INDIAN JOURNAL OF ENGINEERING

19(51), 2022

The Digital Library Web Service System by Using Domain Ontology

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

To Cite:
Aung TT, ThetSu A, Khaing MM, Thant KS, Tin HHK. The Digital Library Web Service System by Using Domain Ontology. *Indian Journal of Engineering*, 2022, 19(51), 196-205

Author Affiliation:
[1,2,3,4,5]Faculty of Information Science, Myanmar
[1,2,3,4]University of Computer Studies, Hinthada, Myanmar
[5]University of Information Technology, Yangon, Myanmar
Email: thetthetaung86.htd@@gmail.com[1], hlainghtakekhaungtin@gmail.com[5]

Peer-Review History
Received: 10 March 2022
Reviewed & Revised: 11/March/2022 to 29/April/2022
Accepted: 01 May 2022
Published: 02 May 2022

Peer-Review Model
External peer-review was done through double-blind method.

ABSTRACT
The proposed system is implemented as the Information Retrieval system by using Domain Ontology. The main point of proposed IR system is the formatting of SPARQL query and context matching process by using SPARQL query. The Ontology-based IR system for Digital Library is implemented based on Service-Oriented Architecture (SOA) by using the XML based web service technology and ASP.NET. The design of this system consists of file storage for documents, one ontology dataset and two types of programming components. They are web service and web application. To show the performance of the system, 33 queries for different properties of documents were tested by using 415 training documents. To evaluate the performance of Ontology-based IR system for Digital Library, precision, recall, and F-measure methods are used. According to the comparison results of precision, recall, and f-measure, the Ontology-based IR system is more accurate in ObjectProperty type and also ObjectProperty is faster than DatatypeProperty in processing time with miliseconds.

Keywords: Web Service System, Digital library, Ontology, XML, ASP.Net, Service-Oriented Architecture.

1. INTODUCTION
The detailed implementation of this information retrieval framework is presented in this paper. Design and use case diagrams of the system, class structure of Ontology Web Language (OWL) is also included in this research paper. The proposed system is implemented as the information retrieval system by using Domain Ontology. The main point of the proposed system is the formatting of the SPARQL query and context matching process by using the SPARQL query. In this system, there are six main steps. In the first step, query preprocessing, which consists of the tokenization and stopwords removal process for the user query, is performed. This system accepts the query and property selected by the user to retrieve relevant documents from Digital Library. In the second step, the tokenized keywords and selected property by the user are transformed to SPARQL query format by the algorithm for the formatting of SPARQL query. In the third step, the context matching process by formatted SPARQL query is performed. This process is used to

DISCOVERY SCIENTIFIC SOCIETY

match the context of documents from Domain Ontology with the formatted SPARQL query. The results of this process are relevant documents by the keywords and property of the document. In the fourth step, relevant documents retrieved by context matching processes are calculated for TF-IDF values and similarity scores by using the VSM (Vector Space Model) and the Dice similarity method respectively. In the next step, retrieved documents are classified according to their similarity scores, and the whole process for retrieving documents is done here. Evaluation of the results of IR is performed in the final step by calculating its precision, recall, and f-measure values. The relevant documents retrieved by SPARQL query are ranked and displayed as the result of our Ontology-based IR system.

The proposed system serves user-friendly, high-performance, and scalable semantic search for information from the digital library. As a result, the Ontology-based IR system is more accurate in searching for ObjectProperty type. Information retrieval by SPARQL query produces exact results; in the case of keyword search, it produces all results containing keywords including non-relevant documents. The exactness and completeness of the IR system are proved by the average value of F-measure which obtains over 95%.

Moreover, the use of Ontology for Digital Library is more flexible and interchangeable than the use of Relational Databases. It provides a chance to extend and define metadata for other resources easily without modifying the implementation. However, this proposed IR model doesn't support to transform the user query in natural language into SPARQL format. And also, it provides to search for only digital documents.

In the rest of the paper, Literature Review presented in section II. The proposed system design, the architecture of the system, and the structure of Digital Library Ontology are described section III. And then, the implementation of programming modules for the proposed system is explained with Graphical User Interfaces in section IV. The experimental results are shown by charts and tables in section V. The conclusion of the research work is drawn in the section VI. In this section, further extensions and limitations that propose some improvements which could be made are presented.

