CITATION ENDORSEMENT FRAMEWORK FOR DIGITAL REPOSITORY

ABDIRIZAK SA'ID GHEDI

MASTER OF SCIENCE IN INFORMATION TECHNOLOGY
UNIVERSITI UTARA MALAYSIA
2016

Permission to Use

In presenting this thesis in fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for review. I further agree that permission for the copying of this thesis in any way, in full or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis. Requests for permission to copy or to make other use of materials in this thesis, in whole or in part should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences

UUM College of Arts and Sciences

Universiti Utara Malaysia

06010 UUM Sintok

Abstrak

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

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

ABSTRACT

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

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

Acknowledgement

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

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

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

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

Table of Contents

List of Tables

List of Figures

List of abbreviations

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

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

CHAPTER ONE

INTRODUCTION

1.1 General Overview

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

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

Most of the librarians and researchers have long used citation management systems as research tools to help scholars organize their work, improve workflows, and ultimately save time. For many years, RefWorks has been the dominant citation management tool in citation management environment. However, a number of competitors now offer citation management systems that are as strong as RefWorks but offer different features to the user (Steeleworthy & Dewan, 2013).

This study proposed citation endorsement framework for digital repository which allows the author who has information about the citation to update the citation information into the repository. The added citation information will be verified by the admin or the author of the citing paper before the citation is counted under that article. This citation endorsement framework helps the academicians to share information about who is citing their work, thus reflect the impact of that article. The citation endorsement is also a platform to manage and preserve the proceeding articles and it allows the users to search the articles in the repository and retrieve the articles from the repository in order to get full value from it.

The citation endorsement framework consists of four entities namely the administrator, the author, the author of the cited paper and the normal user. The admin and the author have the privilege to login the repository and carry on some functions while the author of the cited paper verifies the added citation information before it is indexed in the repository, on the other hand the normal user can only search articles and view the citation and other profile.

1.2 Background of Study

As new information is being published, information sharing has become a general method to gain information in academic environments. The performing of information sharing have generally been defined with variations for example the use of formal vs. informal networks, social vs. technical searching, and person vs. documentary sources (Talja, 2002).

In academic environment, information sharing is one of the important tasks to be carried out by every academician. Typically every academician is keen to share their knowledge and resources with other people. However, the platform needs to be established to allow such interaction. Conferences has been one of the platforms for formal information sharing where academicians share their research findings verbally. Besides verbal presentation, the conferences also publish proceedings that compile the conference materials in to a book format and distribute to the conference participants.

Jamaludin and Ishak (2011) proposed a framework for information sharing that enables academician to share information about their academic collection. The collection of information on academic materials contributed by every academician forms a huge academic repository that is rich of information. This platform also allows other people to search the collection through the search interface. Through the interface detail, information about the collection can be viewed.

The repository that stores digital information is known as digital repository (Marshall, 1997). A digital repository is a mechanism for managing and storing digital content of books, papers, theses, and other works of interest to the institution served. Digital repository may contain information of the media or the digital media object. Digital repository is related to digital library which is a type of information retrieval system which focuses on the collection of digital objects stored as electronic media formats along with revenues for organizing the files and also retrieving the files. The electronic objects can be stored locally, or read remotely through computer networks. The digital repository can be different in scope and size, and can be maintained by individual educators, organization departments, or sometimes allied with an established existing libraries or organizations (Marshall, 1997).

A digital library is a collection of documents in organized electronic forms which are available on the Internet or accessible by computers. A digital library will direct users to electronic collections, such as magazine articles, books, papers, images, sound files, and videos. To date, due to large number of electronic materials, many digital libraries have been developed either for free or commercial purposes. Examples of digital libraries are Mountain West Digital Library (http://mwdl.org/index.php), World Digital Library (http://www.wdl.org/en/), ACM digital library (http://dl.acm.org/) and EBSCOhost (https://www.ebscohost.com/).

1.3 Problem Statement

The existing digital repository models consist of several major functions such as, upload documents, searching and download documents. The author of the articles is also allowed to upload and manage their own articles. Some of the digital repositories have a built in function to trace the citation of the articles indexed in the repository database and the author of the articles cannot update the citation information. This is the limitation of the existing model as the citation information is automatically retrieved from certain databases and the author who has information about the citation cannot update the information into the repository.

However scholar indices have limitations including questionable of accuracy and applicability owing to the difference in spellings. Bibliographic digital libraries such as DBLP and CiteSeer contain a large number of publication metadata records and make these records searchable for academics. A problem occurs when different individuals share the same name. This leads to mismatch problems in which citations to different authors may be mixed together in a single list. Such problems can hinder scientific data gathering, information retrieval and even credit attribution (Tan Kan & Lee, 2006).

Therefore in order to minimize the problem of missed matches, a citation matching system needs to be able to deal with inaccuracies in cited references. This study proposed citation endorsement framework for digital repository which allows the author who has information about the citation to update the citation information into the repository.

1.4 Research Questions

The research questions of this study are:

i. What are the requirements to design citation framework?

ii. How to integrate citation endorsement framework into the digital repository?

iii. How to evaluate the citation endorsement framework in the digital repository?

1.5 Objectives

The aim of this study is to develop citation endorsement a framework for digital repository. Specifically the study aims:

i. To identify citation framework for proceedings articles.

ii. To integrate the citation endorsement framework into the digital repository.

iii. To evaluate the citation endorsement framework in the digital repository.

1.6 Scope

This study focuses on conference proceedings articles as conferences are one of the formal mediums for active information sharing. Digital versions of Knowledge Management International Conference (KMICe) conference proceedings are used as the case study. KMICe conferences provide a platform for international presentation of research findings as well as discussions and sharing of recent advances in the field. This conference brings together leading researchers and developers in a wide variety of areas, with a common interest in improving the state of the art of knowledge

management (KM)**.** The digital repository is equipped with the searching facility which allows users to search the articles.

1.7 Significant of the study

This study is vital in order, to contribute to the development of new citation indexing system that considers the article's owner as a part of the system, also as digital platform to manage proceedings articles, and in order to increase the citation index of the articles by allowing the authors/ articles owners to edit the citation of their articles.

1.8 Organization of the Dissertation

The chapters of this research are organized as follows: Chapter one consists of the introduction of the study, background, problem statement, research questions, research objectives, scope, the significance of the study, while the last but not the least section in the first chapter is the organization of the dissertation. The remainder of this dissertation is organized as following. Chapter 2 presents the background and related works that mainly focuses on digital repository and the information sharing model. Chapter 3 describes the methods and tools, which were used in this research for the design of the digital repository and also the acceptance method that were used to evaluate the expert review of the system. Chapter 4 presents the integration of the digital repository and citation endorsement framework. Chapter 5 presents the thematic analysis results. Chapter 6 presents the contribution of the study, conclusion, limitation and future work.

CHAPTER TWO

LITERATURE REVIEW

2.1 General overview

Information sharing defines the interchange of data between different people, organizations and technologies. There are different kinds of information sharing: information shared by individuals, information shared by organizations and information shared between firmware/software (such as the Internet Protocol (IP) addresses of available network nodes or the availability of disk space). The beginning of extensive distributed networks; cross platform compatibility and normalization of IP protocols have all eased the massive development in worldwide information sharing (Stuckenschmidt & Van Harmelen, 2005).

In old days data were normally kept in storage and often not shared among other individuals because of its limited, non-portable format or the absence of skill to import/export data. Even simple items such as dates were stored in a whole range of different formats making the sharing of such a simple field a potential nightmare. The same applied to an entire range of data, and even if it were well matched it was often not possible to physically transfer the data from one platform to another. Today, these difficulties have all been coded out and information sharing is common between

computer networks. Information sharing has become particularly widespread due to social networking. The twenty first century network models keenly inspire the sharing of information through social networks.

Nowadays people have more chances to share information due to the improvement of the technology for information access. In the past, information sharing was viewed as a business procedure of knowledge markets, where the information consumers and suppliers needed to have the same benefits from the exchange (Stuckenschmidt & Van Harmelen, 2005). Therefore, the motivation for information sharing was considered reputation that was expected the same benefits and trust.

Digital repository which is one of the methods of information sharing consists of multiple knowledge foundations as well as the tools for collection, management, and use of the information. However, the digital repositories are of limited value without an extended and renewable supply of information. The success of information sharing requires that information providers are willing to contribute their information and that information seekers are willing to reusing the well-arranged information.

2.2 Information Storage and Retrieval System

Information storage and retrieval is the systematic procedure of gathering and categorizing data so that they can be located and displayed on request. Computers and data handling systems have made possible the high speed, selective retrieval of large

amounts of information for government, commercial, and academic purposes. There are several basic types of information storage and retrieval systems. Document retrieval systems store entire documents, which are usually retrieved by title or by key words associated with the document. In some systems, the text of documents is stored as data. This permits full text searching, enabling retrieval on the basis of any words in the document. In others, a digitized image of the document is stored, usually on a write once optical disc (Onwuchekwa & Jegede, 2011).

Database systems store the information as a series of discrete records that are, in turn, divided into separate fields (e.g., name, title, and date); records can be searched and retrieved on the basis of the content of the fields. The data are stored within the computer, either in main storage or secondary storage, for ready access. Reference retrieval systems store references to documents rather than the documents themselves. Such systems, in response to a search request, provide the titles of relevant documents and frequently their physical locations. Such systems are efficient when large amounts of different types of printed data must be stored. They have proven extremely effective in libraries, where material is constantly changing (Onwuchekwa & Jegede, 2011).

Information retrieval is the act of finding information resources related to the needed information from a group of information resources. The information search is based on metadata or on full text indexing. Computerized information retrieval systems are used to decrease the information overload (Onwuchekwa & Jegede, 2011). Several public libraries and universities practice information retrieval systems to offer access to

the library collections such as, journals, books and other documents. The Web search engines are the greatest noticeable information retrieval applications.

The information retrieval procedure starts when the user enters the query. A query which is formal statement of information needs into the system, for instance, search sequences in web search engines. The query of information retrieving does not exclusively recognize a particular object in the information collection but it will recognize several items that could be equal to the information retrieval query, perhaps it will recognize the items with different levels of relevancy (Onwuchekwa & Jegede, 2011).

All the information in the database signifies the entity of an object. The user queries are accorded against the information in the database. Depending on the application that the record objects may be, for instance, if the object is text documents, images, an audio, or videos. Mostly the recorded documents are not saved or stored directly in the information retrieval system, but the documents are instead symbolized in the system by document proxies or they are symbolized as metadata.

Almost every information retrieval system calculates a numeric mark on how well each item in the database equals the query and the system ranks the items based on that value. The top ranking items are then presented to the user. The procedure can then be repeated if the user wishes to upgrade the query.

2.2.1 Types of Information Retrieval Systems

There three popular information retrieval systems namely, multimedia information retrieval system, digital library and distributed information retrieval system.

i. Multimedia Information Retrieval

Most of information retrieval systems are modified to work with videos or images collections. The user queries are stated as an example image or as a text , if the query is stated in a text form, in order to find the relevant information the system use the text that is in the image caption or the text of the music description. In this case if the query is an image or video it can be preserved as a digital signal by using the traditional information retrieval technology. The techniques of information retrieval for example the vector's model can be stretched to calculate the resemblance between the two signals, where the structures in the vector space model will not count the amount of frequency in text, but it will total the structures which the digital signal handling procedures removed.

There are several differences between the multimedia information retrieval and the traditional information retrieval system. First, the textual data structure is less compound than the multimedia objects structure. It needs integration of the database of multimedia management system to manage sufficiently, and store the objects of multimedia. Secondly, in order the resemblance degree of user queries to match the multimedia documents and to rank the documents of multimedia that the system

retrieved the similarity measure needs to be extended. Third, the query languages are more complex (Onwuchekwa & Jegede, 2011).

ii. Digital Libraries

In order to store records of files from the library items, normal libraries become one of the first people to use the information retrieval systems so that the users can search the files in the library over the web. The files recorded in the database technology are arranged based to standards such, the title, and a classification number or a few subject headings. As an outcome of the development in electronic distributing, modern libraries are being converted to digital libraries, which make the information available over the web in a digital method (Marshall, 1997).

