Ontology-driven Approach for Knowledge Sharing and Retrieval

Nevin M. Matasyoh[1], George Okeyo[2], Wilson Cheruiyot[3]

[1]Department of Computing, Jomo Kenyatta University of Agriculture and Technology.

P.O BOX 3639-20100, Nairobi, Kenya

[2]Department of Computing, Jomo Kenyatta University of Agriculture and Technology.

P.O BOX 62000-00200, Nairobi, Kenya

[3]Department of Computing, Jomo Kenyatta University of Agriculture and Technology.

P.O BOX 62000-00200, Nairobi, Kenya

Abstract

Knowledge sharing and retrieval is very useful in any research process because it promotes access to reliable research content. Its effectiveness depends on how well the knowledge can be shared and organized for easy access. The techniques that have been used to facilitate knowledge sharing in knowledge intensive organizations are Web 2.0 and Web 3.0. However, Web 3.0 (semantic web) has proven more effective since it enhances the web by providing mechanisms by which computers can process, interpret and connect information to enhance access and retrieval. Though Semantic web has been used to enhance knowledge sharing in various areas such as e-learning, medicine and engineering, it has not been applied in knowledge sharing among research students. This paper, therefore, presents an ontology driven approach for sharing of knowledge in form of research materials among research students. This approach makes use of ontology to structure the research materials for easy access and retrieval by research students. It has proved effective as its scope is wider compared to the other approaches that only allow unidirectional flow of knowledge or limit knowledge contributors. The approach has also proved to be precise. This is because all the documents retrieved are relevant.

Keywords: *Semantic Web, knowledge sharing, knowledge retrieval, ontology, research materials.*

1. Introduction

Knowledge is interpreted as information that resides in an individual's mind [15]. It can be classified as either explicit or tacit. Unlike explicit knowledge which is articulated, written down or published academic knowledge found in books, manuals, papers e.t.c [13], tacit knowledge is more dependent on its holder, attached to a person's mind and deeply grounded in an individual's action and experience [9].

Knowledge sharing is about communicating knowledge within a group of people. These groups may be people in business organizations or learning institutions such as universities. Similarly, knowledge sharing can be seen as the willingness to share whereby knowledge is capable of being used again or repeatedly in the course of its transfer from one party to another [6]. The underlying purpose of all these is to utilize the available knowledge to improve a person or group's performance. As a result, some educational institutions have employed virtual learning in support of knowledge sharing [4].

Knowledge sharing requires collaboration between consumers and contributors of knowledge [14]. The web provides a good environment for these collaborations as research students can interact and share knowledge despite their geographical location. However, the success of knowledge sharing is not in sharing and having it in a repository, but in the mechanisms by which the knowledge can be easily accessed and retrieved. Effective knowledge retrieval systems therefore focus on extracting specific content from a pool of knowledge. This is made possible through the use of web 3.0 techniques (semantic web) that add structure and meaning to what is on the Web

[1] Corresponding author

thus allowing computers to process information, interpret, and connect it to enhance knowledge retrieval [1]. The semantic Web relies heavily on formal ontologies that structure underlying data for the purpose of comprehensive and transportable machine understanding [5].

This paper presents an ontology driven approach for sharing of knowledge among research students. The knowledge domain of the ontology is materials that are used by postgraduate students during research. Various concepts are identified from the knowledge domain and used to create classes in ontology. The ontology is developed using OWL ontology language and then populated from a documents source. The ontology processing is done by Apache Jena Framework for Java [2]. SPARQL queries are processed using ARQ query engine then sent to the ontology, where the results are processed and returned to the user. The accuracy and preciseness of the approach is evaluated using precision and recall. This approach has proved effective as it has a wider scope compared to other ontology based knowledge sharing platforms in place. Unlike other approaches/platforms that limit sources of knowledge or only allow unidirectional flow of knowledge, the knowledge sources used in this approach are not limited. The approach also allows application of semantic web in knowledge sharing among postgraduate research students. Semantic web, in this case, is used to structure and give meaning to research materials for easy access and retrieval. This in turn makes research process much easier and faster.

The rest of this paper is organized as follows: section 2 presents related work, section 3 presents the method used, section 4 presents the ontology while section 5 presents the conclusion and future work.