2. LITERATURE REVIEW

Nowadays, the amount of available information in both printed media and electronic/digital mediums had increased dramatically. Moreover, the number of digital documents had rapidly increased and required easy and accessed mechanized methods. In the information retrieval systems, the information is usually searched by means of a full-text search; every term in the texts of the documents can function as a search key.

Digital libraries (DLs) had become the digital complement of the traditional library structure. 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 states a common word for scientists who want 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. Digital libraries are a set of electronic resources and associated technical capabilities for creating, searching, and using information. They combine the design and gathering of information, which libraries and archives have always done, with the digital illustration that computers have made possible. The main objective of a DL(Digital Library) is to collect, manage, and preserve in perpetuity digital content [1]. The Digital Libraries Federation in 1998 describes digital libraries as: "Digital libraries are organizations that provide resources, including special staff." [2].

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 [3]. Ontologies have aroused the interest of many researchers in Computer Science, being able to highlight main areas: Database, Software Engineering, SW(Semantic Web), IA(Information Architecture), KE(Knowledge Engineering), KR(Knowledge Representation), QM(Qualitative Modeling), LA(Language Engineering), IR(Information Retrieval), and Extraction, KM(Knowledge Management) and Organization, and AI(Artificial Intelligence) as a form of knowledge representation about the world or some part this, describing: individuals, classes, attributes, relationships and events [4].

In the Digital Libraries fields, ontologies can be used to: signify, establish bibliographic descriptions and representation the contents of the document, and share information between users. It's important to note that the usage of ontologies in digital libraries allows us to transfer the profile, the user's browsing conduct to additional digital libraries and catalogs, 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 moved after one base to additional by using the suitable semantic web services because all databases portion a common domain of address 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 project ontology used for digital libraries relevant examples exist such as RDF (Resource Description Framework), in the family of W3C which is used for relating resources; XML (Extensible Markup Language), for describing data, information, and knowledge; OWL(Web Ontology Language), is flattering the normal for relating ontologies and retrieving resources through the web [5].

It also describes relationships linked between ideas in the domain and those ideas. The recent development of the standard is OWL from the World Web Corporation (W3C). Web Ontology Language (OWL) is a language for defining and instance ontologies in the Web. This includes descriptions of classes and their properties and their relationships. The OWL is designed to be used by applications needed to process the content of information rather than to people. It further facilitates the possibility for interpretation by machines of Web content by providing additional vocabulary with formal semantics. OWL is a W3C recommendation [6].

OWL is intended to be used when required to be processed by applications, as opposed to the circumstances required to present the information contained in the documents only to humans. OWL can be used to clearly represent the meaning of terms and the relationships of those terms; the terms and their interactions are called ontology. There are more aids to describe meaning and secrets than XML, RDF, and RDF-S, and so OWL goes beyond these languages by its ability to represent machine-defining content on the Web [7]. OWL ontology consists of three components: Individuals, Properties, and Classes.

3. PROPOSED SYSTEM DESIGN

The Ontology-based IR system for the digital library is implemented on the basis of service-oriented architecture (SOA) by using XML-based web service technology and ASP.NET. The logical architectural style of the system is displayed in Figure 1. The architecture of the proposed system includes file storage for documents, a data machine and two programming components. All functions for the Digital Library web service can be gathered into two modules: Publication Module and Retrieval Module. The functions of the publication module are extracting contexts from documents and saving them to a dataset. The whole IR process of our proposed system is provided by the functions of the retrieval module. In our system architecture, the Digital Library web application just plays in the role of the user interface. Ontology dataset is used to store the extracted context of documents and file storage is used to save documents themselves.