Over the web page, a single interface offers admission to local resources, along with distant admission to the databases in the business, science and civilizations, including newspapers, journals and directories. Very good collections become available over the same gateway not only in text format but in multimedia too. Many traditional libraries carried out digital library plan to reach the ability to exchange and use information and comfort of access and use to the library.

iii. Distributed Information Retrieval Systems

While the group of documents is distributed, an index is made for each partition, but an integrated index is yet required in order to uninterrupted the search for the expressions

in the user's query. One of the special types of distributed information retrieval systems are Peer-to-Peer (P2P) information retrieval systems. In P2P system, the information can be frequent on different computers and there is no central access control. In a P2P system, the servers are self-governing; each server can leave the P2P system or enter the P2P system any time (Dawes, 1996). Great examples of P2P systems are Gnutella and Napster.

2.2.2 Information Retrieval Model

The models of the information retrieval can be used on any text collection. Three best information retrieval models are: The Boolean, the Vector Space, and Probabilistic Model (Onwuchekwa & Jegede, 2011).

i. Boolean Model

Queries are Boolean terms of keywords and when it is needed to implement the Boolean model the document is signified as a set of keyword which connected by NOT, AND, and OR, and in order to specify the scope of these queries we need to include the use of brackets.

There will be no half match output or half ranking in this model, the system output will be a list of documents that are relevant. In the Boolean model, OR = any, AND = all, which makes the model very rigid. It is very hard to manage the number of documents that the system will retrieve because all the documents that fit the user query satisfy the query to the same degree and the system will return all matched

documents, so that makes it hard to rank the documents that retrieved. Additional weakness of the Boolean model information retrieval is that for the users it is not simple for the user to prompt multifarious information retrieval queries (Onwuchekwa & Jegede, 2011).

ii. Vector Space Model

This information retrieval model, which proposed by Salton in 1989, is very successful statistical method. The model generates one sided term vectors for the user query and for each document in the collection. The information retrieval of this model is based on the similarity between the document vectors and the query vector (Onwuchekwa & Jegede, 2011). The similarity of the retrieval of document is based on the frequency amount of the keywords in the document and in the query and the retrieved documents are ranked according to the similarity.

iii. Probabilistic Model

This information retrieval model was proposed by Robertson and Sparck Jones in 1976. This information retrieval model is based on idea called the "ideal answer set". Given a user query, in probabilistic model, there are fixed documents that are relevant document to the user query (Onwuchekwa & Jegede, 2011).

The user query of probabilistic model is a technique for identifying the assets of the answer set, but in the probabilistic model the user doesn't know what the real properties of the answer set are, so that a work has to be made to predict what are the

explanation of the answer set and retrieve a set of documents initially. After that the user looks for the related documents by reviewing the top retrieved documents. By repeating this process the description of the ideal answer set improves because the information retrieval system uses this information to refine the description of the ideal answer set.

2.2.3 Search Engine

Search engines are a software package that search documents for detailed keywords and returns a list of the documents where the keywords were match. A search engine is really a general class of programs; however, the term is often used to specifically describe systems like Google, Bing and Yahoo, their search enables user to search for documents on the World Wide Web (Purcell, Brenner, & Rainie, 2012).

Web Search Engines:

Normally, Web search engines operate by sending out a spider to get as various documents as possible. Then another program called an indexer gets these documents and then, based on the words that each documents contained the indexer creates an index of that documents. Each search engine uses an exclusive algorithm to create its files so that only significant results are returned for each query (Hogan et al., 2011).

Some of the popular search engines are AllTheWeb (www.alltheweb.com/), Google (www.google.com/), MSN Search (search.msn.com/) and Yahoo (search.yahoo.com/). Figure 2.1 shows a snapshot of Google search engine.

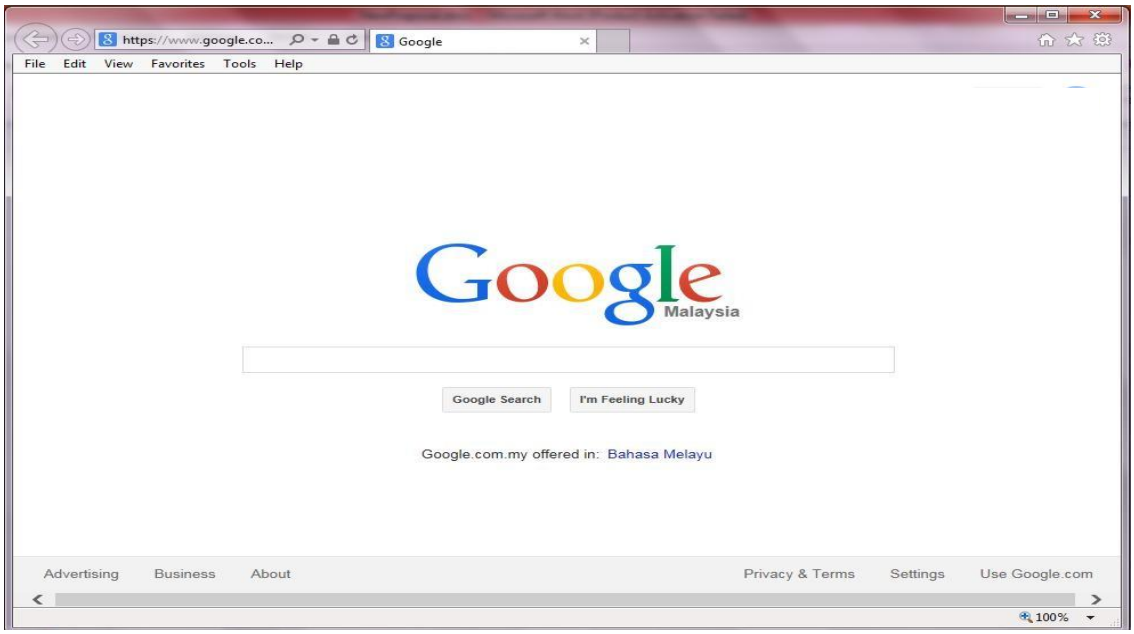


Figure 2.1: **Snapshot of Google search engine**

The meta search engines chains several existing search engines and offer documents related to query of the user. The meta search engines job is to reduce the ranking of the results from the several search engines and removes the duplicates (Hogan et al., 2011).

2.3 Electronic Libraries

Electronic libraries are physical sites or websites that provides 24 hour online access to digitized written materials. This topic discusses the digital repository and the digital library which are two different things. Digital repository makes the intellectual output of an organization or department freely and openly available while on the other hand, digital library is a gateway to electronic resources (Rossi, 2011).

2.3.1 Digital Library

The concept of Digital Library is based on the idea of a library or information service providing virtual electronic access through the Internet to documents and or services. A digital library will direct users to electronic collections, such as research papers or video clips that document in great detail the history of a particular locale or area or that offer a unique thematic perspective on a subject. The main benefit of constructing a digital library is the ability to provide 24 hour, remote access to high demand items from multiple users worldwide (Rossi, 2011).

Example of Digital Library

World Digital Library (WDL): (http://www.wdl.org/en/) The WDL makes it possible to discover, study, and enjoy cultural treasures and significant historical documents on one site, in a variety of ways. Content on the WDL includes books, manuscripts, maps, newspapers, journals, prints and photographs, sound recordings, and films (Rayman, Bertram, & Prom, 2014).

USC Digital Library (USCDL) (http://digitallibrary.usc.edu/cdm/). This digital library is owned by the USC Libraries and its mission is to support the creation, preservation and discovery of knowledge and develop collections and services that help and encourage the academic endeavors of faculty, students and staff. This Digital library offers digital images of drawings, illuminated manuscripts, maps, photographs, posters, prints, rare illustrated books, as well as audio and video recordings. A portion of the

images contained in the USC Digital Library come from the collections of collaborating institutions which, like USC, have valuable archival collections that are of interest to a wide range of people (Kichuk, 2015).

There are many virtual libraries that are owned and managed by the webmaster and the control of the collection is with webmaster instead of the owner of the collection. Therefore some studies (Kichuk, 2015) consider these models as outdated because it is authoritative, making the owner of the collection left without any hold over their own collection. The owner are regarded as having the same status as the ordinary users, where they can only perform routine activities of searching, checking, scheduling, sorting, and printing without any power to edit, add, delete, and other activities that involve changing the entry in the database.

Digital repository and the digital library are two different things. Digital repository makes the intellectual output of an organization or department freely and openly available while the other hand digital library is a gateway to electronic resources (Rossi, 2011). A digital library provides access and meta data. It also maintains cross references between papers.

A digital library offers access and information retrieval from its collections while digital repository offers storage to collections, not access, nor information retrieval from individuals, unless dissemination of such collections are established by the owner of the information and the repository person in charge.

Basically digital library means all kind of library resources are available in digital format and digital repository means repository is an online archive for collecting, preserving, and distributing digital copies of the academic output of an institution, particularly a research institution.

2.3.2 Digital Repository

A digital repository is where digital content, assets, are kept and can be searched and retrieved for later use. A digital repository helps devices to identify, export, import, store and retrieve digital properties. Positioning digital assets into a repository allows staff and institutions to then manage and preserve it, and therefore get maximum value from it. Digital repositories may contain research outputs and journal articles, theses, e-learning items and teaching resources or research data.

A digital repository can keep a widespread collection of materials for a multiplicity of purposes and users. It can help research, learning, and administrative processes. However, repository solutions are most viable and sustainable when they are built on open standards (Kichuk, 2015).

The great benefit of repositories is that they support institutions to develop intelligible and organized approaches to the capture, identification, storage and retrieval of their academic resources. These academic resources go beyond usual publishing regimes, and may include datasets, presentations, learning materials and research works. A managed approach to these resources enhances chances for effective use of existing research,

increases chances for enhanced learning experiences and inspires cooperation within and between different disciplines and groups (Kichuk, 2015).

One example of digital repository is Universiti Utara Malaysia (UUM) Repository (http://repo.uum.edu.my/). The UUM Institutional Repository service is established to host the full text of published research resources created by members of the University. It is part of the UUM Library Archives. The material in the repository is open to be browsed, searched, read or printed by anyone interested in its content.

2.4 Institutional Digital Repository

The mission of institutional repositories is managing digital objects for real task. Institutional repositories creation is a task as well as a chance for academic professionals. It may possibly include a variety of research findings of any organization. The institutional repository is used to ensure that the published work of scholars is available to the academic community even after increases in subscription fees within libraries.

Mostly the research scholars don't provide unrestricted access to their research findings to their colleagues in an organization. Institutional repository offers scholars with a common step so that everyone in the institution can give scholarly materials to encourage cross campus interdisciplinary research. An institutional repository is an online store for collecting, preserving, and distributing digital copies of the knowledgeable findings of an institution (Lynch, 2003).

An institutional repository is an established facility that an institution or university offers to the members of its community for the organization and distribution of digital objects produced by the institution and its community members. The digital objects include materials such as books, theses and academic journal articles. An institutional repository may also contain other digital materials generated by academics, such as organizational documents, learning objects, conference proceedings or course notes.

One of the main objectives for having an institutional repository are to offer open access to institutional research findings by self-archiving it, to create global visibility for an institution's scholarly research, and to record and reserve other institutional digital materials (Lynch, 2003).

2.4.1 Digital Content and Archive

Digital content also known as digital media is any type of content that exists in the form of digital data; digital content is stored on either digital or analog storage in specific formats. The forms of digital content include information that is digitally streamed, broadcast or contained in computer files. The types of digital content include popular media types, while a broader approach considers any type of digital information as digital content. While digital archive is a specially designed system in a controlled operating environment dedicated to the ongoing managed storage of digital content (Ross, 2012).

