effectively searching for *desirable* OER from heterogeneous repositories. This is a major benefit to academics as the system spares them from conducting repeated keyword searches in OER repositories to identify suitable material for use (Figure 3.3). It also allows users to quickly zero in on OER suitable for their needs without reading through all the search results returned by a generic search mechanism such as Google. Table 5.2 summarizes some of the key features of OERScout in contrast to the generic search engines Google, Yahoo! and Bing.

Table 5.2 Key Features of OERScout in contrast to Google, Yahoo! and Bing.

	Key Feature	OERScout	Google	Yahoo!	Bing
1.	Provides a centralised mechanism to search for OER	Yes	Yes	No	No
2.	Searches for only the most *desirable* resources for academic purposes	Yes	No	No	No
3.	Effectively locates and presents resources from the distributed repositories	Yes	No	No	No
4.	Provides a dynamic mechanism instead of a static list of search results which can be used to zero in on the required resources	Yes	No	No	No
5.	Uses autonomously identified keywords for locating the most relevant resources	Yes	No	No	No
6.	Uniformly annotates resources with the relevant keywords to facilitate accurate searching	Yes	No	No	No
7.	Removes human error in the annotation of keywords	Yes	No	No	No

It should be noted that generic search engines such as Google can be configured to search only in a limited set of repositories instead of the whole web. However, the

average OER user does not have the technical knowhow or would not spend the time to configure a generic search engine into one which is specific to a limited set of repositories. Furthermore, it will be a daunting task to add new repositories to the filter on an ongoing basis. Based on these assumptions, the comparison presented in Table 5.2 holds true for the vast majority of OER users.

5.4 Users' Perspective

The user tests, explained in Section 4.4 of the Results chapter, looked at how real-world users react to the complete OERScout concept which incorporates: (i) autonomous keyword identification; (ii) the desirability framework; and (iii) the faceted search approach. The users identified to test the system are experts in the field of OER who have had experience in using generic search engines, native search mechanisms of repositories as well as many of the OER search methodologies discussed in Section 2.4. After the test phase, the users provided qualitative feedback on the advantages of OERScout technology framework and the weaknesses of the current prototype. This feedback is summarized in Table 4.15.

Based on the user feedback, the strengths, weaknesses, opportunities and threats (SWOT) of the OERScout technology framework are shown in Table 5.3. The key strengths of the system include the ease of use, the specific focus on OER, the ability to quickly zero-in on the required resource and the use of desirability. The ability to autonomously identify academic domains in the form of keywords is found to be an advantage of the system. The capability of locating resources from distributed heterogeneous repositories using a central mechanism is found to be among its strengths. The users felt that OERScout will especially benefit academics who are novices to OER.

Table 5.3 SWOT analysis of OERScout based on user feedback.

Strengths	Weaknesses
- Simple, user friendly and intuitive to use - Focuses specifically on OER - Allows drilling down to find the most suitable resources - Generates results quickly based on suggested and related terms - Locates resources which are the most desirable - Provides useful information such as desirability, license type and resource type - Useful for finding resources when developing OER - Autonomous identification of academic domain helps users to pinpoint the right resources - Locates resources from any repository irrespective of the metadata standard used	- Number of resources indexed is small - Generates noise terms - Related terms are not focused - Suggested and related terms lists are too long - Lacks advanced search features - The indication of license is not user friendly - The desirability is not explained to the user - Technology framework limits use
Opportunities	**Threats**
- Provoke discussion about desirability and workflows for finding, adopting, and adapting available OER - Useful for training novice users of OER - Appeals to individuals due to affordability - Helps target OER better and faster for teaching and learning	- Mainstream search engines such Google have larger databases of resources - Some resources might be missed by OERScout while indexing - Change in mindset of the user with respect to faceted search

One of the major weaknesses of the current prototype version is the limited number of resources indexed. This contributes to noise in the identified keywords and results in lengthy lists of suggested and related terms. However, as the number of indexed resources grows, the noise words will reduce giving way to more focused suggested and related terms. The users also suggest that more advanced filters need to be added to the search interface to allow searching for specific file types and licenses, among others. However, the fundamental concept behind the desirability framework is to parametrically identify the most useful resources without the user's intervention. This observation suggests that a change in mindset with respect to search engines needs to

take place before users are accustomed to OERScout. The users also believe that the licensing scheme needs to be explained in non-technical terms such as *"can reuse, redistribute, revise and remix even commercially"* instead of *"CC BY"*. They further suggested that the calculation of the desirability be explained to the user.

The technology architecture used for the prototype is also found to be a limitation of the current system. The Microsoft Windows based client interface limits the users to Microsoft PC consumers. However, real world implementation of the system will be done on a web based platform which will provide wider access regardless of device or operating system. Another limitation is that this version of OERScout is not designed to cluster non-text based materials such as audio, video and animations. This is a drawback considering the fact that more and more OER are now being developed in multimedia formats. However, it is noted from the initial results that the system will accurately index multimedia material using the textual descriptions provided. One more design limitation is its inability to cluster resources written in languages other than English. Despite this current limitation, the OERScout algorithm has a level of abstraction which allows it to be customized to suit other languages in the future.

Considering the opportunities, the system is reported to be thought provoking with respect to finding, adopting and adapting OER. OERScout appeals to the novice OER users in terms of training, affordability, teaching and learning. This in turn will promote further research and development in the field of OER. Analyzing the threats, one of the major threats to OERScout is the scale of the resource databases available to mainstream search engines such as Google. In this respect, users believe that OERScout will be unable to compete with these search engines. However, the users also suggest that OERScout addresses a few focused issues related to OER, and need not be compared to mainstream search engines which are more general in nature. It is also noted that the system will need to continuously update its resource database to ensure

accuracy. Among the threats identified, the change in mindset with respect to this new search approach probably remains the greatest challenge to overcome.

Summary

This chapter has critically analyzed the complete research project in terms of the objectives, methodology, results and the contributions. The first section examined the current issues at hand with respect to OER search in the Asian region. As a result of the empirical research, two key issues were identified, which this project addresses. The second section discussed the desirability conceptual framework, which provides a parametric measure of the usefulness of a resource based on openness, accessibility and relevance. The third section detailed the advantages of the OERScout technology framework, which incorporates: (i) text mining techniques for autonomous keyword tagging of resources; (ii) the desirability framework; and (iii) the faceted search approach. The fourth section probed the views of real-world expert users in terms of advantages of the OERScout framework and the weaknesses of the current prototype.

The next chapter is the concluding chapter which summarizes the complete research project. It further examines the outcomes of the project against the objectives it set out to achieve. Finally, it provides insight into the future direction of this research work.

Chapter 6

CONCLUSION

Chapter 6 : Conclusion

Open Educational Resources (OER) are fast becoming accepted sources of knowledge for teachers and learners throughout the globe. This is especially true in the case of ODL institutions where the teaching and learning philosophy is based on increased access to education. With recent advanced developments in technology as well as the establishment of many high quality OER repositories freely available online, the use and reuse of OER should have become more mainstream practice. However, as it stands today, the use and reuse of OER are still inhibited by a number of technological, social and economic reasons (D'Antoni, 2009).

One technological reason for the slow uptake is the inability to effectively search for useful or *fit-for-purpose* OER from the various heterogeneous OER repositories. With the rapid mushrooming of new and the expansion of existing OER repositories, it has become increasingly difficult to manually trawl each repository to identify OER required for specific teaching purposes. As such, this limitation has become an inhibitor to wider adoption of OER, especially in the Asian region.

When considering the technological limitations, the inability of mainstream searching mechanisms, such as Google, Yahoo! and Bing, to accurately distinguish between an OER and a non-OER becomes a major hurdle. Although the more popular search engines do provide advanced filter criteria to refine the searches, these search engines are not tailored to search for OER which are the most useful in terms of the ability to use, reuse, revise and remix. This limitation forces OER consumers to resort to frequenting the more popular repositories such as Rice Connexions, MIT OCW and Wikieducator to search for suitable OER. However, this too has become a cumbersome and time consuming task due to influx of repositories and the constant expansion of resources volume. It is consequently not feasible to manually keep track of all the

available OER repositories. Also, users have to spend extended hours on these repositories conducting multiple searches using the native search mechanisms to locate the resources they are after. This limits the scope and the variety of OER available to them. Ultimately, even though many of these OER repositories hold a rich selection of material, the user is stuck in a scenario where the use of these materials is not a choice but a lack of options.

Contributing to this issue is the lack of an accepted global metadata standard for OER. Although initiatives such as LRMI are attempting to address this issue, the fundamental fault lies with the metadata itself, as the accuracy and uniformity cannot be completely guaranteed. This results in the accuracy of the search becoming a function of the content creator's ability to accurately annotate a resource with metadata. Therefore the use of human annotated metadata in performing objective searches becomes subjective and inaccurate.

6.1 Research Objectives

This research project has four objectives, as listed in Section 1.1 of the Introduction chapter. The first objective was to **identify user difficulties in searching OER for academic purposes**. Based on the empirical research conducted in the form of literature review and survey study, Section 5.1 outlines the extent of the issues with respect to OER search in the Asian region. This dilemma is of especial concern to academics in the Asian region as Asia would benefit from wider use of open resources. In addition, due to the increasing number of OER initiatives in the region and the generous flow of funding, the volume of regional OER will grow exponentially in the near future. Therefore, the OER search issue in the Asian region adequately represents the wider global problem.

As a result of the survey study, two key issues in the domain of OER search were identified. These issues are:

(i) how can search engines assist in finding resources which are more useful for a particular teaching and learning need?; and

(ii) how does one search for useful OER distributed in heterogeneous repositories using a central mechanism?

The second objective of the project was to **identify the limitations of existing OER search methodologies with respect to locating fit-for-purpose resources from the heterogeneous repositories**. As discussed in Section 2.2 of the Literature Review chapter, the current OER search dilemma is twofold. Firstly, the literature shows that mainstream search engines such as Google are incapable of searching for relevant OER. This is further affirmed through the empirical research discussed in Section 5.3. It is also argued that the large number of resources returned by these search engines deter potential OER users. Therefore, the issue of not being able to assess the

usefulness of an OER for academic purposes negatively impacts the relevance of the search results. The literature also suggests that native search mechanisms of OER repositories are comparatively better at finding relevant resources. However, the major drawback of this methodology is the sheer volume of the repositories available from which to choose. As a result, it is not feasible for users to conduct searches in all of these repositories in order find relevant resources. The second aspect of the dilemma is the heavy dependence of OER search initiatives on user annotated metadata. Section 2.4 provides a detailed overview of existing OER search initiatives which show promise. These are mainly federated search or semantic search initiatives. The use of human annotated metadata in these initiatives for resource federation and ontology development makes the search for relevant OER a function of the human ability to annotate metadata accurately. Furthermore, the non-uniform nature of the metadata and the constantly expanding volume of resources which need to be tagged hinder the progress of these search initiatives beyond the prototype stage.

The third objective was to **conceptualize a framework for parametrically measuring the suitability of an OER for academic use**. The concept of desirability of an OER introduced in Section 3.2 of the Methodology chapter attempts to lessen the pain of OER users with respect to identifying resources which are useful or fit-for-purpose. This usefulness of an OER is derived using the openness, accessibility and relevance attributes unique to the concept of OER.

At present, users who search for OER in specific repositories use native search mechanisms to identify relevant resources. Depending on the algorithms used by these native search mechanisms, the search query will be compared against the metadata of a resource such as title, description and keywords to provide a list of resources which are deemed relevant. However, these search mechanisms do not take into consideration the

level of openness or the technological skills required with respect to use, reuse, remixing and redistribution of a resource. The *D-index*, which is the measure of desirability, is an attempt to factor in the openness and accessibility in addition to the relevance of an OER to provide users a prioritized set of search results which are the most open, accessible and relevant for their needs. Based on the results discussed in Section 5.2, the application of the desirability framework successfully enables search mechanisms to re-order the search results to provide increased access to more useful resources. The *D-index* can be incorporated into any OER repository provided that the necessary metadata for calculation are available. It is most effective when used in portal repositories/content and portal repositories which search multiple heterogeneous repositories to locate OER.

The final objective of the project was to **design a technology framework to facilitate the accurate centralised search of OER from the heterogeneous repositories.** The OERScout technology framework, introduced in Section 3.3 of the Methodology chapter, is a technology framework which uses text mining techniques to annotate OER using autonomously mined domain specific keywords. It has been developed with a view to providing OER creators and users with a centralized search tool to enable effective searching of *desirable* OER for academic use from heterogeneous repositories (the repositories used in this research work are provided in Table 4.17). The benefits of OERScout to content creators include: (i) elimination of the need for manually defining content domains for categorization in the form of metadata; (ii) elimination of the need for publicizing the availability of a repository and the need for custom search mechanisms; and (iii) reach of material to a wider audience. The system benefits OER users by: (i) providing a central location for finding resources of acceptable academic standard. Resources of acceptable academic standard can be defined as resources which have gone through some process of academic Quality Assurance (QA); and (ii) locating

only the most *desirable* resources for a particular teaching and learning need. The ultimate benefit of OERScout is that both content creators and users need only to concentrate on the actual content and not the process of searching for *desirable* OER. The feedback gathered on the technology framework from expert users indicates that the OERScout system is a highly viable solution to the current OER search dilemma.

Considering the aforementioned outcomes, it can be concluded that the project has successfully achieved its research objectives. Furthermore, the two issues identified through the survey study have been convincingly addressed through this research work. The Desirability framework provides a potential solution to the question on how search engines assist in finding resources which are more useful for a particular teaching and learning need; whereas the OERScout technology framework addresses the question on how one could search for useful OER distributed in heterogeneous repositories using a central mechanism.

6.2 Research Contributions

The contributions of this research project are twofold:

1. A major problem in OER search is the difficulty in finding quality OER matching a specific context suitable for academic use. This is due to the lack of a framework which can measure the usefulness of an OER in terms of fit-for-purpose, taking into consideration the key attributes of an OER. **The first contribution of this research project is a conceptual framework which can be used by search engines to parametrically measure the usefulness of an OER taking into consideration the openness, accessibility and relevance attributes**.

   o The advantage of this framework is that, using the well-established four R's and ALMS frameworks, it can restructure search results to prioritize the resources which are the easiest to reuse, redistribute, revise and remix. As a result, academics practicing the Open and Distance Learning (ODL) mode of delivery can locate resources which can be readily used in their teaching and learning.

2. Another major problem encountered in OER search is the inability to effectively search for academically useful OER from a diversity of sources. The lack of a single search engine which is able to locate resources from all the heterogeneous OER repositories further adds to the severity of this issue. **The second contribution of this research project is to develop a novel search mechanism which uses text mining techniques and a faceted search interface to provide a centralised OER search tool to locate useful resources from the heterogeneous repositories for academic purposes**.

   o One of the key advantages of this novel search mechanism is the ability to autonomously identify and annotate OER with domain specific keywords.

- o This removes human error with respect to annotation of metadata as it is done in a consistent and uniform manner by the system. As a result, this novel search mechanism provides a central search tool which can effectively search for OER from any repository regardless of the technology platforms or metadata standards used.

- o Another major advantage of this novel search mechanism is the utilization of the conceptual framework which can parametrically measure the usefulness of an OER in terms of fit-for-purpose. This ability allows the search mechanisms to restructure the search results returned from numerous repositories giving priority to the most open, most accessible and most relevant resources. As a result, academics are able to easily locate high quality OER from around the world which are the best fit for their academic needs.

6.3 Future Work

The translation of OER from English into regional languages has been in effect for many years starting with MIT OCW. However, it is apparent that with increased capacities in creating and re-purposing OER, educators are now concentrating more on creating OER in the local language tailored to the local context to better serve the teaching and learning needs of the region. According to Stacy (2007, p. 1),

"Initiatives involved in translating English OER found that their local stakeholders often sought to not simply import, translate, and reuse these existing OER but to create their own local OER."

This trend is especially prominent in countries such as Indonesia, Malaysia, Vietnam and Philippines where most of the teaching and learning still happens in the national language. However, current mainstream search mechanisms are incapable of searching and locating relevant OER developed in these regional languages as they are even unable to effectively search and locate resources written in the English language. As such there is an increased demand for open standards or protocols, as indicated in the Paris OER Declaration (UNESCO, 2012), which enable the searching and location of specific and relevant OER developed in these regional languages.

Considering the limitation of the current OERScout system with respect to searching resources written in languages other than English, the development of a further extension to OERScout is planned to be designed which will facilitate searching of resources written in Bahasa Melayu, Bahasa Indonesia and Vietnamese. Furthermore, it is my intention to make OERScout available as a public service via www.oerscout.org which would allow academics to search *desirable* OER for their specific teaching and learning needs. I also intend to transfer the system onto a FOSS platform in the spirit of openness and accessibility.

REFERENCES

References

IEEE Learning Technology Standards Committee. (2005). *Final 1484.12.1-2002 LOM Draft Standard*. Retrieved June 7, 2013, from http://ltsc.ieee.org/: http://ltsc.ieee.org/wg12/20020612-Final-LOM-Draft.html

Ebay's Express to take on Amazon. (2006, October 1). *The Sunday Times*.

Abeywardena, I. (2013). Development of OER-Based Undergraduate Technology Course Material: "TCC242/05 Web Database Application" Delivered Using ODL at Wawasan Open University. In G. Dhanarajan, & D. Porter (Eds.), *Open Educational Resources: An Asian Perspective* (pp. 173-184). Vancouver: Commonwealth of Learning and OER Asia.

Anderson, C. (2007). Record relationship navigation: implications for information access and discovery. *In Proceedings of the HCIR Workshop*, (p. 6).

Anido, L. E., Fernández, M. J., Caeiro, M., Santos, J. M., Rodriguez, J. S., & Llamas, M. (2002). Educational metadata and brokerage for learning resources. *Computers & Education, 38*(4), 351-374.

Association of Educational Publishers & Creative Commons. (n.d.). *About the LRMI*. Retrieved May 13, 2013, from Learning Resource Metadata Initiative (LRMI): http://www.lrmi.net/about

Atenas, J., & Havemann, L. (2013). Quality assurance in the open: an evaluation of OER repositories. *INNOQUAL-International Journal for Innovation and Quality in Learning, 1*(2), 22-34.

Baggaley, J. (2013). MOOC rampant. *Distance Education, 34*(3), 368-378.

Balaji, V., Bhatia, M. B., Kumar, R., Neelam, L. K., Panja, S., Prabhakar, T. V., . . . Yadav, V. (2010). Agrotags – A Tagging Scheme for Agricultural Digital Objects. *Metadata and Semantic Research Communications in Computer and Information Science, 108*, 36-45.

Baraniuk, R. G. (2007). Challenges and opportunities for the Open Education movement: A Connexions case study. In T. Iiyoshi, & M. S. Kumar (Eds.), *Opening up education – The collective advancement of education through open technology, open content, and open knowledge*. Cambridge, MA: Massachusetts Institute of Technology Press.

Barritt, C., & Alderman Jr, F. L. (2004). *Creating a reusable learning objects strategy: Leveraging information and learning in a knowledge economy*. California, USA: Pfeiffer.

Barton, J., Currier, S., & Hey, J. (2003). Building quality assurance into metadata creation: an analysis based on the learning objects and e-prints communities of practice. *In 2003 Dublin Core Conference: Supporting Communities of Discourse and Practice - Metadata Research and Applications*. Seattle, Washington.

Bateman, P. (2006). *The AVU Open Educational Resources (OER) Architecture for Higher Education in Africa*. Barcelona: OECD Expert Meeting.

Baumgartner, P., Naust, V., Canals, A., Ferran-Ferrer, N., Minguillón, J., Pascual, M., . . . Schaffert, S. (2007). *Open educational practices and resources: OLCOS Roadmap 2012*. Salzburg, Austria: Open Learning Content Observatory Services.

Ben-Yitzhak, O., Golbandi, N., Har'El, N., Lempel, R., Neumann, A., Ofek-Koifman, S., . . . Yogev, S. (2008). Beyond basic faceted search. *Proceedings of the 2008 International Conference on Web Search and Data Mining* (pp. 33-44). Palo Alto: ACM.

Billsberry, J. (2013). MOOCs: Fad or revolution. *Journal of Management Education, 37*(6), 739-746.

Brasher, A. (2007). A conversion pipeline for audio remixes. *Proceedings of the OpenLearn 2007 conference*. UK.

Brooks, C., & McCalla, G. (2006). Towards flexible learning object metadata. *International Journal of Continuing Engineering Education and Life Long Learning, 16*(1), 50-63.

Buela-Casal, G., & Zych, I. (2010). Analysis of the relationship between the number of citations and the quality evaluated by experts in psychology journals. *Psicothema, 22*(2), 270-276.

Calverley, G., & Shephard, K. (2003). Assisting the uptake of on-line resources: why good learning resources are not enough. *Computers & Education, 41*(3), 205-224.

Cann, A. J. (2007). Embracing Web2.0: online video-beyond entertainment. *Proceedings of the OpenLearn 2007 conference*. UK.

Capra, R. G., & Marchionini, G. (2008). The relation browser tool for faceted exploratory search. *In Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries* (pp. 420-420). ACM.

Casali, A., Deco, C., Romano, A., & Tomé, G. (2013). An Assistant for Loading Learning Object Metadata: An Ontology Based Approach. *Interdisciplinary Journal of E-Learning and Learning Objects (IJELLO), 9*, 11.

Caswell, T., Henson, S., Jenson, M., & Wiley, D. (2008). Open Educational Resources: Enabling universal education. *International Review of Research in Open and Distance Learning, 9*(1), 1-11.

Cechinel, C., Sánchez-Alonso, S., & Sicilia, M. Á. (2009). Empirical analysis of errors on human-generated learning objects metadata. *In Metadata and Semantic Research* (pp. 60-70). Berlin Heidelberg: Springer.

Corcho, O., Fernández-López, M., Gómez-Pérez, A., & López-Cima, A. (2005). Building legal ontologies with METHONTOLOGY and WebODE. Springer Berlin Heidelberg.

Cox, B. (2002, November 1). *It's official, it's called ruby; it's in beta*. Retrieved June 9, 2013, from InternetNews.com: http://www.internetnews.com/ec-news/article.php/1492381/Its+Official+Its+Called+Ruby+Its+in+Beta.htm

Creative Commons. (n.d.). *About The Licenses*. Retrieved June 3, 2013, from creativecommons.org: http://creativecommons.org/licenses/

Crowley, K., Leinhardt, G., & Chang, C. F. (2001). Emerging research communities and the World Wide Web: analysis of a Web-based resource for the field of museum learning. *Computers & Education, 36*(1), 1-14.

D'Antoni, S. (2009). Open Educational Resources: reviewing initiatives and issues. *Open Learning: The Journal of Open, Distance and e-Learning, 24*(1), 3-10.

Daniel, J. (2012). Making sense of MOOCs: Musings in a maze of myth, paradox and possibility. *Journal of Interactive Media in Education, 3*.

Dash, D., Rao, J., Megiddo, N., Ailamaki, A., & Lohman, G. (2008). Dynamic faceted search for discovery-driven analysis. *Proceedings of the 17th ACM conference on Information and knowledge management* (pp. 3-12). Napa Valley: ACM.

DCMI. (n.d.). *DCMI Specifications*. Retrieved June 7, 2013, from http://dublincore.org: http://dublincore.org/specifications/

De la Prieta, F., Gil, A., Rodríguez, S., & Martín, B. (2011). BRENHET2, A MAS to Facilitate the Reutilization of LOs through Federated Search. *In Trends in Practical Applications of Agents and Multiagent Systems* (pp. 177-184). Berlin Heidelberg: Springer.

DeSantis, N. (2012). After leadership crisis fuelled by Distance-Ed Debate, UVa will put free classes online. *Chronicle of Higher Education*.

Devedzic, V. (2004). Education and the Semantic Web. *International Journal of Artificial Intelligence in Education, 14*(1), 39-65.

Devedzic, V., Jovanovic, J., & Gasevic, D. (2007). The pragmatics of current e-learning standards. *Internet Computing, 11*(3), 19-27.

Dholakia, U. M., King, W. J., & Baraniuk, R. (2006). *What Makes an Open Education Program Sustainable? The Case of Connexions*. Retrieved December 27, 2011, from agri-outlook.org: http://www.agri-outlook.org/dataoecd/3/6/36781781.pdf

Dichev, C., & Dicheva, D. (2012). Open Educational Resources in Computer Science Teaching. *Proceedings of the 43rd ACM technical symposium on Computer Science Education* (pp. 619-624). ACM.

Dichev, C., Bhattarai, B., Clonch, C., & Dicheva, D. (2011). Towards Better Discoverability and Use of Open Content. *Proceedings of the Third International Conference on Software, Services and Semantic Technologies S3T* (pp. 195-203). Berlin Heidelberg: Springer.

Donn, K., Plaisant, C., & Shneiderman, B. (1996). Query previews in networked information systems. *Proceedings of the 3rd International Forum on Research and Technology Advances in Digital Libraries* (pp. 120-129). Washington, DC: IEEE Computer Society.

Douce, C. (2007). Creating accesible SCORM content from OpenLearn material. *Proceedings of the OpenLearn 2007 conference*. UK.

Downes, S. (2007). Models for Sustainable Open Educational Resources. *Interdisciplinary Journal of Knowledge and Learning Objects, 3*.

Duval, E., Forte, E., Cardinaels, K., Verhoeven, B., Durm, R., Hendrikx, K., . . . Haenni, F. (2001). The ariadne knowledge pool system. *Communications of the ACM, 44*(5), 72-78.

Farber, R. (2009). Probing OER's huge potential. *Scientific Computing, 26*(1), 29.

Farzan, R., & Brusilovsky, P. (2006). AnnotatEd: a social navigation and annotation service for web-based educational resources. *Proceedings of the E-Learn 2006– World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education.* Honolulu.

Feldman, R., & Sanger, J. (2006). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data.* Cambridge University Press.

Fitzgerald, B. (2006). Open Content Licensing (OCL) for Open Educational Resources. *Proceedings of the OECD Expert Meeting on Open Educational Resources.* Sweden.

Fukuhara, Y. (2008). Current Status of OCW in Japan. *Proceedings of the Distance Learning and the Internet Conference.*

Fulantelli, G., Gentile, M., Taibi, D., & Allegra, M. (2007). The Open Learning Object model for the effective reuse of digital educational resources. *Proceedings of the OpenLearn 2007 conference.* UK.

García, A. M., Alonso, S. S., & Sicilia, M. A. (2008). Una ontología en OWL para la representación semántica de objetos de aprendizaje. *In V Simposio Pluridisciplinar sobre Diseño y Evaluación de Contenidos Educativos Reutilizables (SPDECE).*

Geith, C., & Vignare, K. (2008). Access to Education with Online Learning and Open Educational Resources: Can they close the gap? *Jounral of Asynchonous Learning Networks, 12*(1), 105-126.

Geoffman, W. (1964). On relevance as a measure. *Information Storage and retrival, 2*(3), 201-203.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition, 5*(2), 199-220.

Ha, K. H., Niemann, K., Schwertel, U., Holtkamp, P., Pirkkalainen, H., Boerner, D., . . . Wolpers, M. (2011). A novel approach towards skill-based search and services of Open Educational Resources. *Proceedings of Metadata and Semantic Research* (pp. 312-323). Berlin Heidelberg: Springer.

Hatakka, M. (2009). Build it and they will come?–Inhibiting factors for reuse of open content in developing countries. *The Electronic Journal of Information Systems in Developing Countries, 37*(5), 1-16.

Hearst, M. (2006). Design recommendations for hierarchical faceted search interfaces. *In ACM SIGIR workshop on faceted search*, (pp. 1-5).

Hilton, J., Wiley, D., Stein, J., & Johnson, A. (2010). The four R's of openness and ALMS Analysis: Frameworks for open educational resources. *Open Learning: The Journal of Open and Distance Learning, 25*(1), 37-44.

Hobson, S. D., Horvitz, E., Heckerman, D. E., Breese, J. S., Shaw, G. L., Flynn, J. R., & Jensen, K. (1997). *Patent No. 5,694,559*. Washington, DC: U.S. Patent and Trademark Office.

Hostetter, C. (2006). Faceted searching with Apache Solr. *In ApacheCon US 2006.*

Hylén, J. (2006). Open educational resources: Opportunities and challenges. *Proceedings of Open Education*, (pp. 49-63).

IMS Global Learning Consortium. (2001, September 28). *IMS Learning Resource Meta-Data Information Model [Version 1.2.1 Final Specification]*. Retrieved June 7, 2013, from imsglobal.org: http://www.imsglobal.org/metadata/imsmdv1p2p1/imsmd_infov1p2p1.html

Jones, R. (2007). Giving birth to next generation repositories. *International Journal of Information Management, 27*(3), 154-158.

Joyce, A. (2007). *OECD study of OER: forum report*. UNESCO.

Karam, M., & Zhao, S. (2003). mSpace: interaction design for user-determined, adaptable domain exploration in hypermedia. *In AH2003: Workshop on Adaptive Hypermedia and Adaptive Web Based Systems*. Nottingham.

Knox, J. (2014). Digital culture clash:"massive" education in the E-learning and Digital Cultures MOOC. *Distance Education, 35*(2), 164-177.

Kolowich, S. (2012, November 5). *Pearson's Open Book*. Retrieved May 13, 2013, from INSIDE HIGHER ED: http://www.insidehighered.com/news/2012/11/05/pearson-unveils-oer-search-engine

Koren, J., Zhang, Y., & Liu, X. (2008). Personalized interactive faceted search. *Proceedings of the 17th international conference on World Wide Web* (pp. 477-486). Beijing: ACM.

Kubiszewski, I., Noordewier, T., & Costanza, R. (2011). Perceived credibility of Internet encyclopedias. *Computers & Education, 56*(3), 659-667.

Kumar, M. S. (2009). Open educational resources in India's national development. *Open Learning: The Journal of Open and Distance Learning, 24*(1), 77-84.

Lane, A. (2009). The impact of openness on bridging educational digital divides. *The International Review of Research in Open and Distance Learning, 10*(5).

Larson, R. C., & Murray, M. E. (2008). Open educational resources for blended learning in high schools: Overcoming impediments in developing countries. *Journal of Asynchronous Learning Networks, 12*(1), 85-103.

Leuf, B., & Cunningham, W. (2001). *The Wiki way: Collaboration and sharing on the internet*. Boston: Addison-Wesley Professional.

Levey, L. (2012). Finding Relevant OER in Higher Education: A Personal Account. In J. Glennie, K. Harley, N. Butcher, & T. van Wyk (Eds.), *Open Educational Resources and Change in Higher Education: Reflections from Practice* (pp. 125-138). Vancouver: Commonwealth of Learning.

Lewin, T. (2012). Education site expands slate of universities and courses. *New York Times*.

Lextek. (n.d.). *Onix Text Retrieval Toolkit API Reference*. Retrieved from lextek.com: lextek.com/manuals/onix/stopwords1.html

Little, A., Eisenstadt, M., & Denham, C. (2007). MSG Instant Messenger: social presence and location for the 'ad hoc learning experience'. *Proceedings of the OpenLearn 2007 conference.* UK.

Littlejohn, A., Falconer, I., & Mcgill, L. (2008). Characterising effective eLearning resources. *Computers & Education, 50*(3), 757-771.

McGreal, R. (2010). Open Educational Resource Repositories: An Analysis. *Proceedings of the 3rd Annual Forum on e-Learning Excellence.* Dubai, UAE.

Milojević, S. (2010). Power law distributions in information science: Making the case for logarithmic binning. *Journal of the American Society for Information Science and Technology, 61*(12), 2417-2425.

Moon, B., & Wolfenden, F. (2007). The TESSA OER experience: building sustainable models of production and user implementation. *Proceedings of the OpenLearn 2007 conference.* UK.

Nash, S. (2005). Learning objects, learning object repositories, and learning theory: Preliminary best practices for online courses. *Interdisciplinary Journal of E-Learning and Learning Objects, 1*(1), 217-228.

Ochoa, X., Klerkx, J., Vandeputte, B., & Duval, E. (2011). On the use of learning object metadata: The GLOBE experience. *In Towards Ubiquitous Learning* (pp. 271-284). Berlin Heidelberg: Springer.

OER Africa. (2009). *The Potential of Open Educational Resources: Concept Paper by OER Africa.* Retrieved July 12, 2010, from oerafrica.org: http://www.oerafrica.org/SharedFiles/ResourceFiles/36158/33545/33525/2008. 12.16%20OER%20and%20Licensing%20Paper.doc

Ouyang, Y., & Zhu, M. (2008). eLORM: learning object relationship mining-based repository. *Online Information Review, 32*(2), 254-265.

Pawlowski, J. M., & Bick, M. (2012). Open Educational Resources. *Business & Information Systems Engineering*, 1-4.

Petrides, L., Nguyen, L., Jimes, C., & Karaglani, A. (2008). Open educational resources: inquiring into author use and reuse. *International Journal of Technology Enhanced Learning, 1*(1), 98-117.

Piedra, N., Chicaiza, J., López, J., Martínez, O., & Caro, E. T. (2010). An approach for description of Open Educational Resources based on semantic technologies. *In Education Engineering (EDUCON)* (pp. 1111-1119). IEEE.

Piedra, N., Chicaiza, J., López, J., Tovar, E., & Martinez, O. (2011). Finding OERs with Social-Semantic Search. *Proceedings of the 2011 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1195-1200). Amman, Jordan: IEEE.

Piedra, N., Chicaiza, J., López, J., Tovar, E., & Martinez, O. (2011). Finding OERs with Social-Semantic Search. *Proceedings of the Global Engineering Education Conference (EDUCON)* (pp. 1195-1200). IEEE.