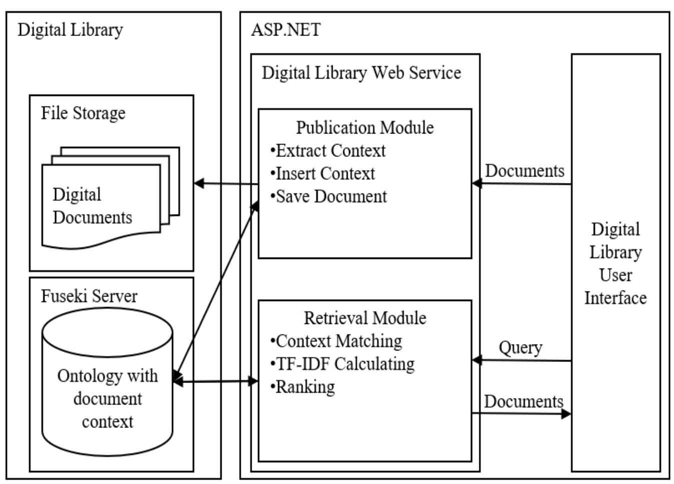


Figure 1. Architecture of the System

The Digital Library web service is implemented by using C# programming language. This web service consists of purposes for journal publication and saving/retrieving of documents. Getting the class structure of Ontology and its instances, saving and manipulating the instances of the specific classes, and extracting the contents of documents are the main functions of the publication module. The functions of the publication module are performed by connecting with the ontology dataset on the Fuseki server. These functions are as follows:

- getOwlClass: getting the whole structure of a specific class including its datatype and object properties from Ontology dataset
- getIndividuals: getting all the instances of a specific class from Ontology dataset.
- getIndividualByName: getting an instance of a specific class by its name from the Ontology dataset.
- setIndividual: saving an instance of a specific class to the Ontology dataset. The name of the instance is programmatically defined by the last inserted ID for this class.

ISSN 2319–7757 EISSN 2319–7765 I OPEN ACCESS

- setIndividualByName: saving an instance of a specific class to the Ontology dataset by a given name.
- updateIndividual: manipulating the properties of an instance of the specific class by name of this instance.
- deleteIndividual: deleting an instance of the specific class by its name from the Ontology dataset.
- isExist: checking the instance of a specific class is exist in our Ontology dataset or not.
- isDocumentExist: checking the specific instance of Document class is exist in our Ontology dataset or not.
- isAuthorExit: checking the specific instance of Author class is exist in our Ontology or not.
- getFileContent: extracting the content from various types of files such as “pdf”, “txt” and “docx”.

Testing the getOwlClass function of publication module of web service with a sample input parameter “Document” is exposed in Figure 2. As a result, the structure of the Document class with fifteen properties is returned by getOwlClass function. Testing the getIndividuals function and its result are shown in Figure 3.