2. Related work

Though web 2.0 techniques have been very useful in enhancing knowledge exchange in interactive virtual communities, they are still not efficient as they have proved to only enhance collaboration but not easy access and retrieval of knowledge. For this reason, semantic web was introduced to structure knowledge for easy processing by machines. The structuring is performed using ontologies which are the pillar of Semantic Web [12].

Semantic techniques have been successfully used in areas such as distance learning, exam systems and medical field for data extraction, integration, and sharing. For instance, DC-THERA directory [11] makes use of metadata and semantic web techniques to organize the knowledge generated in research. The DC-THERA application ontology is defined using the standard ontology language, Web Ontology Language (OWL). The content of the directory is structured using RDFs while the querying of the ontology is done using SPARQL. The directory relies on the ontology to get more specific results from the wide range and general knowledge. Though the ontology is useful in structuring and enabling easy access to knowledge, it is limited in some areas. For instance, its construction, evaluation and maintenance only depend on the members of the European research community. The public cannot contribute to its knowledge. Knowledge flow in DC-THERA is also one way, that is, from DC-THERA to the public and not from the public to DC-THERA. This limits utilization of expert knowledge from non-members of the European Research community.

Similarly, Wiki-I [8] makes use of ontology to organize and structure the innovative ideas shared by engineers during research activities. SPARQL query language is used to construct queries to the knowledge base. It makes use of tags that provide links to the wiki pages associated with the tag names and the origins of the ideas associated with the tag name. Though Wiki-I is more interactive that DC-THERA as it allows any engineer to contribute to the knowledge, it still has some limitations that inhibit bidirectional exchange of knowledge among experts. For instance, only contributors of the knowledge are allowed to evaluate or give opinions on the innovative ideas presented by others. This in turn prevents non-contributors from enhancing or evaluating pre-existing knowledge.

Learning institutions are also knowledge intensive institutions that involve continuous flow of knowledge. The flow is either among staffs, students or between staffs and students in different geographical locations. Easy coordination of activities in such an environment can be made possible using advanced web based applications. For example, the ontology based examination system framework [1] describes an examination ontology that is developed

IJCSI International Journal of Computer Science Issues, Volume 13, Issue 4, July 2016
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org http://dx.doi.org/10.20943/01201604.5967

based on the methontology ontology development. It provides semantically rich question banks to avoid repetition of questions. It also provides a means for scheduling examinations, periods and the personnel in various examination venues. The core component of this framework is an exam ontology that provides a knowledge base for the semantic examination grid. Though the proposed ontology provides a means for integrating several electronic examination applications for easy access and retrieval of information, it has no empirical evidence. The researchers only presented the ontology but its performance was not evaluated.

Though ontologies have been successfully applied in various fields including the ones mentioned above, they have not been applied in sharing and retrieval of knowledge among postgraduate research students. For this reason, this paper seeks to present an ontology driven approach for sharing and retrieval of knowledge among research students. The shared knowledge is in form of research materials that are structured for easy interpretation by machines. The knowledge domain of this ontology is research materials used by research students. Unlike other approaches, such as DC-THERA and Wiki-I, that limit the scope of application to a European community medical research group, this approach allows anyone to access or contribute to the knowledge. The performance of the ontology is also evaluated and documented

3. The ontology driven approach

The purpose of the proposed ontology driven approach is to enable post graduate research students to successfully achieve the objective of their research through efficient use of the knowledge resources they have access to. The ontology is used to structure the shared research materials so that machines can easily interpret the materials and provide meaningful results to researchers.

3.1 Overview of Semantic web

The main aim of semantic web is to make web content machine understandable [3]. Semantic web mainly relies on ontologies for structuring of web content in order to make it machine understandable. Ontologies are therefore best suited for structuring the content shared by researchers to allow easy access and retrieval.

3.1.1 Ontology development languages

Technically, ontology is a text based knowledge base that comprises of knowledge terms. The knowledge terms include the entities, semantic interconnections between the entities and a set of rules of inference about a particular topic or area. Languages used in developing ontologies include OIL (Ontology Inference Layer), DAML (Darpa Agent Markup Language) + OIL and OWL (Web Ontology Language) [10]. OIL is an ontology creation language that extends RDF schema and allows specific description/definition of an entity while OWL is a recently endorsed representation language that contains additional vocabulary to facilitate interpretation of web content. The language used to develop this ontology is the OWL ontology language. This is because it is a powerful language that has greater representational power.

