International Journal on Emerging Technologies 11(4): 474-485(2020)
ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Domain and Schema Independent Question - Answering on Linked Data (QALD): First Order Logic-Based (FOL), Non – SPARQL Knowledge Discovery on Lexical Ontologies

Kaneeka Vidanage[1], Noor Maizura Mohamad Noor[2], Rosmayati Mohemad[3] and Zuriana Abu Bakar[4]
Faculty of Ocean Engineering Technology & Informatics,
University Malaysia Terengganu (UMT), Malaysia.

(Corresponding author: Kaneeka Vidanage)
(Received 29 May 2020, Revised 02 July 2020, Accepted 14 July 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: With the extended use of semantic-web applications, use of ontologies in applied domains such as agriculture, bioinformatics and law rapidly escalated. During the construction of ontologies for the applied domains, the role of the domain specialists is recognised as vital. In collaborative ontology engineering, domain specialists are expected to review the accuracy of the encoded knowledge in each ontology increment, as ontology construction is an iterative and incremental operation. However due to the illiteracy of semantic concepts and unawareness in SPARQL query formation, effective engagement of the domain specialists is constrained. As a mechanism of addressing the aforementioned research gap, QALD prototype has been introduced to interrogate the ontology increments, despite the necessity for being SPARQL literate. Additionally, the prototype is capable of schema and domain-independent operation with no extensive human configuration efforts. All those aspects can be considered as the contributions of this research. Effectiveness of this prototype is quantitatively and qualitatively reviewed.

Keywords: applied ontology; first-order-logic; lexicon; ontology; schema, SPARQL.

I. INTRODUCTION

The semantic web is a domain rich conceptualisation technology, which is both human and computer comprehensible [1, 3]. Hence despite the recent inception, this technology has gained enormous success [6]. It is evident that semantic web is an ideal mechanism to encode specialised human knowledge, in facilitating numerous domain-related application requirements [4]. Therefore, the application of semantic web-based knowledge models has become significantly prominent in non-computing domains such as bioinformatics, agriculture and law [5, 7-8].

In accomplishing the aforementioned requirement of constructing knowledge models for non-computing domains, involvement of the domain specialists is vital, since it is a formation of their cognitive interpretations of a specific domain, which will be modelled via appropriate means of semantic technologies [9-10]. The lack of or absence of knowledge and understanding in semantic concepts vastly hinders the active and effective involvement of the non-computing domain specialists for the process of applied ontology construction. This adversely results in effective grouping of ontologists and domain specialists to collaboratively work towards an efficient and error-free ontology schema [11-13].

Researchers have pointed, the process of ontology construction is incremental and iterative. Therefore, the initial premature version of the ontology increment needs to be reviewed and cross-referenced for its accuracy and compatibility, before progressing further [20-21]. This is where the contribution of non-computing domain specialists becomes crucial. They need to examine the current ontology increment and assure, that the knowledge provided by them are properly encoded within the ontology increment [14 -16]. In reality, this examination can be done via firing multiple SPARQL queries on the current version of the ontology and verifying the accuracy of the results retrieved by referring it against the knowledge of the consultants. However in order to practically fulfil this need, the examiner ought to know the underlying schema of the current ontology version, triple concepts and the syntaxes to construct a proper SPARQL query [24].

However examiner being a non-computer domain specialist, it is not realistic enough to expect such technical skill-sets from the individual or the team [17-19]. This has led towards a bottleneck, hindering the effective and real-time accuracy verification of the ontology increments produced. This will be a strong impeding factor obstructing the effective use of human resources and construction of a high-quality ontology for the domain of concern.

Usually, in collaborative ontology construction projects, VoCamp strategy [46-47] is used to quickly and informally share the basics associated with the domain and technological aspects. Here, in addition to ontologists being enlightened on the basics associated with the domain of concern, domain specialists are also enlightened on the basics associated with the semantic concepts. Usually, multiple VoCamp sessions continue to ensure mutual understanding of the project requirements both by ontologists and domain-specialists [15-16]. As evident in literature, once the mutual understanding state is reached, proper and effective contribution of both the parties can be expected for the collaborative ontology construction project [44-45].

The skill set associated with SPARQL querying, schema understanding and appropriate SPARQL query formulation is a complex skill-set which is infeasible to be transferred via multiple VoCamp sessions to non-computing domain specialists [24-26].