2.4.2 Types of Digital Repository

There are four well known types of digital repository, namely the

i. Research Repositories

This type of repositories is used by organization or researchers in order to capture results. The results that are intended to capture are the publications, including books and other data which considered a worth capturing result which leads to a collection with a different items. Because these items create a record of science, values for deposit and maintenance must be strict. The promoter of the repository is likely to link reporting functions to the deposit mandate, for example, the reporting of grantees to the sponsor or the management of research results in an annual report. Research repositories are likely to contain high quality findings, because the content of the research repositories is peer-reviewed multiple times and the production of the results is well funded. The research repository users who are collaborators, challengers or initiating a new research project are most likely to find the collections of significance (Armbruster & Romary, 2009).

ii. National repository

This type of repositories are developed to record scholarly findings in general and also to support, for instance, teaching and learning in higher education. Indeed, only a national determination will validate the national asset. The national repositories are likely to show the scholarly findings in the national language and focus the publications of noticeable scholars and develop a system for recording dissertations (Armbruster & Romary, 2009).

iii. Subject based repositories

These types of repositories are built by specific community members and are adopted by the wider community. The subject based repositories are essential to scholars so that the natural self-archiving is customary. Subject based repositories are thematically well defined and alert services and usage statistics are meaningful for community users. The subject based repositories as part of a national research library that serves scholarly communication in the national language and supports public policy (Armbruster & Romary, 2009).

iv. Institutional repositories

This is the type of repositories that record the different findings of the institution. While the institution research findings are important among the results, so are works qualification and teaching objects. If the repository captures the whole findings then it's both a library and a showcase, it's a library for holding institutional collection and it's a showcase because of the online open access show and the collection may serve to impress and connect. Furthermore, an institutional repository could have an important function in regional development. It allows firms, public bodies and civil society organizations to understand immediately what kind of expertise is available locally (Armbruster & Romary, 2009).

2.4.3 Features of digital repository

Digital repositories have many features; the following are the main features of Digital Repository (Tansley et al., 2003).

i. **Flexible:** the repositories should be flexible about the format of the data.

ii. **Make data submission easy:** the repository should fit into the copy submission of its partner journals.

iii. **Gives Options:** the repository should give journals the option of making data privately available during peer review.

iv. **Assigns data Digital Object Identifiers (DOIs):** the repository should assign DOIs to data so that researchers can get professional credit over data citation.

v. **Promotes data visibility:** the repository should allow the content to be searched, retrieved and indexed through interfaces.

vi. **Free download:** the repository should allow the content to be downloaded freely and have no legal barriers to reuse.

vii. **Update option:** the repository should have update feature so that the submitters may update data files when corrections or additions are desired.

viii. **Long-term preservation:** by roaming common file formats when older versions become outdated and joining with data observation network for earth promise access to its contents forever.

2.5 Citation Management

Citation management tools allow a user to organize and retrieve information, such as citations for books, articles, and Web sites, by interfacing with library databases. The citation manager then works with word processing software to insert properly formatted footnotes or citations into a paper and create a properly formatted bibliography (Hensley, 2011).

Some example of citation management tools are:

i. **Zotero:** is free and open source reference management software to manage bibliographic data and related research materials (such as PDF files).
ii. **Mendeley:** is a desktop and web program for managing and sharing research papers, discovering research data and collaborating online.
iii. **EndNote:** is a commercial reference management software package, used to manage bibliographies and references when writing essays and articles.

2.5.1 Citation Indexing Services

A citation indexing systems indexes the links between articles that researchers make when they cite other articles. Citation indexes are very useful for a number of purposes, including literature search, evaluation, and analysis of the academic literature. A citation index is a kind of bibliographic database that indexes the citations between publications which allows the user to easily establish which later documents cite which earlier documents. A form of citation index is first found in 12th century Hebrew religious

literature. Legal citation indexes are found in the 18th century and were made popular by citators such as Shepard's Citations. In 1960, Eugene Garfield's Institute for Scientific Information (ISI) introduced the first citation index for papers published in academic journals, first the Science Citation Index (SCI), and later the Social Sciences Citation Index (SSCI) and the Arts and Humanities Citation Index (AHCI). The first automated citation indexing was done by CiteSeer in 1997 (Caragea et al., 2014).

CiteSeer was a public search engine and digital library for scientific and academic papers, primarily in the fields of computer and information science. It is often considered to be the first automated citation indexing system, has a patent on this topic, and was considered a predecessor of academic search tools such as Google Scholar. CiteSeer, like engines and archives, usually only harvest documents from publicly available websites and do not crawl publisher websites. As such authors whose documents are freely available are more likely to be represented in the index (Williams et al., 2014).

CiteSeer was replaced by CiteSeer^{X}, mostly in the fields of computer and information science, and all queries to CiteSeer were redirected. CiteSeer^{X} is a public search engine and digital library and repository for scientific and academic papers primarily with a focus on computer and information science (Teregowda et al., 2010). However, recently CiteSeerX has been expanding into other scholarly domains such as economics, physics and others. Released in 2008, it was loosely based on the previous CiteSeer search engine and digital library and is built with a new open source infrastructure, SeerSuite, and new algorithms and their implementations.

The following are examples of Citation Indexing Systems:

i. Google Scholar

Google Scholar (**http://scholar.google.com/**) is a freely accessible web search engine that indexes the full text or metadata of scholarly literature across an array of publishing formats and disciplines. Google scholar were released in 2004, the Google scholar index includes the most peer reviewed online journals and the largest scholarly publishers and scholarly books (Falagas, Pitsouni, Malietzis, & Pappas, 2008).

ii. Citeseer^{x}

Citeseer^{x} (**http://citeseerx.ist.psu.edu/**) is a scientific literature digital library and search engine that focuses mostly on the literature in computer and information science. Citeseer^{x} aims to increase the distribution of scientific literature and to offer enhancements in functionality, usability, availability, cost, comprehensiveness, efficiency, and timeliness in the access of scientific and scholarly knowledge.

iii. Scopus

Scopus (**http://www.scopus.com/**) is a bibliographic repository containing abstracts and citations for academic articles. Scopus covers around 22,000 titles from over 5,000 publishers. Around 20,000 of those articles are peer-reviewed journals in the field of medical, scientific and social science. Scopus is owned by Elsevier which is academic publishing company and Scopus is available by online (Falagas et al., 2008).

iv. PudMed

PudMed (**http://www.ncbi.nlm.nih.gov/pubmed**) is repository which has been indexing biomedical literature since 1879. PudMed is developed in order to provide health professionals access to information necessary for education, research and health care. In the PudMed collection, over 24 million records represent articles in biomedical literature and a small selection of items from National Center for Biotechnology Information (NCBI) books (Falagas et al., 2008).

v. Web of Science

Web of Science (**http://webofscience.com/**) is an online subscription based citation indexing service. Web of Science is maintained by Thomson Reuters which is a major multinational mass media and information firm that provide a comprehensive citation search. Web of Science gives access to multiple databases that reference cross disciplinary research which allows for detailed exploration of specialized subfields within scientific or an academic discipline (Falagas et al., 2008).

vi. Indian Citation Index

Indian Citation Index (**http://www.indiancitationindex.com/**) is an online citation data which covers peer reviewed journals published from India. It covers major subject areas such as scientific, technical, medical, and social sciences and includes arts and humanities. The citation database is the first of its kind in India. Each of these offers an index of citations between publications and a mechanism to establish which documents cites which other documents (Giri & Das, 2011).

Table 2.1 below is the comparison of five digital repositories, namely, Google Scholar, Citeseer[x], Web of Science, PubMed and Scopus.

Table 2.1 Comparison of five citation indexing systems

	Components	**How they Work**	**Features**	**Scripted Language**
Google Scholar	Search Citations Inclusion Metrics Publishers Libraries	The Google scholar allows the users simple way to search the literature of the scholars, the user can search through many sources for instance, articles, theses, books. Google scholar makes easy for the user to find relevant work across the world of scholarly research (Falagas et al., 2008).	-Search all scholarly literature from one convenient place -Explore related works, citations, authors, and publications -Keep up with recent developments in any area of research -Check who's citing your publications, create a public author profile.	Google Scolar has a very open development environment with many different service APIs. Java, Javascript, C++, Python, Go, Sawzal (a custom logging language), and probably a few other languages are supported. Search is mostly based on C++ and some Python.

Citseer[x]	-Submit and index documents -Authors -Advance Search -Most Cited documents -Venue impact rating	Citeseer[x] attempts to provide resources such as algorithms, data, metadata, services, techniques, and software that can be used to promote other digital libraries. .	- Reference Linking - Citation context - Related documents - Full-Text Indexing - Powerful Search	- XHtml - JavaScript - PHP - Css - MySQL
Scopus	-My Scopus -Alerts -My List -Remote access activation	Scopus is one of the largest citation databases of peer-reviewed literature, for instance, journals, books, and conference proceeding (Falagas et al., 2008).	- Live chat - Elsevier - Discover - Analyze - Search	- Html - Javascript - Jsp - Css - MySQL
PubMed	-Resources -How To -PubMed Tools	PubMed comprises more than 24 million citations for biomedical literature	- Comprehensive search - Journal article	- Html - JavaScript - Css

	-PubMed Tutorials -New and NoteWorthy	from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites (Falagas et al., 2008).	parameters - Secondary ID	- MySQL
Web of Science	-Products and Tools -Benefits and Resources -Training and Support - News and Events	Provides a comprehensive citation search. It gives access to multiple databases that reference cross-disciplinary research, which allows for in-depth exploration of specialized sub-fields within an academic or scientific discipline (Falagas et al., 2008).	- InCites - Journal citation report - Essential science indicators - Endnote - help files - Change search language	- Html - JavaScript - Css - Jsp - MySQL

2.5.2 Citation Extraction Method

Building the tools that collect and organize research literature can offer understanding into the landscape and process of science and help the individual researchers be more resourceful. For example, the analysis of citation graphs between papers can enable automatic collecting for learning developments in scientific sub communities and can support researchers in finding related work (Anzaroot & McCallum, 2013).

Sometimes such bibliographic data is provided in unstructured form, but often the case that data is supplied only in unstructured full text. In the unstructured situation, reference sections of papers must be located, the citations divided from each other, citation fields must be extracted from within each citation, and the citations must be disambiguated. Many citations include fields such as multiple author names, paper title, journal name, volume, number, publisher, and year. Some also include publication status, web address, organization names, thesis indicators, postal addresses, and indication of publication language. Effective analysis requires extracting these fields accurately. Although the task may seem straightforward, truly high accuracy citation field extraction has been indefinable. Real world citation strings are full with wide variety and odd exceptions to common notions about their simplicity. This irregularity makes rule based methods hard, and machine learning methods have become the tool of choice for citation field extraction (Anzaroot & McCallum, 2013). The most widely used labeled data in citation field extraction is the Coriolis Ocean database ReAnalysis (CORA) Field Extraction dataset.

2.6 Collaborative Information Sharing

Collaborative information sharing allows academicians to work at the same time, with the same information integrated from multiple sources. It has the prospective to support the revolution of academic organizations and communication (Zhao & White, 2012). This topic discusses the collaborative digital archive and information sharing model.

2.6.1 Collaborative Digital Archive

The cooperation work between institutions has long been seen as collaboration. Some of the institutions collaborative work has just been labor sharing arrangements where the institutions share group resources to achieve a result (Koh, Gunasekaran, & Rajkumar, 2008). These have accomplished fine work, but the collaboration has been linear; the sharing or cooperation was just a logical way of separating labor. Other activities have relied less on linear collaboration, and crossed boundaries of time. One example of this type of collaborations is the research institution that builds on the results or ideas of a previous generation. This has been an important way to advance knowledge and interchange ideas.