3.1.2 Ontology development tools

Ontologies are normally developed using graphical and integrated ontology authoring tools such as OILed, OntoEdit, Ontolingua and protégé among others [7]. *OILed* is a development tool that supports construction of ontologies using OIL language, *OntoEdit* supports multilingual development while *Ontolingua* enables users to manage, reuse and share ontologies stored on a remote ontology server. It easily imports and exports ontologies constructed using DAML+OIL, OWL etc. Protégé, on the other hand, is a free and the most used development tool that supports rich knowledge models. It provides a development environment that makes use of various plug-ins that support specific knowledge domains. It can be extended to accommodate a number of graphical components such as images, video, graphs and tables among others. The tools allow both development of new ontologies and modifying existing ones.

3.2 Design of the ontology driven approach

This section presents the design of the semantic web approach. The approach has been implemented as a platform that makes use of ontology in structuring content. The specifications of the knowledge domain and the facilities used are discussed in detail.

3.2.1 Description of the platform

This platform makes use of ontology in structuring shared content for easy access and retrieval. It is mainly an application for accessing and querying information that is stored in the OWL ontology located on the web or local system. The content domain is research materials using by postgraduate students. The ontology is populated using the research materials. All the information presented after a query through the application is extracted from the ontology. The ontology processing is done using Apache Jena which is a semantic web framework for Java. The application retrieves information from the ontology using Jena API. Loading and creation of a model of the OWL ontology, content extraction and querying is done by the Jena framework. Though the Jena framework can be used to create ontologies, the OWL ontology in this platform is created using an external editor (protégé) then loaded into the application using Jena. The interface of the application is easy to use and allows users to search through the shared content and get results according to certain specifications.

3.2.2 The conceptual framework of the platform

The framework structure for the platform is based on four main processes: knowledge acquisition, knowledge representation and lastly semantic querying as shown in figure 1.The first process, knowledge acquisition, is what handles the knowledge sharing part. Shared information is stored in a repository (data source). Knowledge representation is done using ontology. The ontology is populated from the data source (database, WWW). A reasoner is run over the OWL files to obtain new OWL files. The OWL files are then queried.

