# Providing Digital Library Users with Proper Search for Information Stored using Information Retrieval system

Thet Thet Aung[1], Khin Shin Thant[2], Myat Mon Khaing[3], April ThetSu[4], Hlaing Htake Khaung Tin[5]

*Faculty of Information Science[1,2,3,4,5]*
*University of Computer Studies, Hinthada, Myanmar[1,2,3,4]*
*University of Information Technology, Yangon, Myanmar[5]*

***Abstract*— In the Digital Library fields, Ontology can be used to organize bibliographic descriptions, represent and expose the contents of the document, and share knowledge between users. In the proposed IR model, preprocessing, context matching, and calculating similarity values steps are included. The algorithm for the formatting of SPARQL query is developed in the context matching step of IR model. Ontology-based IR system for Digital Library is implemented in Service-Oriented Architecture (SOA) by using the XML Web Service technology and ASP.NET. The architecture of the proposed system consists of file storage for documents, one ontology dataset, and two programming components: Digital Library Web Service and Web Application. In this proposed system, Web Ontology Language (OWL) is used to design Ontology for Digital Library using Protégé v3.5 tool. Functions for publication and retrieving of documents are implemented as a web service by using the C# programming language. The user interface is designed and implemented as a web application in ASP.NET platform for consuming the functions of web service. The performance of this research paper, the precision, recall, and F-values are measured and compared. According to the comparison results, the Ontology-based IR system is more accurate in searching for ObjectProperty type. As a result, the proposed system serves user-friendly, high-performance and scalable semantic search for information from the digital library.**

***Keywords:* Ontology, Digital Library, OWL, ObjectProperty, XML, SOA .**

## I. INTRODUCTION

Digital libraries (DLs) had become the digital counterpart of the traditional library system. There are various ways to improve the search technology for accessing documents from DL. In this research, Ontology-based IR system is proposed for Digital Library. Ontologies have the potential to play an important role in DL, because ontology defines a common vocabulary for researchers who need to share information in a domain. The proposed system intends to provide for students to retrieve the relevant information with their concept and to be able to search, read and download the textbooks, old questions (included tutorial, exam, multiple, assignments), journals, thesis papers, reference papers, novels efficiently in the short time.

Information retrieval is the study of helping users to find information that matches their information needs. Technically, IR studies the acquisition, organization, storage, retrieval, and distribution of information. Historically, IR is about document retrieval, emphasizing document as the basic unit [1]. General architecture of the IR system is shown in Figure 1.

In the general architecture of the IR system, the user with information needs issues a query (user query) to the retrieval system through the query operations module. The retrieval module uses the document index to retrieve those documents that contain some query terms (such documents are likely to be relevant to the query); compute relevance scores for them, and then rank the retrieved documents according to the scores. The ranked documents are then presented to the user. The document collection is also called the text database, which is indexed by the indexer for efficient retrieval [1].

Information Retrieval (IR) systems provide populations of users with access to a large collection of stored information. These systems are concerned with the structure, analysis, organization, storage, and searching for such information. Dr. Glöckner [2] described a good IR system is able to accept a user query, understand from the user query what the user requires, search a database for relevant documents, retrieve the documents to the user, and rank the documents according to their relevance.

Information retrieval systems are everywhere: Web search engines, library catalogs, store catalogs, cookbook indexes, and so on. Information retrieval (IR), also called information storage and retrieval (ISR or ISAR) or information organization and retrieval, is the art and science of retrieving from a collection of items a subset that serves the user's purpose [3]