But academic collaboration has been notably difficult to manage, because it has been limited basically to things like co-authored research and conferences. Managing these kinds of projects has always meant coordinating work by mail, by telephone, or by electronic mail (Koh et al., 2008).

2.6.1.1 Scholarly Publishing

According to the Association of Research Libraries (ARL), scholarly publishing is defined as the creation, dissemination, and application of new knowledge which is fundamental to the development of a well-educated community (Cullen & Chawner, 2011). Institutions of higher education exist to fulfill these functions. From the lab to the classroom to industry to the public, the improvement of knowledge through research and teaching is a priceless contribution made by higher education to the public good. Scholarly publishing is the process through which newly discovered knowledge is refined, certified, distributed to, and preserved for researchers, professors, students, and the public (Cullen & Chawner, 2011).

The typical process of the scholarly publishing is (Klingner, Scanlon & Pressley, 2005):

i. The author submits document to academic journal editor
ii. The editor decides whether document has sufficient value in order to be reviewed by editorial board or selected external reviewers
iii. The document will be send back to the author with a denial letter or sent on to reviewers
iv. The reviewers return the document to the editor with remarks and references
v. The editor sends document back to the author with either a denial letter or a demand for revisions
vi. The author resubmits the revised document to the editor
vii. The editor agrees or rejects document
viii. The author provides editing or proofing of final copy before publication

ix. Paper is finally published in journal

The scholar publishing model has detained influence since the beginning of scholarly publishing. It depends on many individuals playing different parts within the general process of publishing scholarly journals (Ramalho, Ana & Carlos, 2005).

2.6.1.2 Open Access

Open access means open online access to peer reviewed scholarly research. Open access is the unrestricted, direct, online availability of research articles, together with the rights to use these articles fully in the digital environment (Eysenbach, 2006). Open access is mostly intended for scholarly journals, but is also provided for a growing number of theses, book chapters, and monographs.

Open access comes in two units, (1) gratis open access, which is free online access, (2) libre open access, which is free online access and some extra usage rights. These further usage rights are often approved over the use of several detailed Creative Commons licenses. Only libre open access is completely submissive with definitions of open access such as the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (Cullen & Chawner, 2011).

According to Cullen & Chawner (2011), there are two methods where authors can offer open access that are (1) by self-archiving their journal articles in an open access repository, which also known as green open access, or (2) by publishing in an open

access journal, known as 'gold' open access. With green open access, authors publish in any journal and then self-archive a type of the article for gratis public use in their institutional repository, in a central repository (such as PubMed Central), or on some other open access website. With gold open access, authors publish in open access journals, which provide immediate open access to all of their articles, usually on the publisher's website. Hybrid open access journals are subscription journals that provide gold open access only for those individual articles for which their authors pay an open access publishing fee.

2.6.2 Active Information sharing model

The word "information sharing" has a long history in the information technology wordlist. The one to one interaction of information between a contributor and receiver is the traditional information sharing that was applied through tons of open and exclusive protocol file formats and protocol message (Stuckenschmidt & Van Harmelen, 2005). Then, electronic data interchange (EDI) that began in the late 1970s, is an effective application of commercial data connections that remains in use up today.

Over the past decade, a various settings and systems have grew on the Internet to apply and gather the information sharing. The original known forms of online information sharing situations date back to the email based conversation lists that precede the Internet (Agrawal, Evfimievski & Srikant, 2003). This original form of information sharing were applied in the form of listservs, majordomo and other unified hub

implementers of group email, these initial forms leveraged groups of email addresses and allowed a restricted form of group discussion.

Recent years, there is improved information sharing done through online information sharing systems that consist of large sources of information and allowing social communications in several methods. One of the advantage of information sharing, beyond being fast and cheap, is that the most communications are documented (Bonito, 2007).

The increase of information sharing environments and systems is however additional strong sign of the elementary need for information sharing systems. There are various stages and choices for sharing information, the social networking online tools are one of the latest improvements in online tools that encourage the information sharing. Information sharing has two helpful things on the performance of both of individuals that are sharing the information because both individual has further information than a private individual.

The benefits of information sharing affect and reduce the environmental discovery and capacity costs resulting to more relevant and timely information for each individual. Uniting the cost profits leads to a shared talented of adjusting to continuously changing networks. All shared information can be the result of speciation or partition of work (Stuckenschmidt & Van Harmelen, 2005).

Figure 2.2 below shows an active information sharing model proposed by Jamaludin & Ishak (2011). This model consists of three entities namely the webmaster (the administrator), the owner (the contributors with full control on their collection) and users (other than contributors, with limited control on the collection). The users could, at the same time, be the owner. These users have full control on their own collection but limited control on the other collections. This model is selected because it is an active information sharing model and also the repository system that is proposed in this study and this model has the same entities.

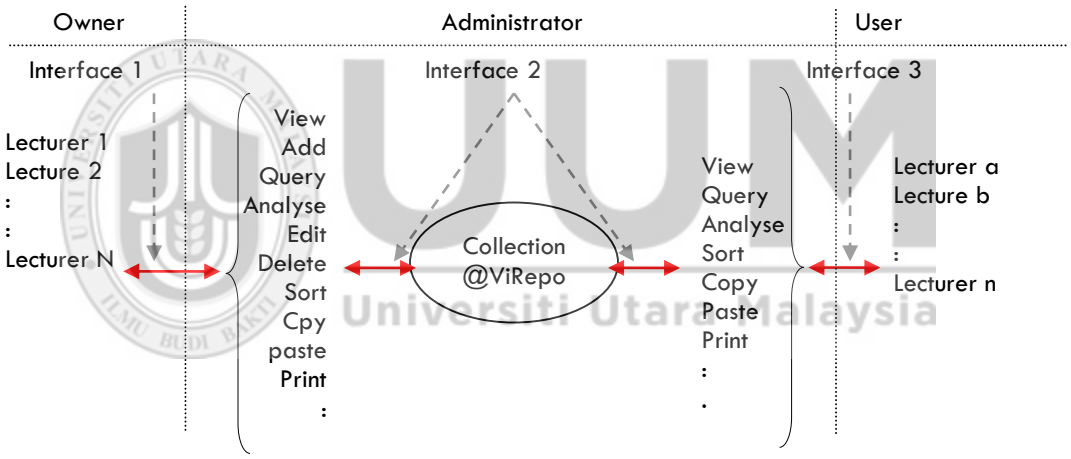


Figure 2.2: **The active information sharing model for *ViRepo* (Jamaludin & Ishak, 2011).**

The model is designed in an attempt to facilitate information distribution and information sharing at minimum effort and cost. A centralized database approach is used in the model enabling any educator in the higher learning institution to participate and manage the database, theorizing that the database is their very own personal library (Jamaludin & Ishak, 2007). ViRepo proved that the model, in actual fact, allows

information access, information sharing, information management, partnership enhancement, and allows an utmost repository control by each and every participating educator (Jamaludin & Ishak, 2011).

This model is unique as it involves the development of a virtual repository where lecturers become the contributor, the user, as well as the owner of the repository. Such a repository enables a systematic information sharing as long as each user keeps the database updated. In this model a lecturer can become a participant, a manager, and also an owner for the virtual library which in return offers another benefit in terms of information access, sharing, and management. Also, each user will obtain a better picture about the overall collection's field, possession status, and type (printed or electronic), which in turns reflects the owner's interest and field of research. This information contributes towards generating a better network between senior and junior lecturers, increase collaboration and teamwork. At the same time, the owner still maintains full control and possession towards their collection in their work space/room.

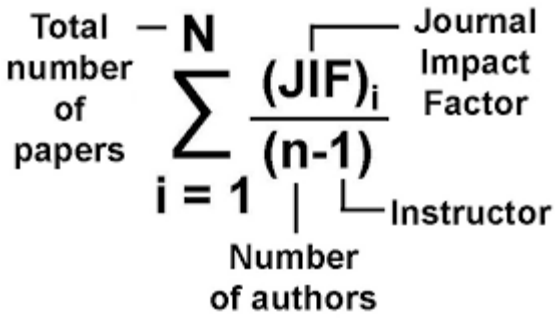
This repository is termed as faculty's repository by the webmaster/administrator, termed as personal library by the owner of the collection, and termed as virtual repository by the users who do not own any material in the collection. Owners and users have login, password, and their identification. Each item in the collection has its owner, types, authors, title, ISBN/ISSN (for books and journals), year published, and other notes that can be edited and updated by the owner.

By modeling and implementing *ViRepo*, it was believed that it would be easy for the department to expand this design into an institutional repository. The repository would in turn, serve as one of the entity under central research information system (CRIS). The ability to categorise, sort, and group items according to types, researchers, fields of research, and team members would greatly assist other researchers in their study (Jamaludin & Ishak, 2011).

2.6.3 H-index and the 10-index

Digital documents stored and indexed in the digital repository have an impact on the authors and the publishers. The impact can be seen in term of h-index. The h-index is an index that attempts to measure both the efficiency and citation impact of the published total output of a researcher. The index is based on the established of the researcher's most cited papers and the number of citations that they have received in other publications. The index can also be applied to the efficiency and impact of a scholarly journal.

The h-index was proposed by a physicist at UCSD named Jorge E. Hirsch in 2005. It was proposed as a tool for determining theoretical physicist's relative quality and is sometimes called the Hirsch index. Since the first publication of the h-index in 2005 the h-index measure has generated a well-known interest. The formula for the H-index is:



(2.1)

The advantage of the h-index is that it combines an assessment of both number of papers and the citation to these papers. A researcher cannot have a great h-index without publishing a significant number of papers. However, these published papers need to be cited by other researchers in order to count for the h-index (Minasny et al., 2013).

On the other hand, the 10-index indicates the number of scholar publications a researcher has written that have at least ten citations from other papers (Minasny et al, 2013). The 10-index was proposed by Google in 2011 as part of their work on Google Scholar. The advantage of 10-index is that it's straight forward and simple to calculate.

Therefore 10-index were included in the repository to show how many articles have more than 10 citations and the h-index function were included in the repository in order to indicate the overall impact of the academic publication for the researchers.

2.7 Summary

Literature on digital repository and citation indexing systems has provided many explanations of digital repository and citation indexing systems. There are many citation indexing systems that index the citation by using algorithm. Yet sometimes there is insufficiency of these algorithms they use because those algorithms have difficulties when indexing the citation information in terms of matching the information.

It is submitted that a comprehensive viewpoint needs to be adopted, one that combines the different theories on citation indexing. To this end a digital repository integrated with citation endorsement framework is developed in the next chapter.

CHAPTER THREE

METHODOLOGY

3.1 General Overview

This chapter explains the methodology used for this study. Four main stages are executed. The first stage is modelling the citation endorsement framework. The second is integrating the citation endorsement in to the digital repository. The last stage in this methodology is the expert evaluation, an interview were conducted on 5 experts to evaluate the effectiveness of the system. Table 3.1 below summarizes the mentioned stages.

Table 3.1: The main stages of the study

Phase	Activity	Outcome	Objectives
Modelling the citation endorsement framework	Model the citation endorsement framework	Developed Citation endorsement framework	To develop citation framework for proceedings articles
Integrating the citation endorsement in	Integrating the designed digital repository with	Developed digital repository with citation endorsement	To integrate the citation endorsement framework into the

to the repository	citation endorsement framework	framework	digital repository
Validating the proposed citation endorsement framework	Expert review evalauation	Collected the feed back of 4 librarians from uum library and 1 Associate Professesor from SOC	To get expert review on the proposed citation endorsement framework

3.2 Developing the Digital Repository Framework

In this phase is where the transition of the design and development of the digital repository is explained. The design of the digital repository includes, the designing the use case diagram and sequence diagram for the system which specifies the functionality of the system and the roles of each user. Then the database is created which consists of five tables, namely article list table which is stored the articles, citation table which is stored the citation of each article and the proceeding table which is stored the proceeding name of Knowledge Management International Conference (KMICe) conferences. Rational Rose was used for the design of the diagram and for the database was created using PhpMyAdmin and the designing of the prototype of the system were used HTML and CSS, and the programming languages which is used for the system are PhP and JavaScript.

