An Ontology-Driven Question Answering System For Computer Network Module

M.I.M.Nowshad
Faculty of Graduate Studies and Research
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
Email: mimnowshad@gmail.com

U.U. Samantha Rajapaksha
Faculty of Graduate Studies and Research
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
Email: Samantha.r@sliit.lk

2021 21st International Conference on Advances in ICT for Emerging Regions (ICter) | 978-1-6654-6686-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICter53630.2021.9774823

***Abstract*—The field of "question and answering" has become a popular area of research in recent years.The main reason for this is that the search through the "question answering" system is found to be more efficient than the normal search. The Question Answering (QA) Systems can be two types in general such as open-domain QA systems and closed-domain QA systems. Search engines are generally open-domain QA systems that are used to search and retrieve the data we need. However, instead of search engines giving accurate and precise answers to the user queries, they often returned the list of links. Then, the user clicks on each link one by one to get the answer. This method of searching can sometimes not give the precise answers to the user queries, or the user may have to spend more time searching for the answer, and hence the users may experience discomfort. This situation can be avoided by using the semantic concept. The normal web data are machine-readable and can be understood by humans, whereas the semantic web information is machine-readable and understandable. Ontology is the main component of the semantic web and it can be described as the structure of knowledge-representation of a particular domain or subject. It clearly describes concepts, roles, instances, and the relationships between them. The Question Answering (QA) system is one of the popular applications of ontology. Here, the QA system is used to extract the precise answer to the user queries from the data repository. The system can be developed using different techniques like NLP with IR, reasoning with the NLP, web-based QA System, ontology-based QA System, and more. This particular question answering system is developed using the ontology model. This ontology-driven question answering (QA) system provides the facility to the users to find accurate and concise answers for their queries in the Computer Network module of ICT Subject.The same questions that are asked in the ontology-based question answering system can be asked in the web-based system. The reason is that both of these methods are used in this system. This will make it easier for users to understand the difference between the two systems.The performance results of both these systems further strengthen the statement.**

***Index Terms*—Ontology, SPARQL, NLP, Computer Network**

I. Introduction

Normally users use search engines to find out the answers to their queries. The search engine is a web program that collects and organizes content according to the user's query. These search engines have billions of websites and web pages that help to answer the queries of different users. It identifies and adds websites to its database using three main techniques such as crawling, indexing, and serving search results. The most relevant and best search results will display as the list from its database. Then the user needs to click the link one by one to get accurate answers for his queries.

The Question Answering Systems (QA-Systems) find out accurate and precise answers for the user queries from its data repository. This data repository can be stored both non-structured and structured data [1]. The QA systems can be developed using different methods such as heuristic approaches, machine learning, and ontology. The researches that are based on the ontology methods are very popular in recent times since the resulting QA systems can benefit from knowledge modeling [2].

The common vocabulary of basic concepts and relations between the concepts can be defined with ontology. This kind of knowledge representation provides more benefit to extracting information about the domain. The question answering system can easily be implemented using the ontology as the ontology consists of structured data [3].

The subject-related QA systems are available in few numbers due to some reasons. One of the reasons can be the inclarity of how the ontology-driven processing and reasoning map with the subject domain. Besides, the QA system is implemented in certain languages such as English, Chinese, and Arabic at a large scale and this situation remains because of the application challenges of the NLP. Techniques such as syntactic analysis and Natural Language Processing were used in a few subject-related approaches of question answering models previously. However, those systems were not very efficient. Therefore, this study intends to expand its provision in this particular area to a certain extent.

This proposed QA system can be utilized in both web and ontology to retrieve the answer for the user's queries. The overall system diagram can be illustrated as in figure I.

This research paper is structured as follows. Section I discussed the introduction of the topic. Section II presents the literature review of previous question answering systems, their technologies, and how this system is different from other systems. Section III describes the adopted methodology of the proposed QA system followed by results and analysis in section IV. Also, a brief conclusion is articulated with the

978-1-6654-6686-8/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: CURTIN UNIVERSITY. Downloaded on August 15,2022 at 05:52:16 UTC from IEEE Xplore. Restrictions apply.