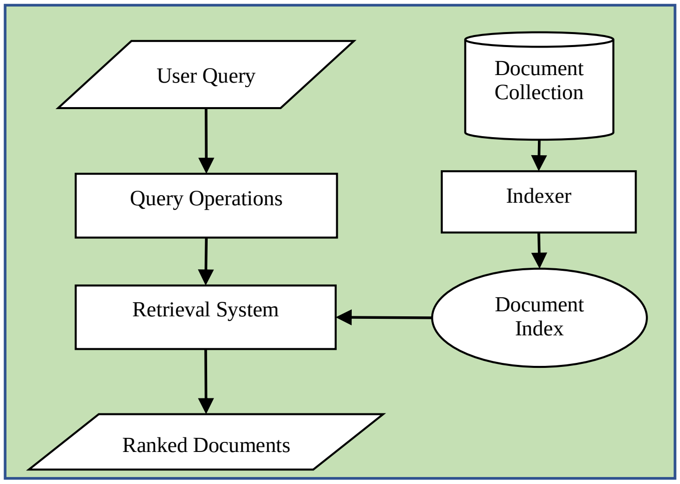


Figure 1. General Architecture of IR System

## II. RELATED WORK

Manual testing is the process of manually testing software for defects. It requires a tester to play the role of an end user whereby they utilize most of the application's features to ensure correct behavior. To guarantee completeness of testing, the tester often follows a written test plan that leads them through a set of important test cases. [4] Manual Testing is a process carried out to find the defects. In this method the tester plays an important role as end user and verify all features of the application to ensure that the behavior of the application. The Manual Testing is very basic type of testing which helps to find the bugs in the application under test. It is preliminary testing, must be carried out prior to start automating the test cases and also needs to check the feasibility of automation testing. The Test Plan is created & followed by the tester to ensure that the comprehensiveness of testing while executing the test cases manually without using automation testing tool. It is not necessary to have knowledge of any testing tool for manual software testing. As the Software testing fundamental always says that “100% Automation is not possible” so the *Manual Testing* is very important. [5]

Digital libraries are a set of electronic resources and associated technical capabilities for creating, searching, and using information. They combine the structure and gathering of information, which libraries and archives have always done, with the digital representation that computers have made possible. The main purpose of a digital library is to collect, manage, and preserve in perpetuity digital content [6].

The Digital Libraries Federation in 1998 defines digital libraries as: "Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities" [7]. Common features of the digital library are as follows:

- providing round the clock services to users, within and without the library environment. Users can access the digital objects at anytime and anywhere i.e. 24 hours and 7 days a week with only a computer and internet connection;
- providing a coherent view of all information contained within a library, no matter its form or format (e.g., text, audio, image and video);
- accessing a digital object by several users at the same time in different locations;
- requiring no large spaces, unlike traditional libraries where physical space is required for the construction and maintenance of the collections [8].

These Common features show the flexibility, portability, and accessibility of Digital libraries. Nevertheless, the principles underlying the functionality of digital libraries were simple, the premise that digital libraries dealing with traditional problems of searching for information delivery to users and to preserve it for posterity. Digital information takes up less space than information on paper and therefore, can help traditional libraries reduce costs is no longer enough anymore, so many more models are defined to meet specific needs that will be stumbling over time and with changing new technologies.

Traditional physical libraries employ metadata in the library catalogs. In digital libraries, metadata is obtained by cataloging resources such as books, periodicals, web pages, digital images, and DVDs, etc. The data is stored in the integrated system, using the MARC metadata standard. The purpose is to direct users to the location of the items and a detailed description of the items. Recently, standards for metadata in the digital libraries include Dublin Core, DDI [9]. Different metadata elements are needed to perform different tasks, for example, author, title and subject support the function of discovery. A DL may require many more forms of metadata than analog for management and use. According to the National Information Standard Organization’s (NISO) publication “Understanding Metadata”, there are three types of metadata [10].

The term ontology has been used for many years, to mean different things like glossaries and data dictionaries, thesauri and taxonomies, controlled vocabulary, schema and data models, and formal ontologies and inference. And also in many areas, such as philosophy, artificial intelligence, knowledge-based systems, it has been used to organize information. There are found in the literature several definitions of ontologies, several types proposed for application in different areas of knowledge, and proposals for building ontologies (methodologies, tools, and languages).

The philosophical field of ontology was not as successful as computer scientists, where they built some large and robust ontology, such as WordNet and Cyc [11]. Ontologies have aroused the interest of many researchers in Computer Science, being able to highlight main areas: Database, Software

Engineering, Semantic Web, Information Architecture, Knowledge Engineering, Knowledge Representation, Qualitative Modeling, Language Engineering, Information Retrieval, and Extraction, Knowledge Management and Organization, and Artificial Intelligence as a form of knowledge representation about the world or some part this, describing: individuals, classes, attributes, relationships and events [12].