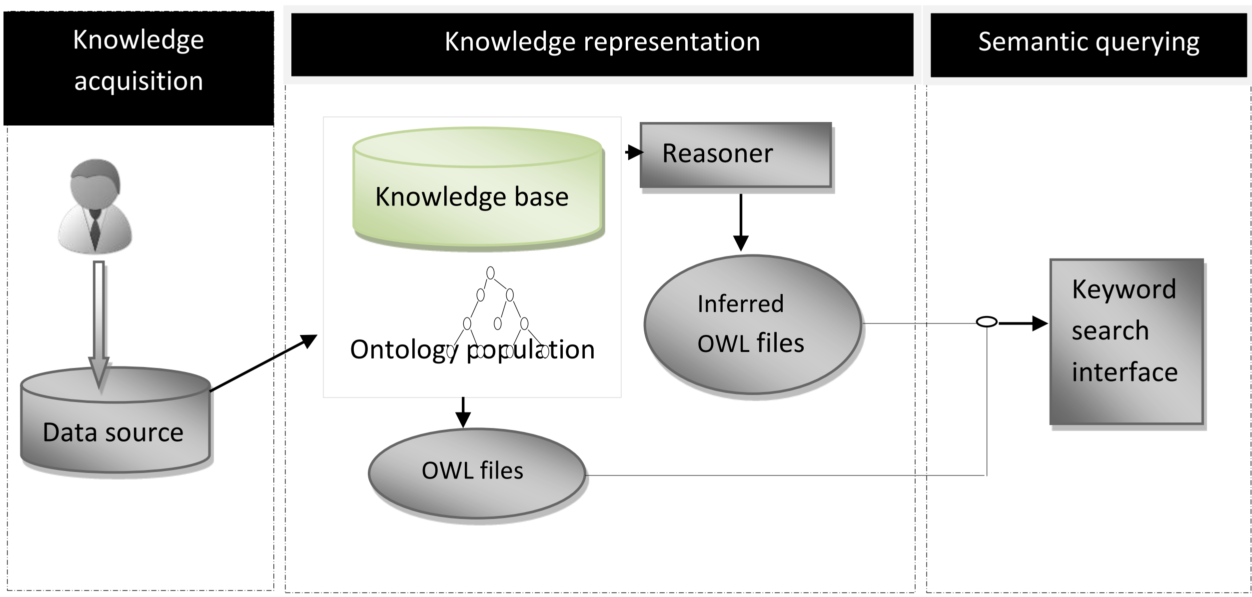

Figure 1: conceptual framework

3.2.3 The platform's knowledge domain

The knowledge domain of the platform is research materials used by research students. There is a lot of research content shared and stored in repositories. The content can be accessed and retrieved but sometimes the results of the retrieval do not satisfy the user. This lack of easy access and retrieval of research content led to the need of implementing this semantic web platform to solve the problem. Concepts such as properties, relations, classes and subclasses can be easily demonstrated in this domain. Most of the properties extracted from the domain such as title, name, publisher etc are common to most of the research materials. The ontology is constructed separately and loaded into the created java application that forms the query interface. It is populated with individuals from the data source after it has been loaded into the java application.

3.2.4 Platform specifications

Representation and storage of information: Information from the domain is represented using ontology. For example, the different kinds of research materials are represented as classes and their properties as either object properties or data properties in the ontology. The properties are also used to form the relationships between various classes.

Ontology processing: The ontology processing is done by a java application that makes use of the Apache Jena Framework for Java. The processing itself is implemented by an API that is provided by the Jena Framework.

Platform development: The application is developed in Netbeans editor using Java programming language. It makes use of Apache Jena, a semantic web framework, to import and create a model of the OWL ontology and query it using SPARQL. The ontology is separately developed using protégé ontology editor then loaded into the application.

3.2.5 Design of the knowledge sharing platform

The platform for sharing and retrieval of knowledge is made up of three main parts: the OWL ontology, query engine (ARQ) and the application interface for interacting with the platform as shown in figure 2. ARQ is a SPARQL processor for Jena. The main information resource for the application is the constructed ontology.

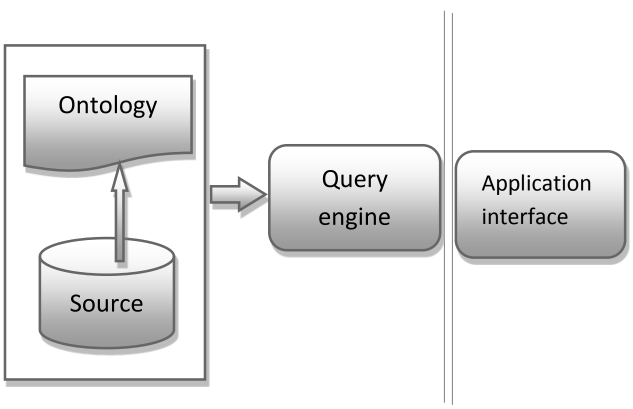

Figure 2: platform structure

3.2.6 The OWL ontology design

The ontology has been constructed in a way that reflects the domain of knowledge in detail. It makes use of various concepts such as class relations, class hierarchies and properties. Figure 3 shows some of the classes in the ontology. Some of the classes are super classes while others are sub classes of some of the super classes. The relations between different classes are formed using object properties. The object properties include; *belongsTo*, *hasAbstract*, *hasAddress*, and *has Author* among others. Datatype properties, on the other hand, describe the characteristics of the instances of classes. They include; *edition*, *year*, *volume*, *name*, *title*, *institution*, and *supervisor* among others. Figure 4 shows some of the relations between various classes of the ontology.