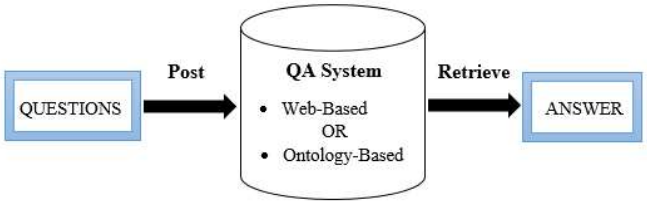

Fig. 1: General System Diagram

future work in Section V.

II. LITERATURE REVIEW

The Question and Answering system has evolved since the days when the "Semantic Web" became popular since the QA system development is a very common application of the semantic web. Semantic provides the meaning of a word in a sentence and semantic analysis is trying to expose the meaning of words and sentences in a real-world environment [4]. This section briefly discusses some Question and Answering systems and their techniques that are previously implemented.

QUASE – "QUestion Answering System for Education" which is a question answering system developed in a closed domain with a finite set of documents for answering natural language questions. This model incorporates several modules such as machine learning, NLP techniques, indexing-based retrieval mechanism, and ontology processing. The system mainly focuses on factoid questions to find different types of answers including human, location, numeric, abbreviation, and entity, and the question classification has been done using the question taxonomy. This QUASE system provides users two options for getting answers such as a correct answer or a nearly correct answer [5].

ADANS - This is an ontology-specific (i.e., Agriculture Domain) semantic question answering system that uses multiple techniques such as NLP, and semantic web technologies to answer the natural language questions of the users. The question is converted into a SPARQL query which is the loosening part of the system. Then, it helps to retrieve the answers within the acceptable time. This QA system is very convenient (i.e., Portable) and can be used for other domains also with small configuration changes [6].

AQUA - This acronym stands for "Automated QUestion Answering" which is a question answering system developed using several technologies such as hand-crafted ontology, Logic, and NLP. This AQUA model has implemented a framework that combines logic queries and information retrieval. The main aim of this system is to find correct textual answers to the user's question quickly [7].

HERO ontology – The acronym HERO stands for "Higher Education Reference Ontology". This HERO ontology has been implemented to overcome the issues of application ontologies. Because most of the application ontologies are too specific and cannot be reusable. The reference ontology of the higher education sector would be an important tool for various stakeholders who have a similar interest in higher education [8].

Authors A.S. Omarbekova et al., [9] has developed a QA system based on the knowledge base. Algebra is the knowledge base that is developed using the protégé editor software. This application-based system is used in smart learning such that students' knowledge on a given subject area can be checked using this system.

PowerAqua – It is also a QA system that uses multiple ontologies to answer the natural language questions from the users. Hence, this system is not limited to a single ontology like other QA systems. This system comprises of three major elements such as "Linguistic Component" which help to analyze the natural language query and translates it into linguistic triples, "Power Map" which identifies the type of ontology from the database, and "Triple Similarity Services" which helps to analyze the ontologies and generate the appropriate results [10].

W. M. D. Weerabahu et al. proposed an "ontology-based QA system for banking domain" which uses the template to generate answers. The answers to Frequently Asked Questions (FAQs) are stored in the ontology and it is considered as the knowledge base of the system. This research was used several techniques such as regular expression pattern creation, and linguistic structure to develop a chat application of digital assistant for the banking domain to answer user queries. The system is used the keyword searching method to execute the ontology. This system had limited data set as it used FAQ as the knowledge base. The system was able to achieve 75 % of the overall accuracy level for twenty test data [11].

Research [12] created an ontology in the music domain for the retrieval of semantic information. The ontology was constructed using the protégé 5.0 software and this ontology was tested using the Descriptive Logic ontology query language.