As evident in literature, researchers have listed three main mechanisms to resolve the complexity associated with ontology verification. Those are knowledge base interfaces, graphical query building interfaces and question-answering systems. Among those three methods, the question-answering systems have been recognised as the most effective and widely used method [26-27]. Therefore, question-answering systems for linked data (QALD) has emerged as a special research niche under the umbrella of the semantic web [23].

Even though most of the existing QALD systems have several critical limitations such as domain dependence, schema-dependence and extensive human involvement for configurations before the usage and high computational processing overload, which can be pointed as a research gap which has not been properly addressed.

Therefore, this research is focused on proposing a novel mechanism to overcome the aforementioned limitations of the QALD systems and extend the practical usage of QALD systems to be user-friendlier, efficient, domain and schema independent.

II. RELATED WORK

[28] has developed a QALD system for a hospital to effectively inquire about important information related to the hospital of concern. The main drawback of this system is, that it is statically mapped only to the hospital ontology. Hence, it cannot work with another domain, making this system domain and schema dependent. Likewise, QALD system proposed by [29] is statically mapped to the Holy Quran for easier accessibility of the required information. Further, QALD system proposed by [30] also had the same limitations as it is statically mapped to the Intangible Cultural Heritage (ICH) domain. A user-friendlier web-based QALD system developed by [22] is also statically mapped to scientific events knowledge graph. Therefore, it can be concluded that all aforementioned QALD systems are bounded with the limitation of domain and schema dependence.

[31] proposed a QALD system called ORAKEL. This system is capable of working with any domain and schema. However prior to being operational, there is an extensive configuration process to be handled in order to overcome the lexical gap. Special software called FrameMapper needs to be used to ensure appropriate mapping of the domain-related lexicons with the relevant semantic regions of the ontology. This needs extensive human effort and has to be done under the careful observation of a human.

QuestIO is another QALD system proposed by [32]. This system is capable of converting the natural language question asked into its SPARQL variant. But the issue raised was the extremely low performance of the system due to excessive processing overload of the internal functionalities on natural language to SPARQL conversions and low precision of the system in investigations conducted. Therefore, it is apparent that, computational overhead and extensive human effort on lexical gap are also firm limitations that need to be resolved.

[23] declare that, in the schema-agnostic operation of a QALD, since the system is not firmly aware of directions and possibilities of how the objects and subjects are linked, it tempts to generate SPARQL queries for all possibilities related with the entities of concern. This is recognised as one of the key bottlenecks which escalate the overall processing overhead. Further, as the usable outcome, among multiple SPARQL query combinations generated, only one will be relevant and the computational processing utilised to generate all remaining query combinations will be an utter waste.

[24-25] have pointed out that, natural language to SPARQL query translation is a complex process with multiple steps, where the processing overhead cannot be fully controlled. As they claim, in this process, as the first step, keywords in the natural language question need to be mapped with the semantic entities of the ontology. Then, the query graph needs to be constructed by linking the relevant semantic entities and relationships by exploring the entire ontology schema. Subsequently, a suitable ranking mechanism needs to be executed in order to select the best SPARQL query being constructed to the natural language question that is being asked.

In an experiment conducted by [26] they have tried to introduce a deep learning model to automate the natural language to SPARQL query generation process. They have introduced this process as a Neural Machine Translation (NMT) approach for SPARQL query generation. However, the resulting outcomes of the experiment were not effective as expected, hence the model training time was close to 20 days. Further, extensive human effort was utilised to manually annotate the dataset used to create the deep learning model. Again with the introduction of a dataset to train the model, the entire attempt will become domain-dependent, as it is not feasible to define a domain-independent dataset.

Researchers have recognised multiple prominent approaches for QALD systems development. Those are rule derivation for SPARQL query generation after database analysis, trained deep learning model-based approaches and template-based methods. The biggest issue associated with all aforementioned mechanisms is that, all of them are domain-dependent and require extensive human effort. Because, database assessment and rule derivation, template-based pattern specific approaches and dataset maintenance for machine learning modules require extensive human involvement and it is infeasible to make those mechanisms domain independent.

Consequently, as argued above, there is a dire requirement for an efficient, domain and schema independent QALD mechanism which can function without extensive human involvement.