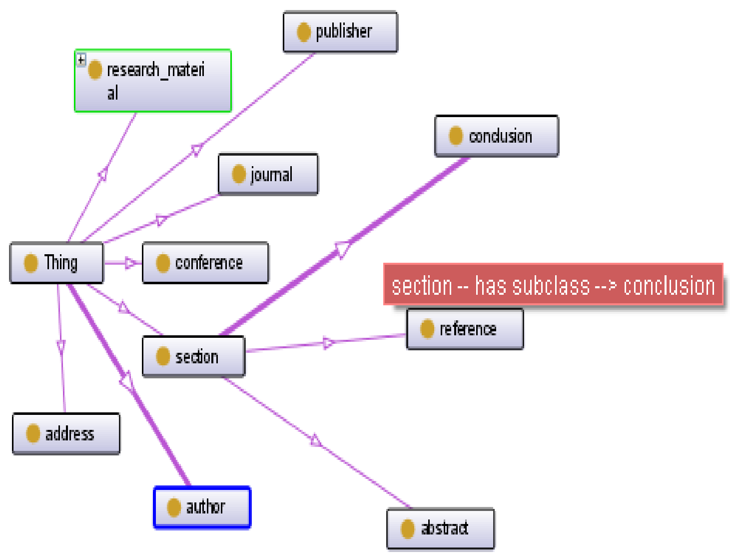

Figure 3 : ontology classes

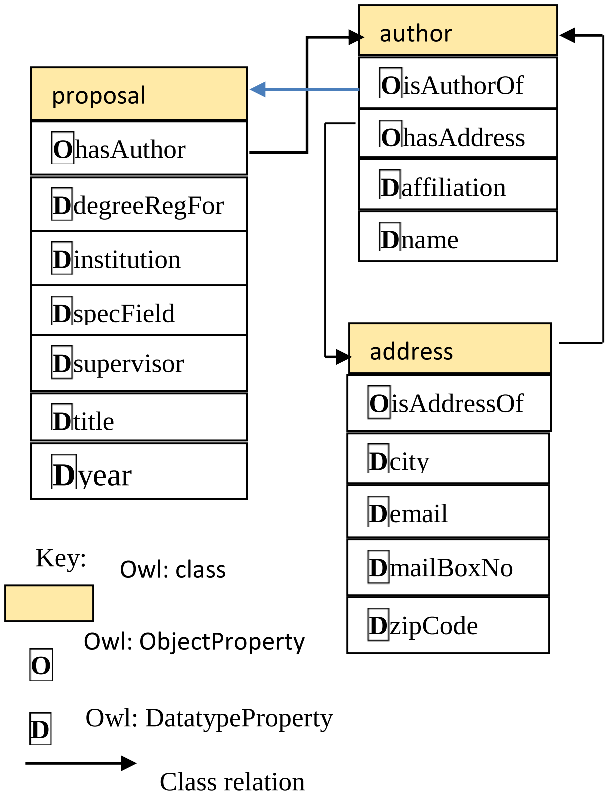

Figure 4: class relations

3.2.7 Query processing

As stated earlier, the query engine used in this case is ARQ query processor for Apache Jena. Apache Jena is a semantic web framework for java. The interface enables users/ clients to send requests to the ontology through the query engine. The query engine processes the query, makes necessary queries to the ontology and returns results to the user through the interface. The user sends request by clicking a particular menu item on the interface.

The interface is a java application that implements query processing and provides user interface for sending queries to the ontology. The different requests sent by a user are processed differently and responses returned depending on the type of query being processed. In this case, the queries are appended on the different menu items found on the interface. Therefore, the type of response to a query depends on the menu item chosen by a user. For example, if user desires to get all the papers under the field of computer science, the query that will be sent to the query engine for processing is as shown in figure 5.

```
//query statement

String queryString =
"PREFIX:<http://www.knowledge_sharing.com/
ontologies/"+ "knowledge_sharing.owl#>"+

"SELECT ?paper WHERE {?p a :paper; :title
?paper ; :specField 'computer science' }";

// execute the queryString to obtain
results

Query query =
QueryFactory.create(queryString);

QueryExecutionqe =
QueryExecutionFactory.create(query,
model);

org.apache.jena.query.ResultSet results =
qe.execSelect();

// Output queryString results

ResultSetFormatter.out(System.out,
results, query);

qe.close();
```

Figure 5: sample query

4. Experiments and Results

4.1 Evaluation Metrics

Empirical evaluation is important to ensure the effectiveness and efficiency of a system. The F-score or F-measure is one of the most commonly used measures in Natural Language Processing, Information Retrieval and Machine Learning applications. F-measure is a weighted harmonic mean of Precision and Recall. Recall, also known as sensitivity or True Positive Rate (TPR), is the frequency by which relevant documents are retrieved by a system. Precision, also known as True Positive Accuracy (TPA) or Positive Predictive Value (PPV), is a form of accuracy that refers to the frequency by which the retrieved documents are relevant [16]. The f-measure combines the two into a single measure used to show the accuracy of a system. The three measures are calculated as follows:

$$\text{Presision}=\frac{\text{True Positives}}{\text{True Positives+False Positives}} \quad (1)$$

$$\text{Recall}=\frac{\text{True Positives}}{\text{True Positives + False Negatives}} \quad (2)$$

$$\text{F-measure}=2*\frac{\text{Recall*Precision}}{\text{Recall+precision}} \quad (3)$$

True positives is the number of relevant documents retrieved, false positives is the number of irrelevant documents retrieved while false negatives is the number of relevant documents not retrieved.