3.3 Integrating the Citation Endorsement Framework into the Repository

The integrating of the digital repository with citation endorsement features section discusses what features are included in the digital repository. After the design of the digital repository for proceeding articles, the system is integrated with citation endorsement framework. The citation endorsement feature that was included in the system after extending are, the add cited by function which allows the authors to add the citation of the article that cited their own articles, the validation of the citation which the author of the cited paper or the admin verifies the citation and the other features that were included in the digital repository are the h-index and 10-index which indicate the impact of the publications for each other. The transition diagram of the features extended is shown in the following interface design and transition diagram.

3.3.1 Interface Design

After developing the digital repository for proceeding article, the system were extended and new information features were added, the information sharing features that were included in the digital repository are, the citation calculation function and displaying the number of citation for each article, and also the add citation function for each article which gives the owner of the article the option to add new citation for his own articles and also the h-index and 10-index of each author. Figure 3.1 below represents the interface of the information sharing features that was included in the digital repository.

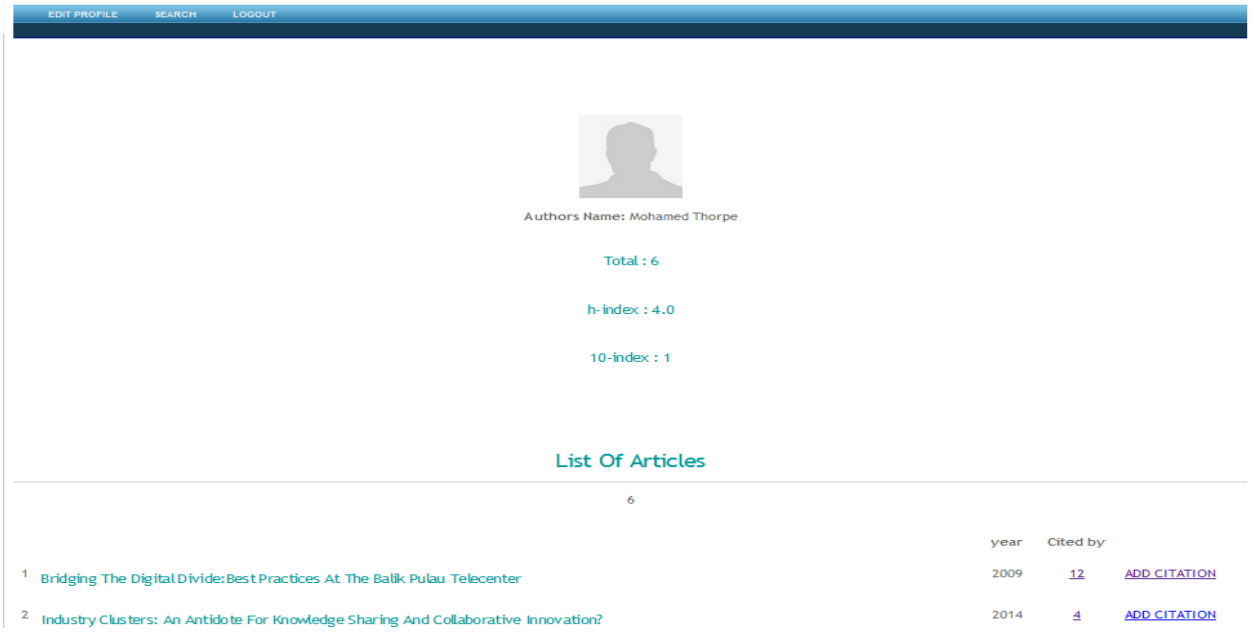


Figure 3.1: **Add citation feature in the digital repository**

3.3.1.1 Flow Chart/Transition Diagram

The following Figure 3.2 is the transition diagram of the new feature that is extended in the digital repository. The new feature which is included in the digital repository is the add citation features which allows the author to add new citations for his/her articles. The flow of the add citation is the author checks the citations and if there is new citation then the author will add the new citation and if there are no new citation that the author wants to add then the flow ends.

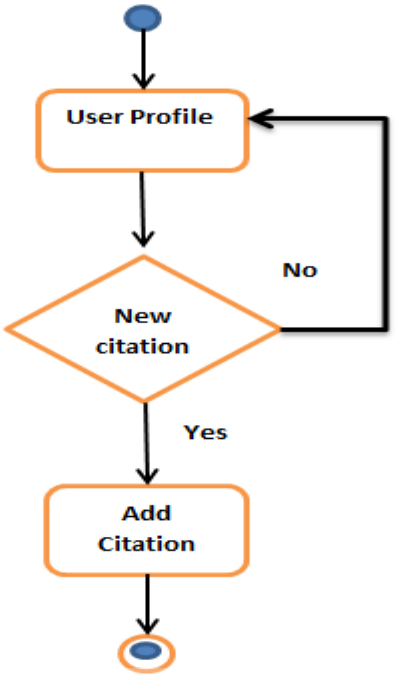


Figure 3.2: **The flow of updating citation information**

3.4 Validating the Proposed Framework

Expert evaluation is conducted in order to determine the expert feedback towards the digital repository system. The expert evaluation was conducted to determine if the system meets the requirement of this study. Since this study focuses on developing a digital repository system, the best way to evaluate the system is by getting the feedback of people who are involved or know better in indexing and citation systems, for example like librarians.

The aim of evaluation of the expert review is to get the feedback of the library officers in Sultanah Bahiyah Library.

3.4.1 Procedure

The procedures of the evaluation of expert review of the digital repository are:

i. The admin assigned username and password for the experts

ii. The experts followed the guideline in the information sheet (Appendix A)

iii. The experts opened the web browser (e.g Google Chrome, Firefox)

iv. The experts typed the URL http://repository.site90.com

v. The experts clicked the login in button and logins the system

vi. Then the experts performed the demonstration of the system.

3.4.2 Materials

The materials that were used for the evaluation of the expert review test are, information sheet (Appendix A), and then the system were put online so that the participants can access the system through web browser. Then an interview was conducted in order to get the expert feedback. The interview was recorded and transcribed as in (Appendix I). The interview questions are adapted from questionnaires for assessing system usability (Tullis & Stetson, 2004).

3.4.3 Participants

The participants of the evaluation of the expert review of the digital repository were librarians from Sultanah Bahiyah Library and Associate Professor from SOC. 4 librarians and 1 associate professor were asked to participate this expert review by giving them

username and password to access the system through online. The participants were selected based on their knowledge towards indexing and citation systems.

3.5. Thematic Analysis Method

This part, thematic analyses were done in order to analyze the interview data or summarize our data in meaningful way. Thematic analyses are very important because if the raw data is simply presented it would be very difficult to visualize what the data were showing. In order to analyze and evaluate the digital repository system, expert evaluation were done which were asked 5 experts to evaluate the digital repository system then their feedback were collected by interviewing. After the collection of the interview data, thematic analyses were used to analyze the interview data. Thematic analysis is the most common form of analysis in qualitative research. It emphasizes pinpointing, examining, and recording patterns (or "themes") within data. Themes are patterns across data sets that are important to the description of a phenomenon and are associated to a specific research question (Fereday & Muir-Cochrane, 2008).

3.6 Summary

To develop a digital repository system with citation endorsement framework for proceeding articles and to integrate with active information sharing framework, Rapid Application Development (RAD) method were used to develop the system prototype and an interviewed method were used in order to get the expert feedback on the proposed citation indexing system then thematic analyses method were used to analyze the expert feedback.

CHAPTER FOUR

CITATION ENDORSEMENT FRAMEWORK

4.1 General Overview

This chapter presents the integrated citation endorsement framework with digital repository. This citation endorsement framework helps the authors to increase the citation index of their publication by updating the citation information of their article into the repository. After the author updates the citation information then verification will be done on that citation before it increases the citation index of that article. The verification of the citation will be done the admin and the author of the cited paper.

4.2 Requirement Analysis

Requirement analysis has been performed to identify user expectations of the proposed repository. The requirement analysis is to make sure that the requirements are assessable, significant and complete. In this study the requirement analysis is used, to identify the features of the digital repository, to identify the actors and their roles and also to identify the functionality of the digital repository. The requirements for this digital repository were gathered by analyzing the existing digital repositories.

4.2.1 Identification of the Digital Repository Features

The main features of the digital repository are, Check the citing of the publications, create a public author profile, display the h-index and 10-index of the author, download the complete document through the web, search the scholarly articles in the repository and verify the citation function after the author adds the new citation.

4.2.2 Identification of the Actors and Roles

In the digital repository, three users are identified which are, the admin, the author, and the normal user. These three users have different roles in the system:

Admin: The admin role in the system is to register articles in the repository and edit article information or delete articles and also to register new members in the system

Author: The author role in the system is to check the citation of his articles and add new citation and also update his user profile

Normal User: The normal user role in the system is to search articles and view citation of the article and also the normal user can download the articles.

4.2.3 Identification of the Functionality

As for the functional requirements, this study enables us to identify the 15 functions being implemented in various repositories. There are only seven popular functions that have been used by all repositories available online. Based on the finding, it's decided to implement only those popular functions in this framework. Those function are operable by three entities namely the webmaster (the administrator), the owner (the contributors

with control on their collection) and users (with limited control on the collection). The users could, at the same time, be the owner. These users have full control on their own collection but limited control on others. Table 4.1 shows the summary of entities and functionality decided for the proposed digital repository.

Table 4.1: The entities, functions, and requirements for the framework

Entity	Function	Requirement for function
Admin	• Create Author record • Add Authors articles • Delete article • Edit Article Information • Maintain the Repository • Verify the citation	Login & Password
Author	• Add Citation • Check citation • Search Article • View other Author's Articles • Update User Profile	Login & Password
Normal User	• Search Article	No need Login

	• View Article Citation • View Publication • View Author Profile • Download Articles	

4.3 Interface Design

This phase focuses on the interface design process which is planning the system specifications in this project; the design stage uses Unified Modeling Language (UML). The unified modeling language consists of few representations such as, Object diagram and Data Flow Diagram (DFD). The designing of the system interface, it is used HTML and CSS, in order to make the interface user friendly and also it is used dolphin tabs for the interface menu. Figure 4.1 below shows the interface of the digital repository.

Figure 4.1: **The interface of the digital repository**

Figure 4.2 is the transition diagram of the digital repository. The digital repository has three users which are the admin, the author and the normal user. At the interface, the members (admin & author) have the privilege to login the system, on the other hand the normal user can only search the articles in the repository using the search function.