The coverage of the examination paper can be determined using the ontology. The research [13] described this study where ontologies are developed to the relevant subjects to calculate the depth. This automated system would help to keep the standard and quality of the examination papers. The examination papers were evaluated using the path of the words which is already saved in the ontology. Also, the feedback of the subject experts was obtained for the evaluation purpose and the accuracy of the ontology-based method is compatible with their values. Semantic of questions and answers can be represented by Ontology. Ontology is a data model that can be used to develop concepts and relationship with concepts [15] [16].Priyankara K.W.T.G.T et al. developed Android based tool to support self-learning for develop cognitive and psychomotor skills [17].Ontology based communication for unknown object was developed Rajapaksha et al. Ontology based semantic representation for unknown object more accurate than other image processing methods [18] [19].

In this particular research, a structured ontology database for the computer network module of ICT subject is constructed

Authorized licensed use limited to: CURTIN UNIVERSITY. Downloaded on August 15,2022 at 05:52:16 UTC from IEEE Xplore. Restrictions apply.

using the protege software. Then, this knowledge base is used to answer the user queries posted by the end-users.

III. METHODOLOGY

Any question answering system comprises mainly three components such as question analysis, document analysis, and answer retrieval. This particular model also complies with that pipeline perfectly.

An end-user enters the question in their native language using the user interface (UI) and gets the precise, and concise answer for their query within a short time. The question analysis part, answer retrieval, and processing part will take place in the middle of the channel.

The following diagram describes the process of the proposed system in detail.

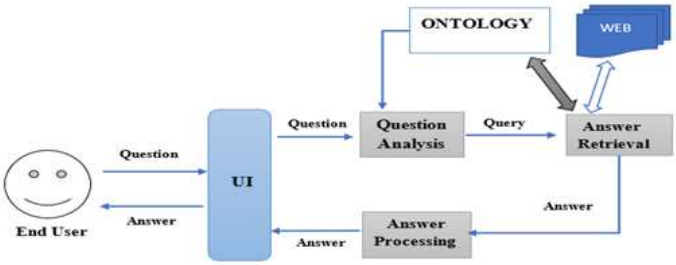

Fig. 2: The Architecture of the Proposed QA System

A. Question Analysis

The question answering process starts with the question analysis phase. Normally, the same question can be asked in different ways in natural language. Besides, the user's input can be the assertive or interrogative sentence. Therefore, the understandings of the semantic relationship of words in a sentence are very important in this case to find a solution to the query. In this stage, the question type also needs to be defined clearly. The below table shows the sample of that question classification.

Question Word	Function	Question Focus
Who	Asking what or which person or people (subject)	Person
Whom	Asking what or which person or people (object)	Person
Whose	Asking about the ownership	Person
Where	Asking in or at what place or position	Location
Year	Asking about the particular date	Date

TABLE I: QUESTION ANALYSIS

1) Pre-processing: The input sentence or question of the user needs to be preprocessed to remove less meaningful words, symbols, and numbers. It is very helpful to find the keyword of the question.

- Tokenization: Each sentence is split into pieces (words) and each piece is called a token. The token can be issued to the sentences also. The Tokenizing process is also changing the question sentence to lowercase, deleting unnecessary characters and symbols in the sentence.
- Stop-word removal: In this process, words that have less meaning, or no meaning are eliminated from the sentence. The deletion will take place when the words are matched with the stop-word dictionary.
- Stemming: Different forms of a word are replaced with its basic root word by eliminating the affixes (e.g., prefix, suffix, etc.).
- Lemmatization: The challenges of the stemming process can be rectified in lemmatization by referring to the dictionary for the word meaning.

B. Build the Ontology

The process of ontology building can take a lot of time, effort, and expert people in that particular domain as it consists of many stages. According to the research [14], the ontology building process should be subjected to a number of paces in practice.