4.2 Experimental Setup

To evaluate the platform performance, a set of queries were prepared as the example shown in the table 1. The table represents search of documents using topics or titles. Q1 gives the authors of the book chapter titled 'Development of a New Mathematical Framework for Seismic Probabilistic Risk Assessment for Nuclear Power Plants'. Q2 gives all the papers about knowledge sharing while Q3 gives all the books on image processing.

Table 1: Evaluation queries

	Queries
Q1	Find authors of a mathematics book chapter titled 'Development of a New Mathematical Framework for Seismic Probabilistic Risk Assessment for Nuclear Power Plants'
Q2	Find papers on 'knowledge sharing'
Q3	Find books on 'image processing'

4.3 Results analysis and discussion

Table 2 shows results of queries that retrieve documents basing on specific topics or titles. The expected number of documents to be retrieved is identified for each of the queries. The results retrieved in each query depend on whether the discipline or category (eg telecommunication engineering, agriculture, computer science, etc) of documents has been specified or not. In order to calculate precision and recall, relevant documents have to be identified from the total number of documents retrieved. According to the results in *Q2* and *Q3*, the category of documents required being included in the query narrows down the search scope thus reducing the number of documents retrieved. Not including the category widens the search scope thus increasing the number of relevant documents that are retrieved. From the results, precision, recall and F-measure for each of the queries are calculated using the given formulas and the results presented in table 3 & 4. The calculations also depend on whether the category of documents required was specified in the query or not. Table 3 shows the results when the category is specified while table 4 shows the results when the category is not specified.

Table 2: retrieved results

	Expected number of documents	Retrieved number of documents			
		category specified		category not specified	
		Total retrieved	Relevant	Total retrieved	Relevant
Q1	5	5	5	5	5
Q2	20	8	8	12	12
Q3	23	12	12	22	22

The results obtained in table 3 & 4 show that exploitation of semantic techniques is fruitful since the platform produces high rates of Precision and Recall. According to the results, Search by specific title increases the precision and recall of the approach thus making it more accurate. Regarding Precision, the results in the table 3 & 4 show that semantic search presents a higher probability of the retrieved documents being relevant. They show that the platform is Precise since the documents retrieved are relevant.

However, recall decreased in some instances such as in Q2 & Q3 of table 3.This can be attributed to the fact that the queries were limited to a specific category of documents. Restricting the search results to a specific category reduces the scope of search thus decreasing recall and increasing precision. Despite the fact that specificity reduces the number of documents retrieved, it increases the preciseness of the approach since the probability of only relevant documents being retrieved increases. In general, the evaluation results show that the approach is more precise. The approach has proved more efficient since only relevant documents are retrieved regardless of the category being specified or not. Though in some cases the f-measure is low, especially when the field is specified, the approach is still more accurate. This can be seen through f-measures of queries that did not specify the category of documents. The high f-measures can be attributed to high precisions and recalls.

Table 3: Evaluation results – category specified

	Precision (%)	Recall (%)	F-measure
Q1	100	100	1
Q2	100	40	0.571
Q3	100	52.174	0.686

Table 4: evaluation results-category not specified

	Precision (%)	Recall (%)	F-measure
Q1	100	100	1
Q2	100	60	0.750
Q3	100	95.652	0.978

Apart from being precise and accurate, the ontology has also proved better than the ones applied in examination system, DC-THERA and Wiki-I. Compared to DC-THERA and Wiki-I, the ontology driven approach in knowledge sharing is open to knowledge from any source. DC-THERA and Wiki-I, only allow unidirectional flow of knowledge while the approach presented in this paper allows bidirectional flow knowledge. Furthermore, unlike the exam system ontology, that was not evaluated, the performance of the ontology presented in this paper is evaluated using software performance metrics. This therefore makes this ontology better than the others.