Piedra, N., Chicaiza, J., Tovar, E., & Martinez, O. (2009). Open Educational Practices and Resources Based on Social Software: UTPL experience. *Proceedings of the 9th IEEE International Conference on Advanced Learning Technologies.*

Pirkkalainen, H., & Pawlowski, J. (2010). Open Educational Resources and Social Software in Global E-Learning Settings. In P. Yliluoma (Ed.), *Sosiaalinen Verkko-oppiminen* (pp. 23-40). Naantali: IMDL.

Pollitt, A. S., Smith, M. P., Treglown, M., & Braekevelt, P. (1996). View-Based Searching Systems--Progress Towards Effective Disintermediation. *Online Information 96 Proceeings*, (pp. 433-441).

Rawsthorne, P. (2007). *Utilizing Open Educational Resources for International Curriculam Development.* Retrieved July 12, 2010, from wikieducator.org: http://directory.wikieducator.org/images/0/01/PeterRawsthorne.OERProgram.pdf

Richards, G. (2007). Reward structure for participation and contribution in K-12 OER Communities. *Proceedings of the 1st Workshop on Social Information Retrieval for Technology-Enhanced Learning and Exchange.*

Sampson, D. (2009). Competence-related metadata for educational resources that support lifelong competence development programmes. *Educational Technology & Society, 12*(4), 149-159.

San Diego, J. P. (2007). Learning from 'OpenLearner-interactions' using digital research techniques. *Proceedings of the OpenLearn 2007 conference.* UK.

Saracevic, T. (1975). Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science, 26*(6), 321-343.

Shelton, B. E., Duffin, J., Wang, Y., & Ball, J. (2010). Linking OpenCourseWares and Open Education Resources: Creating an Effective Search and Recommendation System. *Procedia Computer Science, 1*(2), 2865-2870.

Shokouhi, M., & Si, L. (2011). Federated search. *Foundations and Trends in Information Retrieval, 5*(1), 1-102.

Shum, S. B., & Okada, A. (2007). Knowladge mapping for open sensemaking communities. *Proceedings of the OpenLearn 2007 conference.* UK.

Stacey, P. (2006). *Open For Innovation Strategically using "open" concepts and methods for sustainable development and use of online learning resources in higher education.* Retrieved July 12, 2010, from ares.licef.teluq.uqam.ca: http://ares.licef.teluq.uqam.ca/Portals/10/I2LOR06/11_Open%20for%20Innovation.pdf

Stacy, P. (2007). Open educational resources in a global context. *First Monday, 12*(4). Retrieved 09 30, 2012, from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1769

Tello, J. (2007). Estudio exploratorio de defectos en registros de meta-datos IEEE LOM de objetos de aprendizaje. *In Post-Proceedings of SPDECE 2007 - IV Simposio Pluridisciplinar sobre Diseno, Evaluacion y Desarrollo de Contenidos Educativos Reutilizables*, (pp. 19-21). Bilbao, Spain.

Toikkanen, T. (2008). Simplicity and design as key success factors of the OER repository LeMill. *eLearning Papers*(10), 4.

Tomadaki, E., & Scott, P. J. (2007). Videoconferencing in open learning. *Proceedings of the OpenLearn 2007 conference*.

Tunkelang, D. (2009). Faceted search. In G. Marchionini (Ed.), *SYNTHESIS LECTURES ON INFORMATION CONCEPTS, RETRIEVAL, AND SERVICES* (Vol. 5, pp. 1-80). Morgan & Claypool.

UNESCO. (2012, June 22). *2012 PARIS OER DECLARATION*. Retrieved June 13, 2013, from unesco.org: http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/WPFD2009/English_Declaration.html

Unwin, T. (2005). Towards a framework for the use of ICT in teacher training in Africa. *Open Learning: The Journal of Open, Distance and e-Learning, 20*(2), 113-129.

Vaughan, L. (2004). New measurements for search engine evaluation proposed and tested. *Information Processing and Management, 40*, 677-691.

West, P., & Victor, L. (2011). *Background and action paper on OER*. Report prepared for The William and Flora Hewlett Foundation.

Wiley, D. (2006). *On the Sustainability of Open Educational Resource Initiatives in Higher Education*. Retrieved June 3, 2013, from oecd.org: http://www1.oecd.org/edu/ceri/38645447.pdf

Yamada, T. (2013). Open Educational Resources in Japan. In G. Dhanarajan, & D. Porter (Eds.), *Open Educational Resources: An Asian Perspective* (pp. 85-105). Vancouver: Commonwealth of Learning and OER Asia.

Yergler, N. R. (2010). Search and Discovery: OER's Open Loop. *Proceedings of Open Ed 2010*. Barcelona.

APPENDICES

Appendix A

Abeywardena, I.S., Chan, C.S., & Tham, C.Y. (2013). OERScout Technology Framework: A Novel Approach to Open Educational Resources Search. *International Review of Research in Open and Distance Learning, 14(4),* 214-237.

OERScout Technology Framework: A Novel Approach to Open Educational Resources Search

Ishan Sudeera Abeywardena[1], Chee Seng Chan[2,] and Choy Yoong Tham[1]
[1]Wawasan Open University, Malaysia, [2]University of Malaya, Malaysia

Abstract

The open educational resources (OER) movement has gained momentum in the past few years. With this new drive towards making knowledge open and accessible, a large number of OER repositories have been established and made available online throughout the world. However, the inability of existing search engines such as Google, Yahoo!, and Bing to effectively search for useful OER which are of acceptable academic standard for teaching purposes is a major factor contributing to the slow uptake of the entire movement. As a major step towards solving this issue, this paper proposes *OERScout*, a technology framework based on text mining solutions. The objectives of our work are to (i) develop a technology framework which will parametrically measure the usefulness of an OER for a particular academic purpose based on the openness, accessibility, and relevance attributes; and (ii) provide academics with a mechanism to locate OER which are of an acceptable academic standard. From our user tests, we have identified that OERScout is a sound solution for effectively zeroing in on OER which can be readily used for teaching and learning purposes.

Keywords: OERScout; open educational resources; OER; OER search; desirability of OER; OER metadata

Introduction

Open educational resources (OER) have the potential to become a major source of freely reusable teaching and learning resources, especially in higher education (HE). The UNESCO Paris OER Declaration (2012) defines OER as

> teaching, learning and research materials in any medium, digital or otherwise, that reside in the public domain or have been released under an open license that permits no-cost access, use, adaptation and redistribution by others with no or limited restrictions. Open licensing is built within the existing framework of intellectual property rights as defined by relevant international conventions and respects the authorship of the work.

Claims have also been made by Caswell, Henson, Jenson, and Wiley (2008) that the move towards OER can significantly reduce the costs of learning. Thus, OER has the potential to broaden access and provide equity in education. This is especially important for countries in the Global South.

The recently concluded "OER Asia" study (Dhanarajan & Abeywardena, 2013) surveyed 420 junior to senior academics from public and private HE institutions in nine countries representing a majority of sub-regions in Asia. Based on this study, Abeywardena, Dhanarajan, and Chan (2012) state that 57.4% of the academics feel the lack of ability to locate specific and relevant resources using existing search engines to be a serious inhibitor of the use of OER. It is further pointed out that, in general, academics search and locate OER which are freely available on the Internet. However, many of these resources have not been subjected to academic quality assurance (QA) procedures imposed by degree accrediting organisations such as the Malaysian Qualifications Agency (MQA)[1]. In contrast, institutional and peer-reviewed OER repositories maintain an acceptable level of academic quality of material. These materials can be readily used and reused for teaching purposes. Furthermore, these repositories are equipped with native search mechanisms which facilitate the searching of relevant OER for a particular teaching need. Unfortunately, according to the study, only 43.2% of the academics use native search facilities of OER repositories. On the other hand, generic search engines such as Google, Yahoo!, and Bing are found to be used by 96.9% of the academics for OER search.

From this comparison, it is apparent that many academics depend on generic search mechanisms to locate the required OER for their teaching purposes. As a result, the inability of these generic mechanisms to locate useful OER for a particular teaching need, as will be discussed, has in fact become an inhibitor to the wider adoption of OER for teaching in Asia. In order to overcome this barrier, a centralised search mechanism

---

[1] http://www.mqa.gov.my

which can locate academically useful OER needs to be introduced. As a major step towards solving this issue, in this paper, we propose *OERScout*, a technology framework based on text mining solutions. The objectives of our work are to (i) develop a technology framework which will parametrically measure the usefulness of an OER for a particular academic purpose based on the openness, accessibility, and relevance attributes; and (ii) provide a search mechanism to effectively zero in on OER which are of an acceptable academic standard.

The rest of the paper is structured as follows: The Literature Review section gives an overview of the current solutions available to search for OER; the Methodology section details the proposed method; the Results and Discussion sections provide the expert user test results and discussion respectively; and the Conclusion concludes the work and discusses some future work. Overall, the paper provides a holistic view of the complete project.

Literature Review

Most current OER initiatives are based on established web technology platforms and have accumulated large volumes of quality resources. However, one limitation inhibiting the wider adoption of OER is the current inability to effectively search for academically useful OER from a diversity of sources (Yergler, 2010). This limitation of locating "fit-for-purpose" (Calverley & Shephard, 2003) resources is further heightened by the disconnectedness of the vast array of OER repositories currently available online. As a result, West and Victor (2011) argue that there is no single search engine which is able to locate resources from all the OER repositories. Furthermore, according to Dichev and Dicheva (2012), one of the major barriers to the use and reuse of OER is the difficulty in finding quality OER matching a specific context as it takes an amount of time comparable with creating one's own materials. Unwin (2005) argues that the problem with open content is not the lack of available resources on the Internet but the inability to effectively locate suitable resources for academic use. The UNESCO Paris OER Declaration (2012) states the need for more research in this area to "encourage the development of user-friendly tools to locate and retrieve OER that are specific and relevant to particular needs". Thus, the necessity for a system which could effectively search the numerous OER repositories with the aim of locating usable materials has taken centre stage.

The most common method of searching for OER is to use generic search engines such as Google, Yahoo!, or Bing. Even though this method is the most commonly used, it is not the most effective as discussed by Pirkkalainen and Pawlowski (2010, p. 2) who argue that "searching this way might be a long and painful process as most of the results are not usable for educational purposes".

Alternative methods for OER search can be broadly categorised into federated search and semantic search. Federated search is achieved either by searching across different repositories at runtime or by periodically harvesting metadata for offline searching.

Recent examples of federated search include (i) BRENHET² proposed by De la Prieta, Gil, Rodríguez, and Martín (2011), which is a multi agent system (MAS) which facilitates federated search between learning object repositories (LOR); (ii) OpeScout (Ha, et al., 2011), which copies metadata from existing repositories to create an index of resources accessible through a faceted search approach; (iii) Global Learning Object Brokered Exchange (GLOBE), which acts as a central repository of IEEE LOM educational metadata harvested from various member institutional repositories (Yamada, 2013); and (iv) Pearson's Project Blue Sky (Kolowich, 2012), which is a custom search engine specifically concentrating on searching for OER with an academic focus. Semantic search is derived from semantic web technologies where people are considered as producers or consumers and machines as enablers. Some of the recent semantic search initiatives are (i) the OER-CC ontology which describes various accessibility levels (Piedra, et al., 2010, 2011); (ii) the "Assistant" prototype proposed by Casali, Deco, Romano, and Tomé (2013), which helps users with respect to loading metadata through automation; (iii) the "Folksemantic" project which is a hybrid search system combining OCW Finder and OER Recommender (Shelton, Duffin, Wang, & Ball, 2010); and (iv) "Agrotags", a project concentrating on tagging resources in the agriculture domain (Balaji, et al., 2010). However, despite showing initial promise, only a handful of these solutions have proceeded beyond the prototype stage. Out of these, the ones which have become global players are mainly commercial ventures or global federations backed by philanthropic funding. One reason underpinning the relatively low success rate of these initiatives can be attributed to the current lack of a search methodology which takes into consideration the level of openness, the level of access, and the relevance of a resource for one's needs (Abeywardena, Raviraja, & Tham, 2012). Though one might argue that popular search engines provide advanced facilities to define various filter criteria which would refine the searches, these search engines however are not tailored to effectively locate OER material which are the most useful for a particular academic purpose. As such, OER consumers will need to resort to frequenting OER repositories to search for the resources they are after. Pirkkalainen and Pawlowski (2010) argue that native search mechanisms of repositories are relatively better at locating resources with increased usefulness. However, the problem is which repositories to choose within the large and constantly expanding global pool. Furthermore, users would be spending an extended amount of time on these repositories conducting multiple searches (Figure 1), making it an inefficient method for locating resources.
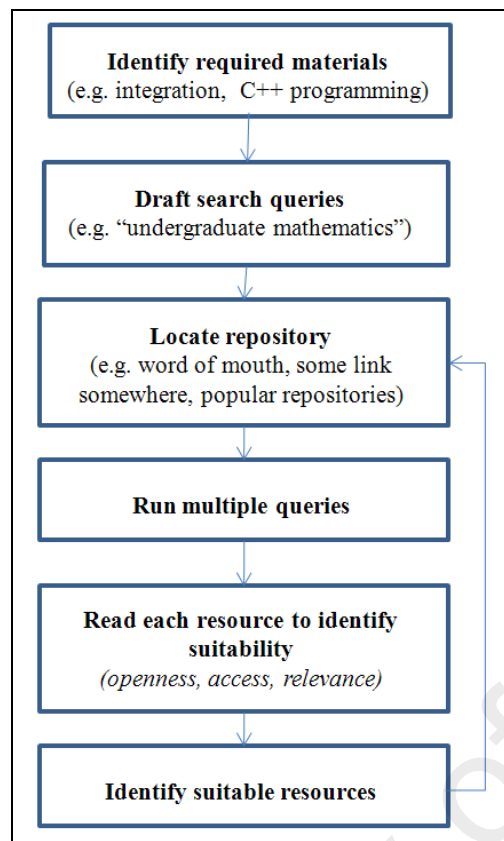
Figure 1. The flow of activities in searching for suitable OER on heterogenous repositories based on personal experience (Abeywardena, 2013). These activities will need to be repeated on multiple repositories until the required resources are located.

Another factor inhibiting effective OER search is the heterogeneity of OER repositories. Within the context of parametric web based search, this disparity can be broadly attributed to (i) the lack of a single metadata standard; (ii) the lack of a centralised search mechanism; and (iii) the inability to indicate the usefulness of an OER returned as a search result.

Metadata provides a standard and efficient way to conveniently characterize educational resource properties (Anido, Fernández, Caeiro, Santos, Rodriguez, & Llamas, 2002). The majority of existing search methodologies, including mainstream search engines, such as Google, work on the concept of metadata for locating educational resources. However, it can be argued that the annotation of resources with metadata cannot be made 100% accurate or uniform if done by the creator(s) of the resource (Barton, Currier, & Hey, 2003; Tello, 2007; Devedzic, Jovanovic, & Gasevic, 2007; Brooks & McCalla, 2006; Cechinel, Sánchez-Alonso, & Sicilia, 2009). Therefore the use of human annotated metadata in performing objective searches becomes subjective and inaccurate. A possible way to overcome this inaccuracy and to ensure uniformity of metadata is to utilise a computer based methodology which would consider the content,

domain, and locality of the resources, among others, for autonomously annotating metadata.

As a solution to these issues , this paper proposes the OERScout technology framework which accurately clusters text based OER by building a *keyword-document matrix* (KDM) using autonomously mined domain specific keywords. The advantage of our work is, using the KDM, the system generates ranked lists of relevant OER from heterogenous repositories to suit a given search query. The contribution of our work is two-fold: Firstly, we propose a technology framework for locating OER, which are useful for academic needs. In this regard, the advantage of OERScout over existing OER search methodologies is the incorporation of the *desirability* framework (Abeywardena, Raviraja, & Tham, 2012) in parametrically measuring the usefulness of an OER with respect to openness, accessibility, and relevance. Secondly, we introduce a novel methodology which allows academics to effectively zero in on OER which can be readily used for their teaching and learning purposes . We strongly believe that the OERScout system will broaden access and provide equity in education, particularly for countries in the Global South such as India, Pakistan, Afghanistan, Myanmar, and Sri Lanka to name a few.

Methodology

As discussed in the Literature Review, mainstream search engines, federated search, and semantic search are the key OER search methodologies adopted at present. However, all of these methodologies depend on human annotated metadata for approximating the usefulness of a resource for a particular need. Given the limitations of human annotated metadata with respect to accurately and uniformly describing resources, the accuracy of search becomes a function of the content creators' ability to accurately annotate resources. Therefore, the OERScout system uses text mining techniques to annotate resources using autonomously mined keywords.

The Algorithm

The OERScout text mining algorithm is designed to "read" text based OER documents and "learn" which academic domain(s) and sub-domain(s) they belong to. To achieve this, a *bag-of-words* approach is used due to its effectiveness with unstructured data (Feldman & Sanger, 2006). The algorithm extracts all the individual words from a particular document by removing noise such as formatting and punctuation to form the *corpus*. The corpus is then *tokenised* into the *list of terms* using the *stop words* found in the Onix Text Retrieval Toolkit[2] as shown in Figure 2.

---

[2] lextek.com/manuals/onix/stopwords1.html
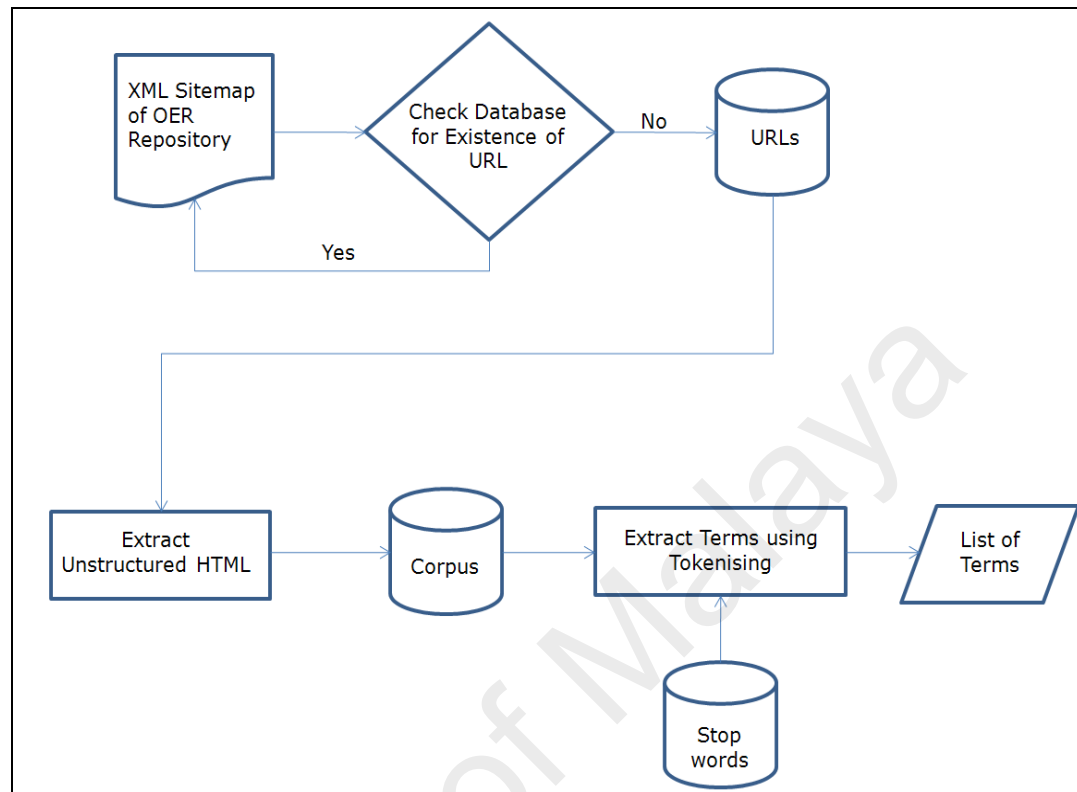
Figure 2. The list of terms is created by tokenising the corpus using the stop words found in the Onix Text Retrieval Toolkit.

The extraction of the content describing terms from the list of terms for the formation of the term document matrix (TDM) is done using the term frequency–inverse document frequency (TF-IDF) weighting scheme. The weight of each term (TF-IDF) is calculated using Equation 1 (Feldman & Sanger, 2006):

$$(TF\text{-}IDF)_t = TF_t \; x \; IDF_t \quad (1)$$

TF_t denotes the frequency of a term t in a single document. IDF_t denotes the frequency of a term t in all the documents in the collection [$IDF_t = Log\,(N/TF_t)$] where N is the number of documents in the collection. The probability of a term t being able to accurately describe the content of a particular document as a keyword decreases with the number of times it occurs in other related and non-related documents. For example the term "introduction" would be found in many OER documents which discuss a variety of subject matter. As such the TF-IDF of the term "introduction" would be low compared to terms such as "operating systems" or "statistical methods" which are more likely to be keywords. Due to the large number of documents available in OER repositories and their document lengths, the TF value of certain words will be quite high. As a result, there will be a considerable amount of noise being picked up while identifying the keywords. However, the large number of documents will also increase the IDF value of words reducing the TF-IDF value which results in the reduction of noise picked up as keywords. As such, the TF-IDF weighting scheme allows the system

to refine its set of identified keywords at each iteration. Therefore, the TF-IDF weighting scheme is found to be suitable for extracting keywords from the OER documents.

Keyword-Document Matrix (KDM)

The keyword-document matrix (KDM), a subset of the TDM, is created for the OERScout system by matching the autonomously identified keywords against the documents as shown in Figure 3.

	Keyword$_1$	Keyword$_2$	....	Keyword$_n$
Document$_1$	√			√
Document$_2$	√		√	
....		√		
Document$_n$		√		√

Figure 3. The keyword-document matrix (KDM), a subset of the TDM, is created for the OERScout system by matching the autonomously identified keywords against the documents.

The formation of the KDM (Figure 4) is done by (i) normalising the TF-IDF values for the terms in the TDM; and (ii) applying the Pareto principle (80:20) empirically for feature selection where the top 20% of the TF-IDF values are considered to be keywords describing 80% of the document.
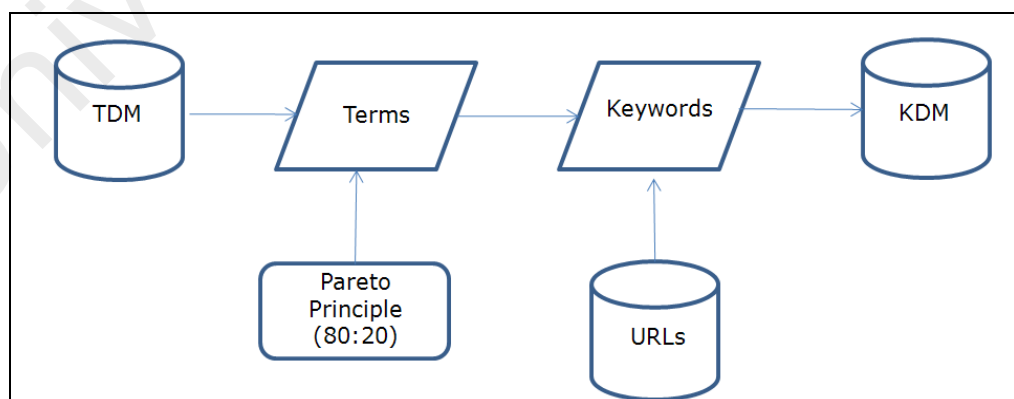
Figure 4. Formation of the KDM by normalizing the TF-IDF values of the terms in the TDM and applying the Pareto principle empirically for feature selection.

The OERScout user interface and algorithm are implemented using the Microsoft Visual Basic.NET 2010 (VB.NET, 2010) programming language. The corpus, list of terms, TDM, and KDM are implemented using the MySQL database platform. The OER resources are fed into the system using sitemaps based on extensible markup language (XML) which contain the uniform resource locators (URLs) of the resources. When implemented, new repositories will be identified for crawling based on referrals by end users. The sitemaps created by the crawlers will be input into the system to be processed. The server tools will continuously run at the server processing new documents and re-visiting processed documents to ensure accuracy. The KDM is accessed by end users through the OERScout Microsoft Windows based client interface. The deployment architecture of OERScout is shown in Figure 5.
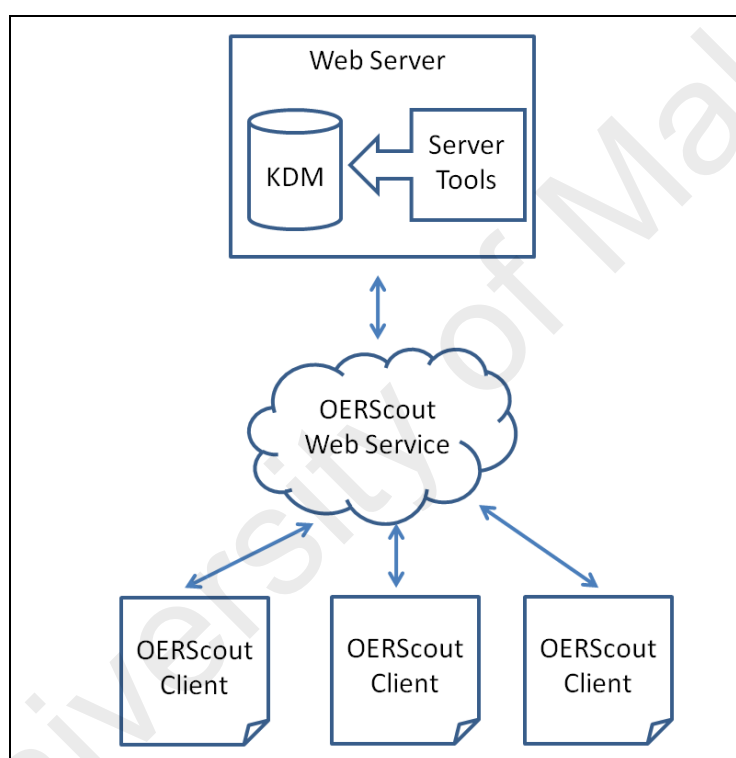
Figure 5. OERScout deployment architecture which has a web server hosting the KDM, a web service for accessing the KDM, and a Microsoft Windows based client interface.

Calculation of the Desirability

The desirability of OER (Abeywardena, Raviraja, & Tham, 2012) is a parametric measure of the usefulness of an OER for a particular academic need based on (i) level of openness, the permission to use and reuse the resource; (ii) level of access, the technical keys required to unlock the resource; and (iii) relevance, the level of match between the resource and the needs of the user. The desirability is calculated using Equation 2 and is denoted as the *D-index*, which is a value between 0 and 1. The higher the D-index, the more desirable an OER is for a particular academic need. The value 256 is used to normalise the access, openness, and relevance parameters. It is the product of the values

16, 4, and 4, respectively, which correspond to the highest value assigned to each parameter.

$$\text{D-index} = (\text{level of access x level of openness x relevance}) / 256 \quad (2)$$

The desirability of each document in the KDM is calculated using the openness, accessibility, and relevance of the document. As suggested by Abeywardena, Raviraja, and Tham (2012), the openness of the document is calculated using the Creative Commons (CC) license of the document (Table 1). A maximum value of 4 is assigned to the most open CC license with respect to permission to reuse, redistribute, revise, and remix (Hilton, Wiley, Stein, & Johnson, 2010). A value of 2 is assigned to the least open license as the CC license starts at the redistribute level.

The accessibility is calculated by extracting the file type of each document as shown in Table 2. This version of OERScout is built only to index documents of type PDF (.pdf), webpage (static and dynamic web pages which include .htm, .html, .jsp, . asp, .aspx, .php etc.), TEXT (.txt), and MS Word (.doc, .docx) as these file types were found to be the most commonly used for text based OER (Wiley, 2006). The value for each file type was calculated based on the ALMS analysis proposed by Hilton, Wiley, Stein, and Johnson (2010) which builds on the parameters (i) **A**ccess to editing tools; (ii) **L**evel of expertise required to revise or remix; (iii) ability to **M**eaningfully edit; and (iv) **S**ource-file access.

The relevance of a document to a particular search query is calculated using the TF-IDF values of the keywords which are stored as additional parameters of the KDM. As shown in Table 3, building on the work by Vaughan (2004), the top 10 search results based on the TF-IDF value are assigned a maximum value of 4, the top 11-20 search results are assigned a value of 3, the top 21-30 results are assigned a value of 2, and search results below 30 are assigned a minimum value of 1.

The D-index of each document is then calculated using Equation 2 and the desirable resources for a particular need are presented to the user in descending order.

Table 1

Openness Based on the CC License (Abeywardena, Raviraja, & Tham, 2012).

Permission	Creative Commons (CC) licence	Value
Reuse	None	1
Redistribute	Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) Attribution-NoDerivs (CC BY-ND)	2
Revise	Attribution-NonCommercial-ShareAlike (CC BY-NC-SA) Attribution-ShareAlike (CC BY-SA)	3
Remix	Attribution-NonCommercial (CC BY-NC) Attribution (CC BY)	4

Table 2

Accessibility Based on the File Type (Abeywardena, Raviraja, & Tham, 2012)

File type	Access (ALMS)				
	A	L	M	S	Value
PDF	Low	High	No	No	1
MS Word	Low	Low	Yes	Yes	8
Webpage	High	Low	Yes	Yes	16
TEXT	High	Low	Yes	Yes	16

Table 3

The Level of Relevance Based on Search Rank

Search rank	Value
Below the top 30 ranks of the search results	1
Within the top 21-30 ranks of the search results	2
Within the top 11-20 ranks of the search results	3
Within the top 10 ranks of the search results	4

One of the key observations made during the calculation of the desirability is that certain OER repositories do not specify or use the CC licensing scheme as the standard for defining the intellectual property rights. However, these repositories explicitly or implicitly mention that the resources are freely and openly available for use and reuse. Due to the inability of the current OERScout system to determine the level of openness of these resources, a value of zero was assigned to any resource which did not implement the CC licensing scheme. As such the desirability of these resources was reduced to zero due to the ambiguity in the license definition. This feature spares the user from legal complications attached to the use and reuse of resources which do not clearly indicate the permissions granted.

Results

The application of the system in a real world scenario was done using the *Directory of Open Educational Resources* (DOER)[3] of the Commonwealth of Learning (COL). DOER is a fledgling *portal OER repository* (McGreal, 2010) which provides an easily navigable central catalogue of OER distributed globally. At present, the OER available through DOER are manually classified into 20 main categories and 1,158 sub-categories. However, despite covering most of the major subject categories, this particular ontology would need to expand by a large degree due to the variety of OER available in an array of subject areas. This expansion, in turn, becomes a tedious and laborious task which needs to be accomplished manually on an ongoing basis. As a possible solution to this issue, a mechanism was needed for autonomously identifying the subject area(s) covered in a particular OER, in the form of keywords, in order for it to be accurately catalogued.

Given this requirement, DOER was used as the training dataset for OERScout. In addition to the resources categorised in DOER, 1,536 resources from the Rice University's Connexions[4] repository were also included in the training dataset due to (i) the large number of OER materials available; and (ii) the relatively high popularity and usage rates. An XML sitemap containing a total of 1,999 URLs belonging to the domains of arts, business, humanities, mathematics, and statistics, science and technology, and social sciences was created as the initial input. The system was run with the initial input and was allowed to autonomously create the KDM. This training process was critical to the functioning of the algorithm as it had to learn a large number of academic domains and sub-domains before being able to accurately cluster resources according to the domain.

On average, each document required 15-90 minutes to be downloaded, read, and learnt by the system depending on the size and file type. The system took approximately five days to process all the documents in the training dataset. Although the training process required a considerable amount of time due to the lack of optimisation and enterprise scale infrastructure, this process takes place as a background operation at the server. Therefore, once the KDM is created, the end user does not experience any delays during the search process.

After completion of the run, the system had processed documents of various size, file types, and licenses from 11 repositories representing many regions of the world (Table 4). It was noted that there was a certain amount of noise in the keywords identified due to the limited number of resources indexed in a given domain. However, with more documents being indexed, the expansion of the list of terms will result in larger IDF values which will decrease the TF-IDF value for noise words. This will result in the algorithm rejecting these noise words as keywords, that is, the reduction of noise.