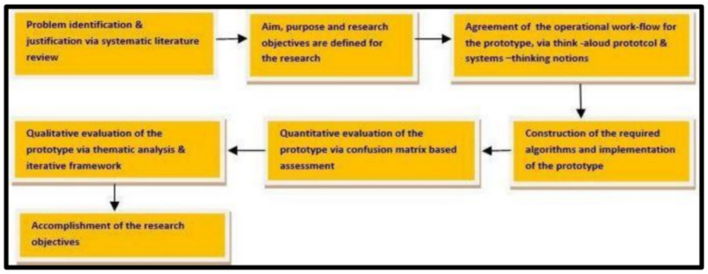

Fig. 1. Phase specific execution of the design science research.

III. METHODOLOGY

Demand for applied ontology construction [5, 7-8], difficulties experienced by the non-computing domain specialists [17-19], shortcomings prevailing in technological assistance [22, 24-25, 28-29] are reviewed in detailed in the aforementioned, introduction and related work sections of the paper. Provision of the technical aide to facilitate the collaborative ontology construction has numerous dimensions to be looked into. But as pointed by researchers [26-27], most prominence is towards the improved question-answering systems for linked datasets (QALD).

Design science research methodology [33] and its phases are used to govern the entire flow of this research, as of its investigative emphasis. Through, intense and extensive systematic literature review [34] the problem of concern is justified. Henceforth, aim and a purpose for the research is defined. Consequently, the blend of the think-aloud protocol [35] and systems thinking [36] notions are used to collectively brainstorm on the problem of concern. Afterwards, initial work-flow sketches resulting from the collective brainstorming sessions are derived to address the problem of concern. Subsequently, algorithms are constructed, depicting the work-flows recognized and implemented using java. Eventually, the blend of thematic analysis [37] and iterative framework [38] are used in combination to evaluate the effectiveness of the proposed solution. Detailed steps associated with each of the phases will be elaborated in the forthcoming, results and discussion, evaluation and conclusion sections. The graphical representation for the execution of design science research methodology in this research is elaborated as in Fig. 1.

Multiple subject experts are involved in the execution of the entire research process, which is depicted in Fig. 1.

IV. RESULTS AND DISCUSSION

The initial step to be accomplished for the successful execution of the proposed QALD system is the knowledge extraction from the ontology file. The developed prototype contains logic to extract knowledge from both Resource Description (RDF) and Ontology Web Language (OWL) formats. Fig. 2, mentioned below illustrates the overall work-flow associated with knowledge extraction from the ontology file formats and saving it in the database schema.

The proposed structure for the database schema is defined in figure 3 below, at an abstract level.

As elaborated in Fig. 2, format-specific logics are defined to extract semantically important sections from the ontology file, one by one and store them appropriately in the database relations, as depicted in Fig. 3. This systematic extraction process is fully automated and it is a one-time execution process, upon uploading the ontology file.

Consequently, the database relations will be filled up with the relevant semantic entities associated information. Saving of this information extracted from the ontology will be useful to several other purposes such as ontology verbalization and visualization requirements, other than the QALD. But, the emphasis of this paper will be limited only to discuss the QALD feature of the prototype.

As the next step of the QALD prototype construction, stored information in the database will be systematically extracted section by section. Afterwards, this information will be converted and written to the Prologue.pl file, in first-order logic format (FOL). This process is also fully automated, functioning without any human intervention. Upon completion of this process, entire ontology and its mappings will be available in its FOL format. The entire process associated with this operation is depicted, in Fig. 4.

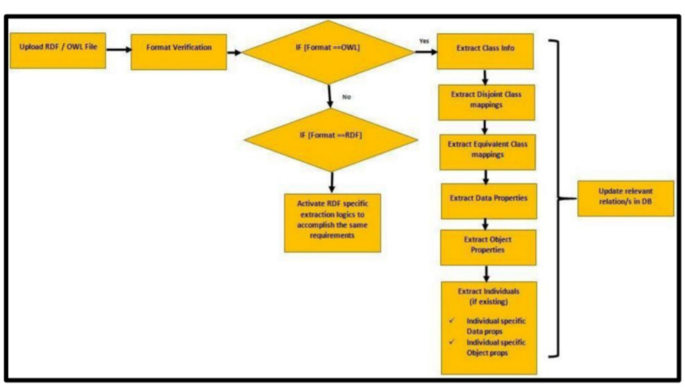

Fig. 2. Knowledge extraction work-flow from the ontology file.