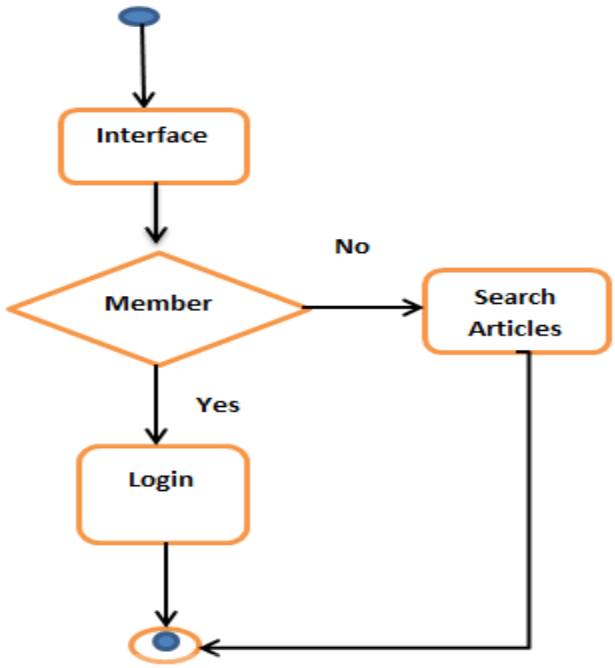


Figure 4.2: **The flow of the Digital repository Interface**

4.4 Modeling of the Citation Endorsement Process

This part focuses the modeling of the citation endorsement process in the digital repository. Within the digital repository context, the actors are involved in different kinds of information activities and processes such as, interface and sharing, searching, retrieval, and creation. The admin adds the proceeding name and then uploads the article in the repository, after the article is uploaded in the repository an email will be sent to the author which gives authenticated username and password. The author logins the system and views the list of article and if there is a new citation information, then the author will add the citation information of the cited papers specially the title of the cited paper and

the URL of the cited paper, then an email will be sent to the author of cited paper to confirm if the information is correct, if the author or the admin confirm the citation information then citation will be counted for the author and the citation number will increase.

Figure 4.3 below presents the framework of the digital repository integrated with citation endorsement framework. This model consists of three entities namely the administrator, the author and the normal user. The admin and the author have the privilege to login the repository and carry on some functions, on the other hand the normal user can only search articles and view the citation and other profile. The admin role, the author role and the author of the cited paper role in the repository are shown in figure 4.5, figure 4.6 and figure 4.7 respectively.

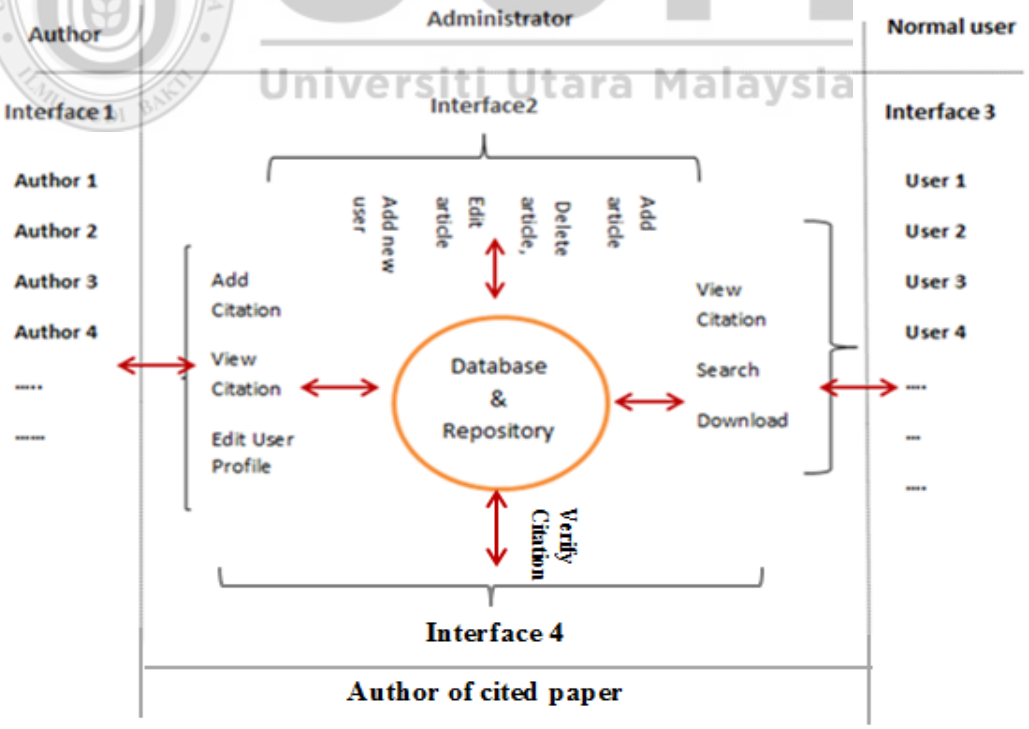


Figure 4.3: **The digital repository with citation endorsement framework**

4.5 Integrating Digital Repository with Citation Endorsement Framework

After finishing the development of the digital repository, the digital repository was integrated with citation endorsement framework. In the digital repository, it allows the admin to upload articles of the authors (UUM lectures) and after uploading the article the system will create a credential of the author who is the owner of the article by giving them a username and password. After the admin creates the record of the author then the author logins the system (digital repository) and they add the citation of the each of their article then citation validation process will start and the author of the cited paper will be sent an email which contains the article information, the article URL and the confirmation link, after the author of the cited paper confirms the citation will be recorded which then allows the other users to check the citation record of the each article.

Figure 4.4 below shows the data flow diagram of the digital repository with the citation endorsement framework. Figure 4.4 explains the role each actor (Admin, author, normal user and author of the cited paper) performs on the system. The role of the admin is add proceeding information and upload articles, the role of the author is to add citation of his own articles, the role of the normal user is to search articles then after article is displayed then the normal user can view the citation information of the article or they can download the article. The role of the author of the cited paper is to verify the added citation when he/she receives the email.

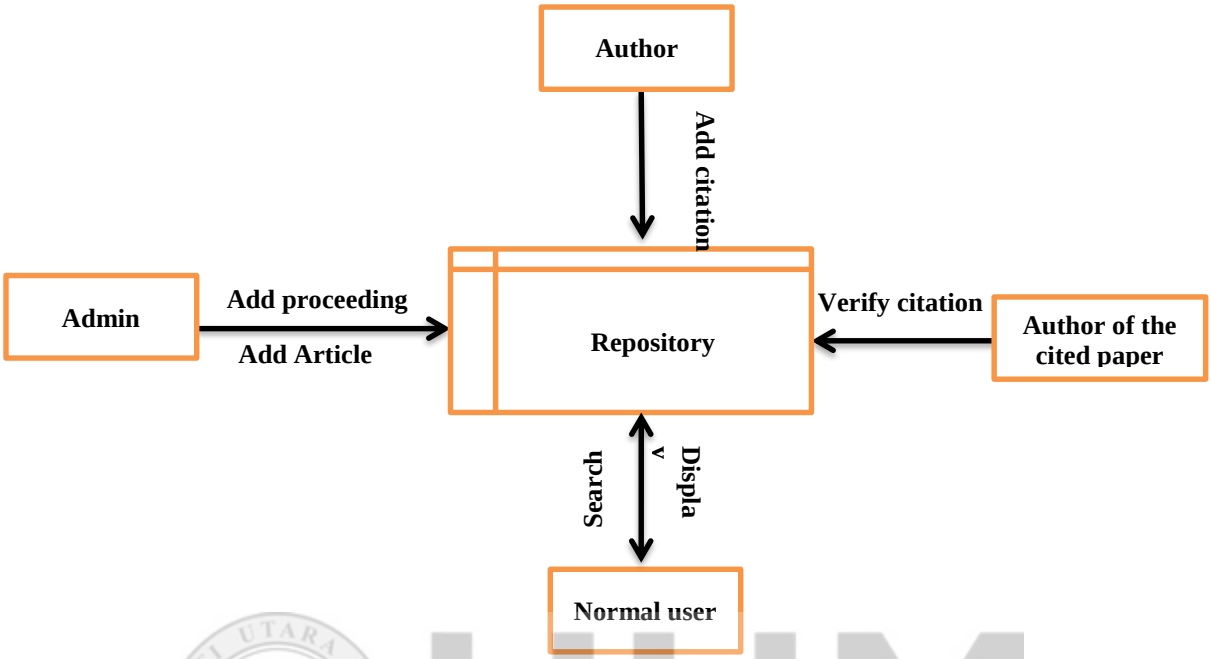


Figure 4.4: **The data flow diagram of the system**

Figure 4.5 below shows the system flow of the admins part. First the admin adds the proceeding information then the admin uploads the article, after the article is saved in the repository then the system sends email to the author of the uploaded article. The email contains a username and password for the author which he/she can use to login in the digital repository system.