---

[3] http://doer.col.org/
[4] http://www.cnx.org

Table 4

Resources Indexed in the KDM Based on the Initial Input

	Repository	Host institution	Region	License	File type	No. resources indexed
1	Connexions	Rice University	USA	CC BY	Webpage	1536
2	OCW Athebasca	Athebasca University	Canada	CC BY	Webpage	07
3	OCW Capilano	Capilano University		CC BY-NC-SA	Webpage	19
4	OCW USQ	University of Southern Queensland	Australia	CC BY-NC-SA	Webpage	10
5	UCT Open Content	University of Cape Town	South Africa	CC BY-NC-SA	Webpage	63
6	OpenLearn	The Open University	UK	CC BY-NC-SA	Webpage	242
7	WikiEducator	COL & Ottago Polytechnic	New Zealand	CC BY-SA	Webpage	38
8	Unow	University of Nottingham	UK	CC BY-NC-SA	Webpage	27
9	TESSA	Multiple African Universities	Africa	CC BY-SA	PDF	15
10	OER AVU	African Virtual University	Africa	CC BY-SA	DOC DOCX PDF	40
11	WOU OER	Wawasan Open University	Malaysia	Various	PDF	02
	Total					1999

In order to test the functionality of the system from a real-world user's perspective, 27 academics who have at least 3-5 years of experience in OER advocacy, creation, use, and reuse were invited to test the system. Out of the 27 experts invited, 19, including six professors, five associate professors, three PhD holders, and four mid career academics, agreed to test the system and provide feedback. This group of users represented Australia, Brazil, Cambodia, Canada, China, Hong Kong SAR, Indonesia, Malaysia, Pakistan, and Vietnam. They comprised of varied backgrounds such as engineering, computer science, electronics, instructional design, distance education, agriculture,

biology, law, and library science. The KDM was made available to this group through the OERScout client interface shown in Figure 6.
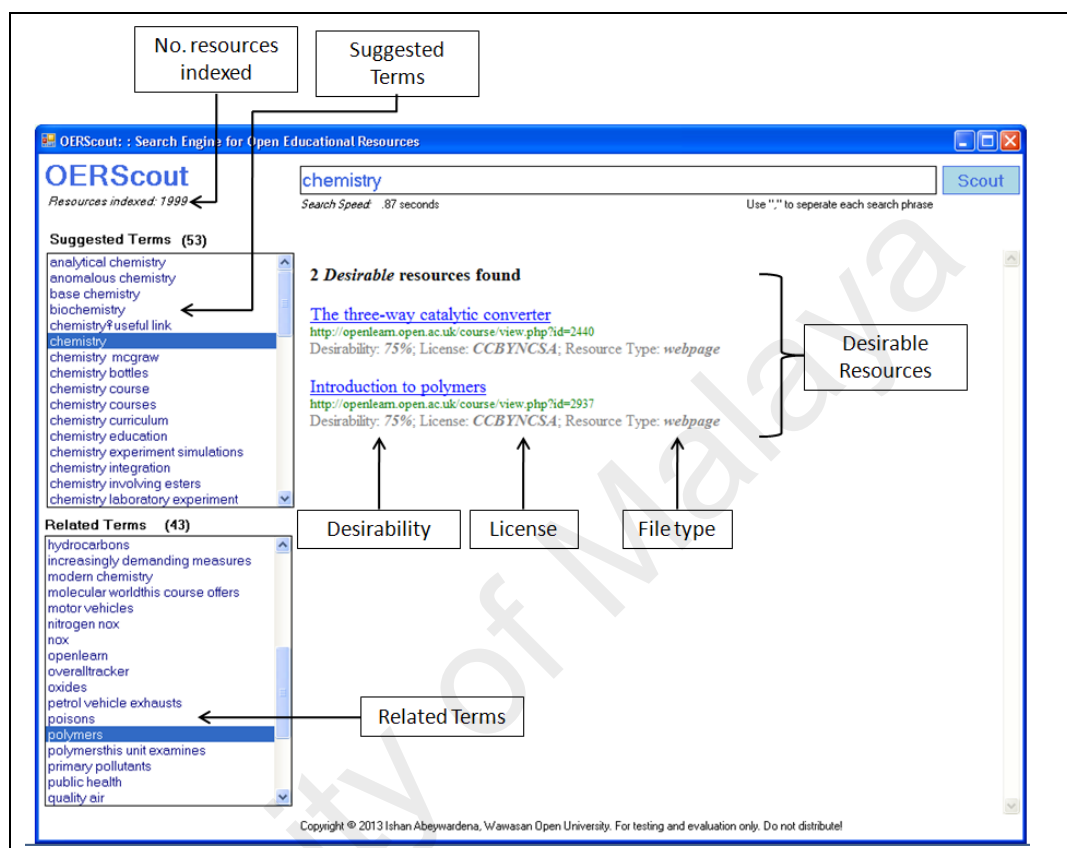
Figure 6. OERScout client interface used for testing the system. The figure shows a search result for resources on "chemistry: polymers".

A comprehensive user manual was provided to the users which outlined how OERScout searched for the most desirable resources. The testing was conducted for a duration of seven days. The users tested the system by searching for OER for their day-to-day academic needs. At the end of the test period, the users provided qualitative feedback through a web based feedback form on various aspects of the OERScout framework. The general feedback which holistically critiques the OERScout technology framework is consolidated in Table 5.

Table 5

Consolidated Feedback Gathered from the OERScout Test Users

	Criteria	Advantages of OERScout	Weaknesses of the prototype
1.	User interface	The user interface is quite simple, friendly, intuitive, un-cluttered and easy to operate. It avoids the hassle of shifting between search modes.	Add advanced search features such as year, language, author and type of resources are not available.
2.	"Faceted search" approach which allows users to dynamically generate search results based on suggested and related terms	The ability to drill down using "faceted search" is very useful. It helps to locate resources faster.	As the number of resources grows the list of suggested and related terms will be quite long. Some noise terms are generated along with the keywords.
3.	Ease of use	It is a powerful tool which allows users to easily locate relevant resources.	The number of resources indexed is quite small.
4.	Relevance of the suggested terms generated according to the search query	The suggested terms are quite relevant and covers the scope of the search adequately.	Some unfamiliar noise words were generated as suggested terms.
5.	Use of related terms to effectively zero in on the resources being searched for	The feature is very useful and performs well. The functionality is similar to a thesaurus used by librarians for cataloging.	Many different terms point to the same resource due to the small dataset. Some terms are not related to the domain. Too many terms are generated.
6.	Usefulness of the resources returned with respect to Openness (the ability to use, reuse, revise and remix)	The use of the CC license to locate the most open resources is a useful feature. The value of this feature will increase along with the increase of quantity and quality of OER available.	The licensing scheme needs to be indicated in a more user-friendly manner.
7.	Usefulness of the resources returned with respect to Access (the ease of reuse and remix of resource type)	The resources returned met the criteria of access with respect to use and reuse. Based on the resource type, users can immediately identify how they can use the resource.	This might not be important as the licensing type defines the reuse and remix capabilities.
8.	Usefulness of the resources returned with respect to Relevance (the match between the results and your query)	Currently quite accurate and very useful.	The small size of the dataset limits the relevance.
9.	Effectiveness with respect to identifying the academic domain(s) of a resource	The autonomous identification of academic domains increases the focus of the search and the quality of the resources returned.	The technology shows promise but the number of domains identified are limited due to the size of the dataset.
10.	Use of the desirability for filtering the most useful resources for ones needs	The desirability framework is an interesting idea which will help in identifying resources appropriate for specific needs.	The concept of desirability needs to be explained to the user through the interface.
11.	Effectiveness with respect to locating desirable resources in comparison to mainstream search engines or native search	A comparison between the OERScout and conventional search engines cannot be made as they serve different purposes. OERScout is	Search engines such as Google have large databases of indexed resources. In this sense they cannot be compared to OERScout.

	engines of OER repositories[a]	much more focused and addresses some key issues in OER search.	
12.	Innovativeness of the technology framework	The technology framework is quite innovative and can bridge the gap between different metadata standards. The simplicity of the user interface complements the scale of innovation.	The scope of the framework needs to be refined. The system needs to be made available as an online service.
13.	How the wider OER community will be benefited	The technology will benefit the wider OER community as a tool for thought provoking discussion on adopting and adapting resources. It will be very beneficial for the novice user with respect to ease of use and affordability.	At the moment it is only a prototype. More resources need to be indexed before it can benefit the community.

Discussion

Empirical Evidence

Figure 6 shows a search conducted for the term "chemistry" on OERScout based on the KDM. In contrast to the static list of search results produced by generic search engines, OERScout employs a "faceted search" (Tunkelang, 2009) approach by providing a dynamic list of *suggested terms* which are related to "chemistry". The user is then able to click on any of the suggested terms to access the most desirable OER from all the repositories indexed by OERScout. Furthermore, based on the selection by the user, the system will provide a list of *related terms* which will enable the user to drill down further to zero in on the most suitable OER for his/her teaching needs. In this particular example (Figure 6), the user has selected "polymers" as the related term to locate two desirable resources from the OpenLearn repository of The Open University which is known to host OER of high academic standard. Furthermore, Figure 7 shows the search results returned by OERScout for the search query on "calculus". The desirable resources returned are from the open course ware OCW Capilano of Canada, OpenLearn of UK, and OER AVU of Africa. As such, it can be seen that OERScout is a more focused and dynamic system for effectively searching for desirable OER. This becomes one of the major benefits to ODL practitioners as the system spares the user from conducting repeated keyword searches in OER repositories to identify suitable material for use. It also allows users to quickly zero in on OER suitable for their needs without reading through all the search results returned by a generic search mechanism such as Google. Table 6 summarises some of the key features of OERScout in contrast to the generic search engines Google, Yahoo!, and Bing.
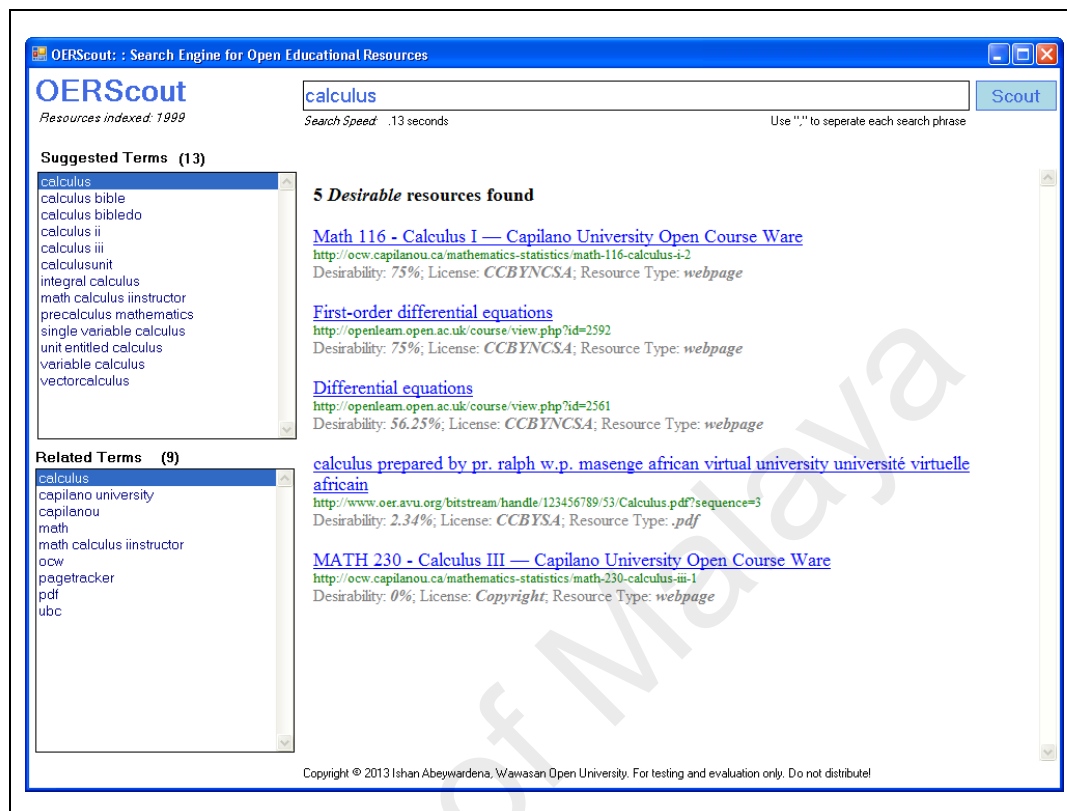
Figure 7. Search results generated by OERScout for the term "calculus". The desirable resources returned are from Capilano University, The Open University, and African Virtual University.

Table 6

Key Features of OERScout in Contrast to Google, Yahoo!, and Bing

	Key Feature	OERScout	Google	Yahoo!	Bing
1.	Provides a centralised mechanism to search for OER	Yes	Yes	No	No
2.	Searches for only the most *desirable* resources for academic purposes	Yes	No	No	No
3.	Effectively locates and presents resources from the distributed repositories	Yes	No	No	No
4.	Provides a dynamic mechanism instead of a static list of search results which can be used to zero in on the required resources	Yes	No	No	No
5.	Uses autonomously identified keywords for locating the most relevant resources	Yes	No	No	No
6.	Uniformly annotates resources with the relevant keywords to facilitate accurate searching	Yes	No	No	No
7.	Removes human error in the annotation of keywords	Yes	No	No	No

User Feedback

Based on the expert user feedback summarised in Table 5, the key strengths of the system include the ease of use, the specific focus on OER, the ability to quickly zero in on the required resource, and the use of desirability in the identification of the resource. The ability to autonomously identify academic domains and locate resources from heterogenous repositories regardless of the metadata standard are also found to be strengths of the system. The users felt that OERScout will especially benefit academics who are novices to OER.

One of the major weaknesses of the current prototype version was the limited number of resources indexed. This contributes to noise in the identified keywords and results in long lists of suggested and related terms. However, as the number of resources indexed grows, the noise words will be reduced giving way to more focused suggested and related terms. The users also felt that more advanced filters need to be added onto the search interface to allow filtering of properties such as file types and licences. However, the fundamental concept behind the desirability framework is to parametrically identify the most useful resources without the user's intervention. This observation suggests that a change in mindset with respect to search engines needs to take place before users can get accustomed to OERScout. The users also felt that the licensing scheme needs to be explained in non technical terms such as "can reuse, redistribute, revise and remix even commercially" instead of "CCBY". They further suggested that the calculation of the desirability be explained to the user.

The technology framework used was also found to be a limitation of the system. The current Microsoft Windows based client interface limits the users to Microsoft PC consumers. However, the real world implementation of the system will be done on a web based platform which will provide wider access regardless of device or operating system. Another limitation is that this version of OERScout is not designed to cluster non-text based materials such as audio, video, and animations which is a drawback considering the growing number of multimedia based OER. However, it is noted from the initial results that the system will accurately index multimedia based material using the textual descriptions provided. One more design limitation is its inability to cluster resources written in languages other than English. Despite this current limitation, the OERScout algorithm has a level of abstraction which allows it to be customised to suit other languages in the future.

Considering the opportunities, the system was found to be thought provoking with respect to finding, adopting, and adapting OER. It also appeals to the novice OER users in terms of training, affordability, teaching, and learning. This in turn will promote further research and development in the field of OER. Analysing the threats, one of the major threats to OERScout is the scale of the resource databases available to mainstream search engines such as Google. In this respect, the users felt that OERScout will be unable to compete with these search engines. However, the users also felt that OERScout addresses a few focused issues related to OER and need not be compared to mainstream search engines which are more general in nature. It is also worth noting

that the system will need to continuously update its resource database to ensure accuracy. Among the threats identified, the change in mindset with respect to this new search approach remains the greatest challenge to overcome.

Based on the above discussion, we strongly feel that the OERScout technology framework addresses the key deficiencies with respect to OER search. In sum, the provision of a centralised system which allows academics to effectively zero in on desirable resources hidden away in heterogenous repositories makes OERScout a viable alternative to existing OER search methodologies.

Conclusion

With more and more OER repositories mushrooming across the globe and with the expansion of existing repositories due to increased contributions, the task of searching for useful OER has become a daunting one. As discussed in the literature, a compounding factor to this current predicament is the inability of present day OER search methodologies to effectively locate resources which are desirable in terms of openness, access, and relevance. As a potential solution to this issue we propose the OERScout technology framework.

OERScout uses text mining techniques to cluster OER using autonomously mined domain specific keywords. It is developed with a view of providing OER creators and users a centralised system which will enable effective searching of desirable OER for academic use. The benefits of OERScout to content creators include (i) elimination of the need for manually annotating resources with metadata used in search; (ii) elimination of the need for publicising the availability of a repository and the need for native search mechanisms; and (iii) reach of material to a wider audience. The system benefits OER users by (i) providing a central location for finding resources of acceptable academic standard; (ii) locating only the most desirable resources for a particular teaching and learning need; and (iii) allowing the user to effectively zero in on the resources they are after. Based on the initial expert user test results, OERScout shows promise as a viable solution to the global OER search dilemma.The ultimate benefit of OERScout is that both content creators and users will only need to concentrate on the actual content and not the process of searching for desirable OER.

It is our intention to make OERScout available as a public service via www.oerscout.org which would allow academics to search desirable OER for their specific teaching and learning needs. We also intend to transfer the system onto a free and open source software (FOSS) platform in the spirit of openness and accessibility. Considering the limitation of the current system with respect to searching resources written in languages other than English, we are currently designing a further extension to OERScout which will facilitate searching of resources written in other languages. Furthermore, we are exploring the possibility of autonomously extracting some important IEEE LOM metadata from OER to provide better recommendations.

Acknowledgements

References

Abeywardena, I. S. (2013). Development of OER-based undergraduate technology course material: "TCC242/05 web database application" delivered using ODL at Wawasan Open University. In G. Dhanarajan & D. Porter (Eds.), *Open educational resources: An Asian perspective* (pp. 173-184). Vancouver: Commonwealth of Learning and OER Asia.

Abeywardena, I. S., Dhanarajan, G., & Chan, C. (2012). Searching and locating OER: Barriers to the wider adoption of OER for teaching in Asia. *Proceedings of the Regional Symposium on Open Educational Resources: An Asian Perspective on Policies and Practice.* Penang, Malaysia.

Abeywardena, I. S., Raviraja, R., & Tham, C. Y. (2012). Conceptual framework for parametrically measuring the desirability of open educational resources using D-index. *International Review of Research in Open and Distance Learning , 13*(2), 104-121.

Anido, L. E., Fernández, M. J., Caeiro, M., Santos, J. M., Rodriguez, J. S., & Llamas, M. (2002). Educational metadata and brokerage for learning resources. *Computers & Education , 38*(4), 351-374.

Balaji, V., Bhatia, M. B., Kumar, R., Neelam, L. K., Panja, S., Prabhakar, T. V., et al. (2010). Agrotags—a tagging scheme for agricultural digital objects. *In Metadata and Semantic Research* (pp. 36-45). Berlin Heidelberg: Springer.

Barton, J., Currier, S., & Hey, J. (2003). Building quality assurance into metadata creation: An analysis based on the learning objects and e-prints communities of practice. *In 2003 Dublin Core Conference: Supporting Communities of Discourse and Practice - Metadata Research and Applications.* Seattle, Washington.

Brooks, C., & McCalla, G. (2006). Towards flexible learning object metadata. *International Journal of Continuing Engineering Education and Life Long Learning , 16*(1), 50-63.

Calverley, G., & Shephard, K. (2003). Assisting the uptake of on-line resources: Why good learning resources are not enough. *Computers & Education , 41*(3), 205-224.

Casali, A., Deco, C., Romano, A., & Tomé, G. (2013). An assistant for loading learning object metadata: An ontology based approach. *Interdisciplinary Journal of E-Learning and Learning Objects (IJELLO), 9*, 11.

Caswell, T., Henson, S., Jenson, M., & Wiley, D. (2008). Open educational resources: Enabling universal education. *International Review of Research in Open and Distance Learning , 9*(1), 1-11.

Cechinel, C., Sánchez-Alonso, S., & Sicilia, M. Á. (2009). Empirical analysis of errors on human-generated learning objects metadata. In *Metadata and Semantic Research* (pp. 60-70). Berlin Heidelberg: Springer.

De la Prieta, F., Gil, A., Rodríguez, S., & Martín, B. (2011). BRENHET2, A MAS to facilitate the reutilization of LOs through federated search. *In Trends in Practical Applications of Agents and Multiagent Systems* (pp. 177-184). Berlin Heidelberg: Springer.

Devedzic, V., Jovanovic, J., & Gasevic, D. (2007). The pragmatics of current e-learning standards. *Internet Computing, 11*(3), 19-27.

Dhanarajan, G., & Abeywardena, I. (2013). Higher education and open educational resources in Asia: An overview. In G. Dhanarajan & D. Porter (Eds.), *Open educational resources: An Asian perspective* (pp. 3-10). Vancouver: Commonwealth of Learning and OER Asia.

Dichev, C., & Dicheva, D. (2012). Open educational resources in computer science teaching. *Proceedings of the 43rd ACM technical symposium on Computer Science Education* (pp. 619-624). ACM.

Feldman, R., & Sanger, J. (2006). *The text mining handbook: Advanced approaches in analyzing unstructured data.* Cambridge University Press.

Ha, K. H., Niemann, K., Schwertel, U., Holtkamp, P., Pirkkalainen, H., Boerner, D., et al. (2011). A novel approach towards skill-based search and services of open educational resources. *Proceedings of Metadata and Semantic Research* (pp. 312-323). Berlin Heidelberg: Springer.

Hilton, J., Wiley, D., Stein, J., & Johnson, A. (2010). The four R's of openness and ALMS Analysis: Frameworks for open educational resources. *Open Learning: The Journal of Open and Distance Learning,25* (1), 37-44.

Kolowich, S. (2012, November 5). *Pearson's open book.* Retrieved from http://www.insidehighered.com/news/2012/11/05/pearson-unveils-oer-search-engine

Lextek. (n.d.). *Onix Text Retrieval Toolkit API reference.* Retrieved from lextek.com: lextek.com/manuals/onix/stopwords1.html

McGreal, R. (2010). Open educational resource repositories: An analysis. *Proceedings of the 3rd Annual Forum on e-Learning Excellence.* Dubai, UAE.

Piedra, N., Chicaiza, J., López, J., Martínez, O., & Caro, E. T. (2010). An approach for description of open educational resources based on semantic technologies. *In Education Engineering (EDUCON)* (pp. 1111-1119). IEEE.

Piedra, N., Chicaiza, J., López, J., Tovar, E., & Martinez, O. (2011). Finding OERs with social-semantic search. *Proceedings of the 2011 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1195-1200). Amman, Jordan: IEEE.

Pirkkalainen, H., & Pawlowski, J. (2010). Open educational resources and social software in global e-learning settings. In P. Yliluoma (Ed.), *Sosiaalinen verkko-oppiminen* (pp. 23-40). Naantali: IMDL.

Shelton, B. E., Duffin, J., Wang, Y., & Ball, J. (2010). Linking opencoursewares and open education resources: Creating an effective search and recommendation system. *Procedia Computer Science, 1*(2), 2865-2870.

Tello, J. (2007). Estudio exploratorio de defectos en registros de meta-datos IEEE LOM de objetos de aprendizaje. *In Post-Proceedings of SPDECE 2007 - IV Simposio Pluridisciplinar sobre Diseno, Evaluacion y Desarrollo de Contenidos Educativos Reutilizables* (pp. 19-21). Bilbao, Spain.

Tunkelang, D. (2009). Faceted search. In G. Marchionini (Ed.), *Synthesis lectures on information concepts, retrieval, and services* (Vol. 5, pp. 1-80). Morgan & Claypool.

UNESCO. (2012, June 22). *2012 PARIS OER Declaration.* Retrieved from unesco.org: http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/WPFD2009/English_Declaration.html

Unwin, T. (2005). Towards a framework for the use of ICT in teacher training in Africa. *Open Learning: The Journal of Open, Distance and e-Learning, 20*(2), 113-129.

Vaughan, L. (2004). New measurements for search engine evaluation proposed and tested. *Information Processing and Management, 40*, 677-691.

West, P., & Victor, L. (2011). *Background and action paper on OER.* Report prepared for The William and Flora Hewlett Foundation.

Wiley, D. (2006). *On the sustainability of open educational resource initiatives in higher education.* Retrieved from oecd.org:http://www1.oecd.org/edu/ceri/38645447.pdf

Yamada, T. (2013). Open educational resources in Japan. In G. Dhanarajan & D. Porter (Eds.), *Open educational resources: An Asian perspective* (pp. 85-105). Vancouver: Commonwealth of Learning and OER Asia.

Yergler, N. R. (2010). Search and ciscovery: OER's open loop. *Proceedings of Open Ed 2010.* Barcelona.

Athabasca University

Appendix B

Abeywardena, I.S., Raviraja, R., & Tham, C.Y. (2012). Conceptual Framework for Parametrically Measuring the Desirability of Open Educational Resources using D-index. *International Review of Research in Open and Distance Learning, 13(2),* 104-121.

Conceptual Framework for Parametrically Measuring the Desirability of Open Educational Resources using D-Index

Ishan Sudeera Abeywardena and **Choy Yoong Tham**
Wawasan Open University, Malaysia
S. Raviraja
University of Malaya, Malaysia

Abstract

Open educational resources (OER) are a global phenomenon that is fast gaining credibility in many academic circles as a possible solution for bridging the knowledge divide. With increased funding and advocacy from governmental and nongovernmental organisations paired with generous philanthropy, many OER repositories, which host a vast array of resources, have mushroomed over the years. As the inkling towards an open approach to education grows, many academics are contributing to these OER repositories, making them expand exponentially in volume. However, despite the volume of available OER, the uptake of the use and reuse of OER still remains slow. One of the major limitations inhibiting the wider adoption of OER is the inability of current search mechanisms to effectively locate OER that are most suitable for use and reuse within a given scenario. This is mainly due to the lack of a parametric measure that could be used by search technologies to autonomously identify desirable resources. As a possible solution to this limitation, this concept paper introduces a parametric measure of desirability of OER named the *D-index*, which can aid search mechanisms in better identifying resources suitable for use and reuse.

Keywords: Open educational resources; OER; desirability of OER; locating suitable OER; use and reuse of OER; D-index

Introduction

Open educational resources (OER) are fast becoming a global phenomenon, which provides hope for bridging the knowledge divide among the masses (Geith & Vignare, 2008). With increased funding and advocacy by governmental and nongovernmental organisations buttressed by generous philanthropy, many OER repositories boasting a large volume of quality resources have mushroomed over the years. With the movement gaining credibility among many an academic community and with the drive toward opening up knowledge for the benefit of the less fortunate taking centre stage (Johnstone, 2005), these repositories have grown rich in knowledge. However, this has in turn given rise to the new challenge of locating resources suitable for use and reuse from the large number of disconnected and disparate repositories available around the globe (Geser, 2007).

As discussed by Hilton, Wiley, Stein, and Johnson (2010) the use and reuse of an OER depends on two factors: the permission and the technologies needed. The authors introduce the four Rs of openness and the ALMS analysis, which can be used to effectively gauge these factors for identifying the most suitable OER for use and reuse. However, at present, all of the three types of OER repositories, which include content OER repositories, portal OER repositories, and content and portal OER repositories (McGreal, 2010), consider only the relevance of a resource to the search query when locating internal and external resources. Thus, the rank of the search result is not a direct indicator of the suitability of a resource as it does not take into consideration the permission nor the technologies needed to successfully use and reuse. This challenge is further heightened by the common use of OER formats such as PDF, which renders resources useless with respect to reuse (Baraniuk, 2007), and the inability of average users to use the available technological tools to remix the resources (Petrides, Nguyen, Jimes, & Karaglani, 2008). Additionally, as resources are constantly added to these repositories (Dholakia, King, & Baraniuk, 2006), a static method of defining the suitability for use and reuse within the metadata becomes an impossible task.

As a possible solution to this issue, this paper introduces the concept of *desirability* of a resource, which parametrically takes into consideration (i) the level of openness with respect to the copyright license, (ii) the level of access with respect to technologies, and (iii) the relevance with respect to search rank. The desirability of an OER is then expressed as the *D-index* which allows search mechanisms as well as users to make informed decisions with respect to the most desirable OER for their needs.

Desirability of an OER

Rationale

In the academic community, the perceived quality of an academic publication or a resource is largely governed by peer review. However, with the present day influx of research publications being made available online, the peer-review mechanism becomes inefficient as not all the experts can review all the publications. As such, an alternative method of measur-

ing the quality of a publication or a resource is needed. According to Buela-Casal and Zych (2010),

> If an article receives a citation it means it has been used by the authors who cite it and as a result, the higher the number of the citations the more utilized the article. It seems to be an evidence of the recognition and the acceptance of the work by other investigators who use it as a support for their own work.

Therefore, at present the number of citations received is widely accepted as an indication of the perceived quality of an academic publication or resource.

As the styles of citation for academic publications are very well established, search mechanisms such as Google scholar (see http://scholar.google.com) have a usable parametric measure for providing an indication of how useful a publication would be for one's academic research. Although there are established styles of citation and attribution for OER as well, these styles are not standardised or widely practiced when using, reusing, remixing, and redistributing OER. As such, it is extremely difficult for a search mechanism to autonomously identify the number of citations or the number of attributions received by a particular OER material. This issue is further amplified as not all the OER repositories available over the Internet are searched and indexed by popular search mechanisms. Providing potential solutions to this issue are systems such as AnnotatEd (Farzan & Brusilovsky, 2006), which uses web-based annotations, use of brand reputation of a repository as an indication of quality, allowing users to review resources using set scales (Hylén, 2005), and the "popularity" in the Connexions repository, which is measured as percentile rank of page views/day over all time. Despite these very specific methodologies, there is still no generic methodology available at present to enable search mechanisms to autonomously gauge the usefulness of an OER for one's teaching and learning needs.

Definition

The usefulness of an OER for a particular teaching or learning need can only be accurately assessed by reading through the content of the resource. As this is quite a subjective exercise due to one's needs differing from another's, it is extremely difficult for a software-based search mechanism to provide any indication of this to a user. This aspect of use and reuse of OER will remain a human function regardless of the improvements in technology. When considering the use and reuse of an OER, there are other aspects of a resource that are fundamental to the usefulness of that particular resource and can be parametrically identified by a software-based mechanism. The first aspect is whether a resource is relevant to a user's needs. This can be assessed by the search ranking of a resource when searched for with a search mechanism. The search mechanism will compare the title, description, keywords, and sometimes the content of the material to find the best match for the search query. The second aspect is whether the resource is open enough for using, reusing, remixing, and redistributing. This becomes important depending on what the user wants to accomplish

with the resource. The third aspect is the accessibility of the resource with respect to technology. If the user cannot easily use, reuse, and remix a resource with available technology, the resource becomes less useful. Therefore, the usefulness of an OER with respect to (i) the level of openness, (ii) the level of access, and (iii) the relevance can be defined as the *d*esirability of an OER, indicating how desirable it is for use and reuse for one's needs. Within the requirement of being able to use and reuse a particular OER, these three parameters can be defined as follows:

1. *level of openness*, the permission to use and reuse the resource;

2. *level of access*, the technical keys required to unlock the resource; and

3. *relevance*, the level of match between the resource and the needs of the user.

As each of these mutually exclusive parameters are directly proportionate to the desirability of an OER, the desirability can be expressed as a three-dimensional measure as shown in Figure 1.
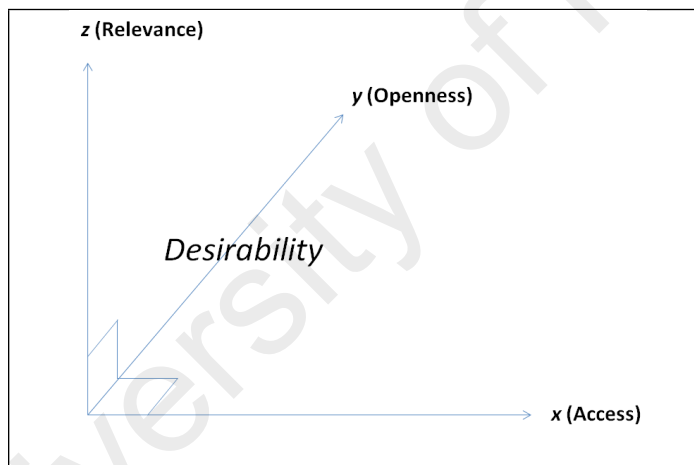
Figure 1. Desirability of an OER.

The Scales

In order to parametrically calculate the desirability of an OER, each of the parameters discussed above needs to be given a numeric value based on a set scale. These scales can be defined in the following ways.

The level of openness can be defined using the four Rs of openness (Hilton, Wiley, Stein, & Johnson, 2010) as shown in Table 1. The four Rs stand for *reuse*, the ability to use all or part of a work for one's own purposes; *redistribute*, the ability to share one's work with others; *revise*, the ability to adapt, modify, translate, or change the form of a work; and *remix*, the ability to combine resources to make new resources. The values 1 to 4 were assigned to the four Rs where 1 corresponds to the *lowest level of openness* and 4 corresponds to the *highest level*.

Table 1

The Level of Openness Based on the Four Rs of Openness

Permission	Value
Reuse	1
Redistribute	2
Revise	3
Remix	4

The level of access was defined on a scale of 1 to 16 using the ALMS analysis (Hilton, Wiley, Stein, & Johnson, 2010), which identifies the technical requirements for localisation of an OER with respect to access to editing tools, level of expertise required to revise or remix, ability to meaningfully edit, and source-file access. As shown in Table 2, the value 1 corresponds to the *lowest accessibility* and value 16 to the *highest accessibility*.

Table 2

The Level of Access Based on the ALMS Analysis

Access (**A**ccess to editing tools | **L**evel of expertise required to revise or remix | **M**eaningfully editable | **S**ource-file access)	Value
Low | High | No | No	1
Low | High | No | Yes	2
Low | High | Yes | No	3
Low | High | Yes | Yes	4
Low | Low | No | No	5
Low | Low | No | Yes	6
Low | Low | Yes | No	7
Low | Low | Yes | Yes	8
High | High | No | No	9
High | High | No | Yes	10
High | High | Yes | No	11
High | High | Yes | Yes	12
High | Low | No | No	13
High | Low | No | Yes	14
High | Low | Yes | No	15

High | Low | Yes | Yes	16

The relevance of a resource to a particular search query can be measured using the rank of the search results. According to Vaughan (2004) users will only consider the top ten ranked results for a particular search as the most relevant. Vaughan further suggests that users will ignore the results below the top 30. Based on this premise, the scale for the relevance was defined as shown in Table 3, where the value 1 is the *least relevant* and value 4 is the *most relevant*.