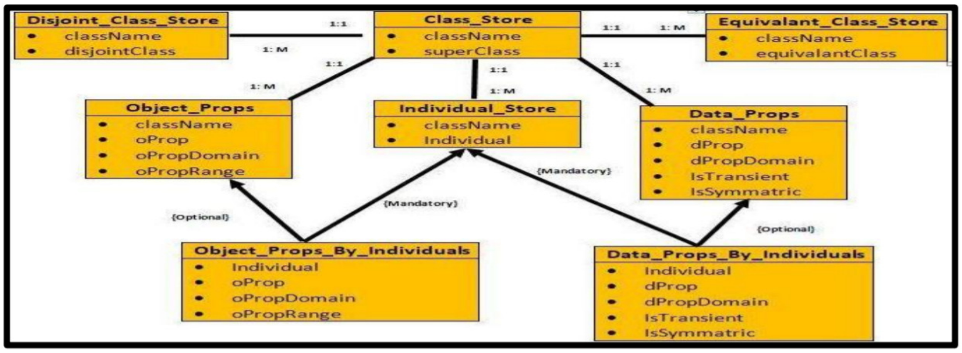

Fig. 3. An abstract representation of the database schema.

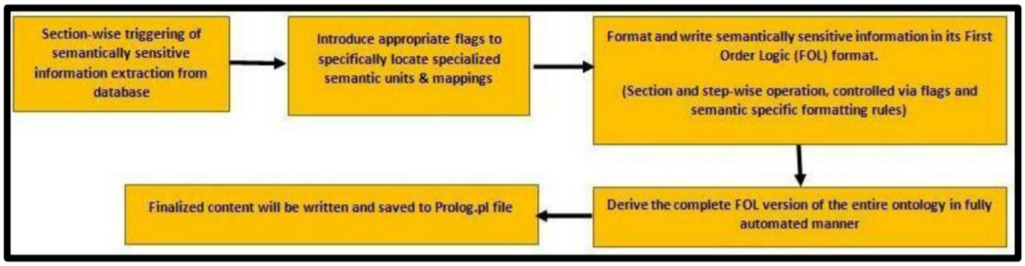

Fig. 4. Work-flow on converting ontology to its FOL version.

A fraction of the generated prologue.pl file contents can be depicted as mentioned in Fig. 5.

```
prolog.pl
class(student).
class(people).
inheritance(student,people).
class(people).
class(thing).
inheritance(people,thing).
class(module).
class(thing).
inheritance(module,thing).
class(cs).
class(module).
inheritance(cs,module).
class(maths).
class(module).
inheritance(maths,module).
class(lecturer).
class(people).
inheritance(lecturer,people).
disjoint(student,lecturer).
individual(student,stu2).
```

Fig. 5. Fraction of the Prologue.pl file containing the FOL version of the ontology.

```
public void RuleAnalyzer(ArrayList words)
{
  if(k==0)
  {
    list.add("classes");
    list.add("super");
    list.add("parent");
    list.add("sub");
    list.add("child");
    list.add("derived");
    list.add("property");
    list.add("propertiese");
    list.add("data");
    list.add("object");
    list.add("individual");
    list.add("individuals");
    list.add("disjoint");
    list.add("disjoint");
```

Fig. 6. Fraction of terms in the bag of words model.

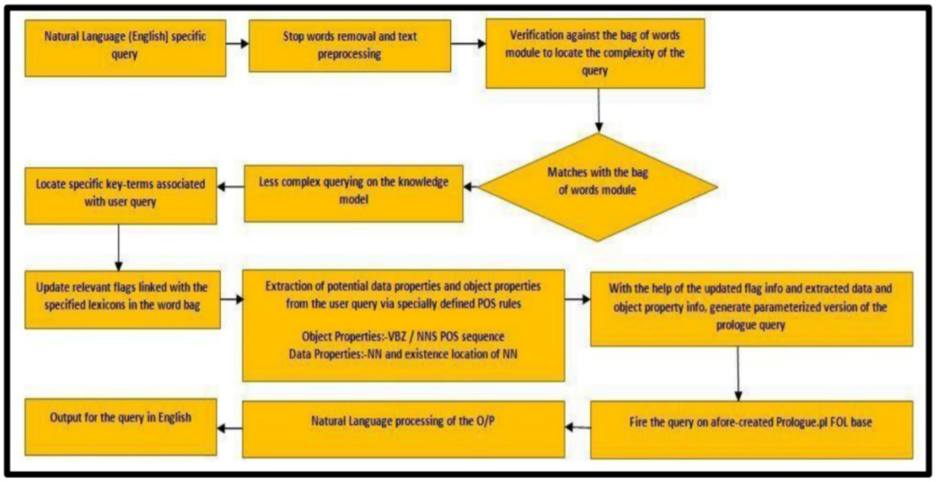

Fig. 7. Operational work-flow for direct questioning instance.