In the Digital Libraries fields, ontologies can be used to: organize bibliographic descriptions, represent and expose the contents of the document, and share knowledge between users. It's important to note that the use of ontologies in digital libraries allows us to transfer the profile, the user's browsing behavior to other digital libraries and databases, so that when a user of a particular DL leaves service to connect to another DL, the user profile (including preferences and navigation behaviour) can be transferred from one base to another by using the appropriate semantic web services because all databases share a common domain of discourse that can be played by rules inference and application logic. For this we have a vast list of ontology languages that allow us to design ontologies according to our needs, however, when it comes to design ontology for digital libraries pertinent examples exist such as RDF (Resource Description Framework), in the family of W3C which is used for describing resources; XML (Extensible Markup Language), for describing data, information, and knowledge; OWL(Web Ontology Language), is becoming the standard for describing ontologies and accessing resources through the web [13].

## III. TYPES OF INFORMATION RETRIEVAL

The typical IR model of the search process consists of three essentials: query, documents, and search results. The goal of an IR system is to retrieve documents containing information that might be useful or relevant to the specific purpose it's being used. Information retrieval systems can also be distinguished by the scale at which they operate. Tasks of information retrieval are as follows [14]:

- Routing and filtering: To direct documents to interested parties
- Multimedia retrieval: To retrieve e.g. images or speech data
- Cross-language Retrieval: To find documents in one language that is relevant to an information need expressed in another language
- Summarization: To capture the essence of a text in fewer words
- Translation: To express in one language the meaning of a document written in another language
- Question-answering: To find text that answers a particular question
- Topic detection: To identify stories that discuss the same topic
- Classification: To assign documents to known classes
- Clustering: To assign documents to previously unknown groupings
- Novelty detection: To determine when a new topic is introduced

There are two good reasons for having models of information retrieval. The first is that models guide to research and provide the means for academic discussion. The second reason is that models can serve as a blueprint to implement an actual retrieval system.

Mathematical models are used in many scientific areas with the objective to understand and reason about some behavior or phenomenon in the real world. A model of information retrieval predicts and explains what a user will find relevant given the user query. The correctness of the model's predictions can be tested in a controlled experiment. In order to do predictions and reach a better understanding of information retrieval, models should be firmly grounded in intuitions, metaphors, and some branch of mathematics [15].

Intuitions are important because they help to get a model accepted as reasonable by the research community. Metaphors are important because they help to explain the implications of a model to a bigger audience. For instance, by comparing the earth's atmosphere with a greenhouse, non-experts will understand the implications of certain models of the atmosphere. Mathematics is essential to formalize a model, to ensure consistency, and to make sure that it can be implemented in a real system. As such, a model of information retrieval serves as a blueprint that is used to implement an actual information retrieval system [16].

An IR model governs how a document and a query are represented and how the relevance of a document to a user query is defined. There are four main IR models: Boolean model, vector space model, language model, and probabilistic model.

Although these models represent documents and queries differently, they used the same framework. They all treat each document or query as a "bag" of words or terms. Term sequence and position in a sentence or a document are ignored. That is, a document is described by a set of distinctive terms. A term is simply a word whose semantics helps remember the document's main themes. The term here may not be a natural language word in a dictionary. Each term is associated with a weight. Given a collection of documents D, let $V = \{t_1, t_2..., t|V|\}$ be the set of distinctive terms in the collection, where it is a term. The set V is usually called the vocabulary of the collection, and |V| is its size, i.e., the number of terms in V. A weight $w_{ij} > 0$ is associated with each term $t_i$ of a document $d_j \in D$. For a term that does not appear in document $d_j$, $w_{ij} = 0$. Each document $d_j$ is thus represented with a term vector, $d_j = (w_{1j}, w_{2j},...,w|V|_j)$, where each weight $w_{ij}$ corresponds to the term $t_i \in V$, and quantifies the level of importance of $t_i$ in document $d_j$. The sequence of the components (or terms) in the vector is not significant. With this vector representation, a collection of documents is simply represented as a relational table (or a matrix). Each term is an attribute, and each weight is an attribute value. In different retrieval models, $w_{ij}$ is computed differently [1].