Table 3

The Level of Relevance Based on Search Rank

Search rank	Value
Below the top 30 ranks of the search results	1
Within the top 21-30 ranks of the search results	2
Within the top 11-20 ranks of the search results	3
Within the top 10 ranks of the search results	4

Calculation

Based on the scales, the desirability of an OER can then be defined as the volume of the cuboid, as shown in Figure 2, calculated using the following formula.

$$desirability = level\ of\ access \times level\ of\ openness \times relevance$$

As a result, the desirability becomes directly proportionate to the volume of the cuboid.
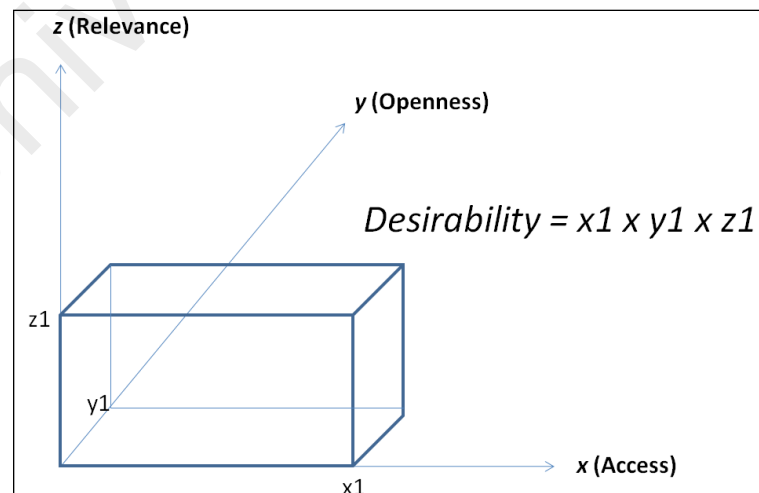
Figure 2. Calculation of desirability.

By normalising the values indicated in Table 1, Table 2, and Table 3 to make the scales uniform for the calculation, the D-index of an OER can be calculated using the following formula.

$$\text{D-index} = (\text{level of access} \times \text{level of openness} \times \text{relevance}) / 256$$

Based on the above calculation, a resource becomes more desirable as the D-index increases on a scale of 0 to 1, where 0 is the *least desirable* and 1 is the *most desirable*.

Verification of Concept

The most commonly used method for locating OER is to use a generic search mechanism such as Google or to use a search mechanism specific to an OER repository such as Connexions (see http://cnx.org/) or Wikieducator (see http://wikieducator.org). However, both of these types of search mechanisms only consider the relevance of the resource either by matching the title and description or the keywords to the search query provided by the user. Therefore, the resources returned as the top search results might not always be the most desirable for use and reuse in a given scenario as they might be less open or less accessible. The D-index is specifically designed to overcome this limitation by taking into consideration the openness and the accessibility of an OER in addition to the relevance to the search query. When applying the D-index to an OER repository, the level of access, discussed in Table 2, needs to be implemented using the file formats of the OER, where their features are mapped against the ALMS. The level of openness, based on the four Rs discussed in Table 1, needs to be measured using the copyright licensing scheme under which the resource was released. The de facto scheme used in most repositories is the Creative Commons (CC) (see http://creativecommons.org/) licensing scheme, which has six derivations based on the level of openness. However, other specific licensing schemes such as the GNU Free Documentation License (see http://www.gnu.org/copyleft/fdl.html) can also be used for this purpose as long as they can be categorised into the four levels of openness constituting the desirability. Table 4 maps the six CC licences to the four Rs of openness. However, it should be noted that the level of openness of the CC licenses starts at the redistribute level.

Table 4

Mapping the CC Licences to the 4 Rs

Permission	Creative Commons (CC) licence	Value
Reuse	None	1
Redistribute	Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) Attribution-NoDerivatives (CC BY-ND)	2
Revise	Attribution-NonCommercial-ShareAlike (CC BY-NC-SA) Attribution-ShareAlike (CC BY-SA)	3
Remix	Attribution-NonCommercial (CC BY-NC) Attribution (CC BY)	4

Methodology

To verify the accuracy of the proposed D-index, experiments were carried out in three widely used OER repositories: OER Commons (see http://www.oercommons.org), Jorum (see http://jorum.ac.uk/), and MERLOT (see http://www.merlot.org/). These repositories were selected for the experiments due to (i) the repositories providing users with native search mechanisms to locate OER available within the repository as well as hosted outside and (ii) the variety of OER available through them in different levels of openness and access. Each repository was searched using the term *calculus* to locate OER on the topic of calculus in mathematics. The term calculus was intentionally selected for these experiments due to the large number of OER written and made available on the topic. Only the top 40 search results from each repository, returned based on relevance, were considered in the experiments as the users tend to ignore results below the rank of 30 (Vaughan, 2004).

Calculation of the D-index

To demonstrate how the D-index was calculated for each search result, a general search was conducted on the OER Commons repository for the term calculus using its native search mechanism. Out of the 165 resources returned as results, three resources at the postsecondary level with different search rank were chosen for comparison as shown in Table 5.

Table 5

Selected Search Results at Postsecondary Level Returned by the OER Commons Search Mechanism for the Search Term Calculus

Resource	Title	Search rank	License	File type
A	Calculus I	2	Creative Commons Attribution-Noncommercial-Share Alike 3.0 (CC BY-NC-SA)	PDF
B	Topics in Calculus	8	Creative Commons Attribution-Noncommercial 3.0 (CC BY-NC)	HTML
C	Calculus I (MATH 151)	23	Creative Commons Attribution 3.0 Unported (CC BY)	MS Word

The search rank, licence, and the file type of each resource in Table 5 was then compared with Table 3, Table 4, and Table 2 respectively to identify the parameters required to calculate the D-index as shown in Table 6.

Table 6

Parameters Required for Calculating the D-index

Resource	Relevance	Openness (four R's)	Access (ALMS)
A	4	3	1 (Low | High | No | No)
B	4	4	16 (High | Low | Yes | Yes)
C	2	4	8 (Low | Low | Yes | Yes)

Looking at Table 6 we can see that the search mechanism has ordered the results according to the relevance where resource A is the most relevant. However, resource A is less open and less accessible when compared with resource B. Table 7 shows how the results would be reorganised when the D-index is applied to the same search results.

Table 7

After Applying the D-index to the Same Search Results Shown in Table 5

Resource	Relevance	Openness	Access	D-index
B	4	4	16	1.00
C	2	4	8	0.25
A	4	3	1	0.05

From the results in Table 7, it can be seen that resource B would be the most desirable OER for use and reuse due to its level of openness and access even though resource A was the most relevant.

Experiment Results

Table 8, Table 10, and Table 12 show the top 10 results returned by the native search mechanisms of MERLOT, JORUM, and OER Commons respectively for the keyword calculus. Table 9, Table 11, and Table 13 show the top 10 results when the D-index is applied to the search results returned by MERLOT, JORUM, and OER Commons respectively.

Table 8

Top Ten Search Results Returned by MERLOT for the Keyword Calculus

Search rank	Title	CC license	File type
1	18.01 Single Variable Calculus	CC BY-NC-SA	PDF
2	Calculus for Beginners and Artists	CC BY-NC-SA	HTML/Text
3	18.01 Single Variable Calculus	CC BY-NC-SA	PDF
4	18.013A Calculus with Applications	CC BY-NC-SA	HTML/Text
5	18.02 Multivariable Calculus	CC BY-NC-SA	PDF
6	Single Variable Calculus	CC BY-NC-SA	PDF
7	Calculus Online Textbook	CC BY-NC-SA	PDF
8	Calculus for Beginners and Artists	CC BY-NC-SA	HTML/Text
9	18.075 Advanced Calculus for Engineers	CC BY-NC-SA	PDF
10	MATH 140 - Calculus I, Summer 2007	CC BY-NC-SA	Protected

Table 9

Top Ten Results when D-index is Applied to the Results Returned by MERLOT

Rank after applying D-index	Original search rank	Title	CC License	File type	D-index
1	2	Calculus for Beginners and Artists	CC BY-NC-SA	HTML/Text	0.75
2	4	18.013A Calculus with Applications	CC BY-NC-SA	HTML/Text	0.75
3	8	Calculus for Beginners and Artists	CC BY-NC-SA	HTML/Text	0.75

4	14	Multivariable Calculus	CC BY	HTML/ Text	0.75
5	19	MATH 10250 - Elements of Calculus I, Fall 2008	CC BY-NC-SA	HTML/ Text	0.56
6	20	18.022 Calculus	CC BY-NC-SA	PDF	0.56
7	22	Single-Variable Calculus I	CC BY	HTML/ Text	0.50
8	25	Single-Variable Calculus II	CC BY	HTML/ Text	0.50
9	15	Highlights of Calculus	CC BY-NC-SA	Video	0.42
10	21	Calculus I	CC BY	HTML/ Text	0.38

Table 10

Top Ten Search Results Returned by JORUM for the Keyword Calculus

Search rank	Title	CC License	File type
1	Introduction to Calculus	CC BY-NC	Video
2	Introduction to Artificial Intelligence - Neural Networks	CC BY-NC-SA	MS Word
3	Calculus (integration) : mathematics 1 level 4	CC BY	Slides
4	Calculus - Income Growth, Consumption and Savings	CC BY-NC	Video
5	Introduction to Econometrics: EC220	CC BY-NC	PDF
6	Further Mathematical Methods	CC BY-NC-SA	XHTML
7	Transient Responses : Laplace Transforms : Electrical and Electronic Principles : Presentation Transcript	CC BY	Slides
8	Calculus - Determining Marginal Revenue	CC BY-NC	Video
9	Film Series Four - Conclusion	CC BY-NC	Video
10	Finding the Optimal Number of Floors in Hotel Construction - Part One	CC BY-NC	Video

Table 11

Top Ten Results when D-index is Applied to the Results Returned by JORUM

Rank after applying D-index	Original search rank	Title	CC License	File type	D-index
1	1	Introduction to Calculus	CC BY-NC	Video	0.75
2	4	Calculus - Income Growth, Consumption and Savings	CC BY-NC	Video	0.75
3	6	Further Mathematical Methods	CC BY-NC-SA	XHTML	0.75
4	8	Calculus - Determining Marginal Revenue	CC BY-NC	Video	0.75
5	9	Film Series Four - Conclusion	CC BY-NC	Video	0.75
6	10	Finding the Optimal Number of Floors in Hotel Construction - Part One	CC BY-NC	Video	0.75
7	13	Maths Solutions	CC BY	HTML/Text	0.75
8	11	Finding the Optimal Number of Floors in Hotel Construction - Part Two	CC BY-NC	Video	0.56
9	12	Finding the Optimal Number of Floors in Hotel Construction - Conclusion	CC BY-NC	Video	0.56
10	14	Mathematical Analysis	CC BY-NC-SA	HTML/Text	0.56

Table 12

Top Ten Search Results Returned by OER Commons for the Keyword Calculus

Search rank	Title	CC License	File type
1	Whitman Calculus	CC BY-NC-SA	HTML/Text
2	Calculus I	CC BY-NC-SA	PDF
3	AP Calculus	CC BY-NC-SA	HTML/Text
4	Applied Calculus	Propritery	HTML/Text
5	A Summary of Calculus	Propritery	PDF
6	Advanced Calculus	CC BY-NC-SA	PDF
7	Multivariable Calculus	Propritery	PDF
8	Topics in Calculus	CC BY-NC	PDF
9	Highlights of Calculus	CC BY-NC-SA	Video
10	Vector Calculus	Propritery	HTML/Text

Table 13

Top Ten Results when D-index is Applied to the Results Returned by OER Commons

Rank after applying D-index	Original search rank	Title	CC License	File type	D-index
1	1	Whitman Calculus	CC BY-NC-SA	HTML/Text	0.75
2	3	AP Calculus	CC BY-NC-SA	HTML/Text	0.75
3	11	Vector Calculus	GNU Free Documentation License	HTML/Text	0.75
4	9	Highlights of Calculus	CC BY-NC-SA	Video	0.56
5	16	Calculus (Student's Edition)	CC BY-NC-SA	HTML/Text	0.56
6	22	Calculus II (MATH 152)	CC BY	HTML/Text	0.50
7	23	Calculus I (MATH 151)	CC BY	HTML/Text	0.50
8	24	Calculus III (MATH 153)	CC BY	HTML/Text	0.50
9	15	Calculus Revisited, Fall 2010	CC BY-NC-SA	Video	0.42
10	21	Calculus (Teacher's Edition)	CC BY-NC-SA	HTML/Text	0.38

Discussion

By comparing Table 8 and Table 9, which show the search results returned by MERLOT, it can be seen that the original top 10 search results (Table 8) only contain resources that are released under the CC BY-NC-SA license. This license significantly restricts the user's freedom with respect to the four Rs. Also six of 10 resources returned are in PDF format, which make them difficult to reuse and remix. It must also be noted that the resource ranked as number 10 is a protected resource, which requires a specific username and password to access. Looking at Table 9 where the results are reranked according to the D-index, it can be seen that eight of 10 resources are in HTML/text formats, which are the most accessible in terms of reuse. Four of 10 resources are available under the CC BY licence, which make them the most open resources in the list. Similarly, by comparing Table 10 and Table 11, we can see that the use of the D-index has reranked the top 10 results so that the most accessible resources are ranked at the top instead of resources that use proprietary software applications. The video resources returned were given an accessibility value of 12 according to the ALMS, where access to editing tools = high; level of expertise required to revise or remix = high; meaningfully editable = yes; and source-file access = yes.

Analysing Table 12 it can be seen that four of 10 results returned by the OER Commons search mechanism are copyright protected. As such these cannot be considered as OER and are the least useful for a user who is searching for open material. A value of 0 for openness was assigned to these resources during the D-index calculation. Furthermore, five of the top 10 results returned by the OER Commons search mechanism were in PDF format. Looking at Table 13, it can be seen that the application of the D-index has reranked the resources to provide eight of 10 HTML/text resources. Also the proprietary content has been replaced with open content released under the CC BY and CC BY-NC-SA licenses. The third-ranked resource which is released under the GNU Free Documentation License was assigned a value of 4 for openness during the calculation of the D-index.

By referring to the above results from the experiments conducted on three widely used OER repositories, it can be concluded that the application of the D-index would greatly improve the effectiveness of the search with respect to locating the most suitable resources for use and reuse.

Application and Limitations

The D-index can be incorporated into any search mechanism of an OER repository provided that the resources in the repository are appropriately tagged with the necessary metadata, such as title, description, keywords, copyright license, and file type. Many OER repositories now require authors to define the basic metadata, such as the title, description, keywords, and copyright license. As such, the use of these parameters to gauge the values for relevance and openness becomes an easier task. However, gauging the access parameter which uses the file type of the OER becomes a much more challenging task as some resources consist of multiple files of multiple formats. This can be rectified by breaking a collection of OER into individual learning objects, which allows software applications to determine the file type of

the individual OER.

A couple of practical limitations can also be identified with respect to the implementation of the D-index in OER repositories. One of these limitations is that the desirability becomes one dimensional due to the copyright license and the file format being fixed in repositories such as Connexions or Wikieducator. As a result, the D-index becomes only a function of the relevance parameter which does not add much value to the existing search mechanism. Therefore, the D-index is best suited for use in portal repositories/content and portal repositories, such as the OER Commons, MERLOT, and JORUM, which have a wide variety of resources of different file types released under various copyright licenses. It will also be quite effective when used with search mechanisms which query multiple repositories to identify resources.

The other practical limitation is the subjectivity of the search algorithms used by the various native search mechanisms, which results in disparity of the search rank. In turn, this disparity results in the relevance parameter becoming a function of the effectiveness of the search algorithm.

Conclusion and Future Work

Open educational resources (OER) are fast becoming accepted sources of knowledge for teachers and learners around the globe. This is especially true in the case of open distance learning (ODL) institutions where the teaching and learning philosophy is based on open access to education. With the recent developments in technology as well as the establishment of many high quality OER repositories freely available online, the use and reuse of OER should have become mainstream practice. However, as it stands, the use and reuse of OER are still inhibited by a number of technological, social, and economic factors. One of the major technological limitations dampening the use and reuse of OER is the inability to effectively locate useful resources for specific teaching and learning needs from the variety of disconnected and disparate repositories available. This gives rise to the challenge of identifying a parametric measure of the usefulness of an OER, which will enable users to effectively identify suitable resources without reading through countless unsuitable ones.

The concept of desirability of an OER introduced in this paper attempts to lessen the pain of OER users with respect to identifying resources that are relevant, open, and accessible for one's particular needs. Currently, users who search for OER in specific repositories use search mechanisms native to the repository to identify relevant resources. Depending on the algorithms used by the native search mechanisms, the search query will be compared against the metadata of a resource such as title, description, and keywords to provide a list of resources which might be of relevance. However, these search mechanisms do not take into consideration the level of openness or the technological skills required with respect to using, reusing, remixing, and redistributing a resource. The D-index is an attempt to factor in the openness and accessibility in addition to the relevance in order to provide OER users a useful set of search results which are appropriate to their needs.

The D-index can be incorporated into any OER repository provided that the necessary metadata for calculation are available. It is most effective when used in portal repositories/content and portal repositories which search multiple disconnected OER repositories to locate relevant material. The greatest benefit of the D-index to teachers and learners is its ability to locate and list the most desirable OER for use and reuse from the numerous combinations of relevance, openness, and access under which OER are released. The authors are currently working on incorporating the D-index into an artificial intelligence (AI) based text mining system named *OERScout* which is used to cluster OER available in all the disconnected repositories based on autonomously identified keywords. The use of the D-index in this clustering process will enable search mechanisms to effectively locate OER which are most desirable for use and reuse.

Acknowledgements

References

Baraniuk, R. G. (2007). Challenges and opportunities for the open education movement: A Connexions case study. In T. Iiyoshi & M. S. V. Kumar (Eds.), *Opening up education – The collective advancement of education through open technology, open content, and open knowledge*. Cambridge, MA: Massachusetts Institute of Technology Press. Retrieved from http://citadel.cnx.rice.edu:8180/risa/docs/presskit/cnxbrochuresposter/baraniuk-MIT-press-chapter-oct07.pdf .

Buela-Casal, G., & Zych, I. (2010). Analysis of the relationship between the number of citations and the quality evaluated by experts in psychology journals. *Psicothema, 22*(2), 270-276.

Dholakia, U. M., King, W. J., & Baraniuk, R. (2006). *What makes an open education program sustainable? The case of Connexions*. Retrieved from http://www.agri-outlook.org/dataoecd/3/6/36781781.pdf.

Farzan, R., & Brusilovsky, P. (2006). AnnotatEd: A social navigation and annotation service for web-based educational resources. *Proceedings: E-Learn 2006–World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education,* Honolulu, Hawaii. Retrieved from http://www2.sis.pitt.edu/~peterb/papers/NRHM-Final-AnnotatEd.pdf

Geith, C., & Vignare, K. (2008). Access to education with online learning and open educational resources: Can they close the gap? *Journal of Asynchronous Learning Networks, 12*(1), 105-126.

Geser, G. (2007). *Open educational practices and resources - OLCOS Roadmap 2012*. Open Learning Content Observatory Services. Salzburg, Austria. 2007. Retrieved from http://www.olcos.org/cms/upload/docs/olcos_roadmap.pdf.

Hilton, J., Wiley, D., Stein, J., & Johnson, A. (2010). The four R's of openness and ALMS Analysis: Frameworks for open educational resources. *Open Learning: The Journal of Open and Distance Learning, 25*(1), 37-44.

Johnstone, S. (2005). Open educational resources serve the world. *Educause Quarterly, 28*(3).

McGreal, R. (2010). Open educational resource repositories: An analysis. *Proceedings: The 3rd Annual Forum on e-Learning Excellence,* Dubai, UAe. Retrieved from http://elexforum.hbmeu.ac.ae/Proceeding/PDF/Open%20Educational%20Resource.pdf.

Petrides, L., Nguyen, L., Jimes, C., & Karaglani, A. (2008). Open educational resources: inquiring into author use and reuse. *International Journal of Technology Enhanced Learning, 1*(1/2).

Vaughan, L. (2004). New measurements for search engine evaluation proposed and tested. *Information Processing and Management 40*, 677–691.

Athabasca University

Appendix C

Abeywardena, I.S., Chan, C.S., & Balaji, V. (2013). OERScout: Widening Access to OER through Faceted Search. *Proceedings of the 7th Pan-Commonwealth Forum (PCF7), Abuja, Nigeria*.

OERScout: Widening Access to OER through Faceted Search

Ishan Sudeera Abeywardena, Wawasan Open University, ishansa@wou.edu.my
Chee Seng Chan, University of Malaya, cs.chan@um.edu.my
V. Balaji, Commonwealth of Learning, vbalaji@col.org

Sub theme: Promoting Open Educational Resources (OER)

ABSTRACT

In recent years, the Open Educational Resources (OER) movement has achieved considerable success within the academic community with respect to advocacy of the concept. As a result, many organisations such as the Commonwealth of Learning (COL), UNESCO and the International Development Research Centre (IDRC), in partnership with academic institutions, have produced large volumes of OER. However, due to the disconnected nature and the constant expansion of volume, many repositories hosting these resources are less frequented or completely ignored by OER users. i.e. only the more popular OER repositories such as Connexions and WikiEducator are frequent stops in the search for academically useful resources. This limitation, in turn, reduces the acces to high quality resources hidden away in isolated repositories hosted by lesser known sources. Furthermore, the time and labour required to trawl these repositories with a view of identifying the most suitable OER is tantamount to creating ones' own material from scratch. As a solution to these issues, this paper discusses how the OERScout technology framework uses a "faceted search" approach to locate the most desirable OER from sources spread throughout the globe. It also highlights how focused searching can greatly improve access to OER readily useable in teaching and learning.

Keywords: OERScout, OER Search, OER Curation, Faceted Search, Access to OER

INTRODUCTION

OER are fast gaining traction amongst the academic community as a viable means of increasing access and equity in education. The concept of OER is of especial significance to the marginalised communities in the Global South where distance education is prominent due to the inability of conventional brick and mortar institutions to cope with the growing demand (Lane, 2009). However, the wider adoption of OER by academics in the Global South has been inhibited due to various socio, economic and technological reasons. One of the major technological inhibitors is the current inability to search for OER which are academically useful and are of an acceptable academic standard.

In his study on "*which inhibiting factors for reuse do content developers in developing countries experience with open content?*" Hatakka (2009) points out that the most inhibiting factor is the inability to locate 'relevant' material for a particular teaching or learning need. The subjects of Hatakka's study attribute the inability to locate relevant material to (i) the inability to locate resources which fit the scope of the course in terms of context and difficulty; (ii) the lack of awareness with respect to how 'best' to search for material on the Internet; and (iii) the inability to choose the most appropriate resources from the large number of resources returned by search engines such as Google. Affirming these statements, Shelton et al. (2010, p. 316) argue

"Well-studied and commercialized search engines like Google will often help users to find what they are seeking. However, if those searching do not know exactly what they are looking for, or they do not know the 'proper' words to describe what it is that they want, the searching results returned are often unsatisfactory".

In an attempt to identify how effective mainstream search engines such as Google are with respect to locating relevant OER, Dichev et al. (2011) of the Winston-Salem State University conducted an experiment by putting Google head to head against native search mechanisms of OER repositories. To make the Google search narrower to OER, the advanced search feature 'free to use, share or modify, even commercially' was used. Alongside Google, native search mechanisms of 12 OER repositories were used to search for material in the computer science domain. The repositories were namely: Connexions, MIT OpenCourseWare, CITIDEL, The Open University, OpenLearn, OpenCourseWare Consortium, OER Commons, Merlot, NSDL, Wikibooks, SOFIA, Textbook Revolution and Bookboon. From their comparison between Google and native OER search mechanisms with respect to locating relevant material, it is apparent that native search mechanisms fair better than Google in terms of locating relevant material.

Commenting further on the inability of mainstream search engines such as Google to effectively locate OER, Pirkkalainen & Pawlowski (2010, p. 24) state that *"… searching this way might be a long and painful process as most of the results are not usable for educational purposes"*. Furthermore, they argue that search mechanisms native to OER repositories are capable of locating resources with an increased relevance. However, the problem is which repositories to choose within the large global pool. Levey (2012, p. 134) relates to this from working in the African 'AgShare' project. She states her experience as

"Despite numerous gateways, it is not always easy to identify appropriate resources. How a resource is tagged or labelled is one problem. Poor information retrieval skills is another. Furthermore, academics are busy".

This inadequacy with respect to searching for OER from a diversity of sources gives rise to the need for new alternative methodologies which can assist in locating relevant resources. Ideally these search tools should return materials which are relevant, usable and from a diversity of sources (Yergler, 2010). Yergler further suggests that the reliance on a full text index and link analysis of mainstream search engines impede the process of discovery by including resources not necessarily educational. As such, *"increasing the relevance of the resources returned by a search engine can minimize the time educators need to spend exploring irrelevant resources"* (Yergler, 2010, p. 2).

The UNESCO Paris OER Declaration (2012) , which is a global non-binding declaration signed by many governments, declares the need for more research into OER search. The recommendation reads

"i. Facilitate finding, retrieving and sharing of OER: Encourage the development of user-friendly tools to locate and retrieve OER that are specific and relevant to particular needs. Adopt appropriate open standards to ensure interoperability and to facilitate the use of OER in diverse media".

This declaration is the culmination of a global effort towards establishing a roadmap for the future development of the OER movement. The above recommendation made with respect to OER search reaffirms the need for new and more effective OER search methodologies within the context of locating relevant material for particular teaching and learning needs.

THE FACETED SEARCH APPROACH

Search engines have undergone rapid evolution in the past decade due to global technological giants such as Google. In his book 'Faceted Search'*,* Daniel Tunkelang (2009) of Google explains how previous search technologies morphed into the faceted search approach. According to Tunkelang, the earliest search engines used the Boolean retrieval model which limited the flexibility and increased the complexity of the search query. Abandoning this method, information retrieval (IR) researches adopted a free-text query approach which provided increased flexibility in creating search queries. This method cast a wide

net to return results based on rank. Although not as accurate as Boolean retrieval, many search engines still follow the free-text query approach incorporating the ranked retrieval framework. Another approach used in searching for information, especially on the World Wide Web (WWW), is the directory approach. The advantage of this approach is the organisation of content based on set taxonomies. This allowed users to navigate categories and sub categories to ultimately arrive at the information they are after. However, Tunkelang highlights that the creators of the taxonomies themselves and the users frequently disagree on the categorisation of the content. i.e. users will have to learn to think like the creators to find the relevant information.

Faceted search is a hybrid search approach which combines parametric search and faceted navigation (Tunkelang, 2009). According to Dash et al. (2008, p. 3) *"First, it smoothly integrates free text search with structured querying. Second, the counts on selected facets serve as context for further navigation".* Marti Hearst of UC Berkley, who was the lead researcher in the popular *Flexible information Access using Metadata in Novel COmbonations* (Flamenco) faceted search project, argues *"a key component to successful faceted search interfaces (which unfortunately is rarely implemented properly) is the implementation of keyword search"* (Hearst, 2006, p. 4)*.* In simpler terms, modern faceted search combines free-text querying to generate a list of results based on keywords which can then be refined further using a Boolean, structured or directory approach. To achieve this functionality, faceted metadata need to be extracted from documents using text mining techniques. A few general strategies are (i) exploit latent metadata such as document source, type, length; (ii) use rule based or statistical techniques to categorise documents into predetermined categories; and (iii) use an unsupervised approach such as terminology extraction to obtain a list of terms from the document (Tunkelang, 2009).

Typical interaction between a faceted search interface and the user is explained by Ben-Yitzhak et al, (2008) as (i) type or refine a search query; or (ii) navigate through multiple, independent facet hierarchies that describe the data by drill-down (refinement) or roll-up (generalization) operations. Koren et al. (2008, p. 477) further explains this interaction as

"The interfaces present a number of facets along with a selection of their associated values, any previous search results, and the current query. By choosing from suggested values of these facets, a user can interactively refine the query."

Ultimately, faceted search allows users to quickly drill down into a more focused set of search results using the initial results set.

SEARCHING FOR OER WITH OERSCOUT

The 'OERScout' technology framework (Abeywardena, Chan, & Tham, 2013) is a comprehensive solution to the current OER search dilemma (Abeywardena & Chan, 2013). It uses text mining techniques to autonomously mine specific keywords which accurately describe the academic domains of a particular OER. In essence, OERScout (i) "reads" textual educational resources; (ii) "understands" the content; and (iii) "recommends" the most useful resources for a particular teaching or learning need. The usefulness of a particular OER is parametrically measured using the 'Desirability' framework (Abeywardena, Raviraja, & Tham, 2012) which takes into consideration the (i) openness; (ii) accessibility; and (iii) relevance attributes of a resource. The system then creates a searchable dataset called the Keyword-Document Matrix (KDM) which is used by the OERScout client interface to effectively search for resources.

In addition to the text mining techniques employed at the server end to create the KDM, the user interface of OERScout equally contributes to the novelty of this solution. The faceted search approach available to the users via the client interface is a far cry from the conventional free text search method where users are presented with a static list of search results spread across hundreds of pages. It is also superior to the directory search method where users are forced to manually drill down multiple layers before arriving at the resources they are after.

The searching for desirable OER using the OERScout interface is threefold. Firstly, the user inputs a search query into the free text search box. Unlike in existing search methodologies where the accuracy of the search query governs the relevance of the search results, OERScout extracts the key terms from the search query by removing stop words to form multiple focused search queries. These queries are then executed on the KDM to generate a list of 'Suggested Terms'. The suggested terms act as the first facet which allows the user to select from a broad list of domains autonomously mined by OERScout. Secondly, the user selects a particular area of interest from the list of suggested terms. This action creates the second facet which lists the 'Related Terms' to the selected suggested term. Thirdly, the user hones in on the exact subject domain he/she is after in the related terms facet to generate a ranked list of desirable resources.
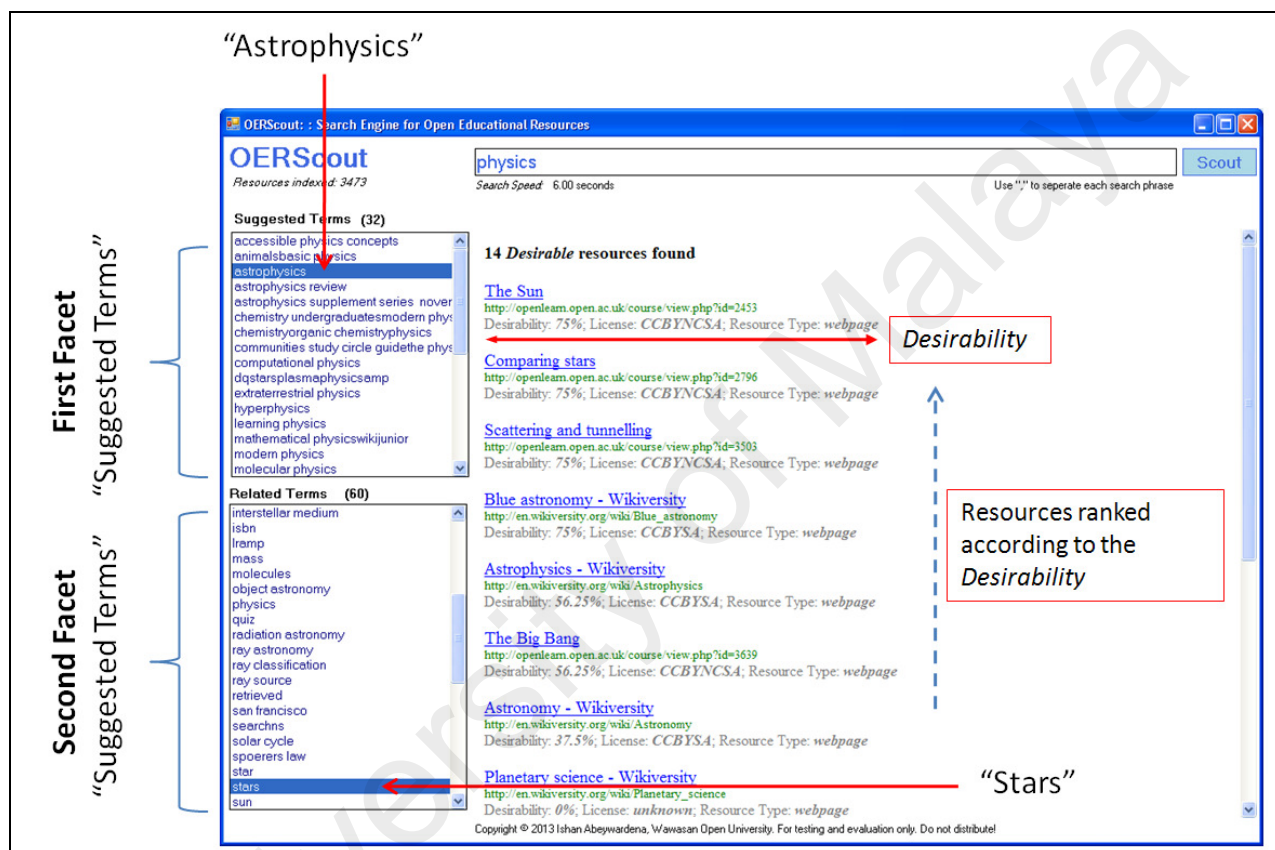
Figure 1 OERScout faceted search user interface. The figure shows a search conducted for *Physics: Astrophysics: Stars*.

Figure 1 shows an example of a faceted search conducted on OERScout to locate resources in "Physics". The suggested terms facet has listed 32 different topic areas identified by the system in the domain of "Physics". According to the selection in the first facet which is "Astrophysics", 60 related topics have been listed in the second facet. Based on the selection in the second facet, a list of desirable resources have been presented to the user which covers the topic "stars". The resources are arranged in descending order of the Desirability. The Desirability, license type and resource types are also indicated to the user to facilitate faster selection.

Referring to Figure 1, the top three resources returned are from the OpenLearn repository of The Open University which is highly reputed for the quality of its academic content. From this example, it is apparent that the OERScout faceted search interface allows users to quickly and effectively hone in on desirable OER required for their teaching and learning needs. It also spares users from reading a large number of resources returned by a search engine to ascertain their usefulness for a particular academic purpose.