Fig. 8. Code snippet elaborating parameterized prologue query operations.

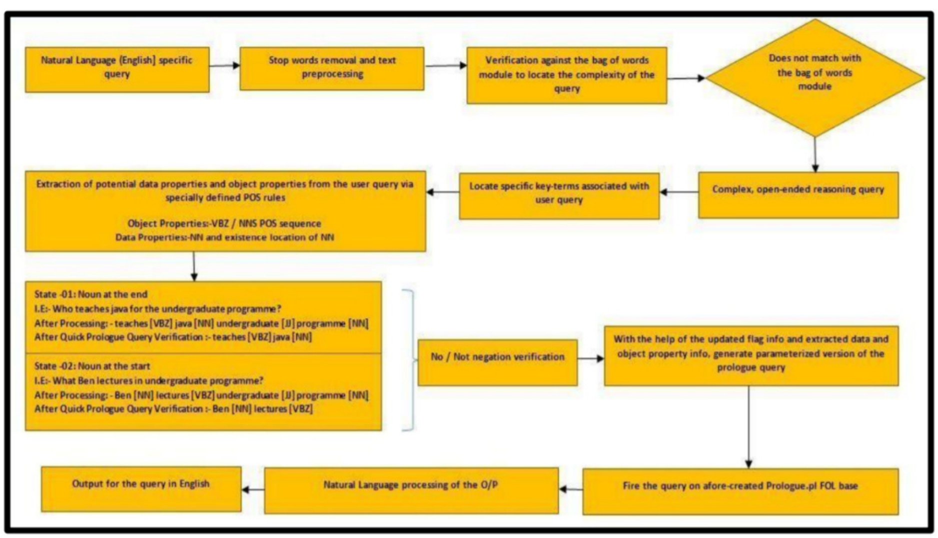

Fig. 9. Operational work-flow for complex reasoning instance.

Fig. 10. Code snippet on the impact of noun location and prologue queryformulation.

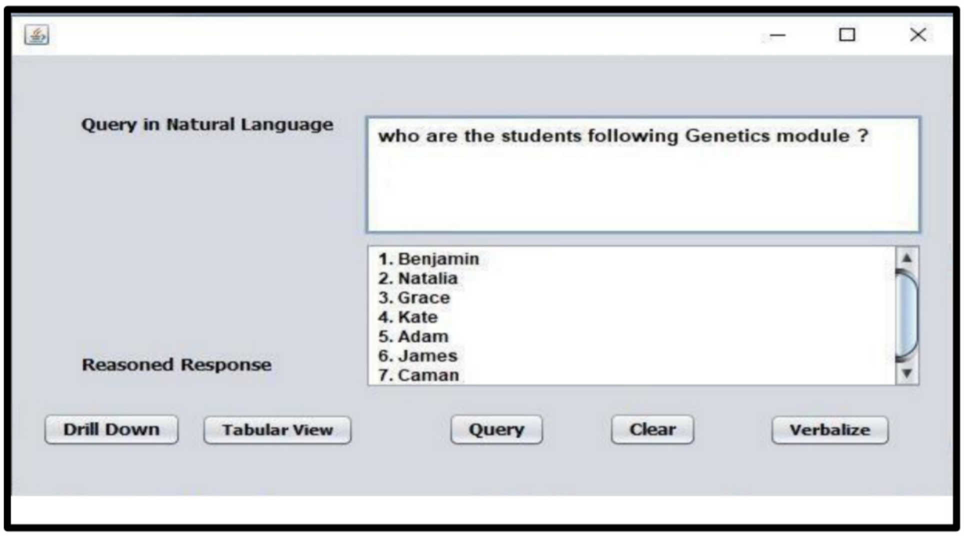

Fig. 11. QALD Interface.