### A. Boolean Model

In Boolean retrieval, a document is represented as a set of terms $d_j = t_1,…,t_k$, where each $t_i$ is a term that appears in document $d_j$. A query is represented as a Boolean expression of terms using the standard Boolean operators: and, or and not. A document matches the query if the set of terms associated with the document stratifies the Boolean expression representing the query. The result of the query is the set of matching documents [17].

### B. Language Model

Statistical language models (or simply language models) are based on probability and have foundations in statistical theory [5]. The basic idea of this approach to retrieval is simple. It first estimates a language model for each document and then ranks documents by the likelihood of the query given the language model. Similar ideas have previously been used in natural language processing and speech recognition.

### C. Probabilistic Model

This family of IR models is based on the general principle that documents in a collection should be ranked by decreasing the probability of their relevance to a query. This is often called the Probabilistic Ranking Principle (PRP). Since true probabilities are not available to an IR system, probabilistic IR models estimate the probability of relevance of documents for a query. This estimation is the key part of the model, and this is where most probabilistic models differ from one another. The probabilistic model is based on probability theory. It can be estimated the relevance of a given document for a user based upon their query.

### D. Vector Space Model

In the vector space model text is represented by a vector of terms. The definition of a term is not inherent in the model, but terms are typically words and phrases. If words are chosen as terms, then every word in the vocabulary becomes an independent dimension in a very high dimensional vector space. Any text can then be represented by a vector in this high-dimensional space. If a term belongs to a text, it gets a non-zero value in the text-vector along with the dimension corresponding to the term. A vector-based information retrieval method represents both documents and queries with high-dimensional vectors while computing their similarities by the vector inner product [18].

## IV. ONTOLOTY BASED INFORMATION RETRIEVAL

In this research paper, designing the Domain Ontology for Digital Library is described. Creating classes and properties of Digital Library is explained using Protégé Tool, which is a free, open-source platform to construct domain models and knowledge-based applications with ontologies. Ontology-based Information Retrieval Model is presented in detail in this section. Using Vector Space Model for ranking the IR results is described. Web Ontology Language (OWL) is used to design ontology for Digital Library. Many ontology editors have been developed to help domain experts to develop and manage ontology, for example, Protégé, OntoEdit, or TopBraid. Protégé [10] is a free, open-source platform to construct domain models and knowledge-based applications with ontologies. Protégé OWL editor: enables users to build an ontology for the Semantic Web, in particular to OWL: Classes (5 subclasses), Properties (16 properties) and Instances.

## V. CALCULATING TF-IDF SIMILARTIY

In this step, the vector space model is used to retrieve more accurate data and rank it which semantically enhances the searching and retrieval process. In the vector space IR model, a document is represented as a weight vector, in which each component weight is computed based on some variation of TF or TF-IDF scheme.

The vector space model is a statistical model for representing text information for Information Retrieval. It is a simple, mathematically based approach that provides partial matching and ranked results. TF-IDF weighting is the most common term weighting approach for vector space model retrieval. The weight of the term in document vector can be determined using the method. The weight of the term is measured how often the term *j* occurs in document *i* (the Term Frequency) and IDF (the Inverse Document Frequency) as shown in Equation 1 and 2.