BENEFITS OF OERSCOUT TO THE COMMUNITY

A number of technological factors have contributed to the current OER search dilemma. The first among these is the inability of mainstream search engines such as Google to effectively locate useful resources for academic purposes. Furthermore, the dependence of these search engines on human annotated metadata and commercial page ranking algorithms force them to give prominence to the widely popular OER repositories such as Wikipedia, WikiEducator and Connexions. However, these repositories might not contain material which are the most desirable for a particular academic need. As such, the use of these search engines limits the access to the wider resource pool available globally. OERScout addresses this issue by autonomously mining metadata for search purposes. As a result, the annotation of resources becomes consistent and uniform. Furthermore, the learning aspect of the algorithm constantly strives to identify the most accurate metadata for a particular OER. When combined with the Desirability framework, OERScout objectively determines the usefulness of a resource based only on the needs of the user. In turn, resources from less popular repositories will be given the same search visibility as the resources from the more popular. This aspect of the system significantly increases the access to the global pool of quality OER.

The repository independence of OERScout is another key benefit to the OER community. Traditionally, content creators would have to archive their material on a repository to enhance searchability. Furthermore, they would have to provide the necessary metadata and comply with the repository's technological requirements. Due to the heterogeneity of these repositories, this task becomes a time consuming and slightly complicated one. Additionally, the heterogeneity of these repositories contributes to the inconsistencies in the search process. These issues once more result in limiting access to OER. In contrast, OERScout is not affected by the heterogeneity of the repositories. It further promotes the decentralisation of resources. As such, content creators can opt to make available their resources via a personal blog, personal website, institutional website or even a cloud space. OERScout will allow users to easily locate these resources through its faceted search interface whereby the visibility of these resources is increased.

In sum, the OERScout technology framework provides a viable solution to the current OER search dilemma. Through the use of the Desirability framework and the faceted search approach, it allows users to locate OER which were previously invisible in the searchsacpe. We see it as a game changer in terms of widening access to desirable OER for academic purposes. The current version of the system is only available as a prototype. We intend to provide a publically accessible faceted search interface in the near future.

REFERENCES

Abeywardena, I.S., Raviraja, R., & Tham, C.Y. (2012). Conceptual Framework for Parametrically Measuring the Desirability of Open Educational Resources using D-index. *International Review of Research in Open and Distance Learning , 13* (2), 104-121.

Abeywardena, I.S., & Chan, C.S. (2013). Review of the Current OER Search Dilemma. *Proceedings of the 57th World Assembly of International Council on Education for Teaching (ICET 2013).* Nonthaburi, Thailand: ICET.

Abeywardena, I.S., Chan, C.S., & Tham, C.Y. (2013). OERScout Technology Framework: A Novel Approach to Open Educational Resources Search. *International Review of Research in Open and Distance Learning , 14* (4), 214-237.

Ben-Yitzhak, O., Golbandi, N., Har'El, N., Lempel, R., Neumann, A., Ofek-Koifman, S., et al. (2008). Beyond basic faceted search. *Proceedings of the 2008 International Conference on Web Search and Data Mining* (pp. 33-44). Palo Alto: ACM.

Dash, D., Rao, J., Megiddo, N., Ailamaki, A., & Lohman, G. (2008). Dynamic faceted search for discovery-driven analysis. *Proceedings of the 17th ACM conference on Information and knowledge management* (pp. 3-12). Napa Valley: ACM.

Dichev, C., Bhattarai, B., Clonch, C., & Dicheva, D. (2011). Towards Better Discoverability and Use of Open Content. *Proceedings of the Third International Conference on Software, Services and Semantic Technologies S3T* (pp. 195-203). Berlin Heidelberg: Springer.

Hatakka, M. (2009). Build it and they will come?–Inhibiting factors for reuse of open content in developing countries. *The Electronic Journal of Information Systems in Developing Countries , 37* (5), 1-16.

Hearst, M. (2006). Design recommendations for hierarchical faceted search interfaces. *In ACM SIGIR workshop on faceted search*, (pp. 1-5).

Koren, J., Zhang, Y., & Liu, X. (2008). Personalized interactive faceted search. *Proceedings of the 17th international conference on World Wide Web* (pp. 477-486). Beijing: ACM.

Lane, A. (2009). The impact of openness on bridging educational digital divides. *The International Review of Research in Open and Distance Learning , 10* (5).

Levey, L. (2012). Finding Relevant OER in Higher Education: A Personal Account. In J. Glennie, K. Harley, N. Butcher, & T. van Wyk (Eds.), *Open Educational Resources and Change in Higher Education: Reflections from Practice* (pp. 125-138). Vancouver: Commonwealth of Learning.

Pirkkalainen, H., & Pawlowski, J. (2010). Open Educational Resources and Social Software in Global E-Learning Settings. In P. Yliluoma (Ed.), *Sosiaalinen Verkko-oppiminen* (pp. 23-40). Naantali: IMDL.

Shelton, B. E., Duffin, J., Wang, Y., & Ball, J. (2010). Linking OpenCourseWares and Open Education Resources: Creating an Effective Search and Recommendation System. *Procedia Computer Science , 1* (2), 2865-2870.

Tunkelang, D. (2009). Faceted search. In G. Marchionini (Ed.), *SYNTHESIS LECTURES ON INFORMATION CONCEPTS, RETRIEVAL, AND SERVICES* (Vol. 5, pp. 1-80). Morgan & Claypool.

UNESCO. (2012). *Paris OER Declaration.* Paris.

Yergler, N. R. (2010). Search and Discovery: OER's Open Loop. *Proceedings of Open Ed 2010.* Barcelona.

Appendix D

Abeywardena, I.S., & Chan, C.S. (2013). Review of the Current OER Search Dilemma. *Proceedings of the 57th World Assembly of International Council on Education for Teaching (ICET 2013), Thailand.*

Review of the Current OER Search Dilemma

Ishan Sudeera Abeywardena[1]
School of Science and Technology
Wawasan Open University
54 Jalan Sultan Ahmad Shah, Penang 10050, Malaysia
ishansa@wou.edu.my

Chee Seng Chan[2]
Faculty of Computer Science and Information Technology
University of Malaya
50603, Kuala Lumpur, Malaysia
cs.chan@um.edu.my

Open Educational Resources (OER) are fast gaining traction amongst the academic community as a viable means of increasing access and equity in education. The concept of OER is of especial significance to the marginalised communities in the Global South where distance education is prominent due to the inability of conventional brick and mortar institutions to cope with the growing demand. However, the wider adoption of OER by academics in the Global South has been inhibited due to various socio, economic and technological reasons. One of the major technological inhibitors is the current inability to search for OER which are academically useful and are of an acceptable academic standard. Many technological initiatives have been proposed over the recent past to provide potential solutions to this issue. Among these are OER curartion standards such as GLOBE, federated search, social semantic search and search engines such as DiscoverEd, OCW Finder, Pearson's Project Blue Sky. The research discussed in this paper is carried out in the form of literature review and informal interviews with experts. The objective of the study is to document the extent of the OER search issues contributing to the slow uptake of the concept of OER. This review paper discusses the current OER search dilemma and the impact of some of the key initiatives which propose potential solutions.

Keywords: Open Educational Resources, OER, OER Search, OER Search Technologies

1. Introduction

With the new drive towards accessible and open information, Open Educational Resources (OER) have taken centre stage after being first adopted in a UNESCO forum in 2002. OER can be defined as "*web-based materials, offered freely and openly for use and re-use in teaching, learning and research*" (Joyce, 2007). Although many countries have, in theory, embraced the concept of OER, it is still to become mainstream academic practice due to various inhibitors. One such inhibitor is the inability to effectively search for OER which are academically useful and are of an acceptable academic standard.

With the dramatic changes taking place in Higher Education (HE) within the past 10 years, academics have had to adopt new cost effective approaches in order to provide individualised learning to a more diverse student base (Littlejohn, Falconer & Mcgill, 2008). In this context, OER has the potential to become a major source of freely reusable teaching and learning resources, especially in higher education, due to active advocacy by organisations such as UNESCO, the Commonwealth of Learning (COL), Organisation for Economic Co-operation and Development (OECD); and the International Council of Distance Education (ICDE).

Despite the fact that OER were initially limited to text based material and are still predominantly in text based formats, they are not restricted by the media types or the file types used. Many modern OER are released as images, movie clips, animations, datasets, audio clips, podcasts, among others, providing rich multimedia based material for use and reuse. These multimedia resources are made available through large repositories such as YouTube[1] (video), Flickr[2] (images) and iTunesU[3] (podcasts) under the *Creative Commons* (CC) licensing scheme.

According to McGreal (2010), modern OER repositories can be classified into three categories:

- *Content repositories* – hosts content internally within the repository.
- *Portal repositories* – provides searchable catalogues of content hosted in external repositories.
- *Content and portal repositories* – hosts content internally in addition to providing catalogues of content hosted externally.

Within these three types of repositories, Wiki, *"…a software tool that promotes and mediates discussion and joint working between different users…"* (Leuf & Cunningham, 2001), plays a central role in the present day OER arena. Projects such as WikiEducator, Wikibooks, Wikimedia Commons and Wikiversity are among the popular Wiki based OER repositories. Another widely used repository is *Rhaptos* developed by Rice University. This repository hosts the popular Connexions OER repository which allows the easy creation, use and re-use of text based learning objects (LO). The *Rhaptos* platform is currently also being used by other repositories such as Vietnam Open Educational Resources (VOER) under FOSS licenses. When considering institutional OER repositories, the popular DSPACE[4] repository systems is the most

---

[1] http://www.youtube.com/
[2] http://www.flickr.com/
[3] http://www.apple.com/education/itunes-u/
[4] http://www.dspace.org/

commonly used due to its compatibility with existing library systems and protocols. However, DSPACE only acts as a repository of content and does not provide features which facilitates reuse and remix of resources.

The attribute common to all of these repositories is the use of metadata for resource curation. These metadata are defined according to established standards such as Dublin Core Metadata Initiative (DCMI) and IEEE Learning Object Metadata (IEEE LOM). However, one of the key concerns regarding OER curation is the standardisation of metadata across repositories and ensuring the integrity of the metadata defined by content creators. The manual cataloguing of OER has also become an issue due to the human resources required to keep up with the constant expansion in OER volume. However, new technology platforms and initiatives are currently being developed which will eventually lead to viable solutions to these issues. This paper briefly introduces some promising innovations which claim to provide long term solutions to the current OER search dilemma. The rest of the paper discusses the current OER search dilemma and looks at some promising innovations currently in development.

2. The Current Dilemma

Over the recent past, many global OER initiatives have been established by organisations such as UNESCO, COL and the United Nations (UN) to name a few. Among these initiatives are the 'Education for All' initiative by the UN and World bank (Geith & Vignare, 2008), the Open e-Learning Content Observatory Services (OLCOS) (Geser, 2007), OER Africa (OER Africa, 2009), the African Virtual University (AVU) (Bateman, 2006), China's Open Resources for Education (CORE) (Downes, 2006), Japan's Open Courseware Consortium (JCW) (Fukuhara, 2008), Teacher Education for Sub-Saharan Africa (TESSA) (Moon & Wolfenden, 2007), the European educational digital library project 'Ariadne' (Duval et al., 2001), eVrest which links Francophone minority schools across Canada (Richards, 2007) and the Blended Learning Open Source Science or Math Studies Initiative (BLOSSOMS) (Larson & Murray, 2008). A great majority of these OER initiatives are based on established web based technology platforms and have accumulated large volumes of quality resources which are shared openly. However, one limitation inhibiting the wider adoption of OER is the current inability to effectively search for academically useful OER from a diversity of sources (Yergler, 2010). This limitation of locating *fit-for-purpose* (Calverley & Shephard, 2003) resources is further heightened by the disconnectedness of the vast array of OER repositories currently available online. As a result, West & Victor (2011) argue that there is no single search engine which is able to locate resources from all the OER repositories. Furthermore, according to Dichev & Dicheva (2012), one of the major barriers to the use and re-use of OER is the difficulty in finding quality OER matching a specific context as it takes an amount of time comparable with creating one's own materials. Unwin (2005) argues that the problem with open content is not the lack of available resources on the Internet but the inability to effectively locate suitable resources for academic use. The Paris OER Declaration (2012) states the need for more research in this area as *"encourage the development of user-friendly tools to locate and retrieve OER that are specific and relevant to particular needs"*. Thus, the necessity of a system which could effectively search the numerous OER repositories with the aim of locating usable materials has taken centre stage.

The most common method of OER search is generic search engines such as Google, Yahoo! or Bing (Abeywardena, Dhanarajan & Chan, 2012). Even though this method is the most commonly used, it is not the most effective as discussed by Pirkkalainen & Pawlowski (2010) who argue that "…*searching this way might be a long and painful process as most of the results are not usable for educational purposes*". As possible alternatives, many methods such as Social-Semantic Search (Piedra et al., 2011), DiscoverEd (Yergler, 2010) and OCW Finder (Shelton et al., 2010) have been introduced. Furthermore, semantic web based alternatives such as Agrotags (Balaji et al., 2010) have also been proposed which build ontologies of domain specific keywords to be used for classification of OER belonging to a particular body of knowledge. However, the creation of such ontologies for all the domains discussed within the diverse collection of OER would be next to impossible. Furthermore, Abeywardena, Raviraja & Tham (2012) state that despite all these initiatives there is still no generic methodology available at present to enable search mechanisms to autonomously gauge the usefulness of an OER taking into consideration (i) the level of openness; (ii) the level of access; and (iii) the relevance; of an OER for ones needs. As such, new innovations need to take place to address the present technological issues hampering the growth of the OER movement.

3. Some Promising Innovations

As discussed earlier, there are many research initiatives exploring various technological angles trying to provide long term solutions to the current OER search dilemma. Among these research projects, there are a few experimental or prototype initiatives which provide great promise on a global scale.

Pearson's Project Blue Sky

One of the more exciting technologies unveiled recently is the *Blue sky* project (Kolowich, 2012) by the global publishing giant Pearson. This custom search engine specifically concentrates on searching for OER with an academic focus. The platform allows instructors to search for e-book chapters, videos and online exercise software from approximately 25 OER repositories distributed worldwide. However, it gives precedence to e-book material published under Pearson. Irrespective of this possible bias towards its own products, Associate Professor David Wiley states that "*the more paths to OER there are in the world, the better*" (Kolowich, 2012).

GLOBE

Another promising initiative is the Global Learning Object Brokered Exchange (GLOBE) initiative which uses a federated search approach to solving the OER search dilemma. The GLOBE consortium, which was founded in 2004, has now grown to 14 members representing America, Asia, Australia, Europe and Africa. GLOBE acts as a central repository of IEEE LOM educational metadata harvested from various member institutional repositories. Users are provided with a single sign-on query interface where they can search for resources across repositories, platforms, institutions, languages and regions. As of February 2012 the total number of metadata harvested available through globe is 817,436 (Yamada, 2013). The consortium is currently expanding its reach to more institutions worldwide. One limitation however is the standardisation, harvesting and tagging of the constantly expanding volume of resources.

LRMI

Among the highly anticipated initiatives is the Learning Resource Metadata Initiative (LRMI) launched by the Association of Educational Publishers and *Creative Commons*. This project aims to build a common metadata vocabulary for educational resources. This *common metadata framework* is used for uniform tagging of web based learning resources. According to the official website of the project, it believes that "*Once a cricital mass of educational content has been tagged to a universal framework, it becomes much easier to parse and filter that content, opening up tremendous possibilities for search and delivery*" (http://www.lrmi.net/about retrieved May 13, 2013). The inclusion of LRMI into schema.org, a joint project by Bing, Google and Yahoo! looking at standardising metadata, is an early indication of the potential global impact.

Desirability Framework

The *desirability* of OER, proposed by Abeywardena, Raviraja & Tham (2012), is a parametric measure of the usefulness of an OER for a particular academic need. This framework provides a breakthrough in the parametric measure of the usefulness of OER by search engines taking into consideration (i) *level of openness*: the permission to use and reuse the resource; (ii) *level of access*: the technical keys required to unlock the resource; and (iii) *relevance*: the level of match between the resource and the needs of the user. By calculating the *D-index*, the measure of *desirability*, for a particular set of OER search results, search engines can better present OER which are more suitable for use and reuse in a given academic scenario. The relative simplicity of the *desirability* framework allows it to be easily incorporated into any existing OER search mechanism.

OERScout

In contrast to the large scale projects such as Blue Sky, GLOBE and LRMI, *OERScout* (Abeywardena et al., 2012) is a relatively small research project which looks at providing a solution to the OER search dilemma by autonomously generating metadata for a particular resources. The novelty and innovation of this project can be largely attributed to the clustering and text mining approaches used in the design to "read" text based OER, "understand" them and tag them using autonomously mined domain specific metadata. This approach eliminates the need for manually tagging resources with human defined metadata. Thus, *OERScout* provides a viable solution to tackle the need for increased human resources due to the exponential expansion in OER volume. *OERScout* also incorporates the *desirability* framework and a faceted search approach which allows users to quickly zero-in on the most suitable resources. Many experts believe that the technological concepts behind *OERScout* would be a game changer challenging the traditional norms of OER search.

4. Conclusion

Open Educational Resources (OER) are fast gaining traction in the academic community as a viable solution to educating the masses. However, despite the fact that many governmental, non-governmental and philanthropic organisations have heavily promoted the OER movement, it is

still to become mainstream practice in many countries and regions. One limitation hindering the spread of OER is the current dilemma with respect to OER search. Based on the literature, no search engine exists at present which has a keen focus on locating OER distributed worldwide. Providing some hope are initiatives such as Pearson's Project Blue Sky, GLOBE and LRMI which looks at solutions to this issue on a global scale. In addition, there are other ambitious research projects such as the *desirability* framework and *OERScout* which look at breaking the norms in conventional OER search to provide game changing solutions. With more and more research interests growing in this area, the future of OER seem to be positive.

Acknowledgements

References

Abeywardena, I.S., Raviraja, R., & Tham, C.Y. (2012). Conceptual Framework for Parametrically Measuring the Desirability of Open Educational Resources using D-index. *International Review of Research in Open and Distance Learning, 13(2),* 104-121.

Abeywardena, I. S., Tham, C.Y., Chan, C.S., & Balaji. V. (2012). OERScout: Autonomous Clustering of Open Educational Resources using Keyword-Document Matrix. *Proceedings of the 26th Asian Association of Open Universities Conference, Chiba, Japan.*

Abeywardena, I. S., Dhanarajan, G., & Chan, C.S. (2012). Searching and Locating OER: Barriers to the Wider Adoption of OER for Teaching in Asia. *Proceedings of the Regional Symposium on Open Educational Resources: An Asian Perspective on Policies and Practice, Penang, Malaysia.*

Balaji, V., Bhatia, M. B., Kumar, R., Neelam, L. K., Panja, S., Prabhakar, T. V., Samaddar, R., Soogareddy, B., Sylvester, A. G., & Yadav, V. (2010). Agrotags – A Tagging Scheme for Agricultural Digital Objects. *Metadata and Semantic Research Communications in Computer and Information Science 108*, 36-45.

Bateman, P. (2006). The AVU Open Educational Resources (OER) Architecture for Higher Education in Africa. *OECD Expert Meeting, Barcelona.*

Calverley, G., & Shephard, K. (2003). Assisting the uptake of on-line resources: why good learning resources are not enough. *Computers & Education*, *41*(3), 205-224.

Dichev, C., & Dicheva, D. (2012). Open Educational Resources in Computer Science Teaching. *SIGCSE'11, February 29–March 3, 2012, Raleigh, NC, USA*.

Downes, S. (2006). Models for Sustainable Open Educational Resources. *National Research Council Canada*.

Duval, E., Forte, E., Cardinaels, K., Verhoeven, B., Durm, R. V., Hendrikx, K., Forte, M. W., Ebel, N., Macowicz, M., Warkentyne, K., & Haenni, F. (2001). The ariadne knowledge pool system. *Communications of the ACM, 44(5),* 72–78.

Fukuhara, Y. (2008). Current Status of OCW in Japan. *Proceedings of the Distance Learning and the Internet Conference*.

Geith, C., & Vignare, K. (2008). Access to Education with Online Learning and Open Educational Resources: Can they close the gap?. *Jounral of Asynchonous Learning Networks, 12(1)*.

Geser, G. (2007). Open educational practices and resources: OLCOS Roadmap 2012. *Open Learning Content Observatory Services. Salzburg, Austria.*

Joyce, A. (2007). OECD Study of OER: Forum Report, *OECD*. Retrieved December 12, 2011 from http://www.unesco.org/iiep/virtualuniversity/forumsfiche.php?queryforumspages_id=33.

Kolowich, S. (2012). Pearson's Open Book. *Inside Higher ED*. Retrived May 13, 2013 from http://www.insidehighered.com/news/2012/11/05/pearson-unveils-oer-search-engine .

Larson, R. C., & Murray, E. (2008). Open Educational Resources for Blended Learning in High Schools: Overcoming Impediments in Developing Countries. *Journal for Asynchronous Learning Networks, 12(1),* 85-103.

Leuf, B., & Cunningham, W. (2001). The Wiki way: Collaboration and sharing on the internet, *Boston: Addison-Wesley Professional*.

Littlejohn, A., Falconer, I., & Mcgill, L. (2008). Characterising effective eLearning resources. *Computers & Education*, *50*(3), 757-771.

McGreal, R. (2010). Open Educational Resource Repositories: An Analysis. *Proceedings: The 3rd Annual Forum on e-Learning Excellence, 1-3 February 2010, Dubai, UAE.*

Moon, B., & Wolfenden, F. (2007). The TESSA OER experience: building sustainable models of production and user implementation**.** *Proceedings of the OpenLearn 2007 conference*.

OER Africa. (2009). The Potential of Open Educational Resources: Concept Paper by OER Africa**.** *Retrieved July 12, 2010 from http://www.oerafrica.org/SharedFiles/ResourceFiles/36158/33545/33525/2008.12.16%20OER%20and%20Licensing%20Paper.doc* .

Pirkkalainen, H., Pawlowski, J. (2010). Open Educational Resources and Social Software in Global E-Learning Settings. *In Yliluoma, P. (Ed.) Sosiaalinen Verkko-oppiminen. IMDL, Naantali, 23–40.*

Richards, G. (2007). Reward structure for participation and contribution in K-12 OER Communities. *Proceedings of the 1st Workshop on Social Information Retrieval for Technology-Enhanced Learning and Exchange.*

Shelton, B. E., Duffin, J., Wang, Y., Ball, J. (2010). Linking OpenCourseWares and Open Education Resources: Creating an Effective Search and Recommendation System. *Procedia Computer Science, 1(2), 2865-2870.*

Unwin, T. (2005). Towards a Framework for the Use of ICT in Teacher Training in Africa. *Open Learning 20*, 113-130.

West, P., Victor, L. (2011). Background and action paper on OER. *Report prepared for The William and Flora Hewlett Foundation.*

Yamada, T. (2013). Open Educational Resources in Japan. *Open Educational Resources: An Asian Perspective*. *Commonwealth of Learning and OER Asia,* 85-105.

Yergler, N. R. (2010). Search and Discovery: OER's Open Loop. *In Open Ed 2010 Proceedings: Barcelona: UOC, OU, BYU.*

2012 Paris OER Declaration.

About the LRMI. *Retrieved May 13, 2013 from* http://www.lrmi.net/about.

Appendix E

Abeywardena, I. S., Tham, C.Y., Chan, C.S., & Balaji. V. (2012). OERScout: Autonomous Clustering of Open Educational Resources using Keyword-Document Matrix. *Proceedings of the 26th Asian Association of Open Universities Conference, Chiba, Japan.*

OERScout: Autonomous Clustering of Open Educational Resources using Keyword-Document Matrix

Ishan Sudeera Abeywardena[a], Choy Yoong Tham[b], Chee Seng Chan[c] and Venkataraman Balaji[d]

[ab] School of Science and Technology, Wawasan Open University, 54 Jalan Sultan Ahmad Shah, Penang, 10050, Malaysia. e-mail: [a] ishansa@wou.edu.my , [b] cytham@wou.edu.my

[c] Faculty of Computer Science and Information Technology, University of Malaya, 50603, Kuala Lumpur, Malaysia. e-mail: cs.chan@um.edu.my

[d] Commonwealth of Learning (COL), 1055 West Hastings Street, Suite 1200, Vancouver, BC V6E 2E9, Canada. e-mail: vbalaji@col.org

Sub-theme:
Open Educational Resources (OER) and ODL

Abstract

The Open Educational Resources (OER) movement has gained momentum in the past few years. With this new drive towards making knowledge open and accessible, a large number of OER repositories have been established and made available online throughout the globe. However, despite the fact that these repositories hold a large number of high quality material, the use and re-use of OER has not taken off as anticipated due to various geographic, socio and technological limitations. One such technological limitation is the present day inability to effectively search and locate OER materials which are specific and relevant to a particular academic domain. As a first step towards a possible solution to this issue, this paper discusses the design and development of a clustering algorithm which accurately clusters text based OER materials by building a Keyword-Document Matrix (KDM) using autonomously identified domain specific keywords. This algorithm is the first phase of a larger technology framework named "OERScout" which is a new methodology for effectively searching and locating desirable OER for academic use.

Keywords: OERScout, Open Educational Resources, OER, OER searching and location, Text mining algorithms, Document clustering, Autonomous keyword identification

1 Introduction

With the new drive towards accessible and open information, Open Educational Resources (OER) have taken centre stage after being first adopted in a UNESCO forum in 2002. OER can be defined as "*web-based materials, offered freely and openly for use and re-use in teaching, learning and research*" (Joyce, 2007) which are heavily dependent on technology and the internet to be accessible by the masses. According to Farber (2009) "*Just as the Linux operating system and other open source software has become a pervasive computer technology around the world, so too might OER materials become the basis for training the global masses*" which clearly outlines the significance of OER as a global movement. The move towards OER has also helped reduce significantly the costs of production, reproduction and distribution of course material (Caswell, Henson, Jenson & Wiley, 2008) especially as initiatives such as MIT OpenCourseWare (OCW), Rice University Connexions and the Commonwealth of Learning (COL) funded Wikieducator project are sharing high quality educational resources under the Creative Commons (CC) license which enables institutions and individuals globally to adapt and re-use material without developing them from scratch. This is especially important for countries in the Global South such as India which has 411 million potential students, out of which only 234 million enter school at all, less than 20% reach high school and less than 10% graduate (Kumar, 2009).

Over the recent past, many global OER initiatives have been established by organisations such as UNESCO, COL and the United Nations (UN) to name a few. Many of these initiatives are based on established web based technology platforms and have accumulated large volumes of quality resources which are shared with the masses. However, the use of diverse and disparate technology platforms in these projects entails the inability to effectively trawl and located OER using generic search methodologies. This is affirmed by Abeywardena, Raviraja and Tham (2012) who state that there is no generic methodology available at present to enable search mechanisms to autonomously gauge the *desirability* of an OER which is a function of (i) the level of openness; (ii) the level of access; and (iii) the relevance; of an OER for ones needs. Thus, the necessity for a methodology which could effectively trawl and search the numerous disconnected and disparate OER repositories with the aim of locating *desirable* materials has taken center stage as the problems with open content is not the lack of available resources on the Internet but the inability to locate suitable resources for academic use (Unwin, 2005).

OERScout is a technology framework which aims to accurately cluster text based OER by building a Keyword-Document Matrix (KDM) using autonomously mined domain specific keywords. Using the KDM, the system accurately generates lists of specific and relevant OER from the distributed repositories to suit a given search query. In this context, *specific* denotes the suitability of an OER for a particular teaching need. For example, an OER on physics from the final year syllabus of a physics degree would not be suitable for a high school physics class. *Relevant* denotes the match between the content of the OER and the content needed for a particular teaching need. For example, physical chemistry is not relevant for a teaching need in organic chemistry. This paper, which is organised under the headings *methodology*, *pilot tests*, *discussion* and *conclusion*; discusses how *OERScout* benefits the Open Distance Learning (ODL) community, who are arguably the largest group of OER creators and consumers (Abeywardena, 2012), by providing a centralised system for effectively searching and locating specific and

relevant OER materials from the disconnected and disparate repositories scattered across the globe.

2 Methodology

The *OERScout* text mining algorithm is designed to "read" text based OER and "learn" which academic domain(s) and sub-domain(s) they belonged to. To achieve this, a *bag-of-words* approach is used due to its effectiveness when used with unstructured data (Feldman & Sanger, 2006). The algorithm extracts all the individual words from a particular document by removing noise such as formatting and punctuations to form the *corpus*. The *corpus* is then *Tokenised* into the *List of Terms* using the *stop words* found in the Onix Text Retrieval Toolkit[1]. The extraction of the content describing terms from the *List of Terms* for the formation of the *Term Document Matrix* (TDM) is done using the *Term Frequency–Inverse Document Frequency* (TF-IDF) weighting scheme. The weight of each term (TF-IDF) was calculated using the following formula (Feldman & Sanger, 2006):

$$(TF\text{-}IDF)_t = TF_t \text{ x } IDF_t$$

TF_t denotes the frequency of a term *t* in a single document. IDF_t denotes the frequency of a term *t* in all the documents in the collection [$IDF_t =$ Log (N/TF_t)] where *N* is the number of documents in the collection. The probability of a term *t* being able to accurately describe the content of a particular OER as a keyword decreases with the number of times it occurs in other related and non-related materials. For example the term "introduction" would be found in many OER which discuss a variety of subject matter. As such the TF-IDF of the term "introduction" would be low compared to a term such as "operating systems" or "statistical methods" which are more likely to be keywords. As the TF-IDF weighting scheme takes the inverse document frequency into consideration, it was found to be suitable for extracting the keywords from an OER.

The Keyword-Document Matrix (KDM), which is a subset of the TDM, is created for the *OERScout* system by matching the autonomously identified keywords against the documents. The formation of the KDM is done by (i) normalising the TF-IDF values for the terms in the TDM; and (ii) applying the Pareto principle (80:20) where the top 20% of the TF-IDF values are considered to be keywords describing 80% of the OER (Figure 1).
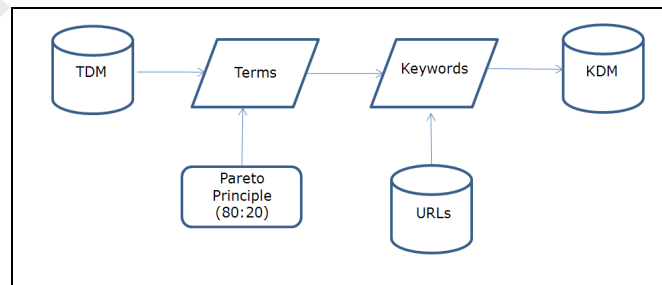
Figure 1 Creation of the KDM

---

[1] lextek.com/manuals/onix/stopwords1.html

The *OERScout* algorithm is implemented using the Microsoft Visual Basic.NET 2010 (VB.NET 2010) programming language. The corpus, *List of Terms*, TDM and KDM are implemented using the Microsoft SQLServer 2008 database platform. The OER resources are fed into the system using sitemaps based on extensible markup language (xml) which contain the uniform resource locators (URLs) of the resources.

3 Pilot Tests

Two pilot tests were conducted to test the functionality of the system. As the first test case, the Rice University's OER repository Connexions[2] was used due to (i) the large number of diverse OER materials available; (ii) the relatively high popularity and usage rates; and (iii) the availability of the OER materials in text format. An xml sitemap containing 1238 URLs belonging to the domains of arts, business, humanities, mathematics and statistics; science and technology; and social sciences was created as the initial input. The system was run with the initial input and was allowed to autonomously create the KDM. The average time taken for *OERScout* to extract terms from an OER and update the KDM was found to be approximately two minutes as the OER were in HTML format. After the completion of the pilot test, the system had created 1013 clusters in the KDM with an average density of 1.23 resources per cluster. It was also noted that 1238 resources had contributed 141901 new terms. An example of the KDM is shown in Figure 2.
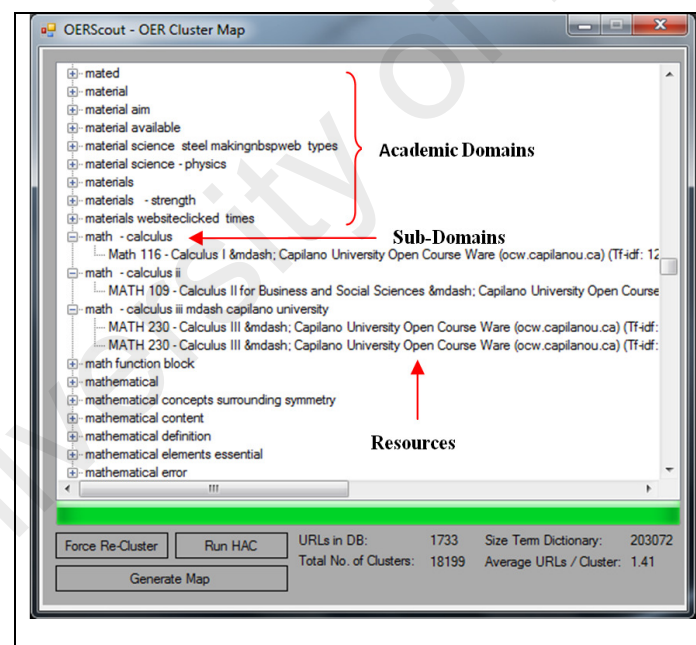
Figure 2 Example of the cluster map generated using the KDM

The second test was conducted on the Directory of Open Educational Resources (DOER)[3] of the COL. DOER is a fledgling *portal OER repository* (McGreal, 2010) which provides an easily navigable central catalogue of OER scattered across the globe. At present the OER available through DOER are manually classified into 20 main categories and 1158 sub-categories.