- Decide the domain and the ontology scope – The process of ontology development begins with deciding the domain and identifying its scope. This research covers only one specific subject module called Computer Network. The user can get short concise answers from the ontology.
- Reuse the existing ontologies – At this point, the ontology previously created in the same domain area is taken into account as it is easier to modify the existing ontology to suit one's needs than to create a new ontology. However, for this research work, the ontology is created newly since the previous ontology in this domain does not exist.
- List out key jargon - All the essential terms that are used in this particular ontology are listed out at this stage. These terms are collected to form a comprehensive list that does not bother about the overlap of the concepts, relation among the terms, and properties of the concepts.
- Demarcate classes and hierarchy – The terms that are listed in the ontology are organized hierarchically. Then, the class hierarchy should be created. This can be implemented using multiple methods such as the top-down approach and bottom-up approach. We have used the "top-down development approach" that starts with the definition of the most general concepts in the domain and subsequent specialization of the concepts.
- Demarcate classes-slots properties - The competency questions from Step-1 cannot be answered using the classes alone because it does not provide adequate information to answer such queries. The internal structure of the concepts needs to be described before defining the classes.
- "Define the facets of the slots" – Here, "the facets are added to the properties and these facets include value type, allowed values, the number of the values (cardinality), and other features of the values the slot can take".
- "Build the instances" – At is this pace we used to create class instances in the hierarchy. The factors such as "choosing a class, creating an instance of that class, and filling in the slot values" are required to define an instance.

Any object is a subclass of a Thing in the protégé as the Thing is the superclass to all. The Network module has several

Authorized licensed use limited to: CURTIN UNIVERSITY. Downloaded on August 15,2022 at 05:52:16 UTC from IEEE Xplore. Restrictions apply.

sub-modules such as IP address, network devices, reference models, topologies, security concerns, protection methods, and so on.

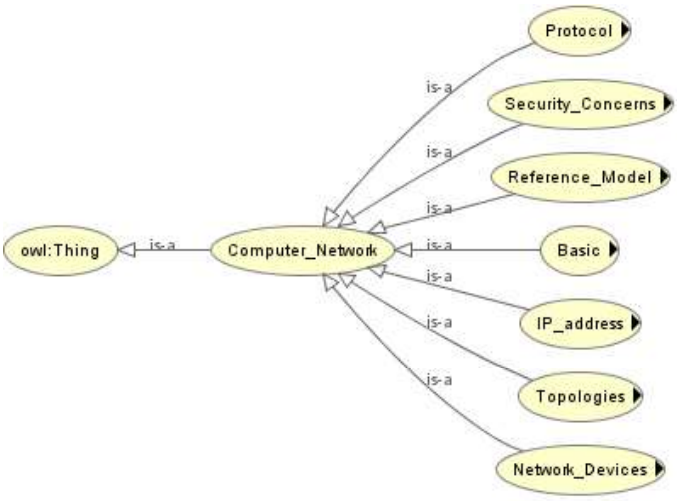

Fig. 3: Graphical view of Classes

The ontology of the Computer Network module consists of 49 classes that describe the concepts of the domain. Apart from this, the ontology has object properties that describe the relationship between the two classes and the data properties which describe the relationships between instances and data values. The ontology of this QA system comprises 24 object properties and 46 data properties.

C. Dataset

The ontology of the Computer Network module is created using the Protege software. The Computer Network is a module in ICT subject that taught in Grade 12 school curricular which was developed by the National Institute of Education of Sri Lanka. This data can be accessed from their official website http://www.nie.lk.

There are a total of 13 modules in the ICT subject in the Advanced Level curriculum. The computer network is one of them. This particular module also includes many sub-units such as network types, topology, network devices, IP address, and security concerns, and so on. Therefore, this ontology is built based on the computer network module syllabus.

D. Answer Retrieval

The answer extraction is carried out based on the keyword property. Which class or individual consists of the answer for the user queries can be detected easily using the keyword property.