The weight equation for the term within document is as follows:

$$w_{ij} = tf_{ij} \times idf_i \tag{1}$$

where,

$w_{ij}$ = weight of the term $t_i$ in document $d_j$

$tf_{ij}$ = the normalize term frequency (TF) of term $t_i$ in document $d_j$

$idf_i$ = the inverse document frequency (IDF) of term $t_i$

$$tf_{ij} = \frac{f_{ij}}{\max\{f_{1j}, f_{2j}, ...., f_{|v|j}\}} \tag{2}$$

where, $f_{ij}$ = the raw frequency count of term $t_i$ in document $d_j$

$$idf_i = \log \frac{N}{df_i} \tag{3}$$

where,

$df_i$ = number of document in which term $t_i$ appears at least once

$N$ = the total number of documents in the system

A query $q$ is represented in exactly the same way as a document. The weight equation for the term within query is as follows:

$$w_{iq} = \left[0.5 + \frac{0.5\, f_{iq}}{\max\{f_{1q}, f_{2q}, \ldots\ldots, f_{|v|q}\}}\right] \times \log \frac{N}{df_i} \quad (4)$$

where,

$w_{iq}$ = weight of the term $t_i$ in query vector $q$

$f_{iq}$ = the raw frequency count of term $t_i$ in query vector $q$

The similarity between query $q$ and $j^{th}$ document retrieved by context matching process is calculated by Dice similarity method as shown in Equation 5.

$$Dice(d_j, q) = \frac{2\left|\sum_{i=1}^{|v|} w_{ij} \times w_{iq}\right|}{\sum_{i=1}^{|v|} w_{ij}^2 + \sum_{i=1}^{|v|} w_{iq}^2} \quad (5)$$

where,

$Dice(d_j, q)$ = the dice similarity between document $d_j$ and query $q$

$w_{ij}$ = weight of the term $i$ within document $d_j$

$w_{iq}$ = weight of the term $i$ within query

Example of TF-IDF and Dice Similarity calculation is explained with step by step in below.

Table 1. Input Query:

| | |
|---|---|
| Keyword: | "classification of query" |
| In Property: | "title" |

Table 2. Retrieved Documents by Context Matching Process:

| Resource | Value |
|---|---|
| dl:Document1 | "Web Query Classification System using NoSQL Graph Database" |
| dl:Document2 | "Query Classification based Information Retrieval System" |
| dl:Document3 | "Performance Comparison between Keyword based and Classification based Information Retrieval System" |

**STEP-1:** Term Extraction (Tokenization and Stopword Removal) from resources. The extracted terms from resources are shown in Table 3. In total 19 terms from three documents are extracted. Among these, 7 terms with underline are duplicated, so 12 are received.

Table 3. Extracted Terms from Documents

| Document1: | Document2: | Document3: |
|---|---|---|
| 1. web | | 1. performance |
| 2. query | 1. <u>query</u> | 2. comparison |
| 3. classification | 2. <u>classification</u> | 3. keyword |
| 4. system | 3. information | 4. <u>classification</u> |
| 5. nosql | 4. retrieval | 5. <u>information</u> |
| 6. graph | 5. <u>system</u> | 6. <u>retrieval</u> |
| 7. database | | 7. <u>system</u> |

**STEP-2:** Term Frequency (TF) Calculating for each Term in each Document. In this step, the Term Frequency for extracted terms from resources is calculated by Equation 2. These TF values are described in Table 4.

Table 4. Term Frequency Result for Extracted Terms

| ID | Term | Document1 | Document2 | Document3 |
|---|---|---|---|---|
| 1 | Web | 1/1 = 1 | | |
| 2 | Query | 1/1 = 1 | 1/1 = 1 | |
| 3 | Classification | 1/1 = 1 | 1/1 = 1 | 1/1 = 1 |
| 4 | System | 1/1 = 1 | 1/1 = 1 | 1/1 = 1 |
| 5 | Nosql | 1/1 = 1 | | |
| 6 | Graph | 1/1 = 1 | | |
| 7 | Database | 1/1 = 1 | | |
| 8 | Information | | 1/1 = 1 | 1/1 = 1 |
| 9 | Retrieval | | 1/1 = 1 | 1/1 = 1 |
| 10 | Performance | | | 1/1 = 1 |
| 11 | Comparison | | | 1/1 = 1 |
| 12 | Keyword | | | 1/1 = 1 |

**STEP-3:** Inverse Document Frequency (IDF) Calculating for each Term in each Document. The IDF values for terms are calculated by Equation 3. and shown in Table 5.