---

[2] http://www.cnx.org
[3] http://doer.col.org/

However, despite covering most of the major subject categories, this particular ontology would need to expand by a large degree due to the unlimited variety of OER available in a kaleidoscope of subject areas. This expansion, in turn, becomes a tedious and laborious task which needs to be accomplished manually on an ongoing basis. As a possible solution to this issue, a mechanism was needed for autonomously identifying the subject area(s) covered in a particular OER, in the form of keywords, in order for it to be accurately catalogued. Given this requirement DOER was used as the training dataset for the second pilot test of *OERScout*. This training process was critical to the functioning of the algorithm as it had to learn a large array of academic domains and sub-domains before being able to accurately cluster resources according to the domain. After completion of the second test, the system had processed 2598 resources of file types HTML, PDF, TEXT and MS Word from a multitude of OER repositories. On average, each resource required approximately 15-90 minutes to be read and learnt by the system due to the size and the format of the documents. The creation of the KDM required approximately 12-24 hours each time.

3 Discussion

Generic search methodologies such as Google, Yahoo! and Bing are the most widely used search mechanisms for locating OER (Abeywardena & Dhanarajan, 2012). Even though this method is the most commonly used, it is not the most effective as discussed by Pirkkalainen and Pawlowski (2010) who argue that "*searching this way might be a long and painful process as most of the results are not usable for educational purposes*". Despite semantic web based alternatives such as Agrotags (Balaji et al., 2010) which build ontologies of domain specific keywords to be used for classification of OER belonging to a particular body of knowledge, the creation of such ontologies for all the domains discussed within the diverse collection of OER would be next to impossible. As such, the *OERScout* system was developed to use clustering techniques instead of semantic web techniques to enable OER to be clustered based on autonomously identified keywords.
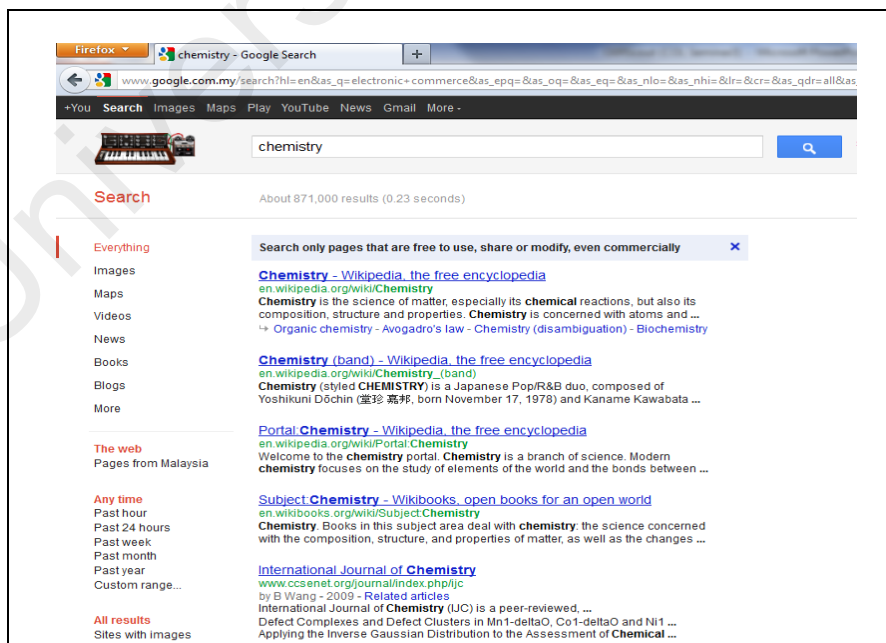
Figure 3 Google "Advanced Search" results for OER on Chemistry (24th May 2012)

Figure 3 shows an advanced search conducted on Google[4] for the term "chemistry" specifically searching for resources which are *free to use, share or modify, even commercially*. This example confirms the statements made in literature as the first three results are from Wikipedia[5] which is an encyclopedia of user created learning objects rather than a repository of pedagogically sound educational material. Furthermore, the fifth result is a non-OER source. According to Vaughan (2004) users will only consider the top ten ranked results for a particular search as the most relevant. Vaughan further suggests that the users will ignore the results below the top 30 ranks. As such, generic search methodologies such as Google are currently inapt at locating specific and relevant OER for a particular teaching need.

Figure 4 shows a search result for the term "chemistry" on *OERScout* conducted on the KDM created during the second pilot test. Contrary to the static list of search results produced by typical search engines, *OERScout* provides an autonomously identified dynamic list of *Suggested Topics* which are related to "chemistry". The user is then able to click on any of the suggested topics to access specific and relevant OER, identified in the KDM, from all the repositories indexed by *OERScout*. Furthermore, based on the selection by the user, the system will provide a list of *Related Topics* which will enable the user to drill down further to identify the most suitable OER for his/her teaching needs. As such, it can be seen that *OERScout* is a centralised system which is much more dynamic and effective in locating specific and relevant OER from the disconnected and disparate repositories. This becomes one of the major benefits to ODL practitioners as the system spares the user from conducting countless keyword searches in the OER repositories in order to identify suitable material for use. It also allows content creators to quickly isolate the OER suitable for their needs without reading through all the search results returned by a typical search mechanism such as Google.
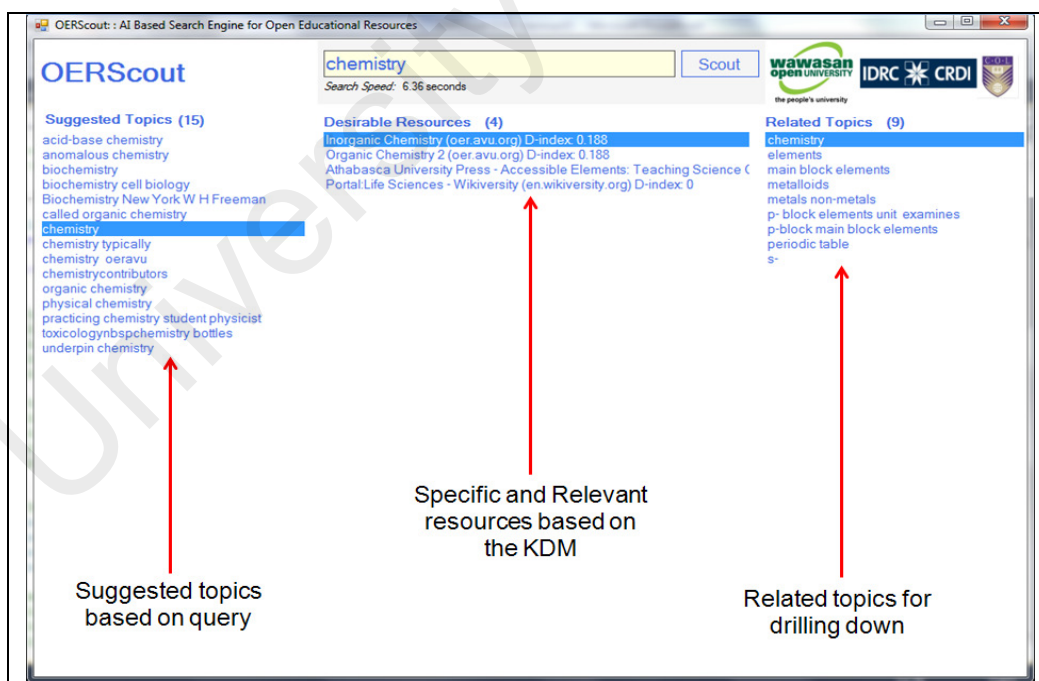
Figure 4 OERScout search result for OER in Chemistry

[4] http://www.google.com.my
[5] http://www.wikipedia.org

This first version of *OERScout* is unable to cluster non-text based materials such as audio, video and animations which is a major drawback considering the fact that more and more OER are now being developed in multimedia formats. However, it was noted from the pilot tests that the system will accurately cluster multimedia based material using the text based descriptions provided. Another limitation is its inability to cluster resources written in languages other than English. Despite this current limitation, the *OERScout* algorithm has a level of abstraction which allows it to be customised to suit other languages in the future.

4 Conclusion

Open Educational Resources (OER) is a phenomenon which is rapidly gaining acceptance and credibility in the academic community as a potent tool for teaching and learning. With more and more OER repositories mushrooming across the globe and with the expansion of existing repositories due to increased contributions, the task of searching and locating specific and relevant OER has become a daunting one. This is further heightened due to the disconnectedness and disparity among the various OER repositories which are based on a number of technological platforms. Another hurdle to the searching and location of OER is the inability of current mainstream search technologies to effectively locate OER material for academic use. As such, each OER repository has to be searched using its own native search methodologies in order to locate the necessary OER. This again has had a discouraging effect on the OER practitioner as the number of repositories available is substantial and growing.

OERScout is a text mining algorithm used for clustering OER using autonomously mined domain specific keywords. It was developed with a view of providing OER creators and users a centralised system which will enable effective searching and location of specific and relevant OER for academic use. The benefits of *OERScout* to the content creators include (i) elimination of the need for manually defining content domains for categorisation in the form of metadata; (ii) elimination of the need for publicising the availability of a repository and the need for building custom search mechanisms for them; and (iii) more visibility and reach of material to a wider audience. The system benefits OER users by (i) providing a central location for finding resources scattered across the globe hidden in high volume repositories; and (ii) locating only the most specific and relevant resources. The ultimate benefit of *OERScout* is that both content creators and users now only need to concentrate on the actual content and not the searching and location of specific and relevant OER.

The next version of *OERScout* will enable ODL practitioners to effectively locate the most *desirable* OER for academic use based on parametric measures of (i) *openness* calculated using the Creative Commons license; (ii) *accessibility* calculated using the accessibility of the file format; and (iii) *relevance* calculated using the KDM.

Acknowledgements

References

Abeywardena, I. S. (2012). A report on the Re-use and Adaptation of Open Educational Resources (OER): An Exploration of Technologies Available. *Commonwealth of Learning,* 51. Retrieved August 11, 2012 from
 http://www.col.org/resources/publications/Pages/detail.aspx?PID=411.

Abeywardena, I. S., & Dhanarajan, G. (2012). OER in Asia Pacific: Trends and Issues. *Keynote address of the Policy Forum for Asia and the Pacific: Open Education Resources organised by UNESCO Bangkok and Commonwealth of Learning (COL), 23rd April 2012, Thailand.* Report available at http://www.unescobkk.org/education/ict/online-resources/databases/ict-in-education-database/item/article/oer-in-asia-trends-and-issues/.

Abeywardena, I.S., Raviraja, R., & Tham, C.Y. (2012). Conceptual Framework for Parametrically Measuring the Desirability of Open Educational Resources using D-index. *International Review of Research in Open and Distance Learning, 13(2),* 104-121.

Balaji, V., Bhatia, M. B., Kumar, R., Neelam, L. K., Panja, S., Prabhakar, T. V., Samaddar, R., Soogareddy, B., Sylvester, A. G., & Yadav, V. (2010). Agrotags – A Tagging Scheme for Agricultural Digital Objects. *Metadata and Semantic Research Communications in Computer and Information Science 108*, 36-45.

Caswell, T., Henson, S., Jenson, M., & Wiley, D. (2008). Open Educational Resources: Enabling universal education. *International Review of Research in Open and Distance Learning 9*(1), 1-11.

Farber, R. (2009). Probing OER's huge potential [Electronic Version]. *Scientific Computing 26*(1), 29-29.

Feldman, R., & Sanger, J. (2006). The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. *Cambridge University Press*.

Joyce, A. (2007). OECD Study of OER: Forum Report, *OECD*. Retrieved December 12, 2011 from http://www.unesco.org/iiep/virtualuniversity/forumsfiche.php?queryforumspages_id=33.

Kumar, M. S. V. (2009). Open educational resources in India's national development. *Open Learning: The Journal of Open and Distance Learning 24*(1), 77-84.

McGreal, R. (2010). Open Educational Resource Repositories: An Analysis. *Proceedings: The 3rd Annual Forum on e-Learning Excellence, 1-3 February 2010, Dubai, UAE, Retrieved December 27, 2011 from http://elexforum.hbmeu.ac.ae/Proceeding/PDF/Open%20Educational%20Resource.pdf*.

Pirkkalainen, H., Pawlowski, J. (2010). Open Educational Resources and Social Software in Global E-Learning Settings. *In Yliluoma, P. (Ed.) Sosiaalinen Verkko-oppiminen. IMDL, Naantali, 23–40*.

Unwin, T. (2005). Towards a Framework for the Use of ICT in Teacher Training in Africa. *Open Learning 20*, 113-130.

Vaughan, L. (2004). New measurements for search engine evaluation proposed and tested. *Information Processing and Management 40*, 677–691.

Wolfenden, F. (2008). The TESSA OER Experience: Building sustainable models of production and user implementation. *Journal of Interactive Media in Education*. Retrieved December 9, 2011 from http://jime.open.ac.uk/2008/03/.

Onix Text Retrieval Toolkit API Reference. Retrieved December 19, 2011 from http://www.lextek.com/manuals/onix/stopwords1.html.

Appendix F

Abeywardena, I. S., Dhanarajan, G., & Chan, C.S. (2012). Searching and Locating OER: Barriers to the Wider Adoption of OER for Teaching in Asia. *Proceedings of the Regional Symposium on Open Educational Resources: An Asian Perspective on Policies and Practice, Penang, Malaysia.*

SEARCHING AND LOCATING OER: BARRIERS TO THE WIDER ADOPTION OF OER FOR TEACHING IN ASIA

Ishan Sudeera Abeywardena[a], Gajaraj Dhanarajan[b] and Chee Seng Chan[c]

[a]*School of Science and Technology, Wawasan Open University, 54 Jalan Sultan Ahmad Shah, Penang, 10050, Malaysia. e-mail: ishansa@wou.edu.my*

[b]*Institute for Research and Innovation, Wawasan Open University, 54 Jalan Sultan Ahmad Shah, Penang, 10050, Malaysia.*

[c] *Faculty of Computer Science and Information Technology, University of Malaya, 50603, Kuala Lumpur, Malaysia.*

Abstract

Open Educational Resources (OER) are fast becoming a global phenomenon which could potentially provide free access to knowledge for the masses. Since the inception of this concept, governmental and non-governmental grants alongside generous philanthropy have given rise to a vast array of OER repositories all over the world. With this movement gaining momentum, more and more of the learned community have started contributing resources to these OER repositories making them grow exponentially rich in knowledge. However, despite the availability of a large number of OER repositories, the use and re-use of OER are yet to become mainstream in many regions and institutions. One reason for this slow uptake is the inability to effectively search and locate desirable OER using the available search methodologies as it would be next to impossible to trawl through all the disconnected and disparate repositories manually. The findings discussed in this paper are part of a broader study into the OER landscape in the Asian region concentrating mainly on China, Hong Kong, India, Indonesia, Japan, South Korea, Malaysia, Philippines and Vietnam where close to five hundred and eighty academics from public, private not-for-profit and private for-profit institutions participated. This research paper discusses how Asia fares with respect to searching and locating desirable OER and whether it is truly a barrier to the wider adoption of OER for teaching in the region.

Keywords: Desirability of OER, Open Educational Resources, OER, Searching and Locating OER, OER in Asia, Barriers to OER.

1 Background

The Open Educational Resources (OER) movement has gained much momentum recently as a relatively new global phenomenon which is capable of bridging the knowledge divide. With increased funding and advocacy by governmental and non-governmental organisations coupled with generous philanthropy, OER are fast becoming mainstream in many academic circles. However, even though the number of OER repositories has grown exponentially over the years, boasting rich archives of quality OER in various disciplines, the wider adoption of OER in teaching still remains low especially in the Asian region where the necessity for OER is much higher.

One limitation inhibiting the wider adoption of OER is the current inability to effectively search and locate relevant and usable OER from a diversity of sources (Yergler, 2010). This inability is further heightened by the disconnectedness and disparateness of the vast array of OER repositories currently available online as no single search engine is still able to locate resources from all the OER repositories (West and Victor, 2011). According to Dichev and Dicheva (2012) one of the major barriers to the use and re-use of OER is the difficulty of finding quality OER matching a specific context as it takes an amount of time comparable with creating one's own materials.

The most common method for searching and locating OER is to use generic search engines such as Google, Yahoo! or Bing. Even though this method is the most commonly used, it is not the most effective as discussed by Pirkkalainen and Pawlowski (2010) who argue that "*searching this way might be a long and painful process as most of the results are not usable for educational purposes*". As possible alternatives to this method, many methods such as Social-Semantic Search (Piedra et al., 2010), DiscoverEd (Yergler, 2010) and OCW Finder (Shelton et al., 2010) have been introduced. However, Abeywardena, Raviraja and Tham (2012) state that despite all these initiatives there is still no generic methodology available at present to enable search mechanisms to autonomously gauge the *desirability* of an OER which is a function of (i) the level of openness; (ii) the level of access; and (iii) the relevance; of an OER for ones needs.

Knowing the issue of the inability to search and locate *desirable* OER, this research paper discusses how this inability is affecting the wider adoption of the use and re-use of OER in the Asian region and presents a set of recommendations which would improve the effectiveness of the search and location of specific, relevant and quality OER. The paper is structured into four key sections under the headings methodology, findings, discussion and recommendations.

2 Methodology

A regional group of researchers (collaborators) from China, Hong Kong, India, Indonesia, Japan, South Korea, Malaysia, Philippines and Vietnam, who are currently active in the OER arena, jointly developed a survey instrument consisting of seventy nine independent items which would be used to elicit an understanding of the OER landscape in the Asian region with respect to (i) the use of digital resources; (ii) the use of OER; and (iii) the understanding of copyright from both an individual as well as an institutional perspective. The survey was conducted using hardcopies and an online version over a period of twelve months by the collaborators where approximately five hundred and eighty responses were gathered from academics who has had some exposure to the concept of OER. The responses were then consolidated and split into two cohorts according to (i) individuals who have experience in OER; and (ii) competent authorities of institutions who can comment holistically on the institution's practice of OER. The resulting data was analysed using the open source statistical analysis software package PSPP and was published by Abeywardena and Dhanarajan (2012). The findings discussed in this research paper are part of the first cohort which concentrated on the individuals' perspective.

3 Findings

For the purposes of this particular research paper, the analysis of the data only concentrates on four hundred and twenty responses (N=420) from eleven countries which represent the various Asian regions as shown in Figure 1.
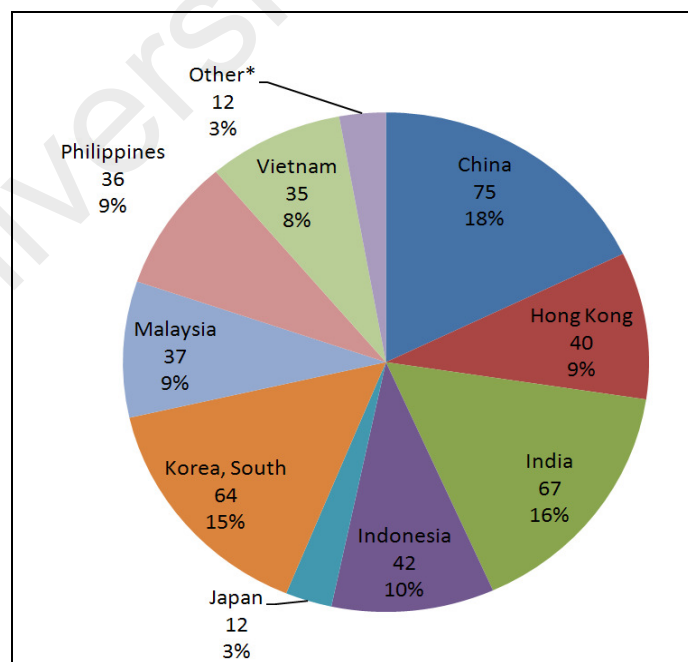
Figure 1 Participant profile

The cohort comprises of academics from 312 (74.30%) public, 63 (15%) private not-for-profit and 45 (10.7%) private for-profit institutions. The extent of the use of OER by the participants in their teaching is shown in Figure 2 and their attitudes towards using OER in their teaching are highlighted in Table 1.
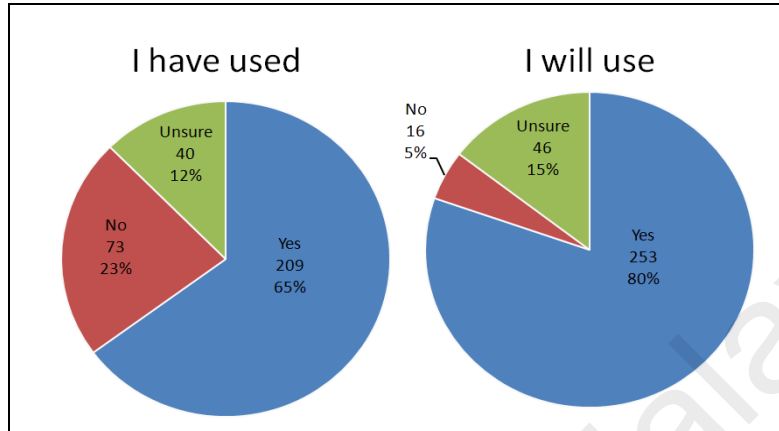
Figure 2 Use of OER in teaching

Table 1 Attitudes towards using OER in teaching

	Agree	Disagree	Neutral	N
Reusing OER is a useful way of developing new courses	77% (240)	3.5% (11)	19.5% (61)	100% (312)
Exploring the available OER worldwide will enhance my teaching and raise standards across the University	79.8% (249)	1.9% (6)	18.3% (57)	100% (312)

To understand the OER downloading habits of the participants, they were asked whether they predominantly download OER from OER repositories or whether they freely download them from the internet using search engines (Figure 3).
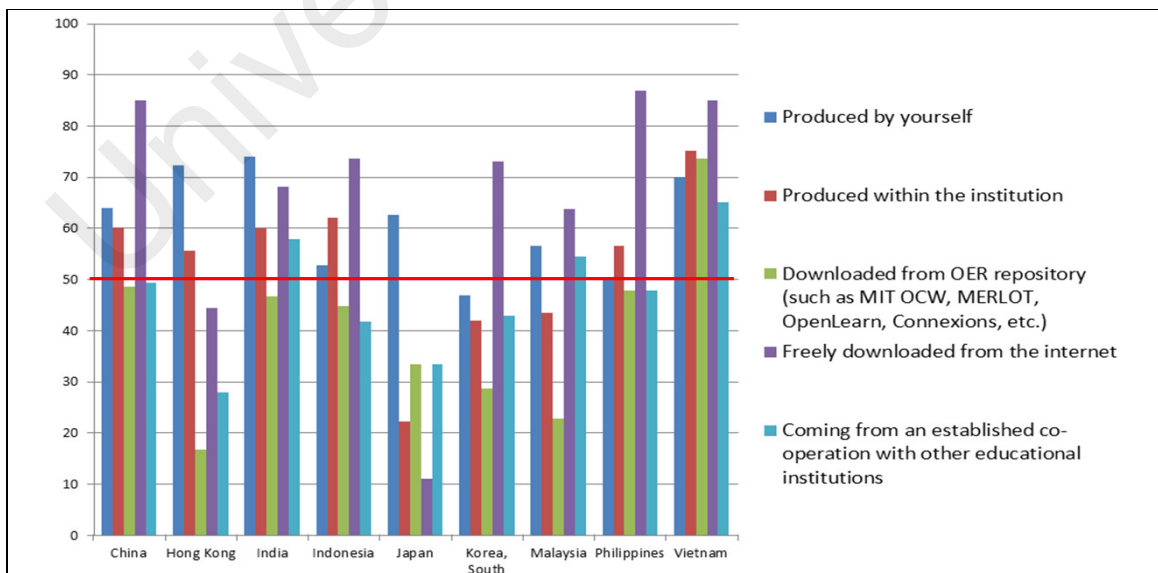
Figure 3 OER downloading habits

4

Table 2 shows the extent of use of the available search methodologies for locating OER according to the respondents who have used OER in their teaching before (Figure 2). This cohort also mentioned that they locate OER through other means such as by word of mouth from colleagues, through Wikipedia and through face-to-face networking in addition to the common methodologies mentioned in the survey instrument.

Table 2 Extent of use of available search methodologies for locating OER

	Use less	Use more	N
Generic search engines such as Google, Yahoo, Bing etc.	3.1% (6)	96.9% (189)	100% (195)
Specific search engines such as Google Scholar	31.1% (60)	68.9% (133)	100% (193)
Wikieducator Search facilities	51.8% (99)	48.2% (92)	100% (191)
Specific search facilities of OER repositories such as OCW, Connexions etc.	56.8% (108)	43.2% (82)	100% (190)
Any other methods for locating OER	66.7% (50)	33.3% (25)	100% (75)

When asked what barriers they consider to be significant to the use of OER, 64% of the participants who had used OER before in their teaching mentioned that the lack of awareness of the university OER repository and other OER repositories was a major barrier. 56.6% of the same cohort mentioned that the relevance of the available OER to their teaching is also one of the barriers for wider use of OER.

Table 3 shows how the participants felt with respect to the lack of ability to locate specific, relevant and quality OER for teaching. In this context (i) *specific* denotes the suitability of an OER for a particular teaching need. For example, an OER on physics from the final year syllabus of a physics degree would not be suitable for a high school physics class; (ii) *relevant* denotes the match between the content of the OER and the content needed for a particular teaching need. For example, physical chemistry is not relevant for a teaching need in organic chemistry; and (iii) *quality* denotes perceived academic standard of an OER for a particular teaching need.

Table 3 The importance of locating specific, relevant and quality OER for teaching

	Unimportant	Important	Neutral	N
Lack of ability to locate specific and relevant OER for my teaching	20.5% (63)	57.4% (176)	22.1% (68)	100% (307)
Lack of ability to locate quality OER for my teaching	13.8% (42)	67.6% (207)	18.6% (57)	100% (306)

4 Discussion

This research paper is underpinned by the hypothesis that the inability to effectively search and locate desirable OER using current technologies is posing a barrier to the adoption of OER for teaching in the Asian region. The nine countries identified in Figure 1 are representative of the majority of sub-regions in Asia (Table 4).

Table 4 Representation of Asian sub-regions

	Country	Region
01	China	East Asia
02	Japan	
03	Hong Kong	
04	South Korea	
05	Malaysia	South East Asia
06	Philippines	
07	Indonesia	
08	Vietnam	
09	India	South Asia

Out of the academics who had participated in the survey, 65% had used OER from other academics in their teaching and 80% mentioned that they will use OER in their teaching in the future. This shows that the use of OER is gaining popularity and wider acceptance in the Asian region. Additionally, referring to Table 1, the attitudes towards the use of OER is also taking a positive turn as 77% of the participants found OER to be a useful way of developing courses while 79.8% agreed that OER will improve the standard of their teaching. However, even though the use of OER and the attitudes towards it are improving, 57.4% of the academics found that the lack of ability to locate specific and relevant OER was an important inhibitor towards the use of OER. Furthermore, as shown in Table 3, 67.6% of the academics felt that the lack of ability to locate quality OER was also an issue worth consideration.

In order to identify the reason behind academics not being able to locate desirable OER for their teaching, the mode of searching and locating OER needs to be scrutinised. Looking at Figure 3, it is apparent that most of the time academics search and locate OER which are freely available on the internet as opposed to using specific OER repositories which maintain a certain level of quality. Furthermore, these repositories are equipped with native search mechanisms which facilitate the location of more specific and relevant OER for a particular teaching need. However, as shown in Table 2, only 43.2% of the academics use specific search facilities of OER repositories. Therefore, the lack of use of dedicated OER repositories and their tailored search mechanisms for locating OER has indeed become an inhibitor with respect to searching and location of specific, relevant and quality OER. 64% of the same cohort mentioned that the lack of awareness of the existence of such repositories was the key contributor to this current situation.

Looking at Table 2, it can be seen that generic search engines such as Google, Yahoo! and Bing are used almost all the time for searching and locating OER compared with the specific search mechanisms such as Google Scholar or the native search mechanisms of OER repositories. From this comparison, it is apparent that many academics depend on generic search mechanisms to locate the required OER for their teaching purposes. However, the inability of these generic mechanisms to locate desirable OER for a particular teaching need, as highlighted in literature, has in fact become an inhibitor to the wider adoption of OER for teaching in Asia.

5 Conclusions and Recommendations

Open Educational Resources (OER) are fast becoming a global movement which could potentially bridge the knowledge divide between the masses. Even though there are a large number of rich OER repositories located across the globe, the uptake with respect to use and re-use of OER in teaching has been slow due to a number of reasons. One such reason is the current inability to effectively search and locate specific, relevant and quality OER from the various disconnected and disparate OER repositories. With the rapid mushrooming of new OER repositories and the expansion of the existing, it has become highly infeasible to manually trawl each repository to identify OER required for specific teaching purposes. As such, this limitation has become an inhibitor to wider adoption of OER especially in the Asian region.

When considering the technological limitations, the inability of mainstream searching mechanisms, such as online search engines, to accurately distinguish between an OER and a non-OER material becomes a major hurdle. Although one might argue that the most popular search engines do provide the advanced facilities to define various filter criteria which would refine the searches, these search engines are not tailored to easily and effectively locate OER material which are the most suitable for a specific purpose. As such the OER consumers will need to resort to frequenting the more popular OER repositories such as Rice Connexions, MIT OCW or Wikieducator to search for the OER material they are after. However, this too has become a cumbersome and time consuming task as the number of repositories and the volume of each repository keeps on expanding. Thus it becomes an infeasible affair to keep track of all the OER repositories available. Also, users would be spending quite a number of hours on these popular but disconnected OER repositories conducting multiple searches using the native search mechanisms; and by so doing limit the scope as well as the variety of OER material available to them. Ultimately, even though many of these popular OER repositories hold a rich selection of material, the user is stuck in a scenario where the use of these materials is not a choice but a lack of options.

Another factor inhibiting the effective searching and location of specific, relevant and quality OER is the disparateness and disconnectedness of present day OER archives. Within the context of parametric web based searching mechanisms, the terms *specific*, *relevant* and *quality* denote key parameters which need to be considered seriously. *Specific* refers to the uniqueness of a piece of information which is returned as a result of an online search. This parameter is important with respect to ensuring that only a minimum number of instances of a piece of OER material are presented to the user. The term *relevant* refers to the standardisation of metadata which will facilitate more accurate searches. *Quality* stands for the desirability of OER material. As such, the disparateness and the disconnectedness of OER repositories can be broadly attributed to (i) the lack of adoption of a standardised method for defining metadata; (ii) the lack of a centralised search mechanism which will identify and locate OER from all of these disconnected repositories; and (ii) the inability to indicate the desirability of an OER returned as a search result.

Considering the lack of a standardised method for defining metadata for OER, it can be argued that the definition of metadata cannot be made one hundred percent accurate or uniform for all OER resources if done by the creator(s) of the resource. Therefore the use of human defined metadata in performing objective searches becomes subjective and inaccurate. A possible solution to overcome this inaccuracy and to ensure uniformity of metadata would be to utilise a computer based methodology which would consider the content, domain and locality of the OER material, among others, for autonomously defining uniform metadata.

The authors are currently involved in a pilot project named "OERScout" which uses artificial intelligence (AI) techniques combined with text mining algorithms to cluster OER from the various disconnected and disparate repositories by autonomously identifying keywords which best describe the content of the OER. This system looks at categorising all the OER from the repositories with an aim to providing accurate recommendations of desirable OER based on a particular curriculum provided by an academic.

Acknowledgements

References

Abeywardena, I. S., & Dhanarajan, G. (2012). OER in Asia Pacific: Trends and Issues. *Keynote address of the Policy Forum for Asia and the Pacific: Open Education Resources organised by UNESCO Bangkok and Commonwealth of Learning (COL), 23rd April 2012, Thailand.* Report available at http://www.unescobkk.org/education/ict/online-resources/databases/ict-in-education-database/item/article/oer-in-asia-trends-and-issues/

Abeywardena, I.S., Raviraja, R., & Tham, C.Y. (2012). Conceptual Framework for Parametrically Measuring the Desirability of Open Educational Resources using D-index. *International Review of Research in Open and Distance Learning, 13(2),* 104-121

Dichev, C., Dicheva, D. (2012). Open Educational Resources in Computer Science Teaching. *SIGCSE'11, February 29–March 3, 2012, Raleigh, NC, USA. Retrieved December 25, 2011 from* http://myweb.wssu.edu/dichevc/Research/SIGCSE2012_DichevDicheva.pdf.

Piedra, N., Chicaiza, J., López, J., Tovar, E., Martinez, O. (2011). Finding OERs with Social-Semantic Search. *Proceedings: 2011 IEEE Global Engineering Education Conference (EDUCON), April 4 - 6, 2010, Amman, Jordan. Retrieved December 25, 2011 from* http://www.psut.edu.jo/sites/EDUCON/program/contribution1482_b.pdf.

Pirkkalainen, H., Pawlowski, J. (2010). Open Educational Resources and Social Software in Global E-Learning Settings. *In Yliluoma, P. (Ed.) Sosiaalinen Verkko-oppiminen. IMDL, Naantali, 23–40.*

Shelton, B. E., Duffin, J., Wang, Y., Ball, J. (2010). Linking OpenCourseWares and Open Education Resources: Creating an Effective Search and Recommendation System. *Procedia Computer Science, 1(2), 2865-2870.*

West, P., Victor, L. (2011). Background and action paper on OER. *Report prepared for The William and Flora Hewlett Foundation. Retrieved December 25, 2011 from* http://www.oerknowledgecloud.com/sites/oerknowledgecloud.com/files/Background_and_action_paper_on_OER.pdf.