N=TP + TN + FP + FN

	Predicted :No	Predicted :Yes	
Actual :No	TN=297	FP=2	(TN+FP)=299
Actual :Yes	FN=3	TP=13	(FN+TP)=16
	(TN+FN)=290	(FP+TP)=15	

Sensitivity = TP / (TP +FN)
=13 / (13 + 3)
= 0.81

Specificity = TN / (TN +FP)
=297 / (297 +2)
=0.99

Precision =TP / (TP +FP)
=13 / (13 +2)
=0.86

Accuracy = (TP + TN) / (TP + TN + FP + FN)
= (13 + 297) / (13 + 297 + 2 + 3)
=0.98

F-Measure = 2 (Precision x Sensitivity) / (Precision + Sensitivity)
= 2(0.86 x 0.81) / (0.86 + 0.81)
=0.83

Fig. 12. Confusion matrix derived for thoughts ontology verification for phase –II.

Provision of answers from the ontology to the natural language question occurs in two states. The first state is direct questioning from the ontology. A bag of words model (i.e. Fig. 6) containing common keywords associated with ontology querying is constructed. If the user's natural language question includes any of the terms located in the bag of words model, that query is considered as a direct questioning occurrence. In the triggering of a direct questioning scenario, work-flow depicted in Fig. 7 will be functional.

As depicted above in Fig. 7, upon the entry of the natural language question, first the text processing will take place and remaining lexicons in the user's query will be cross-referenced against the contents of the bag of words model defined in Fig. 6. This model contains lexicons associated with basic and straightforward questions which can be asked from an ontology, even without any idea about its underlying schema.

Direct questions such as "what are the classes of this ontology?", "what are the super/subclasses of this ontology? and etc" can be easily fired despite any understanding of the underlying schema. Henceforth, after viewing the results returned for those questions, more awareness about the schema can be derived, which paves the path to execute more specific queries such as "what are the individuals derived from the subject class ?" and "what are the data /object properties associated with Genetics individual derived from the subject class ? "

If the user query contains basic lexicons defined in the bag of words model, that query is recognised as a less complex direct questioning occurrence. Depending on the existence of specified terms in the user query after text processing, relevant flag attributes will be updated such as negation verifiers and object/data property conformance. Afterwards, collective governance, based on the information extracted from the flag structures, are utilized to trigger a parameterized version of a prologue query to be fired at the

'Prologue.pl', the file generated earlier. As necessary in certain instances, chained reasoning prologue queries are defined to conduct linked reasoning to derive the final result.

Afterwards, upon completion of the required natural language processing steps, the final answer is presented to the user in English. Some of the steps mentioned above are elaborated in the code snippet extracted and depicted in Fig. 8.

The second state of operation is a complex reasoning instance, where the lexicons in the user query do not match with the contents defined in the bag of words model. In the event of a complex reasoning operation, the work-flow defined in Fig. 9 will be operational.

As depicted in Fig. 9, during the complex reasoning work-flow operation, part of speech (POS) tagging based rules are defined to locate object properties and data properties. As an outcome of extensive verifications done with the experts involved for this research, it is identified that, data properties are always defined as nouns (i.e: first_name, last_name, salary, address) which are used to describe entities. However the object properties are used to describe relationships that take place in-between entities (i.e. actions), which are often recognized as certain form of a verb.

Apart from that, another important observation noticed was the change of interpretations, in natural language query, depending on the location of its nouns existence. Here, this observation is associated with, only the noun of concern, after text pre-processing and removal of the stop words.

The noun of concern can be easily recognized via a quick prologue verification query. In the occasion of locating multiple nouns, even after completion of text pre-processing phases, a simple prologue query can be fired on each noun existing and it will return true, only for the noun of concern, after inferencing against the 'prologue.pl' file being generated. Through this means, the noun of concern can be confirmed. Henceforth, the location of the noun of concern needs to be investigated as it is at the beginning or the end. Depending on its location, variables defined in the chained prologue query needs to be interchanged to derive the accurate results. This instance is clearly defined in Fig. 10.

As elaborated above, the definition of X and Y variables can be examined in the chained prologue queries.

In the example of noun at the end (i.e. who teaches java [NN]?) noun, 'java' is defined as the object (i.e. X) of the oProp triple sequence. However when the noun is at the beginning (i.e. what John [NN] teaches?), noun 'John' is defined as the subject (i.e. X) of the oProp triple sequence.