Table 5. Inverse Document Frequency Result for Extracted Terms

| ID | Term | Inverse Document Frequency (IDF) |
|---|---|---|
| 1 | Web | Log(3/1) = 0.47712 |
| 2 | Query | Log(3/2) = 0.17609 |
| 3 | Classification | Log(3/3) = 0 |
| 4 | System | Log(3/3) = 0 |
| 5 | Nosql | Log(3/1) = 0.47712 |
| 6 | Graph | Log(3/1) = 0.47712 |
| 7 | Database | Log(3/1) = 0.47712 |
| 8 | Information | Log(3/2) = 0.17609 |
| 9 | Retrieval | Log(3/2) = 0.17609 |
| 10 | Performance | Log(3/1) = 0.47712 |
| 11 | Comparison | Log(3/1) = 0.47712 |
| 12 | Keyword | Log(3/1) = 0.47712 |

**STEP-4:** TF-IDF (weight) Calculating for each Term in each Document. The TF-IDF values for terms are calculated by Equation 1 and shown in Table 6.

Table 6. TF-IDF Result for Extracted Terms

| ID | Term | Document1 | Document2 | Document3 |
|---|---|---|---|---|
| 1 | Web | 1 x 0.477 = 0.477 | | |
| 2 | Query | 1 x 0.176 = 0.176 | 1 x 0.176 = 0.176 | |
| 3 | Classification | 1 x 0 = 0 | 1 x 0 = 0 | 1 x 0 = 0 |
| 4 | System | 1 x 0 = 0 | 1 x 0 = 0 | 1 x 0 = 0 |
| 5 | Nosql | 1 x 0.477 = 0.477 | | |
| 6 | Graph | 1 x 0.477 = 0.477 | | |
| 7 | Database | 1 x 0.477 = 0.477 | | |
| 8 | Information | | 1 x 0.176 = 0.176 | 1 x 0.176 = 0.176 |
| 9 | Retrieval | | 1 x 0.176 = 0.176 | 1 x 0.176 = 0.176 |
| 10 | Performance | | | 1 x 0.477 = 0.477 |
| 11 | Comparison | | | 1 x 0.477 = 0.477 |
| 12 | Keyword | | | 1 x 0.477 = 0.477 |

**STEP-5:** TF-IDF (weight) Calculating for each Term in the keyword. The weight values for extracted terms from keyword are calculated by Equation 4 as follows.

$\text{weight}_{(\text{classification, keyword})} = (0.5 + (0.5 \times (1/1))) \times \log 3/3 = 0$

$\text{weight}_{(\text{query, keyword})} = (0.5 + (0.5 \times (1/1))) \times \log 3/2 = 0.17609$

**STEP-6:** Dice Similarity Calculating. The similarity between keywords and retrieved documents by context matching process is calculated by Dice Equation 5 as shown in below and the similarity results are shown in Table 7.

Table 7. Similarity Results for Documents

| Id | similarity score |
|---|---|
| Document1 | 0.06377 |
| Document2 | 0.5 |
| Document3 | 0 |

**STEP-7:** Ranking Documents by Similarity Score. After calculating the similarity values between documents and keywords, the retrieved documents are ranked by score. The ranked result is shown in Table 8.

Table 8. Ranked Results for Documents

| Id | similarity score | Remark |
|---|---|---|
| Document2 | 0.5 | Most Relevant Document |
| Document1 | 0.06377 | |
| Document3 | 0 | |

## VI. CONCLUSIONS

The designing the Domain Ontology for Digital Library is described. Web Ontology Language and Protégé editor are used to design Ontology for Digital Library. Five subclasses: Document, Author, Category, DocumentType and FileType are defined in our Ontology. Four object type properties and twelve data type properties are also defined for the specific class. The designed Ontology plays a significant role in our IR model to define the common vocabulary and structure for resources. In this proposed IR model, preprocessing, context matching and calculating weight values steps are included. All the steps of proposed IR model are explained in detail in this research.