Yergler, N. R. (2010). Search and Discovery: OER's Open Loop. *In Open Ed 2010 Proceedings: Barcelona: UOC, OU, BYU. Retrieved December 25, 2011 from* http://hdl.handle.net/10609/4852 .

Appendix G

Abeywardena, I. S., Dhanarajan, G., & Lim, C.K. (2013). Open Educational Resources in Malaysia. In G. Dhanarajan & D. Porter (Eds.), *Open Educational Resources: An Asian Perspective*. *Commonwealth of Learning and OER Asia (ISBN 978-1-894975-61-2),* 119-132.

Open Educational Resources in Malaysia

Ishan Sudeera Abeywardena, Gajaraj Dhanarajan and Choo-Khai Lim

Abstract

Open educational resources (OER) are a relatively new phenomenon in the Malaysian higher education (HE) sector. Although there have been "lone rangers" strongly advocating the use of OER in the country, many HE institutions, including Wawasan Open University, Open University of Malaysia and Asia e University, are yet to make use and reuse of OER a mainstream practice. There also seems to be reticence over making content freely available to the nation or the region, as well as an absence of policy directions. Notwithstanding, some of these institutions, urged on by individual staff, are taking a serious look at adopting an institutional policy on OER and digital resources. A prime example of this new movement is the OER-based, self-directed open and distance learning course material developed by Wawasan Open University as a pilot project leading to an institutional policy on the use and reuse of OER.

Under a grant from the International Development Research Centre of Canada through an umbrella study on Openness and Quality in Asian Distance Education, a team of collaborators from various Asian countries developed an extensive survey instrument to identify the Asian landscape of digital resources and OER. In Malaysia, the instrument was officially made available to 15 public, private not-for-profit and private for-profit HE institutions. A total of 43 valid responses were received from individuals who are using digital resources/OER, as well as institutional authorities who commented on the institutional stand on OER.

This report summarises the findings from the survey responses gathered from Malaysia and provides an overview of the Malaysian HE landscape with respect to digital resources and OER use.

Keywords: *OER, open educational resources, Malaysia, OER Asia, OER Malaysia, open educational resources Malaysia, open educational resources Asia*

Overview of Higher Education

Malaysia is a middle-income country with a population of about 27 million. It is multi-ethnic, multilingual and multireligious. Its economy is mixed, and whilst agriculture and natural resources, including petroleum, have underpinned the economy in the past, over the last two decades manufacturing and services, including tourism, have become the main economic drivers. Malaysia's economic growth (GDP) in the year 2011 stood at about 6 per cent, and its per capita income in 2010 was about USD 14,744.36, each below a number of its Asian neighbours, such as Taiwan, Hong Kong, Japan, Korea and Singapore. The present government seems determined to move out of the middle-income economic tier by the end of this decade and is investing quite extensively in building its human capital. Over the last ten years, some 20 per cent of the national budget has been spent on education. As a result, the participation rates in basic and secondary education are well above the 95 per cent point, whilst the participation rate in higher education (HE) is around 30 per cent of the age cohort (Table 8.1).

Table 8.1: Percentage of the population aged 19–24 enrolled in tertiary education[1]

Year	Population	Enrolment	%
1970	1,420,687	8,633	0.6
1980	1,624,274	26,410	1.6
1990	2,028,100	58,286	2.9
2000	2,626,900	211,484	8.1
2005*	3,353,600	649,653	19.4
2007*	3,474,200	847,485	24.4

* Aged 18–24 (*Source:* Ministry of Education, Pembangunan Pendidikan, 2001–2010)
Source: Department of Statistics and Ministry of Education, educational statistics;. Ministry of Higher Education website.

Post-secondary education in Malaysia is amongst the growth areas in the education sector. "Post-secondary" refers to education past grades 11 or 12 and includes pre-university courses (largely in public institutions) or technical/vocational courses leading to certificates and diplomas from colleges, universities and other HE institutions. Post-secondary studies take the form of pre-university courses such as grades 12 and 13, matriculation programmes, and technical and vocational courses leading to certificates and diplomas. Post-introductory university courses lead to baccalaureate degrees after four years of study. Post-graduation universities also offer programmes of study leading to master's and doctoral qualifications. The Malaysian Qualifications Framework (Table 8.2) precisely defines these programmes' hierarchy of qualifications and expectations in terms of entry behaviour, as well as the length of study required. Programmes of study leading to all of the above-mentioned qualifications are offered in public and private universities, university colleges and overseas branch campuses in a wide range of subject areas. Modes of delivery include single-mode conventional and distance teaching institutions, as well as those functioning as dual-mode institutions with on- and off-campus studies through correspondence and eLearning facilities.

---

[1] Table from Fernandez-Chung, 2010.

Table 8.2: The Malaysian Qualifications Framework[2]

MQF Levels	Sectors			Lifelong Learning
	Skills	Vocational and Technical	Higher Education	
8			Doctoral Degree	Accreditation of Prior Experiential Learning (APEL)
7			Masters Degree	
7			Postgraduate Certificate & Diploma	
6			Bachelors Degree	
6			Graduate Certificate & Diploma	
5	Advanced Diploma	Advanced Diploma	Advanced Diploma	
4	Diploma	Diploma	Diploma	
3	Skills Certificate 3	Vocational and Technical Certificate	Certificate	
2	Skills Certificate 2			
1	Skills Certificate 1			

Table 8.3: Overview of Malaysian higher education, 1967, 1999, 2007[3]

	1967	1997	2007
Public universities	1	10	20
Private universities and university colleges	0	0	33[a]
Foreign branch campuses	0	0	4
Private colleges and HE institutions	2	690*	488[b]
Polytechnics	0	8	24
Community colleges	0	0	37
Students	4,560	550,000*	873,238
Post-graduates	398	?	45,888
Foreign students	n/a	4,500	47,928
Malaysian students studying abroad	n/a	30,000*	54,915
Population aged 18–24	n/a	?	3,474,200

[a] Excluding local branch campuses
[b] Including local branch campuses
Sources: *Lee, 2004; Fernandez-Chung, 2006; 1967 data: Interim Report to the Higher Education Advisory Council, 1974; 1997 data: Ministry of Education; 2007 data: Ministry of Higher Education

---

[2] Table from Malaysian Qualifications Agency, 2012.
[3] Table from Fernandez-Chung, 2010.

The post-secondary sector is made up of some 20 public and 32 private universities. In addition, there are some 450 colleges and six branch campuses of offshore universities (primarily British and Australian). These numbers are expected to increase as Malaysia opens up the private education space to international participation. Scores of investors in the education sector, from almost all of the English-speaking countries, are lining up to establish colleges and universities in Malaysia. Table 8.3 captures an overview of the Malaysian HE sector (data available up to the 2007–2008 academic year).

If Malaysia's desire to escape the middle-income economic tier is to be achieved, it has to greatly improve educational attainment levels for the population in general and its workforce in particular. Currently, semi-skilled individuals comprise the bulk of the labour force; unskilled labour is mostly imported from neighbouring countries, and those with post-secondary and university-level education are relatively few (Table 8.4).

Table 8.4: Number of employed persons by highest certificate obtained, 1985, 1990, 2000, 2001, 2005 and 2008[4]

Year	Total ($\times 10^3$)	Diploma		Degree	
		N ($\times 10^3$)	%	N ($\times 10^3$)	%
1985	5,653.4	150.8	2.7	120.2	2.1
1990	6,685.0	216.8	3.2	165.8	2.5
2000	9,269.2	535.1	5.8	471.3	5.1
2001	9,357.0	564.5	6.0	533.9	5.7
2005	10,045.4	840.7	8.4	733.5	7.3
2008	10,659.6	786.1	7.4	874.1	8.2

Source: Labour Force Survey, 1985–2008

To increase its educated workforce supply, the country needs to expand the HE sector at an even faster rate than it has done over the past ten years (Table 8.5). Expansion is also expected to meet another of the nation's goals: to become a major HE hub for the region by the year 2012. This expansion is firmly embedded in the National Higher Education Strategic Plan (NHESP) launched with much fanfare by the country's prime minister in 2007. The plan envisages a number of goals and objectives. The major goals are:

- Ensuring access to higher education for diverse groups of students, talents and abilities, based on meritocracy in diversity, irrespective of ethnic origin, gender, social status or physical capability.

- Ensuring that no qualified applicant is denied a place in tertiary education for financial reasons.

- Ensuring equity in higher education through various programmes, open entry criteria, improvement in infrastructure and expansion of information and communication technology (ICT) use.

---

[4] Table from Fernandez-Chung, 2010.

Table 8.5: Expansion in enrolment by educational level, 1985–2008[5]

	1985	1990	1995	2000	2005	2008	Increase in enrolment (%) 1985–2008	Annual rate of increase (%) 1985–2008
Primary	2,191,676	2,447,206	2,827,627	–	3,137,280	3,154,090	30.5	1.3
Secondary*	1,251,447	1,366,068	1,589,584	–	2,217,749	2,310,660	45.8	2.0
Tertiary**	64,025	99,687	146,581	–	463,582	921,548	93.1	4.1
Total	3,507,148	3,912,961	4,563,792	–	5,818,611	6,386,298	45.1	2.0

* Figures include Form Six.
** Figures include enrolment in pre-university and matriculation courses in higher education institutions.
Sources: Ministry of Education; Ministry of Higher Education

ICT in Higher Education

One of the Critical Agenda Projects under the NHESP is the promotion and expansion of eLearning. The use of ICT in HE has kept pace with the development of ICT awareness and investments by both the public and private sectors since the mid-1980s. Massive progress was achieved with the creation of the Multimedia Super Corridor (MSC) in 1996. This is a long-term strategic initiative (1996–2020) involving a partnership between the Malaysian government (as the chief architect of the vision) and the private sector (as the main drivers for its implementation). The intention is to build a competitive cluster of local ICT companies and a sustainable ICT industry (www.mscmalaysia.my). Basically, the MSC is a dedicated corridor (15 kilometres wide and 50 kilometres long) that stretches from the Kuala Lumpur city centre in the north to the new Kuala Lumpur International Airport in the south. Besides offering ICT initiatives, the corridor attracts global ICT companies to relocate their multimedia industries in Malaysia and undertake innovative research and development (R&D) whilst developing new products and technologies for export, keeping this corridor as their base. In other words, the MSC becomes a base for local entrepreneurs to transform themselves into world-class companies. The MSC was further buttressed by ancillary organisations such as the Malaysian Institute of Microelectronic Systems, which assisted in developing a whole range of provisions and protocols to support R&D efforts in ICT-related fields, helped in creating legislative instruments in association with the Ministry of Science, Technology and Innovation, organised dialogue platforms and generally became the backbone of the intellectual repository on matters relating to ICT.

For the first 30 years of ICT growth, Malaysia concentrated on building the right infrastructure to support ICT growth in the country. During the 1980s, most of the ICT infrastructure investment went into provision of basic telephony services to rural and urban people; concerted efforts were also made to increase access to mobile and fixed-line services for a wider segment of the population. One of the key initiatives during this period was the privatisation of the state-owned telecommunication provider, Telekom Malaysia, which helped improve the market reach of telecommunication services. In the last ten years, policy consolidation and further improvement of the infrastructure has also been undertaken, including increased access to the Internet and related services.

[5] Table from Fernandez-Chung, 2010.

Investments into wired and wireless technologies and services through increased privatisation efforts have also continued. This has resulted in expanded broadband services throughout the country, although the conquest of "the last mile" continues to be a challenge; however, there is hope that this will be achieved by 2020 (Kuppusamy et al., 2009). The outcome of all these initiatives is a country well endowed with ICT provisions, infrastructure, legal frameworks, sufficient and adequate technical skills, as well as knowledge to exploit the benefits of the digital revolution (Table 8.6).

Table 8.6: Selected ICT indicators

Indicators	2000 ($\times 10^6$)	2005 ($\times 10^6$)	2010 ($\times 10^6$)
Fixed telephone lines	4.6	4.4	–
Mobile phone subscriptions	5.0	19.5	24
Ownership of personal computers	2.2	5.7	11.5
Internet subscriptions	1.7	4.1	??

Source: Ninth Malaysia Plan (2006–2010), p. 135

From the beginning, ICT provisions for education have been at the centre of these efforts, with the consequence that by the late 1990s ICT-based learning environments were being introduced in Malaysian schools. Portals like MySchoolNet were created to help teachers and students access web-based resources through a variety of technologies. Further encouragement for the use of digital resources came with the creation of a cluster of "smart schools", as well as free or easy provision to own personal computers, tax incentives to connect to the Internet and extensive efforts at training teachers. HE institutions, which have a great deal of autonomy in how they develop policies and practices relating to the application of ICT, were also provided with funding, especially in the public sector, to support the establishment of ICT infrastructure on campuses throughout the country and to induct and train staff.

Despite all of these provisions, as well as policies by the institutions themselves to promote the use of ICT to teach and learn, the impression is that the take-up is slow to modest (Embi, 2011). It is in this context that our study was carried out.

Digital and Open Educational Resources

The cohort of respondents for the survey consisted of academics at various stages in their careers (Table 8.7), teaching at various levels (Table 8.8). Thirty-seven valid responses were gathered from individual users' perspectives and six responses were gathered from an institutional perspective.

Table 8.7: Respondent profile

Participant title	Institution's status			
	Public	Private not-for-profit	Private for-profit	Total
Prof.	1	0	0	1
	(100%)	(0.0%)	(0.0%)	(100%)
Dr.	3	2	3	8
	(37.5%)	(25.0%)	(37.5%)	(100%)
Mr.	1	6	10	17
	(5.9%)	(35.3%)	(58.8%)	(100%)
Ms.	2	4	5	11
	(18.2%)	(36.4%)	(45.5%)	(100%)
Total	7	12	18	37
	(18.9%)	(32.4%)	(48.6%)	(100%)

Table 8.8: Level of teaching

Participant title	Level of teaching		
	Undergraduate	Post-graduate	High school
Prof.	–	1	–
Dr.	7	3	–
Mr.	13	4	1
Ms.	10	2	2
Total	30	10	3

Use of Digital Resources

Through the analysis of the data shown in Figure 8.1 it was identified that digital readers (e.g., Adobe Acrobat reader), online class discussions, images or visual materials (drawings, photographs, art, posters, etc.) and news or other media sources were the most widely used types of digital resources. Digital facsimiles of ancient or historical manuscripts, personal online diaries (e.g., blogs) and maps were the least used.

Figure 8.1: Types of digital resources

A: Digital readers (e.g., Adobe Acrobat reader)

B: Online class discussions (including archived discussions)

C: Images or visual materials (drawings, photographs, art, posters, etc.)

D: Online reference resources (e.g., dictionaries)

E: Online or digitised documents (including translations)

F: Data archives (numeric databases, e.g., census data)

G: Digital film or video

H: News or other media sources and archives

I: Course packs

J: Curricular materials and websites that are created by other faculty and/or other institutions (e.g., MIT OpenCourseWare, World Lecture Hall, MERLOT)

K: E-book readers (e.g., Kindle)

L: Other

M: Government documents in digital format

N: Simulations or animations

O: Audio materials (speeches, interviews, music, oral histories, etc.)

P: Digital facsimiles of ancient or historical manuscripts

Q: Personal online diaries (e.g., blogs)

R: Maps

Search engines/directories (e.g., Google, Yahoo!), personal collections of resources, and online journals were identified as the best sources for finding digital resources (Figure 8.2), whilst incorporating digital resources into lectures/online lectures and using them in project-based or problem-based assignments were found to be the most popular uses (Figure 8.3). However, the majority of the respondents agreed that the use of digital resources would not help them get promoted or obtain tenure. They also pointed out that they do not want students to copy or plagiarise material from the Web. Half of the respondents felt that the use of digital resources distracts from the core goals of teaching.

Figure 8.2: Sources of digital resources

A: Search engines/directories (e.g., Google, Yahoo!)

B: My own personal collection of digital materials

C: Online journals (e.g., JSTOR)

D: Public (free) online image databases

E: Other

F: Campus image databases from my own institution (e.g., departmental digital slide library)

G: Portals that provide links or URLs relevant to particular disciplinary topics

H: Library collections (digital)

I: Media sites (e.g., NPR, New York Times, CNN, PBS)

J: Online exhibits (e.g., from museums)

K: Commercial image databases (e.g., Saskia, AMICO)

Figure 8.3: Use of digital resources

A: Presented during/incorporated in my lectures/class (e.g., images, audio, MIT lecture, etc.)

B: Posted directly on my course website

C: Used in tests and quizzes

D: Presented in my online lectures

E: Linked from my course website

F: Assigned to students to create their own digital portfolios and/or multimedia projects

G: Assigned for student research projects or problem-based learning assignments

H: Assigned to students for review and/or study

I: Presented in the context of an online discussion

J: Other

The respondents felt that more support was needed for them to fully harness the potential of digital resources in teaching and learning. Some of the areas in which support was needed included finding digital resources, assessing the credibility of digital resources, evaluating the appropriateness of resources for teaching goals and interpreting copyright laws and/or securing copyright permissions.

Use of OER

Contrary to the belief that the use of OER is not widespread, 70 per cent of the respondents mentioned that they have used OER in their teaching at some point during their career. Although 13 per cent had not used OER before, 86 per cent mentioned that they would in the future; 17 per cent were unsure whether they had used OER, indicating that more advocacy and capacity-building needs to take place in the country.

OER produced by teachers themselves, produced within the institution, freely downloaded from the Internet and coming from co-operation with other institutions were the main sources for use. Surprisingly, OER downloaded from repositories such as MIT OpenCourseWare, MERLOT, OpenLearn and Connexions were not widely used in Malaysia.

It was encouraging to see that 74 per cent of the respondents were producing OER as learning objects or as part/full courses and programmes (Figure 8.4). This could be due to the support the respondents are getting from the institutions in terms of use and production of open content and open source software. However, as shown in Figure 8.5, there seems to be a lack of co-operation with other educational institutions when it comes to producing and exchanging OER.

Figure 8.4: Production of OER
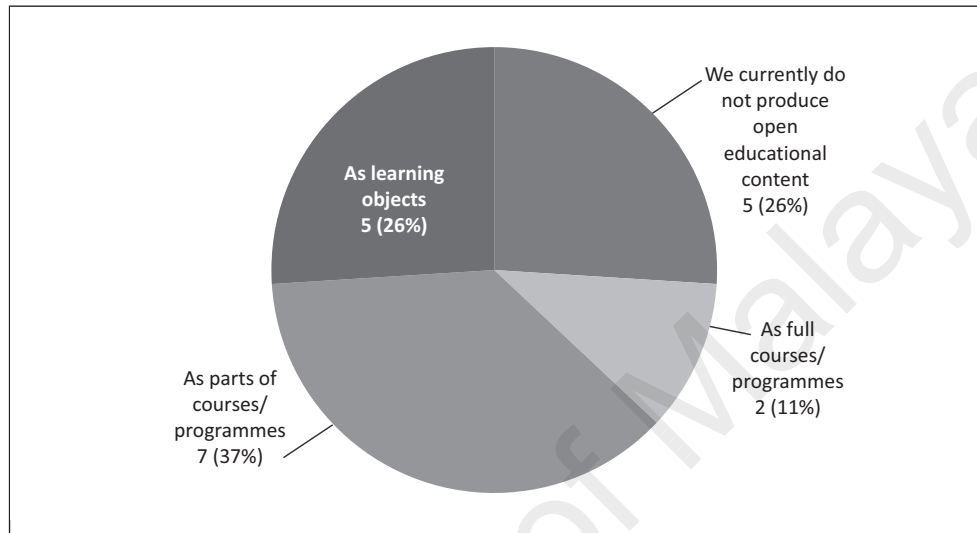
Figure 8.5: Co-operation with educational institutions

Co-operation with other educational institutions for *exchanging* OER

Yes, in other parts of the country
2
(9%)

Yes, in the same region/state
1
(5%)

Yes, internationally
4
(18%)

No
15
(68%)
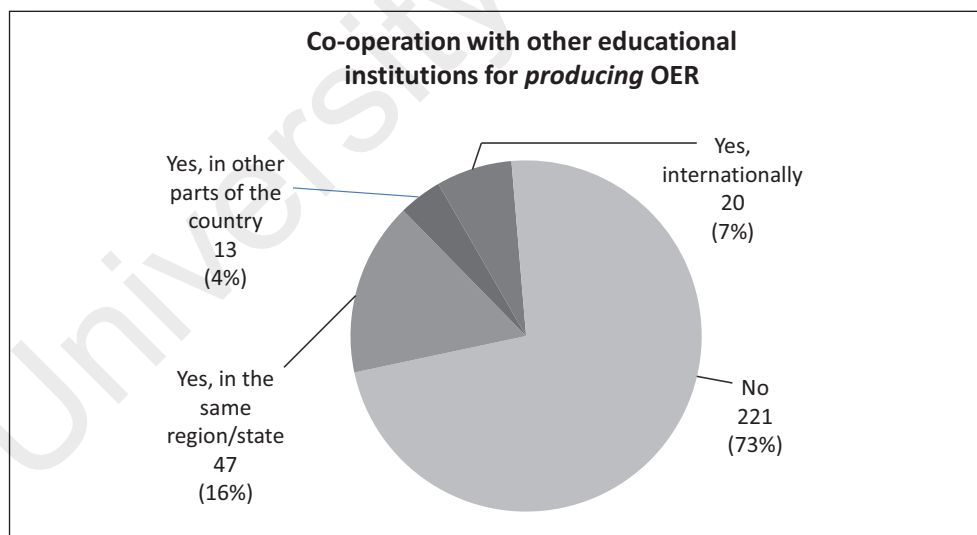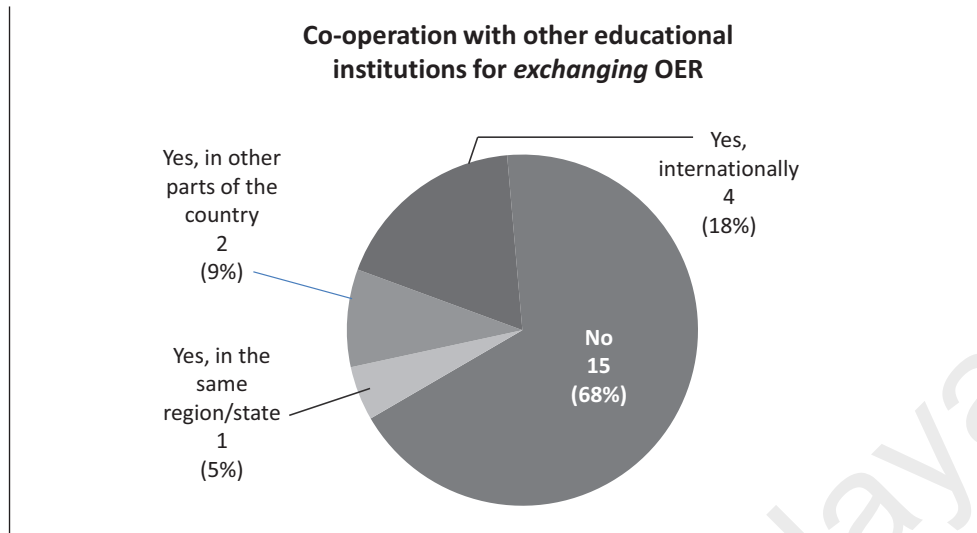
The major identified barriers to the use of OER were lack of awareness, lack of skills, lack of time, lack of ability to locate specific and relevant OER, lack of ability to locate quality OER, lack of interest in pedagogical innovation amongst staff members and lack of support from the management level. Figures 8.6 and 8.7 provide more details about the concerns the respondents had with respect to producing and using OER, respectively. The respondents also highlighted that the lack of rewards and recognition for staff devoting their time to OER-based activities was a major deterrent. However, they commented that infrastructure such as hardware, software and access to computers was not an issue.

Figure 8.6: Concerns about producing OER

130
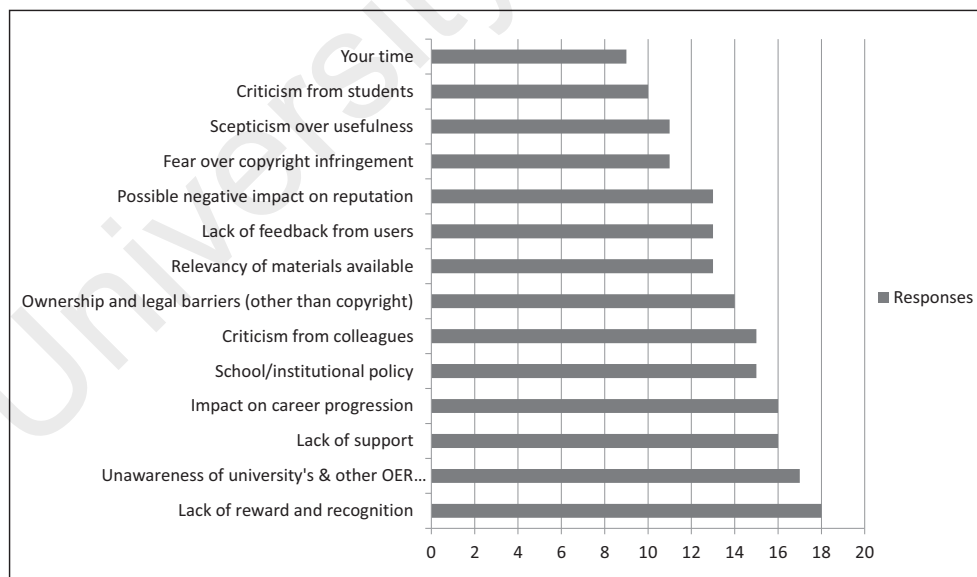
Figure 8.7: Concerns about using OER
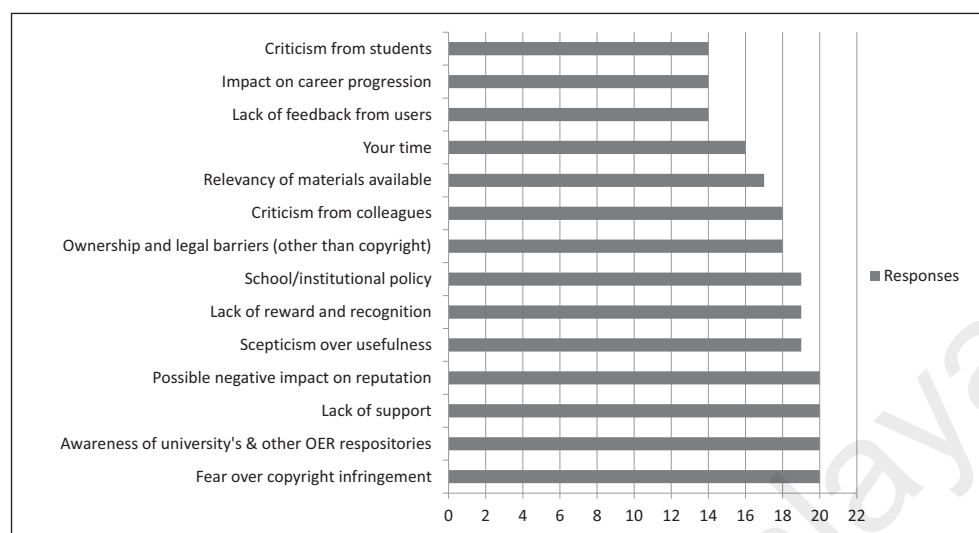
The attitudes towards the use of OER were generally positive. The respondents agreed that OER do not help other institutions copy their best ideas. They also agreed that publishing OER would not stop students from attending lectures. However, they were concerned about how others would use the material they had produced. They were also concerned about the damaging effect that poorly developed OER could have on an institution's reputation.

Regarding copyright and licensing, 51.4 per cent of the respondents understood the word "copyright" and 82 per cent had used open content licences. Only six per cent had used Creative Commons licences, even though 44 per cent had heard of them. Major concerns were expressed with respect to remixing different resources legally, publishing material that incorporated unlicensed third-party content, discovering materials that could be legally used and publishing material created.

From an institutional perspective, four of the six respondents mentioned that their institutions do not have a policy on the creation and use of OER and that fewer than five per cent of the staff were engaged in OER-related activities. They also mentioned that even though the use of OER material was encouraged over the use of "copyright" protected material, there was no mechanism to reward or recognise these attempts.

Conclusion and Recommendations

At present Malaysia is placing great emphasis on building a knowledge community by increasing the number of citizens with access to higher education. In this roadmap, ICT funded and nurtured by the government play a major role. With more and more digital resources being developed and made available for use, the question arises whether the academic community is ready to undertake the responsibility of using these resources in their teaching and learning activities.

In general, Malaysian academics seem comfortable with locating, identifying and using digital resources in their day-to-day teaching and learning. However, further support is needed, especially at the institutional level, to facilitate capacity-building in this area. OER, a subset of digital resources, are fast becoming

mainstream practice amongst academics. It is encouraging to see that the majority of academics who participated in this study were knowledgeable about OER, had used them at some point in their careers and were willing to use them more in the future. One area of concern, however, is the lack of co-operation between academic institutions when producing and exchanging OER. This culture of collaboration between institutions needs to be established to harness the full potential of open content.

Special concerns were expressed with respect to copyright and the management of copyright. Even though academics had been exposed to open content licences such as those provided by Creative Commons, there was still a degree of trepidation with respect to using material licensed in this manner. More capacity-building is needed at an institutional as well as national level to familiarise users with the benefits and limitations of open content licensing.

From an institutional perspective, fewer than five per cent of staff are engaged in activities related to OER. As such, most institutions do not have an institutional policy on OER. This in turn has discouraged many staff from undertaking OER-based activities on a day-to-day basis, as there are no rewards or recognition for their efforts. One of the key actions to promote greater adoption of OER in Malaysia would be for institutions to establish policies encouraging the wider use and reuse of open content.

Acknowledgements

References

Economic Planning Unit, Prime Minister's Department (2006). Ninth Malaysia Plan (2006–2010). Putrajaya, Malaysia. Retrieved from http://www.pmo. gov.my/dokumenattached/RMK/RM9_E.pdf

Embi, M. A. (2011). e-*Learning in Malaysian institutions of higher learning: Status, trends and challenges*. Keynote address at the International Lifelong Learning Conference (ICLLL). 14–15 November 2011, Kuala Lumpur.

Fernandez-Chung, R. M. (2010). Access and equity in higher education, Malaysia. Unpublished paper presented at Higher Education and Dynamic Asia Workshop, Asian Development Bank, Manila, June 2010.

Kuppusamy, M., Raman, M., & Lee, G. (2009). Whose ICT investment matters to economic growth — private or public? The Malaysian perspective. *The Electronic Journal of Information Systems in Developing Countries, 37*(7), 1–19.

Malaysian Qualifications Agency. (2012). *Malaysian Qualifications Framework: Point of reference and joint understanding of higher education qualifications in Malaysia*. Retrieved from http://www.mqa.gov.my/portal2012/dokumen/ MALAYSIAN%20QUALIFICATIONS%20FRAMEWORK_2011.pdf

Appendix H

Dhanarajan, G., & Abeywardena, I. S. (2013). Higher Education and Open Educational Resources in Asia: An Overview. In G. Dhanarajan & D. Porter (Eds.), *Open Educational Resources: An Asian Perspective. Commonwealth of Learning and OER Asia (ISBN 978-1-894975-61-2),* 3-18.

Higher Education and Open Educational Resources in Asia: An Overview

Gajaraj Dhanarajan and Ishan Sudeera Abeywardena

Abstract

Higher education has experienced phenomenal growth in all parts of Asia over the last two decades. This expansion, coupled with a diversity of provisions, has meant that more and more young Asians are experiencing tertiary education within their own countries. Notwithstanding this massive expansion of provisions, equitable access is still a challenge for Asian countries. There is also concern that expansion will erode quality. The use of digital resources is seen as one way of addressing the dual challenges of quality and equity. Open educational resources (OER), free of licensing encumbrances, hold the promise of equitable access to knowledge and learning. However, the full potential of OER is only realisable by acquiring: (i) greater knowledge about OER, (ii) the skills to effectively use OER and (iii) policy provisions to support its establishment in the continent's higher education milieu.

Keywords: Asia, higher education, digital resources, open educational resources, OER awareness, policies, practices, benefits and barriers

Higher Education in Asia

The last three decades has seen a rapid increase in the provision of higher education in almost all parts of greater Asia — from the Korean peninsula in the east to the western borders of Central Asia. Nowhere has this increase matched the growth seen in South, South East and Far East Asia. Universities, polytechnics, colleges and training institutes with a variety of forms, structures, academic programmes and funding provisions have been on an almost linear upward progression (Table 1.1).

Table 1.1: Number of higher education institutions in selected countries[1]

Country	Three- to four-degree & post-graduate schools	Two- to four-year undergraduate schools	Two- and three-year diploma schools	Short certificate schools	Professional and technical schools
Cambodia	69	9	-	-	-
PRC	1,237	1,264	1,878	-	-
India	504	28,339	-	-	3,533
Indonesia	480	3967	162	-	-
Laos	34	-	11	-	-
Malaysia	57	488	24	37	-
Philippines	1,710	-	114	30	-
South Korea	197	152	-	-	-
Sri Lanka	15	16	-	-	-
Thailand	102	32	19	-	-

In addition to governments, private for-profit and not-for-profit organisations, public–private partnerships, international agencies and intergovernmental agencies have been participating in and financially supporting this growth. With the arrival of and access to the Internet, World Wide Web and a huge range of fast and intelligent information and communication technologies (ICT), many individuals have also been prepared to share their life experiences and knowledge with others through YouTube, Flickr, Wikieducator and other similar tools. Consumers of education have themselves become producers of education. The growth in Asia reflects the growth in many other parts of the world, which was experiencing increased participation from 28.6 million in 1970 to about 152.7 million in 2007, at a rate of increase of almost 4.6 per cent per year (UNESCO, 2009). Between 1990 and 2005, about 98 million Asians had experienced one or another form of tertiary education in a variety of institutions, ranging from technical colleges to universities (Table 1.2).