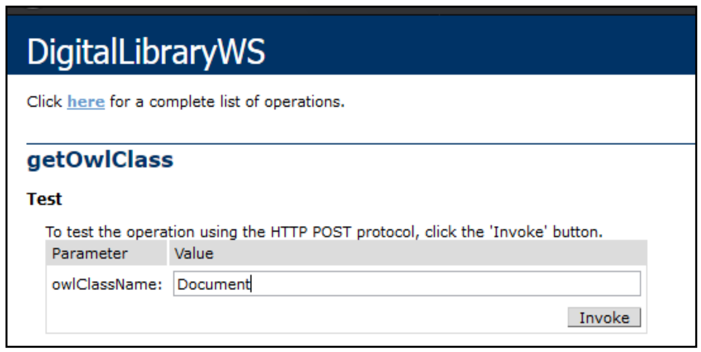


Figure 2. Testing the getOwlClass Function of Web Service

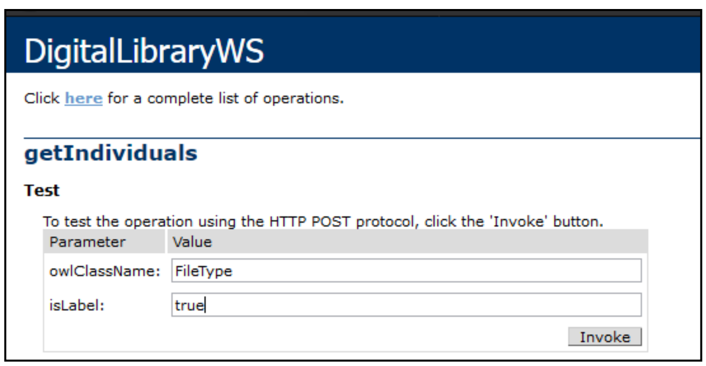


Figure 3. Testing the getIndividuals Function of Web Service

4. IMPLEMENTATION OF THE SYSTEM

The consumer interface is designed and implemented as a web program in ASP.NET platform for testing the functionality of web services. Two types of roles for the user: Admin Role and User Role are defined in web applications. Admin can edit all the resources of Information Retrieval structure for Digital Library, such as management of user information, the publication of documents to Ontology dataset and manipulation of their information. The users can search for digital documents by keywords and property of documents. This application consists of five menus: Home, Search, Result, Publish, and Administration. All of these menus are available only for authenticated users. The admin and users must be login to our Digital Library web application by the “Login” page as shown in Figure 4.

The “Result” menu is designed and implemented for displaying the results of IR in detail. These results consist of precision, recall, f-measure. The results for all tested queries are shown on this page. As the result, tested queries are grouped by type of properties: DatatypeProperty and ObjectProperty. The “Result” pages with DataTypeProperty and with ObjectProperty of the web application is shown in Figure 5.

ISSN 2319–7757 EISSN 2319–7765 I OPEN ACCESS

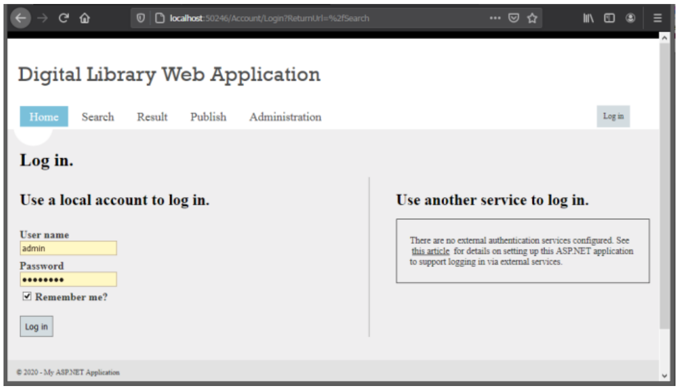
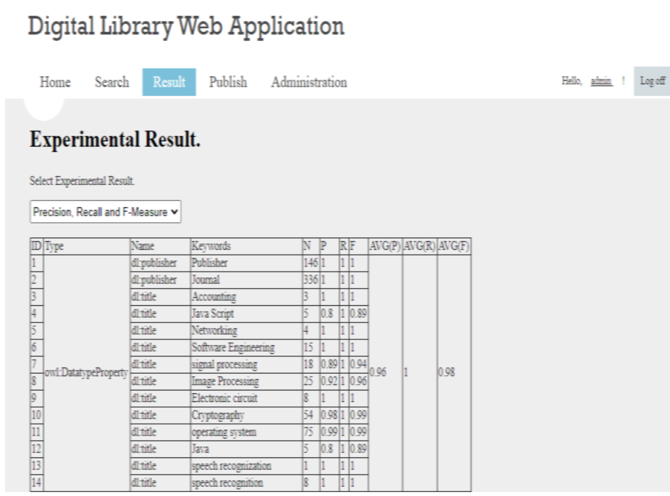


Figure 4. Login Page of Digital Library Web Application



ID	Type	Name	Keywords	N	P	R	F	AVG(P)	AVG(R)	AVG(F)
1	owl:DatatypeProperty	dl:publisher	Publisher	146	1	1	1	0.96	1	0.98
2		dl:publisher	Journal	336	1	1	1			
3		dl:title	Accounting	3	1	1	1			
4		dl:title	Java Script	5	0.8	1	0.89			
5		dl:title	Networking	4	1	1	1			
6		dl:title	Software Engineering	15	1	1	1			
7		dl:title	signal processing	18	0.89	1	0.94			
8		dl:title	Image Processing	25	0.92	1	0.96			
9		dl:title	Electronic circuit	8	1	1	1			
10		dl:title	Cryptography	54	0.98	1	0.99			
11		dl:title	operating system	75	0.99	1	0.99			
12		dl:title	Java	5	0.8	1	0.89			
13		dl:title	speech recognization	1	1	1	1			
14		dl:title	speech recognition	8	1	1	1			

Figure 5. Result Page with DataTypeProperty of Digital Library Web Application

5. EVALUATION OF THE SYSTEM

To show how the system works, 33 questions were tested for various documents with 415 training documents (.doc, .pdf,.txt) containing different types of files. These testing queries are related to Object and Datatype Properties. The training documents are collected from the Google search engine.