## REFERENCES


[1] Thet Thet Aung, Myat Mon Khaing, Khin Shin Thant, Hlaing Htake Khaung Tin, "Senitment Analysis On Who Southeast Asia Region Organization (Who Sero) User Comment Review And Opinion Mining", International E-Conference In Association with International Engineering Journal For Research& Development, ICIPPS June 2020, E-ISSN NO:- 2349-0721, IMPACT FACTOR : SJIF - 6.549, pp. 314-321.

[2] C. Xia, X. Wang, "Graph-Based Web Query Classification", Proceedings of the 12th Web Information System and Application Conference IEEE, (2016) 241-244.

[3] Ontology for Digital Library, Master Thesis, International Master in Digital Library Learning, (2015).

[4] Thet Thet Aung, Khin Shin Thant, Myat Mon Khaing, "Finding Out the Possible Benefits Using Social Media for Higher Education During the Epidemic Covid-19", University Journal of ICT in Multidisciplinary Issues on Arts & Science, Engineering, Economics and Education, June, UJIM 2020, Hinthada, Myanmar, ISBN 978-99971-0-872-2, Volume 1, Issue 1, pp. 59-63.

[5] Khin Shin Thant, Thet Thet Aung, Myat Mon Khaing, Hlaing Htake Khaung Tin, "Comparison Of Loan Finance System Of Fullerton Myanmar And Woori Finance Myanmar", International E-Conference In Association with International Engineering Journal For Research& Development, ICIPPS June 2020, E-ISSN NO:- 2349-0721, IMPACT FACTOR : SJIF - 6.549, pp. 339-344.

[6] Bethesda, MD, Metadata Types and Functions, NISO. (2004) Understanding Metadata.

[7] Dan, Munteanu. (2007). Vector space model for document representation in information retrieval. The Annals of "Dunarea de Jos", University of Galati Fascicle III Electrotechnics Electronics Automatic Control and Informatics, 43-46, December 2007.

[8] Dr. Varshil Bhagaji Dashrath, Role of Metadata in Digital Resource Management, International Journal of Digital Library Services (IJODLS), 4(3): 209-217, 2014.

[9] Khin Shin Thant, Thet Thet Aung, Hlaing Htake Khaung Tin, "Exploring Better Teaching Methods By Surveying from Computer Science Students (Case Study for Software Engineering)", International Journal of Research & Innobation in Applied Science, IJRIAS May 2020, EISSN 2454-6194, Volume V, Issue V, pp. 93-96.

[10] Protégé, A free, open-source ontology editor and framework for building intelligent systems.

[11] Isah, A.; Mutshewa, A.; Serema, B. & Kenosi, L. (2013). Digital Libraries: Analysis of Delos Reference Model and 5S Theory. Journal of Information Science Theory and Practice 1(4): 38-47, 2013.

[12] D. Hiemstra, "Information Retrieval Models", Information Retrieval: Searching in the 21st Century, (2009) 1-18.

[13] Nilesh Shewale , Dr. J. Shivarama (2018). Ontology based digital library search system for enhanced information retrieval in engineering domain, International Journal of Advanced Research,

Ideas And Innovations In Technology (IJARIIT), Vol. 4, No. 5, ISSN: 2454-132X

[14] Apache Software Foundation (2020). SPARQL Tutorial. Access in https://jena.apache.org/tutorials/sparql.html

[15] Deborah L. McGuinness, Frank. (2004). OWL Web Ontology Language-Overview, W3C Recommendation, 10 February 2004. From https://www.w3.org/TR/owl-features/ access in 23/01/2020

[16] Khaing Thazin New, Lwin Sandar Soe, Thet Thet Aung, "Effective Concept and Components of 4ps Marketing Information System", International Journal of Research and Scientific Innovation (IJRSI), June 2020, ISSN No. 2321-2705, Volume VII, Issue VI, pp. 30-33.

[17] B. Liu, "Web Data Mining", 2nd Edition, Springer-Verlag Berlin Heidelberg, New York, (2011).

[18] K. M. Risvik, "Scaling Internet Search Engines: Methods and Analysis", Ph.D. Thesis, Department of Computer and Information Science, Norwegian University of Science and Technology, (2004) 1-172.