When the end-users post their questions in natural language, then they will be changed into machine understandable instructions. During this process, several things will be carried out such as Stemming, Stop-word Removal, Tokenizing, Post Tagging, Keyword identification, and SPARQL query formation, etc.

SPARQL query formation consists of three processes such that keyword association, predicate identification, and property identification. These three processes related to how the SPARQL queries were built, which were used to draw answers from the ontology knowledge base.

In the SPARQL queries formation, all keywords play a crucial role as the keywords are associated with certain instances. The identification of the predicates is based on the keywords that are identified already in the previous stage. This process is important especially when an instance has more than one keyword. Then the property identification needs to be done as it is related to what kind of information will be displayed as the answer to a question.

After the SPARQL query is formulated, then that queries needs to be executed in the server called Apache Jena Fuseki which is a standalone, SPARQL server. The SPARQL query retrieves the short answers from the local ontology databases to the questions which are posed by the users.

E. User Interface

The user interface is designed using the python Tkinter package with the question search text field and the button. It connects the ontology and the user. The answer to the question will be displayed on the same screen. This user interface comprises also a drop-down list button where the user can choose the domain whether ontology-based or web-based.

F. Performance Evaluation

The performance of the proposed system can be evaluated using some parameters such as Precision, Recall, and F-Measure. The way of computing these figures using the mathematical equations are given below:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (1)$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (2)$$

$$F - Measure = \frac{2XPrecisionXRecall}{Precision + Recall} \quad (3)$$

IV. RESULTS AND ANALYSIS

Both QA systems are tested using some sample questions. Forty-five sample questions that are relevant to the computer networking module are posted in both systems and recorded the answers to compare with the real dataset answers.

The web-based QA system did not give precise and concise answers to the domain-specific questions most of the time. This system uses web resources to answer the user's questions. However, these web resources comprise a large amount of information and hence it consumes more time to answer the user queries compared to the ontology QA system.

The following results are obtained for the web-based QA system;

Authorized licensed use limited to: CURTIN UNIVERSITY. Downloaded on August 15,2022 at 05:52:16 UTC from IEEE Xplore. Restrictions apply.

Results	Quantity
Correct Answer	11
Nearly correct answer	06
Couldn't find the answer	23
Incorrect answer	05

TABLE II: WEB-BASED RESULTS

Here in this QA model, the non-comprehensive answers are considered to be the nearly correct answers. For instance, when the question is asked like "What is the computer network?" and the system returned the answer as "a network of computers". This answer is not a comprehensive one, hence we take this kind of answer as the Nearly or Partially correct answer.Furthermore, the system sometimes did not provide the answers at all and says to try something else. In this scenario, we considered this situation as "Couldn't Find the Answer". Otherwise, the system returned complete incorrect answers to some questions.

On the other hand, the ontology-based QA system provides very accurate answers to most of the sample questions. Here, the domain-specific knowledge base (e.g. Ontology) is used to retrieve the answers to the questions. Also, the ontology-based QA system takes a very short time to return the answers.

However, the ontology-based QA system's results were categorized into three groups such as Correct Answer, Incorrect Answer, and No Answer. These results also can be tabulated as follows:

Results	Quantity
Correct Answer	39
Incorrect Answer	03
No Answer Returned	03

TABLE III: WEB-BASED RESULTS

Based on these results, the performance of the systems can be evaluated. The aforementioned equations can be used to compute the Accuracy, Precision, Recall, and F-Measure. Then, the results can be tabulated to compare the difference. This outcome also provides the reliability of different QA systems to the users.

Performance Measure	Web-Based	Ontology-Based
Accuracy	38 %	87 %
Precision	0.4	0.93
Recall	0.8	0.93
F-Measure	0.5	0.93

TABLE IV: PERFORMANCE EVALUATION RESULT

The accuracy is expressed as a percentage and the value of the other performance parameters is found to be between 0 and 1. Value 1 is considered to be optimum and 0 is considered to be worst.

The figures which are tabulated in the above table can be graphically shown to better understandings as below figure 4.