To assess the presentation of Ontology-based IR structure for Digital Library, precision, recall, and F-measure methods are used as shown in Equations 4.1, 4.2, and 4.3.

Precision (P)

$P = TP / (TP+FP)$ (4.1) (4.1)

Recall (R)

$R = TP / (TP+FN)$ (4.2) (4.2)

F-Measure (F)

$F = 2 * [(P*R) / (P+R)]$ (4.3) (4.3)

Where TP denotes the number of relevant documents in retrieved documents. FP is the number of non-relevant documents in retrieved documents. FN denotes the number of relevant documents in non-retrieved documents. Precision is the ability to retrieve top-ranked documents that are most relevant. The recall is the aptitude of the search to discovery all of the relevant substances in the corpus. This means that the precision is the exactness and the recall is the completeness of the IR system. The f-measure is just a combination of the exactness and completeness of the system. The precision, recall, and f-measure values of experimental results for the ObjectProperty are shown in Table 1.

ISSN 2319–7757 EISSN 2319–7765 I OPEN ACCESS

Table 1. Precision, Recall and F-measure Results for ObjectProperty

Property Name	Keywords	No Retrieved	P	R	F
dl:hasAuthor	Information Security	1	1	1	1
dl:hasAuthor	Khin	8	1	1	1
dl:hasAuthor	Kirti Rajadnya	1	1	1	1
dl:hasAuthor	John	11	1	1	1
dl:hasAuthor	aye	6	1	1	1
dl:hasAuthor	Aung	9	0.78	1	0.88
dl:hasAuthor	Giftlin Sherin	1	1	1	1
dl:hasAuthor	hlaing	2	1	1	1
dl:hasAuthor	myo	5	0.8	1	0.89
dl:hasCategory	system analysis and design	121	1	1	1
dl:hasCategory	data mining	88	1	1	1
dl:hasCategory	Unified ModelingLanguage	23	1	1	1
dl:hasCategory	artificial intelligence	68	1	1	1
dl:hasCategory	Human computer Interaction	110	1	1	1
dl:hasCategory	Natural language processing	63	1	1	1
dl:hasCategory	digital signal	5	1	1	1
dl:hasCategory	Embedded system	121	1	1	1
dl:hasCategory	Data structure	88	1	1	1
dl:hasCategory	Cloud Computing	24	1	1	1
dl:hasCategory	Data warehouse	88	1	1	1
AVERAGE			0.98	1	0.99

In the above table, the precision (P), recall (R), and f-measure (F) values for four ObjectProperty of documents are shown. The recall for all properties is 1 and the average precision for all properties is 0.98. The average F-measure value is 0.99. According to these results, the exactness and completeness of Ontology-based IR systems in ObjectProperty is over 98%. The precision, recall, and f-measure values of experimental results for the DatatypeProperty are shown in Table 2.

Table 2. Precision, Recall and F-measure Results for DatatypeProperty

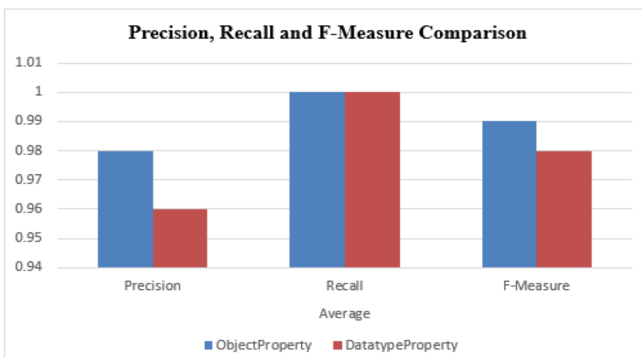
Property Name	Keywords	No Retrieved	P	R	F
dl:publisher	Publisher	146	1	1	1
dl:publisher	Journal	336	1	1	1
dl:title	Accounting	3	1	1	1
dl:title	Java Script	5	0.8	1	0.9
dl:title	Networking	4	1	1	1
dl:title	Software Engineering	15	1	1	1
dl:title	signal processing	18	0.89	1	0.9