5. Conclusions and Future Work

This paper presented an ontology driven approach for sharing and retrieval of knowledge among research students. First, a survey was done that showed most of the techniques used in sharing of knowledge are web 2.0 and web 3.0 (semantic web). Web 2.0 techniques have proven to support collaboration while semantic web structures the shared content for easy processing by machines. Semantic web has been used in various areas such as medicine, engineering and e-learning among others to structure knowledge for easy access and retrieval. It is therefore, the most appropriate technique to use to structure research materials for easy access and retrieval by research students. An ontology is developed and its performance evaluated using precision and recall. This approach proves better for it does not limit the sources of research materials that are used to populate the ontology. The wide scope of research materials makes the approach more interactive and allows bidirectional flow of knowledge in a research environment. Currently, the query construction is restricted to selection boxes on the interface. This limits the expressiveness of resulting queries.

Future work should explore how to use natural language processing (NLP) in query construction to improve expressivity of queries.

References

[1] F. Adebayo, T. Akinwale, and A. Sofoluwe, “A Conceptual Framework for an Ontology-Based Examination System”, *(IJACSA) International Journal of Advanced Computer Science and Applications*, Vol. 2, No. 5, 2011, pp. 36-42.

[2] Apache Jena – A free and open source java framework for building semantic web and linked data applications, retrieved May 2, 2016 from http://jena.apache.org/

[3] T. Berners-Lee, J. Hendler, and O. Lissila, “The Semantic Web”, *Scientific American*, Vol. 284, No. 5, 2001, pp. 34-43.

[4] Kinshuk, I. Y. Chen, and N. Chen, “Examining the Factors Influencing Participants' Knowledge Sharing Behavior in Virtual Learning Communities”, *Educational*

Technology & Society, Vol. 12, No. 1, 2009, pp. 134-148.

[5] K. Giri, "Role of ontology in semantic web", *DESIDOC Journal of Library & Information Technology*, Vol. 31, No. 2, 2011, pp. 116-120.

[6] S. Hassan, and O. Ishaq, "Determinants of Knowledge Sharing Using Web Technologies among Students in Higher Education", *Journal of Knowledge Management, Economics and Information Technology*, Vol. 4, No. 2, 2014, pp. 1-22.

[7] G. Kalem, "Semantic web application: ontology-driven recipe querying", *Atilim University: computer Engineering master's Thesis*, 2005.

[8] D. Monticolo, I. Lahoud, V. Hilaire, and S. Gomes, "A Semantic Wiki to Support Knowledge Sharing in Innovation Activities", *IAENG Transactions on Engineering Technologies*, Vol. 9, No. 16, 2013, pp. 217-230.

[9] S. Panahi, J. Watson, and H. Partridge, "Towards tacit knowledge sharing over social web tools", *Journal of knowledge management*, VOL. 17, NO.3, 2013, pp. 379-397.

[10] K. N. Shamsi, and Z. I. Khan, "Development of an E- Learning System Incorporating Semantic Web", *International Journal of Research in Computer Science*, Vol. 2 No. 5, 2012, pp. 11-14.

[11] G. Michaela, A. Splendiani, J.M. Austyn, C. Duccio, C. Scognamiglio, and M. Brandizi, "Knowledge sharing and collaboration in translational research, and the DC-THERA Directory", *Briefings in Bioinformatics*, Vol. 12, No. 6, 2011, pp. 562-575.

[12] M. M. Taye, "Understanding Semantic Web and Ontologies: Theory and Applications", *Journal of Computing*, Vol. 2, No. 6, 2010, pp. 182-192.

[13] J. Y Ting, and M.S Majid, "Knowledge-sharing patterns of undergraduate students in Singapore", *Library review*, Vol.56, No.6, 2007, pp.485-494.

[14] J. H. Yang, Y.L. Chen, "A social network-based system for supporting interactive collaboration in knowledge sharing over peer-to-peer network", *International Journal of Human-Computer Studies*, Vol. 66, No. 1, 2008, pp. 36 – 50.

[15] Y. Yao, Y. Zeng, N. Zhong, and X. Huang, "Knowldge Retrieval", *International Conference on Web Intelligence*, Vol. 30, 2007, pp. 729-735.

[16] R. Marijan, and R. Leskovar, "A library's information retrieval system (In) effectiveness: case study", *Library Hi Tech*, Vol. 33, No. 3, 2015, pp. 369 – 386.