V. CONCLUSION AND FUTURE WORK

In this calculation, our ontology-based system gives the highest value for all performance indicators than the other one.

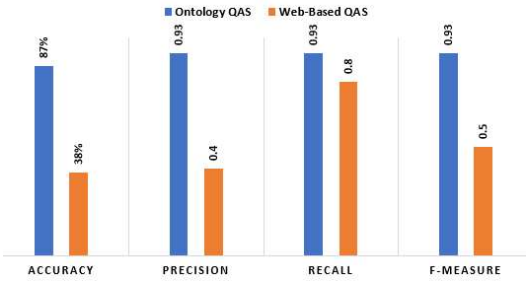

Fig. 4: The Comparison of the Performance

Therefore, the ontology-based question answering system can be the optimum system for users to find the answers to their queries.

The user can enter questions in the natural language completely or just use the keyword in both QA systems. The ontology-based system will be provided the same answer in both situations. However, in the domain-independent QA model, this situation is seen to be completely negative. For example, if acronyms are used in some query, it will not answer the query. But the system answers when it is queried in its full form.

The "Web-Based QA system" consumes more time compared to the ontology-based system as it consists of many resources. The domain-specific or ontology-based QA system has higher accuracy than the open-domain QA system. Especially when the user asks a very specific question in any particular domain (e.g., Computer Network Module), then the open-domain system returned several incorrect answers or could not find the answer. Because the web-based system is generally able to provide correct answers for common general questions. However, an ontology-based QA system gives very precise answers in that particular domain. The performance matrix of the system further confirms this outcome.

This ontology can be extended to accommodate other modules like Operating Systems, DBMS, and Data Communication, etc., in the ICT subject in the future. This can be done with small modifications of the current ontology. Otherwise, separate ontologies can be developed for each module and integrated all of them to answer the particular user query.

Besides, this ontology-based QA system can be implemented in other local languages of Sri Lanka like Sinhala or Tamil. Then, most users can greatly benefit from those QA systems. This QA model can be incorporated with some machine learning algorithms to improve its efficiency and usage across multiple dynamic knowledge-base

REFERENCES

[1] D. Kotkov, J. Veijalainen, and S. Wang, "How does serendipity affect diversity in recommender systems? A serendipity-oriented greedy algorithm," Computing, vol. 102, no. 2, pp. 393–411, 2020, doi: 10.1007/s00607-018-0687-5.

[2] W. Franco et al., "Ontology-based question answering systems over knowledge bases: A survey," ICEIS 2020 - Proc. 22nd Int. Conf. Enterp. Inf. Syst., vol. 1, no. April, pp. 532–539, 2020, doi: 10.5220/0009392205320539.

Authorized licensed use limited to: CURTIN UNIVERSITY. Downloaded on August 15,2022 at 05:52:16 UTC from IEEE Xplore. Restrictions apply.

[3] S. M. and P. C. R. M., Athira P., "Architecture of an Ontology-Based Domain- Specific Natural Language Question Answering System - india-Notes.pdf." International Journal of Web and Semantic Technology (IJWesT) Vol.4, No.4, October 2013, p. 9, 2013, [Online]. Available: https://docplayer.net/298326-Architecture-of-an-ontology-based-domain-specific-natural-language-question-answering-system.html.

[4] V. Mallamma and M. Hanumanthappa, "Semantical and Syntactical Analysis of NLP," Int. J. Comput. Sci. Inf. Technol., vol. 5, no. 3, pp. 3236–3238, 2014.

[5] V. Mishra and N. Khilwani, "QUASE: AN Ontology-Based Domain Specific Natural Language Question Answering System," Int. J. Recent Technol. Eng., vol. 8, no. 4, pp. 261–268, 2019, doi: 10.35940/ijrte.d6773.118419.

[6] M. Devi and M. Dua, "ADANS: An agriculture domain question answering system using ontologies," Proceeding - IEEE Int. Conf. Comput. Commun. Autom. ICCCA 2017, vol. 2017-Janua, pp. 122–127, 2017, doi: 10.1109/CCAA.2017.8229784.