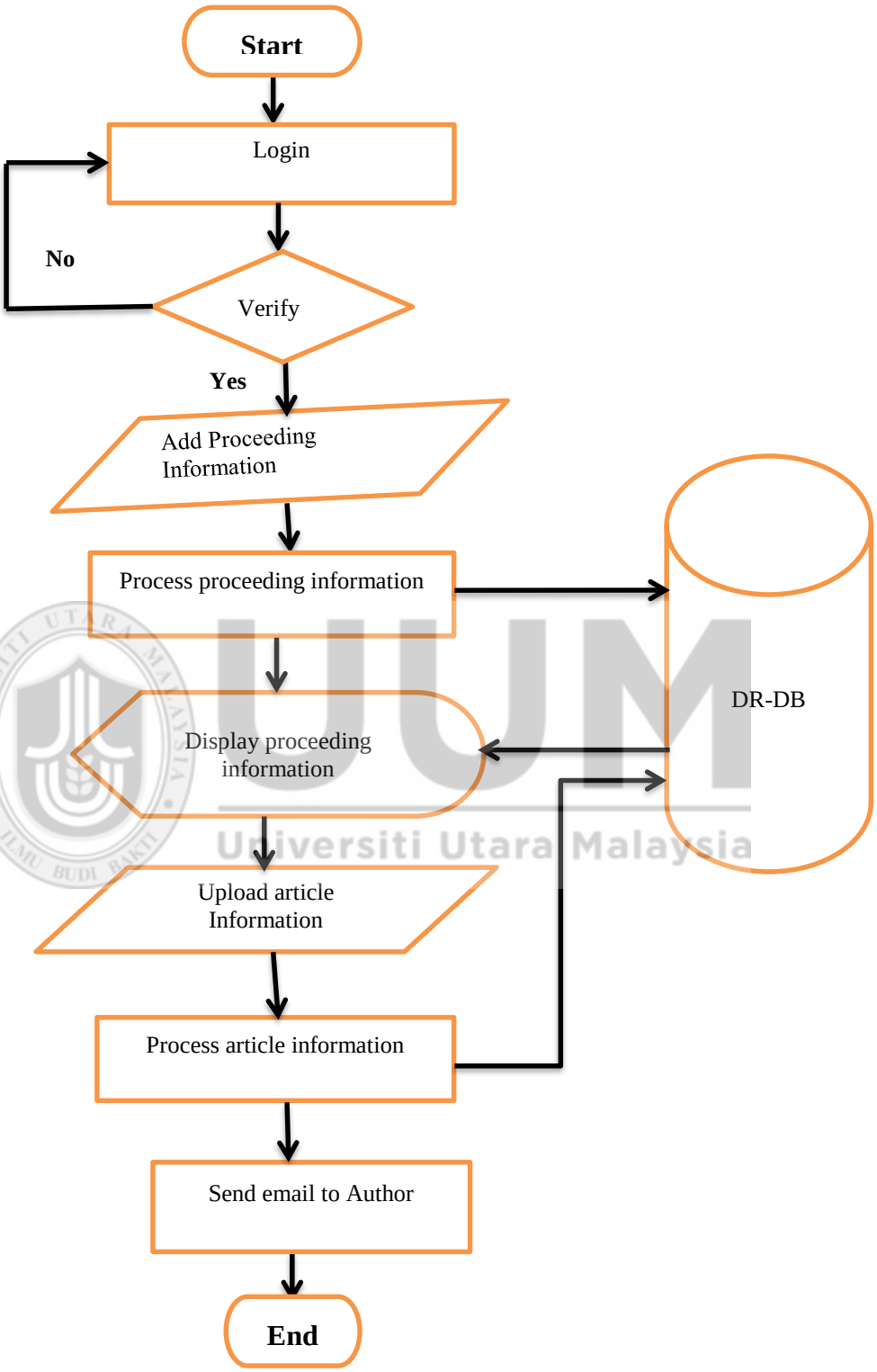


Figure 4.5: **The system flow of the admin**

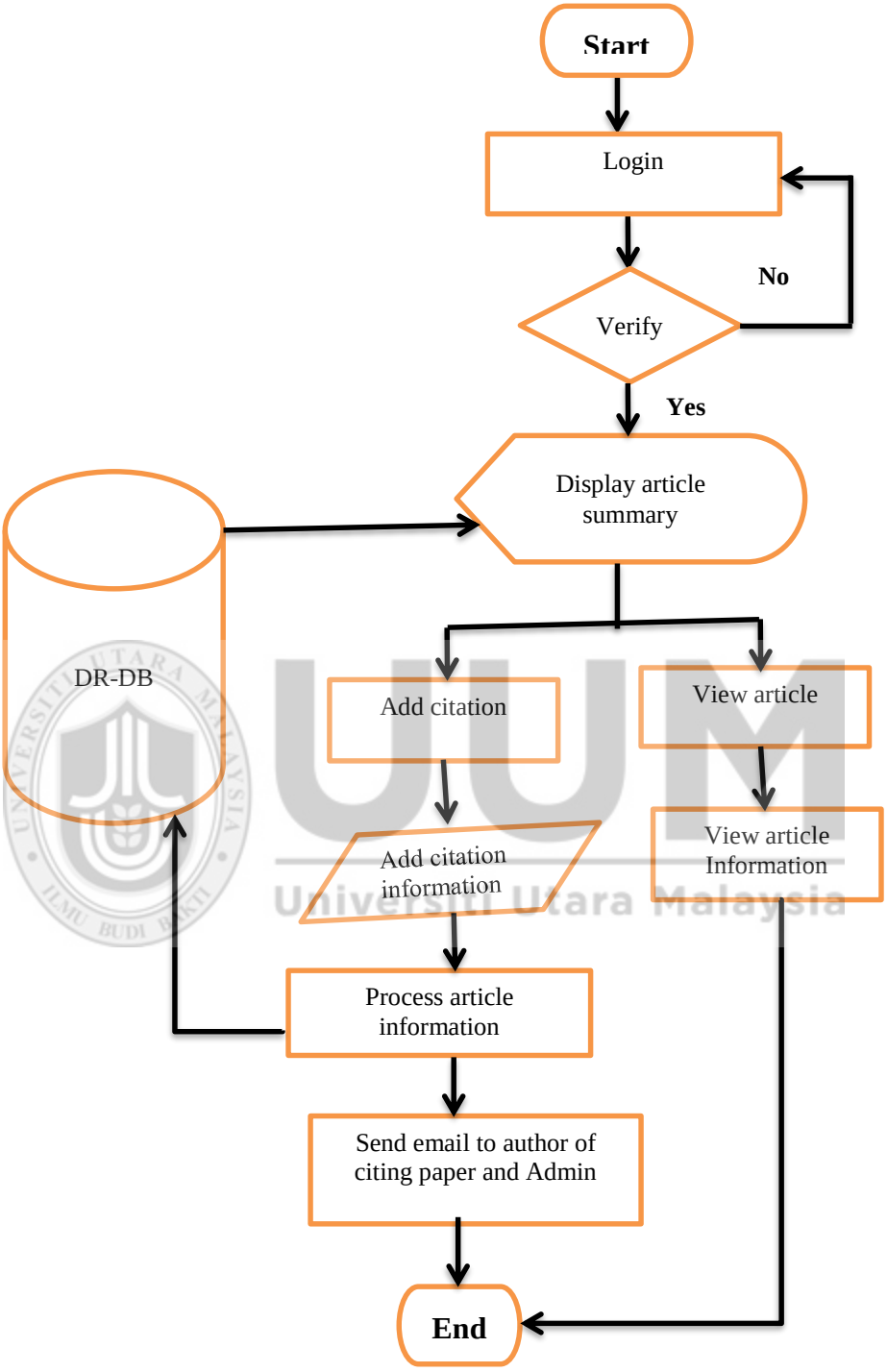


Figure 4.6: **The system flow of the author**

Figure 4.6 above shows the system flow of the author part. First the author login in the repository system by using the username and password he/she received in the email. Then the author adds the citation information in his/her own articles. After the new citation is added in the repository, the system sends an email to the author of the cited paper then the citation record will be saved under the article if the author of the cited paper or the admin verifies the citation.

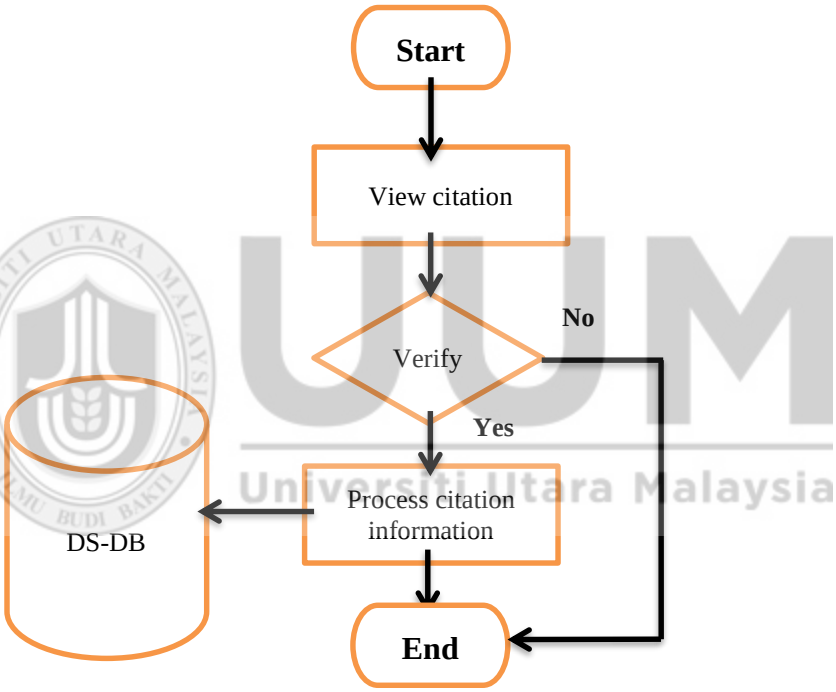


Figure 4.7: **The system flow of the author of the cited paper**

Figure 4.7 above explains the role of the author of the cited paper. First the author receives an email containing the citation information, the URL of the article and the confirmation link. After the author verifies the citation then citation record will be stored in the repository database.

Figure 4.8 below shows the system login interface of the digital repository. The login page contains three fields, the first two fields are the username and password fields while the third field is user type field, the user type field have two option that the users can choose, the first option is 'Admin' which is for the administrator of the system and the other option is 'Author' which is for the author to choose in order to login the system.

Figure 4.8: **The login interface of the system**

Figure 4.9 below shows the author profile page. After the author login the system by using his/her login username and password the system redirects the author to his/her personal profile which contains the summary of the articles that the author published. There is also Add citation feature in the author profile page which allows the authors to add the citation to their own articles.

Authors Name: Abdi rizak ghedee

Total : 2

h-index : 2.0
10-index : 0

List Of Articles

2

		year	Cited by	
1	A Study Of The Relationship Between Transformational Leadership, Empowerment And Organizational Commitment	2010	2	ADD CITATION
2	Industry Clusters: An Antidote For Knowledge Sharing And Collaborative Innovation?	2014	2	ADD CITATION

Figure 4.9: **The author profile page of the system**

Figure 4.10 below is the form that the admin uses to upload the articles into the system. The form contains 13 fields which the admin needs to fill in the information about the article. After the admin fills in the form then the admin clicks the 'upload article' button then the article will be stored in the digital repository. After the article is uploaded, an email will be sent to the author of the article, the email notifies the author that hi/her article were uploaded in the repository. The email also contains a username and password which the author can use in order to login the system.

Figure 4.10: **The form that is used to upload articles into the system**

Figure 4.11 below is the add citation form which allows the author to add the new citation record into the repository. The form contains 6 fields which 5 of the 6 fields the author needs to fill in the information of the cited paper. After the author adds the citation, an email will sent to the author of the cited paper to verify the citation.

Figure 4.11: **The form that is used to add citation into the system**

4.6 Summary

This chapter was presented the integration framework of the digital repository with the citation endorsement framework. In this study a citation endorsement framework with digital repository is proposed as the solution of the information sharing problem in academic environment. The presented system was implemented to store academic proceeding articles and also to allow the owner of the articles to add new citations for his own articles and also to show the h-index of his publications and 10-index, which indicates how many of his articles, have more than 10 citations.

After the author adds the citation information an email will be sent to the author of the cited article, the email contains the citation information, the URL of the article and the confirmation link. After the author of the cited article verifies the citation then the citation will be recorded and saved under the article and if the citation is not verified then that citation will not be calculated.

The presented repository also allows the normal users to search the articles in the repository and check the citation of the articles or download the articles from the repository. Several research issues remain for future work. A researcher might analysis other existing repositories and come out with additional features in his new repository.

CHAPTER FIVE

EXPERT EVALUATION

5.1 General Overview

This chapter presents the result of the expert interview towards their evaluation of the digital repository with the citation endorsement framework. 5 experts were interview in order to get their feedback towards the proposed digital repository with citation endorsement framework.

5.2 Result of the Thematic Analysis

Table 5.1 below explains how long the experts have been on the duty. The five experts that evaluated the digital repository have been on the duty between 11 to 30 years, the majority of the experts have been on the duty between 11 to 14 years, while one of the experts have been on the duty 20 years. However only one expert have been on the duty 30 years.

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

	Are you familiar with digital repositories?
Expert 1	11 years
Expert 2	12 years.
Expert 3	30 years
Expert 4	14 years
Expert 5	20 years

Table 5.2 below shows the expert familiarity of digital repositories. All of the experts were familiar with digital repositories and some of the experts were in charge of digital repositories however some of the experts said that they are familiar with digital repository only as a user and that they are not involved the management of a digital repository.

Table 5.2: The familiarity of the experts towards digital repositories

	Are you familiar with digital repositories?
Expert 1	Yes, because I am in charge of digital repository.
Expert 2	Yes.
Expert 3	Yes, I suppose I do use digital repositories.
Expert 4	Yes, but I am not really involve in the management of it.
Expert 5	Yes**,** but only as a user**.**

Table 5.3 below shows the answer of the experts when asked how well they know the indexing and citation systems like, Scopus, Google Scholar and Web of Science. All of the experts answered that they know the indexing and citation system, some of the experts answered they only know how to do search in these indexing and citation systems, while another expert answered that they know these indexing and citation systems but not so depth. However one of the experts answered that he/she doesn't know these indexing and citation systems very well.

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

	How well do you know the indexing and citation systems? Example Scopus, Web of science, Google scholar?
Expert 1	I am familiar with these systems because I used it but I am not familiar with their whole function.
Expert 2	Not so depth but I know Scopus, web of science and Google scholar.
Expert 3	Not very well.
Expert 4	As a librarian I only know how to do searching in Scopus, Google Scholar and Web of Science.
Expert 5	I am bit familiar with indexing and citation system but as a user only, how the systems work I don't know.

Table 5.4 below shows the answer of the experts when they were asked what their first impression towards this repository system was. All of the experts answered that they think this is good repository system. One of the experts answered that this system helps the university to record the publication done at the university in terms of citation and that would help the ranking of the publication. However one of the experts answered that it's just simple and small scale system to capture citation.

Table 5.4: The answer of experts on their first impression of this repository system

	What's your first impression of this Repository system?
Expert 1	I think it is good system because it is good for academicians to know the achievement of their publication.
Expert 2	Good innovation.
Expert 3	I think this is a good system whereby it helps the university to record the publication done at the university in terms of citation that would help the ranking of the publication.
Expert 4	I think it is good because it can do one single search because no need to have many fields. Like Google it has single search and it can retrieve all the data and you get a full text from there.
Expert 5	My first impression is just a simple system and small scale system to capture the citation.