Table 1.2 is also illustrative of high levels of termination in higher education by millions of young people who, despite being qualified to meet the challenges of higher education, are unable to fulfil their aspirations. The gap between demand for and supply of higher education still continues to be high. Further exacerbating this situation is that those failing to gain admission into higher education are often from the marginalised segments of a nation's population.

Unequal access to higher education on the basis of gender, economic and social status, location of residence and poor prior schooling all continue to challenge many Asian nations. Countries such as Cambodia, Laos, India, Indonesia, Pakistan and Vietnam have low participation rates for the 17–24 age cohort. Further, policies on widening participation in higher education will also require serious regard for many other groups besides those described so far. These other groups include challenged and displaced persons, migrant labourers, immigrants and the elderly. Many international conventions and covenants provide a framework for countries to consider. As of June 2009, only India, the Philippines and Bangladesh had ratified conventions, whilst others are moving slowly on this front, even though countries like Malaysia have policies in place to facilitate access for challenged persons.

[1] Data extrapolated from Asian Development Bank, 2012.

Table 1.2: Upper secondary gross, graduation and tertiary entry ratios (Asian Development Bank, 2012)

Country	Secondary gross enrolment ratio	Upper secondary gross graduation/ completion (ISCED 3A)	Gross entry ratio into tertiary (ISCED 5A)
Cambodia	23a	7.5e, f	
China	72a	33	14
India	52a	28	13c
Indonesia	58a	31	17
Laos	27b	5.3c, f	
Malaysia	82a		26c
Philippines	72a, c	64	
South Korea	102a	62	61
Sri Lanka	56.6f	28.3c, f	21.2c
Thailand	82a	40	20
Vietnam	25.5a, b	12.5c	

ISCED = International Standard Classification of Education. ISCED 3A = upper secondary level of education; programmes at level 3 are designed to provide direct access to ISCED 5A. ISCED 5A = first stage of tertiary education; programmes are largely theoretically based and are intended to provide sufficient qualifications for gaining entry into advanced research programmes and professions with high skills requirements.
Sources: (a) UNESCO, 2009 (data from [b] 2005, [c] 2006, [d] 2001); (e) not segregated under ISCED; (f) Barro & Lee, 2010.

Besides this normal age cohort, many other groups are also seeking or requiring access to higher education. The biggest amongst these are adults who wish to return to learning. For many of these adults, higher education was denied them earlier. Their return to study requires facilitation which in an already supply-poor situation presents difficulties. Not facilitating or incentivising such returnees is not only a social denial, but also economically counterproductive. Malaysia presents such a situation. The country aspires to be high-income in another decade. To support that aspiration, it requires an adult workforce of highly skilled and knowledgeable citizens. Currently, of its 12 million citizens in the workforce, more than 80 per cent have less than a secondary school education. This is a serious concern, given the country's ambition. Policy initiatives will be required to increase participation. Countries such as Malaysia recognise this dilemma and are actively pursuing policies to widen participation. This may not be the case all across Asia. Special policies include creating alternate pathways of entry, part-time studies, distance education, special financial incentives and arrangements, recognition of workplace training and according of academic credit for such training through policy instruments promoting lifelong learning. South Korea, like its other OECD counterparts, has long been a leader in such arrangements. The Philippines, Indonesia, Thailand, India and China all have enculturised lifelong learning or are moving towards doing so.

Besides "balancing the continued expansion of access with greater attention to equity" (Asian Development Bank, 2011), higher education in Asia is also

challenged by other concerns. According to a recently published study by the Asian Development Bank (2011), these include the following:

- Maintaining and improving education quality, even in the face of serious financial constraints.

- Increasing the relevance of curriculum and instruction at a time of rapid change in labour market needs.

- Increasing and better utilising the financial resources available to higher education.

In many development circles in Asia, ICT has been viewed if not as a panacea then at least as having the potential to address many of the above challenges. In an earlier report on the role of ICT in education, the Asian Development Bank (2009) went on to declare:

> ICT has the potential to "bridge the knowledge gap" in terms of improving quality of education, increasing the quantity of quality educational opportunities, making knowledge building possible through borderless and boundless accessibility to resources and people, and reaching populations in remote areas to satisfy their basic right to education. As various ICTs become increasingly affordable, accessible, and interactive, their role at all levels of education is likely to be all the more significant in making educational outcomes relevant to the labor market, in revolutionizing educational content and delivery, and in fostering "information literacy".

Many Asian nations have been investing in ICT for the last four decades or so, and some of these countries (e.g., South Korea, Japan, Singapore, Malaysia) have ICT infrastructures that rank amongst the best in the world; on the other hand, in many Asian countries ICT developments are somewhat modest, or even inadequate to support the needs of higher education. Notwithstanding, there is a clear appreciation of the role that ICT, especially digitised learning resources, can play in expanding access and improving the quality of education.

Use of Digitised Educational Resources in Asian Higher Education

During the last 40 years, Asian nations have developed an affinity for the use of ICT to serve education in a variety of ways. These technological tools have been employed to deliver education in various sectors and at various levels. Institutions have been using both low and high technologies, and many that have been using the former, such as analogue broadcast radio and television and print, have been gradually moving in tandem with the evolution of the latter, i.e., from the analogue to the digital realm using the Internet, the World Wide Web and multimedia resources. Amongst a few, pedagogy has also evolved along with the technologies, albeit not at the same pace. Of the new pedagogies, distance education or open distance education has proven to be especially attractive to policy makers and budget-conscious administrators, as well as a segment of learners who look for a much more self-directed and flexible learning environment. But increasingly, eLearning, virtual campuses and online courses are also being delivered, especially in ICT-rich environments like South Korea and

Japan. The availability of new technologies has also created opportunities in other Asian countries to embed digital resources in their courses delivered on- or offline. However, the use of digital resources for teaching or learning is not uniform across or within nations. A number of factors either enable or hinder such use. In a recent study conducted with the support of a grant from the International Development Research Centre of Canada, researchers found, through a survey of some 580 academic staff from ten Asian countries (South Korea, Japan, China, Hong Kong, the Philippines, Indonesia, Vietnam, Malaysia and India), the following.

Access to ICT infrastructure and digital infrastructure

What was seen as a major impediment even as recently as the turn of the millennium is no longer viewed as a major challenge. Reliable electricity, available and affordable appliances, the skills to install, maintain and use appliances, and access to the Internet (albeit at a higher connection cost and smaller bandwidth) are there for most workers in higher education. Urban populations, both staff and students, have easier access to ICT infrastructure, but with the increasing availability of mobile devices and telephones the urban–rural imbalance is somewhat mitigated. Infrastructural resources besides the availability of personal computers and mobiles also include access to the Internet, the World Wide Web, email, presentation software and in some cases electronic libraries (Figure 1.1).

Figure 1.1: The availability of ICT infrastructure in selected Asian countries
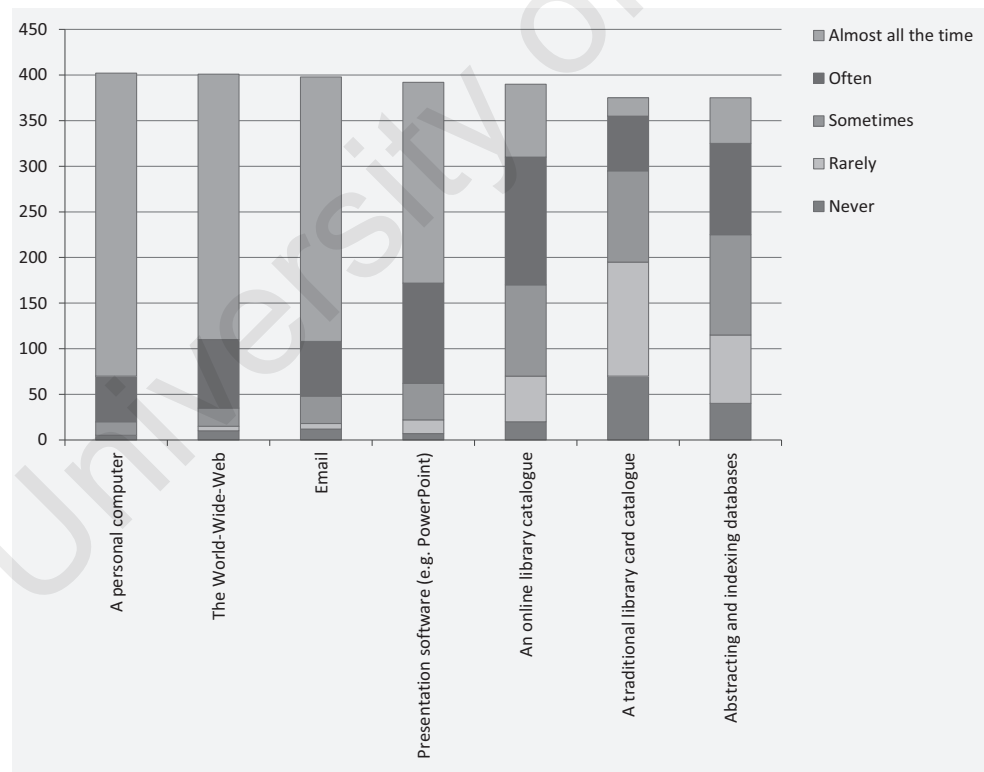
Sources of digital resources

As Table 1.3 indicates, almost all academic staff use the popular search engines (Google, Yahoo!, Safari and Bing). A few build and maintain their own personal collections and/or use media sources, such as CNN, BBC or local television and radio channels. There is limited use of resources from museums, professional organisations and commercial databases (probably a reflection of the cost to access these resources).

Table 1.3: Sources of digital resources (after Dhanarajan & Abeywardena, 2012)

Sources of digital resources	Use (%)					
	Almost all the time	Often	Sometimes	Rarely	Never	N
Search engines/directories (e.g., Google, Yahoo!)	54.38	32.47	9.54	2.32	1.29	388
My own personal collection of digital materials	30.59	39.85	17.48	9.77	2.31	389
Public (free) online image databases	23.31	34.27	27.53	9.55	5.34	356
Online journals (e.g., via JSTOR)	21.43	28.06	27.3	15.82	7.4	392
Library collections (digital)	16.41	27.95	29.23	17.69	8.72	396
Campus image databases from my own institution (e.g., departmental digital slide library)	13.44	22.22	28.17	18.35	17.83	387
"Portals" that provide links or URLs relevant to particular disciplinary topics	13.04	33.25	36.32	11.51	5.88	391
Media sites (e.g., NPR, New York Times, CNN, PBS)	10.97	25.59	32.64	19.58	11.23	383
Other	5.56	11.11	18.52	12.04	52.78	108
Online exhibits (e.g., from museums)	3.66	10.44	25.85	32.11	27.94	383
Commercial image databases (e.g., Saskia, AMICO)	2.86	9.61	24.16	27.01	36.36	385

Use of digital resources

Table 1.4 shows that depending on residential locations and bandwidth availability, academics *mostly* accessed a range of resources, such as: digital readers (e.g., Adobe Acrobat); images or other visual materials, such as drawings, photographs and art posters; online reference materials; digitised documents; digital film or video; and course packs. The least accessed resources included data archives; audio materials, such as speeches and oral interviews; online diaries; government documents; and simulations or animations.

Table 1.4: Types of digital resources and their frequency of use (after Dhanarajan & Abeywardena, 2012)

Types of digital resources	Use (%)					
	Almost all the time	Often	Sometimes	Rarely	Never	N
Digital readers (e.g., Adobe Acrobat)	30.4	34.2	21.3	8.0	6.1	395
Images or visual materials (drawings, photographs, art, posters, etc.)	26.8	41.3	23.3	7.3	1.5	400
Online reference resources (e.g., dictionaries)	24.2	40.9	25.0	7.1	2.9	396
Online or digitised documents (including translations)	17.3	34.9	23.4	16.3	8.0	398
Online class discussions (including archived discussions)	15.9	25.8	27.4	16.6	14.3	391
Digital film or video	15.4	33.9	35.7	10.6	4.3	395
News or other media sources and archives	15.3	35.1	32.3	13.0	4.3	393
Course packs	14.7	20.4	35.6	16.2	13.1	388
Curricular materials and websites that are created by other faculty and/or other institutions (e.g., MIT OpenCourseWare, World Lecture Hall, MERLOT)	13.8	29.4	33.3	15.3	8.3	398
Other	13.3	20.5	25.8	9.3	31.1	151
E-book readers (e.g., Kindle)	10.3	19.6	19.57	22.83	27.72	368
Data archives (numeric databases, e.g., census data)	9.16	23.4	31.6	20.6	15.3	393
Audio materials (speeches, interviews, music, oral histories, etc.)	7.9	23.5	35.4	22.0	11.1	395
Personal online diaries (e.g., blogs)	6.9	18.9	27.0	27.3	19.9	392
Government documents in digital format	6.6	21.1	33.84	21.37	17.05	393
Simulations or animations	5.37	26.6	34.2	23.3	10.5	391
Maps	3.8	12.2	33.9	29.4	20.8	395
Digital facsimiles of ancient or historical manuscripts	2.3	6.9	16.0	26.7	48.2	394

Factors inhibiting the use of digital resources

Two types of barriers seem to dissuade individuals, especially teachers, from using digital resources: technical and attitudinal. The technical barriers include: needing technical support to search and find digital resources; locating and clearing copyright; setting up technical infrastructure (computers, connections); installing appropriate software; evaluating the quality of resources; integrating resources into learning management systems; and using learning management systems (Table 1.5). The attitudinal barriers mostly arise from (i) apprehension about the quality of the digital resources, the context of their creation and the appropriateness of the resources to buttress the curriculum, (ii) lack of confidence in learners' skills to use digital resources and (iii) anxieties over issues relating to plagiarism (Table 1.6).

Table 1.5: Technical barriers to the use of digital resources (after Dhanarajan & Abeywardena, 2012)

Barriers	Extremely important	Very important	Somewhat important	A little important	Not at all important	N	Percentage
Support with interpreting copyright laws and/or securing copyright permission	35.60%	38.90%	16.20%	6.40%	2.80%	388	92.40%
Support with finding digital resources	35.00%	42.20%	13.80%	5.40%	3.60%	391	93.10%
Support with assessing the credibility of digital resources	34.60%	41.30%	15.40%	5.40%	3.30%	390	92.90%
Support with obtaining or setting up technical infrastructure (servers, computers, smart classrooms, etc.)	31.30%	38.20%	20.40%	6.70%	3.40%	387	92.10%
Support with evaluating the appropriateness of resources for my teaching goals	27.50%	38.00%	19.00%	11.60%	3.90%	389	92.60%
Support with gathering, organising, and maintaining digital materials	26.50%	45.50%	16.20%	7.70%	4.10%	389	92.60%
Support with digitising existing resources	26.00%	39.70%	22.90%	7.30%	4.20%	385	91.70%
Support with integrating resources into a learning management system (e.g., Moodle, Sakai)	24.90%	33.40%	23.10%	12.40%	6.20%	386	91.90%
Support with training students to find or evaluate digital resources	24.00%	39.80%	25.10%	7.80%	3.40%	387	92.10%
Support with importing resources into a course website or a database	21.80%	36.40%	23.40%	13.50%	4.90%	385	91.70%
Support with learning how to use a learning management system (e.g., Moodle, Sakai)	20.00%	42.10%	19.00%	12.20%	6.80%	385	91.70%
Support with creating my own website	19.30%	32.00%	27.60%	14.70%	6.40%	388	92.40%

Table 1.6: Non-technical barriers to the use of digital resources (after Dhanarajan & Abeywardena, 2012)

Barriers	Strongly agree	Somewhat agree	Somewhat disagree	Strongly disagree	N	Percentage
They cannot substitute for the teaching approaches I use	13.60%	26.90%	33.80%	25.80%	361	86.00%
I don't have time to use digital resources	11.80%	24.60%	33.00%	30.60%	382	91.00%
Digital resources are difficult for me to access	9.70%	20.20%	35.20%	34.90%	381	90.70%
Digital materials can be presented outside their original context	8.30%	24.50%	41.90%	25.30%	363	86.40%
They are irrelevant to my field	7.70%	23.10%	35.60%	33.50%	376	89.50%
Using them distracts from the core goals of my teaching	5.60%	22.70%	40.60%	31.00%	374	89.00%
Students don't have the information literacy skills to assess the credibility of digital resources	5.40%	25.10%	37.60%	31.90%	367	87.40%
I don't want my students to copy or plagiarise material from the Web	4.20%	21.90%	42.70%	31.20%	356	84.80%

Factors enabling or encouraging academic staff to use digital resources

These factors relate either to pedagogical reasons (Table 1.7) — such as a desire to be current in knowledge, access to content not available in the local institution, and availability of sophisticated media, digital resources and supporting research — or to personal reasons (Table 1.8), including "exciting" learners about new ways of learning and engaging in critical thinking, providing learners with current knowledge from primary sources, supporting learner creativity and enabling learning flexibility by allowing content to be available 24/7. Also emerging amongst innovators are many novel opportunities that new digitised resources present. These include collaborating in and sharing of curriculum, learning materials and associated tools/technologies. In parallel to technological advancements has been a desire of many to share — especially learning materials — free of legal and logistical restrictions. The rearrangement of licensing protocols and regulations, such as via the family of Creative Commons provisions, is encouraging Asian academics to explore a range of activities, including participation in the global open educational resources (OER) movement.

Table 1.7: Pedagogical reasons (after Dhanarajan & Abeywardena, 2012)

Factors	Strongly agree	Somewhat agree	Somewhat disagree	Strongly disagree	N	Percentage
It helps me get students excited about a topic	57.30%	36.10%	5.90%	0.80%	393	93.60%
It improves my students' learning	54.50%	39.50%	5.90%	0.00%	387	92.10%
It helps me let students know the most up-to-date (or most current) developments in the subject	54.40%	37.90%	7.20%	0.50%	388	92.40%
It helps me provide students with a context for a topic	52.40%	44.00%	3.10%	0.50%	391	93.10%
It allows me to integrate primary source material into the course	45.50%	44.70%	9.00%	0.80%	387	92.10%
It allows my students to be more creative	42.50%	46.40%	9.80%	1.30%	386	91.90%
It is more convenient for my students and their schedules	40.50%	42.60%	14.60%	2.30%	383	91.20%

Table1.8: Personal reasons (after Dhanarajan & Abeywardena, 2012)

Factors	Strongly agree	Somewhat agree	Somewhat disagree	Strongly disagree	N	Percentage
It saves me time	39.50%	37.10%	16.40%	7.00%	385	91.70%
It provides access to resources that we don't have at our college	39.10%	46.10%	12.20%	2.60%	386	91.90%
It allows me to do things in the classroom that I could never do otherwise	36.40%	47.30%	11.40%	4.90%	385	91.70%
It allows me to stay up to date with my colleagues	35.70%	35.90%	20.60%	7.80%	384	91.40%
It helps me to teach critical thinking skills	35.10%	41.00%	19.10%	4.90%	388	92.40%
It helps me to integrate my research interests into my course	34.10%	49.40%	14.50%	2.10%	387	92.10%
I like or feel very comfortable with the new technologies	30.60%	48.10%	17.70%	3.60%	385	91.70%
It helps me to teach information literacy (i.e., evaluating the online materials for themselves)	29.90%	47.90%	18.00%	4.10%	388	92.40%
I enjoy having my teaching practices and course materials available to anyone in the world who would like to use them	29.70%	43.00%	19.90%	7.40%	377	89.80%
The administration (deans, chairs, provost) encourages me to use digital resources more	20.80%	32.80%	26.60%	19.80%	384	91.40%
It may help me get promoted or get tenure	10.70%	25.10%	35.50%	28.70%	383	91.20%

Pursuing OER

Open educational resources are increasingly being promoted by enthusiasts as a solution, amongst many others, to overcome the challenges of access, quality and cost in providing or participating in higher education, all over the world. Whilst in many parts of the developed world cost has often been cited as a reason to seriously consider OER as an alternative to expensive textbooks, skyrocketing tuition fees and inflexible learning opportunities within conventional systems, in the developing world inequitable access to learning, especially at the tertiary level — both formal and non-formal — has been presented as an argument to buttress the case. Conceiving of OER purely in terms of access, cost and quality is perhaps limiting, as there are other more profound reasons to assert a place for OER in higher education.

Even though ideas relating to OER have been in circulation, globally, over the last decade or so, developments in the poorer Asian nations have been slow. Similarly, and despite the contemporary international debate and dialogue, knowledge of OER and their value amongst members of the larger Asian academic community as well educational policy makers is modest at best. Even in countries where there is familiarity, such as Japan, China and India (all of which already have some kind of arrangements to share digitised course content through consortium arrangements),[2] discernible gaps exist regarding understanding and application in many of the following aspects:

- Detailed knowledge of OER as a practice.
- Knowledge of user needs.
- Knowledge of usage levels amongst various user groups.
- The characteristics of organisations successfully using OER.
- A knowledge of and compliance with standards.
- The range of technological assets required to benefit from OER.
- The human capacities needed to develop and manage OER.
- Other contextual factors (e.g., bandwidth).

Notwithstanding the above, a number of national and institutional initiatives are ongoing, ranging from the big to the tiny. Some examples of OER activity in the formal academic sector, described in the present volume, are: India's NPTEL (National Programme on Technology Enhanced Learning), the efforts by a consortium of the Indian Institutes of Technology (Chapter 17); Beijing Open University's non-formal educational courses (Chapter 1); formal degree programmes at the Virtual University of Pakistan (Chapter 8); South Korea's provision of employment-related training programmes (Chapter 6); Vietnam's efforts at producing translated versions of academic texts as open textbooks (Chapter 10); and formative efforts by Malaysia's Wawasan Open University (Chapter 11). In the non-formal sector, Indonesia's Open University is building a community of teachers to share learning resources through its teacher education forum (Chapter 18); a commercial publisher in the Philippines is putting together on a free-to-use basis historical and cultural documents about the Philippines (Chapter 13); and in India an international development agency, ICRISAT (International Crops Research Institute for the Semi-Arid Tropics) has created a suite of learning objects on agriculture and climate sciences, and made it available to farmers, extension workers and academics as OER (Chapter 12).

---

[2] www.ocwconsortium.org

There are any number of reasons why participation in an OER movement is beginning to happen (Table 1.9). It is still early days to predict how well a culture of producing, sharing, using and reusing OER will develop in most parts of Asia. At best, it is a development in progress, and at worst, it could be perceived as yet another techno-fad. Institutions and individuals who produce, access and use OER clearly perceive benefits, despite some difficult barriers. Survey findings from nine Asian countries regarding perceptions of benefits and barriers are presented in Tables 1.9 and 1.10.

Table 1.9: Perceived benefits of accessing and using OER (after Dhanarajan & Abeywardena, 2012)

Benefits	1 Very important	2	3	4	5 Unimportant	N	Percentage
Gaining access to the best possible resources	72.30%	21.00%	5.40%	0.60%	0.60%	314	74.80%
Promoting scientific research and education as publicly open activities	47.50%	34.90%	11.90%	3.80%	1.90%	318	75.70%
Bringing down costs for students	45.40%	29.30%	16.10%	6.60%	2.50%	317	75.50%
Bringing down costs of course development for institutions	42.40%	30.10%	15.20%	6.60%	5.70%	316	75.20%
Providing outreach to disadvantaged communities	44.00%	28.20%	17.70%	7.60%	2.50%	316	75.20%
Assisting developing countries	37.80%	26.70%	21.30%	9.80%	4.40%	315	75.00%
Becoming independent of publishers	27.60%	23.70%	28.80%	12.20%	7.70%	312	74.30%
Creating more flexible materials	47.20%	33.20%	12.00%	3.20%	4.40%	316	75.20%
Conducting research and development	50.30%	27.40%	15.60%	4.80%	1.90%	314	74.80%
Building sustainable partnerships	41.50%	27.50%	21.10%	6.10%	3.80%	313	74.50%

Table 1.10: Barriers to producing and utilising OER (after Dhanarajan & Abeywardena, 2012)

	1 Very important	2	3	4	5 Unimportant	N	Percentage
Lack of awareness	51.00%	29.90%	9.90%	3.80%	5.40%	314	74.80%
Lack of skills	30.60%	40.80%	17.20%	5.40%	6.10%	314	74.80%
Lack of time	24.20%	30.60%	24.20%	9.70%	11.30%	310	73.80%
Lack of hardware	17.30%	24.70%	25.00%	15.10%	17.90%	312	74.30%
Lack of software	18.70%	28.80%	23.40%	13.60%	15.50%	316	75.20%
Lack of access to computers	19.50%	19.20%	13.40%	16.00%	31.90%	313	74.50%
Lack of ability to locate specific and relevant OER for my teaching	23.60%	33.70%	22.30%	11.30%	9.10%	309	73.60%
Lack of ability to locate quality OER for my teaching	27.90%	39.60%	18.80%	8.40%	5.20%	308	73.30%
No reward system for staff members devoting time and energy	25.60%	31.10%	22.80%	7.40%	13.10%	312	74.30%
Lack of interest in pedagogical innovation amongst staff members	28.60%	32.80%	22.80%	7.70%	8.00%	311	74.00%
No support from management level	27.40%	28.10%	21.80%	11.90%	10.90%	303	72.10%

Awareness and knowledge of OER

To those who are ardent advocates of OER, benefits of utilising these free resources are familiar. However the higher education community in Asia is large, diverse and relatively conservative in its attitudes towards teaching and learning. Awareness as well as knowledge-building, amongst both teachers and policy makers, is critical for the acceptance and integration of resources for teaching. Such awareness is currently very low — recent advocacy efforts by UNESCO and the Commonwealth of Learning (COL) through their joint declaration on OER (UNESCO & COL, 2012) are helpful, but OER need to be popularised; greater efforts at knowledge-building, especially amongst policy makers and institutional management, have to be enhanced. Such knowledge-building has to be comprehensive and current — those in decision-making positions must be aware of what OER exist, in what contexts and how they have been used, how to gain access to them, what technologies and skills are required for teachers and learners alike to access them, and the pedagogical and economic benefits of OER.

Table 1.11: Familiarity with and awareness of OER (after Dhanarajan & Abeywardena, 2012)

| Country | Familiarity and awareness | | | |
	Yes	No	Unsure	Total (N)
China	40	21	11	72
	55.60%	29.10%	15.30%	100.00%
Hong Kong	8	9	2	19
	42.10%	47.40%	10.50%	100.00%
India	25	14	9	48
	52.10%	29.20%	18.80%	100.00%
Indonesia	27	7	4	38
	71.10%	18.40%	10.50%	100.00%
Japan	5	4	0	9
	55.60%	44.40%	0.00%	100.00%
Malaysia	16	3	4	23
	69.60%	13.00%	17.40%	100.00%
Philippines	20	1	3	24
	83.30%	4.20%	12.50%	100.00%
South Korea	46	10	6	62
	74.20%	16.10%	9.70%	100.00%
Vietnam	15	4	1	20
	75.00%	20.00%	5.00%	100.00%

Purpose of OER

The international debate on a purpose for OER in the higher education milieu continues to engage scholars passionately. Such debate also encompasses more recent arguments around massive open online courses, or MOOCs, and their range of analogues. What was once considered a straightforward purpose for OER — i.e., resources such as "courses, course materials, content modules,

collections, and journals . . . [as well as] tools for delivering educational content, e.g., software that supports the creation, delivery, use and improvement of open learning content, searching and organisation of content, content and learning management systems, content development tools, and on-line learning communities meant to be used for education",[3] not necessarily for academic credit — is no longer the case. As technology innovations progress, new agendas have become part and parcel of OER dialogues; MOOCs are a recent innovation that have confused the open space for consumers and academics alike.

In the context of developing Asia, it may be useful to promote OER with an unambiguous clarity of purpose, such as that OER improves cost-free access to up-to-date and current information relating to content, reduces the cost of curriculum transformation, assists in designing employment-relevant curriculum, supports flexible ways of delivering curriculum and facilitates inter-institutional collaboration and co-operation in content development and sharing.

Policies on OER

In many parts of Asia, government policy support can accelerate the adoption of innovations in education. Governments have it in their powers, through a variety of instruments, to support innovation or retard it. Asian governments could discourage OER production, use, reuse and distribution in a number of ways, including: (i) restricting the free flow of information, (ii) limiting access to search engines, (iii) limiting financial support for adopting innovations, (iv) limiting the extent to which curriculum and content can be explored at the delivery end and (v) discouraging the use of Creative Commons licences. At the last count, some eleven countries in Asia had established national affiliates. Some of the affiliates are active, whilst others are not.

Besides policy support at government levels, such support or lack thereof at institutional levels also places limitations on the extent to which OER can play an effective role. Familiarity with the purpose and benefits of OER as well as comprehensive knowledge of copyright matters play a role in encouraging academic staff to engage in OER-related activities. Recent studies indicate that whilst there is sufficient familiarity, at a surface level, with copyright legislation and Creative Commons licensing in at least 300 of the academics surveyed, fewer had in-depth knowledge of both (Dhanarajan & Abeywardena, 2012). Institutional policies to incentivise, through recognition and rewards, the production and use of OER are also somewhat thin in most Asian institutions.

Table 1.12: Policy matters (after Dhanarajan & Abeywardena, 2012)

Institutional policy items	Yes	No	Total [N]
Knowledge of copyright	63 [97%]	24 [3%]	65
Knowledge of CC licences	41 [63%]	24 [37%]	65
Provisions for sharing, collaborating in and using OER	13 [18%]	58 [82%]	71
Provisions for incentivising OER participation	25 [35%]	46 [65%]	71
Provisions for staff development	29 [42%]	40 [58%]	69

---

[3] http://en.wikipedia.org/wiki/Open_educational_resources

Skills at using the technologies buttressing OER

Adequate national ICT infrastructures, such as telephony, access to computers, adequate bandwidth, freedoms relating to using the Internet, exploring the WWW for content through search engines, as well as knowledge of and skills to use a range of appropriate software are all important prerequisites for greater participation in OER-related activities. As mentioned earlier, most Asian nations have adequate ICT provisions. Skills to use computers and access to the Internet are also adequate; however, the limited availability of bandwidth and appropriate software to access, remix, reuse and redistribute content requires further and additional investment. The poorer nations and their institutions (especially in the rural areas) are somewhat handicapped in this aspect. Until all the technologies buttressing OER are freely and easily available, many developing Asian countries will not be in a position to benefit from the full potential of OER for a little whilst to come.

Conclusion

Whilst interest in and the production, distribution and use of OER are still very much in the early stages of development in most parts of Asia, OER's potential value to improve the quality of curriculum, content and instruction, facilitate academic collaboration and enhance equitable access to knowledge resources cannot be overstated. Marshall Smith, in an unpublished paper (2011), articulated this elegantly:

> Knowledge should be universal but is unequally and unfairly distributed and OER will help to overcome the gaps. A second narrative emphasize[s] the opportunity for users to become producers by having the opportunity to change and adapt OER for their purposes. This same narrative [holds] that OER [provide] new opportunities for teachers and other non-technical people to become producers of totally new open content and tools. A third narrative holds that OER [have] the potential to transform opportunities for learning and teaching by providing opportunities for students to learn on their own for free and from others (peers, mentors) on the networks and in the crowd, and to potentially get credit for the learning.
>
> All of these narratives are still operable. A fourth narrative is about fulfilling the first three in the developed world and, more importantly, in the developing world. This is the narrative of implementation, helping to create appropriate technical infrastructure including the necessary tools such as platforms and Creative Commons licences to construct quality open materials, making it possible for OER to be easily accessed and used, and supporting local communities, government and NGOs in their efforts to use OER effectively. This is the narrative of our times — it will not be a smooth road but the opportunities that it may provide are worth it.

It is in pursuit of especially the fourth narrative that educators and their political masters need to invest efforts in OER, which have the potential to serve a potpourri of multiple purposes in Asian higher education.

Acknowledgements

References

Asian Development Bank. (2009). *Good practice in information and communication technology for education*. Manila: Asian Development Bank. Retrieved from http://www.adb.org/sites/default/files/pub/2009/Good-Practice-in-ICT-for-Education.pdf

Asian Development Bank. (2011). *Higher education across Asia: An overview of issues and strategies*. Manila: Asian Development Bank. Retrieved from http://www.adb.org/sites/default/files/pub/2011/higher-education-across-asia.pdf

Asian Development Bank. (2012). *Access without equity? Finding a better balance in higher education in Asia*. Manila: Asian Development Bank. Retrieved from http://www.adb.org/sites/default/files/pub/2012/access-without-equity.pdf

Barro, R., & Lee, J. W. (2010). A new data set of educational attainment in the world, 1950–2010. NBER Working Paper No. 15902. Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w15902

Dhanarajan, G., & Abeywardena, I. S. (2012). A study of the current state of play in the use of open educational resources in the Asian region. Unpublished report of a project on open access and quality in Asian distance education.

Smith, M.S. (2011). Open educational resources: Opportunities for the developing world. Unpublished report.

UNESCO. (2009). *Global education digest: Comparing education statistics across the world*. Montréal: UNESCO Institute of Statistics. Retrieved from http://www.uis.unesco.org/Library/Pages/DocumentMorePage.aspx?docIdValue=80&docIdFld=ID

UNESCO & COL. (2012). 2012 Paris OER Declaration. Retrieved from http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/Events/Paris%20OER%20Declaration_01.pdf