[7] M. Vargas-Vera, E. Motta, and J. Domingue, "AQUA: An Ontology-Driven Question Answering System.," Directions Quest. Answering, 2003, [Online].Available: http://www.aaai.org/Papers/Symposia/Spring/2003/SS-03-07/SS03-07-009.pdf.

[8] L. Zemmouchi-Ghomari and A. R. Ghomari, "Process of building reference ontology for higher education," Lect. Notes Eng. Comput. Sci., vol. 3 LNECS, no. July 2013, pp. 1595–1600, 2013.

[9] A. S. Omarbekova, G. S. Nurgazinova, A. A. Sharipbay, A. Barlybayev, and G. T. Bekmanova, "Automatic formation of questions and answers on the basis of the knowledge base," Proc. 2017 Int. Conf. Eng. Technol. ICET 2017, vol. 2018-Janua, pp. 1–3, 2018, doi: 10.1109/ICEngTechnol.2017.8308192.

[10] V. Lopez, E. Motta, M. Sabou, and M. Fernandez, "PowerAqua: A multi-ontology based question answering system–v1," OpenKnowledge Deliv. D, vol. 8, p. 4, 2007, [Online]. Available: http://technologies.kmi.open.ac.uk/poweraqua/okdeliverable/OKD8 4PowerAquavfinal.pdf

[11] W. M. D. Weerabahu, M. A. G. A. S. Amalika, H. K. C. Dulakshi, and T. C. Sandanayake, "Ontology-Based Question Answering System for Banking Domain," no. Reiter 1995, pp. 205–211, 2007.

[12] N. Kaur, "Semantic Information Retrieval using String Ontology in Music Domain using Protege5 . 0," pp. 47–52.

[13] Y. T. Prematilaka, K. Banujan, and B. T. G. S. Kumara, "Ontology-Based Approach to Determine the Coverage of Examination Papers," in 2020 International Conference on Decision Aid Sciences and Application, DASA 2020, 2020, pp. 613–617, doi: 10.1109/DASA51403.2020.9317235.

[14] G. Ashour, A. Al-Dubai, and I. Romdhani, "Ontology-based course teacher assignment within universities," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 7, pp. 720–728, 2020, doi: 10.14569/IJACSA.2020.0110787

[15] S. K. Rajapaksha and N. Kodagoda, "Internal structure and semantic web link structure based ontology ranking," *in 2008 4th International Conference on Information and Automation for Sustainability*, 2008, pp. 86–90

[16] U. Rajapaksha and H. Fernando, "Ontology matching and ranking: issues and research challenges in semantic web application development," *in ITRU Research Symposium,* 2009.

[17] K. W. T. G. T. Priyankara, D. C. Mahawaththa, D. P. Nawinna, J. M. A. Jayasundara, K. D. N. Tharuka and S. K. Rajapaksha, "Android based e-Learning solution for early childhood education in Sri Lanka," 2013 *8th International Conference on Computer Science Education*, 2013, pp. 715-718,

[18] U. U. S. Rajapaksha and C. Jayawardena, "Ontology based Optimized Algorithms to Communicate with a Service Robot using a User Command with Unknown Terms," 2020 2nd *International Conference on Advancements in Computing (ICAC)*, 2020, pp. 258-262

[19] H. M. Y. L. Waruna Bandara, D. S. Wijesekera, H. M. T. D. Bandara Herath, D. L. Kodagoda and S. Rajapaksha, "Methodology for Coping with Uncertain Information Contained in Natural Language Instructions in a Robotic System," 2020 2nd *International Conference on Advancements in Computing (ICAC)*, 2020, pp. 234-239

Authorized licensed use limited to: CURTIN UNIVERSITY. Downloaded on August 15,2022 at 05:52:16 UTC from IEEE Xplore. Restrictions apply.