Table 5.5 below shows the answer of the experts when asked if they believe this repository system has all the necessary function and capabilities they expected a digital repository to have. Most of the experts answered that they believed it is good repository while one of the experts suggest if this repository system have collaboration with existing library it can be functional. However one of the experts believed a few more features can be included, such G-index and impact factor.

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

	Do you believe this system has all the necessary functions and capabilities you expected a digital repository to have?
Expert 1	From what I see I think it is good system.
Expert 2	Yes but I think if this system have collaboration with library it can be functional.
Expert 3	Well from what I seen I think still there is some more features you can add**.**
Expert 4	Yes, because this system allows the user to get the full text of the articles.
Expert 5	I think a few more features can be included, such as may be you can include G-index and may be impact factor.

Table 5.6 below shows the answer of the experts towards citation indexing feature in this digital repository. All of the experts answered the feature is good and it is beneficial for the academicians to add their own citation, while one of the experts answered the feature is ok and it seems very simple feature with email verification by the author and admin will confirm the citation. However one of the experts recommended that the feature can be improved.

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

	How do you see the citation indexing feature in this digital repository?
Expert 1	I think it is beneficial for the academicians.
Expert 2	Good.
Expert 3	I think so far is ok, it seems to be very simple just add the citation details then you can see the email verification by the author and the admin will confirm the citation.
Expert 4	I think it is good, the citation is really clear and the author can do the citation by themself.
Expert 5	The features can be improved by manipulation the data I mean let's say you can sort or arrange by higher citation or higher h-index of the articles.

Table 5.7 below shows the answer of the experts when they were asked if they think its good ideas to allow the author to add their own citation. All of the experts answered that it is a good idea to allow the author to add their own citation. However one of the experts believed it is good idea only if the citation author added is genuine and the administrator can verify that it is correct citation before it is added.

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

	Do you think its good idea to allow the authors to add their own citation? If yes, why? If No, why not?
Expert 1	Yes, because the author can control the citation of their articles.
Expert 2	Yes, because author is able to know how many citations their articles have.
Expert 3	Well I think it is good that the author can add their citation provided it is genuine and the administration can verify that its correct citation before it is added.
Expert 4	As I said before I think it is good idea to allow the author to add their own citation.
Expert 5	It is good to give the author option to upload their own citation but it takes a lot of job.

Table 5.8 below shows the answer of the experts when they were asked how they see the citation verification process of the repository system. All of the experts believed it is good process, one of the experts answered it is a good process because the system alerts the author by sending email while another expert answered its good process because the admin and the author can verify the citation. However one of the experts answered in terms of authentication it is a good process but the admin has a lot of work to do.

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

	How do you see the citation verification process of the repository system?
Expert 1	I think the verification process is good because the admin and the author can verify the citation.
Expert 2	Good
Expert 3	Well from what I seen the process now was very fast but in real time I don't know how long it will take the author to confirm but I think so far it is working.
Expert 4	It is good because the system alerts the author by sending email.
Expert 5	In term of authenticity is good but the admin have a lot of work to do.

Table 4.9 below shows the answer of the experts when asked if there is other feature that they could recommend to have been added to this repository system. Some of the experts said for them it is ok and they would not add another feature, while one expert suggested that may be the layout should be more interesting display. However one expert suggested that he/she thinks other features can be added like G-index and impact factor.

Table 5.9: The expert's recommendation of other features

	Is there other feature that you could recommend to have been added to this repository system?
Expert 1	At the moment I have no recommendation.
Expert 2	Librarian to have access at citation prolix and articles.
Expert 3	May be the layout should be more interesting display.
Expert 4	For me it is ok.

Expert 5	As I said before I think other features can be added like G-index and impact factor.

Table 5.10 below shows the answer of the experts when they were asked if they could change one thing about this system, what they would change. Some of the experts answered they would not change anything from the repository system, while one of the experts said he/she would link the repository system to Scopus, Web of Science or Google Scholar. However one of the experts said in order to make the system more globally look like he/she would allow other proceeding to be uploaded not only KMICe proceedings.

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

	If you could only change one thing about this system, what would you change? Why?
Expert 1	I would not change anything.
Expert 2	Link to Scopus, web of science or Google scholar.
Expert 3	So far no comment on that.
Expert 4	Nothing to change
Expert 5	In order to make the system more globally look like the input should be open to other proceedings not only KMICe.

Table 5.11 below shows the answer of the experts when they were asked what aspect of the system that they don't like or disagree. All of the experts answered that there is

nothing they don't like or disagree from the system. However some of the experts believed that there is still some improvement that can be done on the system.

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

	What aspects of this repository system that you don't like or disagree? Why?
Expert 1	Actually the system is good i don't have further comment.
Expert 2	No comment.
Expert 3	I think there is still some work you can do.
Expert 4	So far I agree with it.
Expert 5	Generally I like the system but still need some improvement.

Table 5.12 below shows the answer of the experts when they were asked about their expectation of missing things in the repository system. Some of the experts answered that they don't think that this system is missing anything because this system can be center to all of them to access conference proceeding.

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

	What is your expectation of missing things in repository system? Can you suggest any?
Expert 1	Sometimes the conference papers URL is changed or the server is down so it will be good to have permanent URL like DOI.
Expert 2	I think this system is not missing anything because this system can be on stop center to all of us to access conference proceedings.
Expert 3	I think the system should have abstract display option.

Expert 4	For me there are no missing things.
Expert 5	I think you should give more option in the searching field, I mean the searching field should have various options I mean you can search by institution, combination of subject and year. Let us say I want to search certain subject with certain year. You should give more option in the search field.

Table 5.13 below shows the answer of the experts when they were asked if there anything on the repository system that they particularly like. Most of the experts answered yes there is something they like about the repository, one of the expert mentioned that he/she likes that the author can add their own citation, while another expert mentioned he/she like that the author profile they have option to upload their picture so that the author can be recognized, while another expert mention that he/she likes that the system is fast and everything can be retrieved very quickly. However one of the experts answered that there is nothing they he/she particularly like about the repository system.

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

	Is there anything on the repository system that you particularly like? If yes, what in particular do you like about the repository system?
Expert 1	What I like about the system is that the author can add their own citation.
Expert 2	No.
Expert 3	Well is quit fast and everything can be retrieved very quickly.

Expert 4	I think the citation indexing is clear and the arrangement of the indexing is not confusing the user.
Expert 5	I like this one you have the author profile you give the option of their picture so you can now the author

In an attempt to identify the user evaluation of the proposed citation indexing system, five experts were interviewed in order to get their expert feedback. Four of the experts are librarian officers from Sultanah Bahiyah Library of Universiti Utara Malaysia and one expert is Associate Professor at UUM School of Computing. Based on the result of the expert feedback, the conclusion can be made that using this citation indexing system good idea while allowing the author to add citation in their own articles. However the author of the citing paper and the administrator will confirm whether the citation is correct or incorrect.

5.3 Discussion

Digital repository is proposed as the solution of the information sharing problem in academic environment in this study. The presented repository was implemented to store conference proceeding article and to allow the author of the article to update the citation information. The presented repository not only allows the author to update the citation information, but shows the h-index and 10-index of the author which indicates the impact the author's publication.

Expert evaluation was conducted on the proposed digital repository with the citation endorsement framework. Four librarians and one associated professor from SOC were asked to evaluate the proposed digital repository with the citation endorsement framework. Figure 5.1 shows how long the experts that evaluated the proposed framework were on the duty. The least experienced expert was on duty 11 years and the most experienced expert was on duty 30 years. Figure 5.2 shows that all the experts are familiar with digital repositories.

Several research issues remain for future work. A researcher could apply other index measures to enhance the information sharing in academic environment. The digital repository can be improved with the use of other index measures like G-index; the G-index is the biggest number such that, altogether the top G articles received at least G square citations.

5.4 Summary

This chapter presents the result of the thematic analysis of the expert interview towards the proposed citation endorsement framework for digital repository. Four librarians and one associated professor from SOC were asked to evaluate the proposed digital repository with the citation endorsement framework. Based on the result of the expert feedback, the conclusion can be made that using this citation indexing system is good idea while allowing the author to add citation in their own articles.

CHAPTER SIX

CONTRIBUTION AND FUTURE WORK

6.1 General Overview

Digital repository framework for conference proceeding articles is developed and now the proceeding article can be stored and retrieved from the repository. The repository is integrated with enhanced citation endorsement framework which is the citation indexing that allows the article owners to update the citation information of their articles. The repository is included H-index and 10-index which indicates the impact of their publication. Also the repository is included citation validation feature which process the new added citation by sending an email to the author of the cited paper which asks to verify the new citation and if the author of the cited paper verifies the citation then citation will be stored in the repository database and the citation will be counted for that article. After developing the digital repository with the citation endorsement framework, then expert review evaluation was done on the developed citation endorsement framework which five experts were asked to do system demonstration and the expert feedback were collected through interview.

6.2 Contribution of the study

This study contributes a citation endorsement framework which allows the author of the article as part of the system; it allows the author to update the citation information of his article. This study also contributes a digital platform to manage proceeding articles and by storing the proceeding articles in the repository makes easy to manage and preserve that articles and it allows the users to search the articles in the repository and they can retrieve the articles from the repository in order to get full value from it. Also this study helps the academicians to increase the citation indexing of their article by allowing the author of the article to update the citation information of their articles and as a result the impact of the author's articles will be increased. However the citation information that the author updates will be verify by the admin and the author of the cited paper before the system creates the record of that citation.

6.3 Revisit the Research Objectives

This research project has three objectives, the first two objectives are, to identify citation framework for proceedings articles and to integrate the citation endorsement framework into the digital repository and the third objective is evaluate the citation endorsement framework in the digital repository.

The first milestone was developing a digital repository prototype as a citation framework for the proceeding articles, and the second milestone was integrating the citation endorsement framework into the digital repository and third milestone was evaluating the

citation framework in the digital repository. After the repository system was integrated with citation endorsement framework, five experts were asked to do system demonstration on the proposed repository framework and then their feedbacks were collected through interview.

Based on the result of the expert feedback, the conclusion can be made that using this citation indexing system is good idea while allowing the author to add citation in their own articles. However the author of the citing paper and the administrator will confirm whether the citation is correct or incorrect.

6.4 Conclusion

This digital repository system with citation endorsement framework was developed as the solution of the information sharing problem in academic environment in this study. The developed repository was implemented to store academic proceeding article and also to allow the owner of the article to add new citations for his own articles. The developed system was implemented to show the h-index of author's publications and 10-index, which indicates many of his articles have more than 10 citations. The developed digital repository is included citation validation process which verifies if the added citation is correct or incorrect. The developed repository also allows the normal users to search the articles in the repository and check the citation of the article or download the article from the repository. Several research issues remain for future work. A researcher can analysis other existing repositories and come out with additional features in his new repository.

6.5 Limitation

Although the research has reached its aims, there were some unavoidable limitations. First, because of the limit time, the expert review of the study was conducted only on small size of experts. Therefore, to generalize the result for larger groups, the evaluation of the new developed framework should have involved more experts at different levels. Secondly, the new developed framework is for KMICe proceedings only, therefore to help the users get more benefits from the new framework, the study should have extended framework to include other proceedings and articles. Finally, although the study has reached its aims, other existing repositories can be analyzed and come out with additional features that can be included in the digital repository system.

6.6 Future Work

New features are implemented in this study. The implemented features are: the citation index, which allows the author to add new citation in his articles, the h-index feature, which allows the author to see the impact of his publication, and the 10-index features which calculates the number of articles that the author publishes that have more than 10 citations, however in future work, in order to generalize the result for larger groups, the evaluation of the framework can be involved more experts at different levels and also in future work the framework can be extended to include other proceedings and articles. Also the researchers can research and develop frameworks for cooperatives repository networks and services which is the inclusion of the use of third party service providers.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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