ISSN 2319–7757 EISSN 2319–7765 I OPEN ACCESS

dl:title	Image Processing	25	0.92	1	1
dl:title	Electronic circuit	8	1	1	1
dl:title	Cryptography	54	0.98	1	1
dl:title	operating system	75	0.99	1	1
dl:title	Java	5	0.8	1	0.9
dl:title	speech recognization	1	1	1	1
dl:title	speech recognition	8	1	1	1
AVERAGE			0.96	1	0.98

In the above table, the precision (P), recall (R), and f-measure (F) values for four DatatypeProperty of documents are shown. The average precision for all properties is 0.96 and the recall for all properties is 1. The average F-measure value is 0.98. According to these results, the exactness and completeness of Ontology-based IR systems in DatatypeProperty is 96%.

The average results of Ontology-based IR system for ObjectProperty and DatatypeProperty are compared and described with the bar chart in Figure 6. According to the evaluation consequences of precision, recall, and f-measure, the Ontology-based IR system is more precise in ObjectProperty type because the standards for this property are all examples of an OWL class.



	Precision	Recall	F-Measure
Object Property	0.98	1	0.99
Datatype Property	0.96	1	0.98

Figure 6. Comparison Results of Precision, Recall and F-Measure

To assess the presentation of proposed system, the processing time of IR is compared with traditional IR system. The processing time of both proposed IR and traditional IR system is recorded in database for each tested query. And then the average value of processing time for both IR systems is calculated and grouped by type of query property. The unit of processing time in our experiment is in milliseconds. The average processing time results for the ObjectProperty are shown in Table 3.

Table 3. Average Processing Time Results for ObjectProperty

PropertyName	Keywords	Processing Time (ms)	
		Proposed-IR	Traditional-IR
dl:has Author	Information Security	526	1378
dl:has Author	Khin	528	1196
dl:has Author	Kirti Rajadnya	515	1829
dl:has Author	John	568	1492

ISSN 2319–7757 EISSN 2319–7765 I OPEN ACCESS

dl:has Author	aye	501	1340
dl:has Author	Aung	532	1395
dl:has Author	Giftlin Sherin	239	959
dl:has Author	hlaing	465	856
dl:has Author	myo	623	1630
dl:hasCategory	system analysis and design	528	1262
dl:hasCategory	data mining	500	1300
dl:hasCategory	Unified ModelingLanguage	457	1383
dl:hasCategory	artificial intelligence	499	1335
dl:hasCategory	Human computer Interaction	646	6639
dl:hasCategory	Natural language processing	744	6449
dl:hasCategory	digital signal	2500	2795
dl:hasCategory	Embedded system	2660	2925
dl:hasCategory	Data structure	399	1484
dl:hasCategory	Cloud Computing	634	1377
dl:hasCategory	Data warehouse	326	1730
AVERAGE		720	2038

As a result, the minimum processing time of proposed IR system for ObjectProperty queries is 239 milliseconds and the maximum is 2500 milliseconds. The maximum processing time of traditional IR system for ObjectProperty queries is 720 milliseconds. According to the comparison result of average processing time which shown in Table 4.3, the proposed IR system is more than two times faster than the traditional IR system in finding for ObjectProperty type queries.

The average processing time results for the DatatypeProperty are shown in Table 4. As a result, the minimum processing time of proposed IR system for DatatypeProperty queries is 233 milliseconds and the maximum is 2660 milliseconds. The maximum processing time of traditional IR system for DatatypeProperty queries is 6639 milliseconds. The average value of processing time for our proposed system is 720 milliseconds and traditional IR system is 2038 milliseconds. According to the comparison result of average processing time which shown in Table 4.4, the proposed IR system is more than three times faster than the traditional IR system in searching for DatatypeProperty type queries.

The processing time comparison result of both IR systems for ObjectProperty and DatatypeProperty queries are described with the bar chart in Figure 4.19. The average processing time of proposed IR system for ObjectProperty queries is 499 milliseconds and DatatypeProperty queries is 610 milliseconds. According to this evaluation outcomes, the proposed Ontology-based IR system is faster in ObjectProperty query than the DatatypeProperty query because the values for this property are all instances of an OWL class.

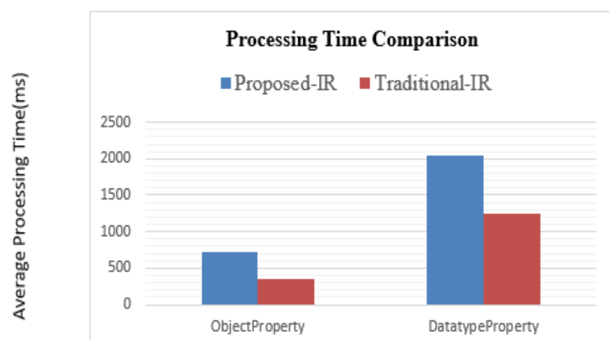
Table 4. Average Processing Time Results for DatatypeProperty

PropertyName	**Keywords**	**Processing Time (ms)**	
		Proposed-IR	**Traditional-IR**
dl:publisher	Publisher	233	1664
dl:publisher	Journal	239	1648
dl:title	Accounting	257	1619
dl:title	Java Script	510	1668
dl:title	Networking	280	1104
dl:title	Software	241	891

ISSN 2319–7757 EISSN 2319–7765 I OPEN ACCESS

	Engineering		
dl:title	signal processing	440	973
dl:title	Image Processing	337	943
dl:title	Electronic circuit	351	1109
dl:title	Cryptography	365	965
dl:title	operating system	276	1150
dl:title	Java	250	1491
dl:title	speech recognization	856	1090
dl:title	speech recognition	266	1058
AVERAGE		350	1241

The average processing time of Ontology-based IR system for ObjectProperty and DatatypeProperty are compared and described with the bar chart in Figure 7. According to the comparison results of Proposed-IR and Traditional-IR, the Ontology-based IR system with objectProperty is faster than in Datatype Property type.



	ObjectProperty	DatatypeProperty
Proposed-IR	720	2038
Traditional-IR	350	1241

Figure 7. Processing Time Comparison for Proposed and Traditional IR

5. CONCLUSION AND FUTHER EXTENSIONS

The proposed system presents the implementation of Ontology-based information retrieval for Digital Library. This system introduces a system that users can use to retrieve digital documents from the Ontology dataset. The ontology method is used to represent the context model based on digital library resources. Ontology acting a key role in the evolution of digital libraries. In interoperability at the semantic level, context-sensitive query processing over heterogeneous information resources requires the matching of concepts. The system presents the available varied information bases and recovers the accuracy of information retrieval using semantic web technology. In addition, the system can help users to reduce the consuming time for surfing the information they wanted. The proposed system is tested by using only the dataset with document resources. The dataset can be extended with multimedia resources, such as video, audio, and others, by modifying the Digital Library Ontology. Obtaining a better result in the configuring of SPARQL query is a motivation for further research work such as the development of an algorithm to transform the natural language query to SPARQL.

Funding

This study has not received any external funding.

Conflict of Interest

The author declares that there are no conflicts of interests.

Data and materials availability

All data associated with this study are present in the paper.

REFERENCES AND NOTES

1. 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, pp. 339-344.
2. Bethesda, MD, Metadata Types and Functions, NISO. (2004) Understanding Metadata; https://marciazeng.slis.kent.edu/ metadatabasics/types.htm Accessed in 04/08/2020
3. Protégé, A free, open-source ontology editor and framework for building intelligent systems. https://protege.stanford.edu Accessed in 04/08/2020.
4. 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.
5. D. Hiemstra, "Information Retrieval Models", Information Retrieval: Searching in the 21st Century, (2009) 1-18.
6. Berners-Lee, T. (1997). Axioms of Web Architecture: Metadata. http://www.w3.org/DesignIssues/Metadata.html 05/032014
7. Salton, Gerard. Introduction to Modern Information Retrieval. McGraw-Hill

ISSN 2319–7757 EISSN 2319–7765 I OPEN ACCESS