This observation of 'location of the noun of concern' is tested for quite a large number of natural language queries on several different lexical based ontology structures and it`s evident, adjustment of the noun as the subject and object are vital in obtaining accurate results for the natural language query of the user.

Apart from that, the existence of no/ not negations is also assessed to ensure accurate prologue query formation to derive correct results. Once all these processing steps are completed, finalised prologue query will be fired on the Prologue.pl file, which is auto-generated as per the work-flow explained in Fig. 4 above. Eventually, the result is presented in natural language to the end-user as depicted in Fig. 11.

One of the most important advantage of this mechanism is that, it has overcome the complex process associated with converting the natural language query into it` SPARQL variant, which could have high potential to be erroneous. Therefore, the intricate and multi-facet process associated with converting, natural language question to its SPARQL variant is entirely replaced by the auto conversion of the semantic web knowledge model into its first-order logic (FOL) format, with no human involvement and manual configuration at all. This has made the evaluation and accuracy verification of the approach also simple and straightforward, as English, grammar and SPARQL query mapping needs are not necessary to be reviewed, as that process is completely replaced. Many scientists have commended the efficacy of FOL format, due to its simplicity, unambiguity and structuralism [48-49]. Further, once the ontology is converted into its FOL format, the requirement to be aware of the schema is also spontaneously resolved. Hence, all these novel features resulting from this research can be pointed out as significant contributions to the domain of semantic web and QALD.

V. EVALUATION

Evaluation stage comprises of both quantitative and qualitative phases. In the quantitative phase, QALD prototype is exposed to five different ontologies in different domains and of various scales. Confusion matrix is used while true positive, true negative, false positive and false negative aspects are evaluated by verifying the accuracy of the returned results for the natural language queries fired against the knowledge of stakeholders. Returned results are logged in a matrix as mentioned in Fig. 12.

Hence forth, via following the similar process, by cross-referencing the accuracy of the results, precision, sensitivity, specificity, accuracy and F-measure calculations are derived. Consolidated outcomes of the test results are depicted in below Table 1.

In Phase-I testing, seven direct queries are fired at each of the above mentioned existing ontologies and results are verified. In phase-II after some assessment of the schemata of the ontology, five complex queries are fired at each and results are verified. The tests result depicted in Table 1 above, conveys the prototype's ability to function even with existing ontologies as well, despite the schemata and domain.

Among these ontologies tested, thoughts.rdf is a custom developed lexical ontology as a partial accomplishment of a cognitive behavioural therapy (CBT) assessment tool developed to assist consultant psychiatrists in providing psychotherapy for patients taking treatment for official stress via analysing patients' typed text-based internal-self-talk.

This is the ontology, which is used for the qualitative evaluation phase of the QALD prototype as well. Therefore it has been inferred as highly important to obtain a qualitative verification of this prototype by the stakeholders (i.e. a total of 09 consultant psychologists and psychiatrists) involved for the construction of the thoughts ontology. Details associated with thoughts ontology will not be discussed in this research, since it is out of the scope of this paper. However a snapshot of the taxonomy (i.e. Fig. 13) of the ontology and screen captures of the protégé version of the ontology (i.e. Fig. 14, 15) are provided for the verification means.

Table 1: Quantitative test results.

Ontology Name	Phase –I [Direct Querying]	Phase – II [Reasoned Querying]
Wine.rdf (W3.org,2020)	Sensitivity-0.85 Specificity-0.84 Precision-0.93 Accuracy-0.85 F Meassur-0.88	Sensitivity-0.78 Specificity-0.93 Precision-0.88 Accuracy-0.88 F Measure- 0.82
Shakesphear.owl (Chiba, 2017)	Sensitivity-0.88 Specificity-0.99 Precision-0.88 Accuracy-0.97 F Measure- 0.88	Sensitivity-0.73 Specificity-0.97 Precision-0.89 Accuracy-0.85 F Meassure-0.80
Pizza.owl (Protege, 2020)	Sensitivity-0.84 Specificity-0.98 Precision-0.91 Accuracy-0.98 F Measure- 0.87	Sensitivity-0.83 Specificity-0.98 Precision-0.88 Accuracy-0.98 F Measure- 0.85
Diabetics.rdf (El-Sappagh et.al, 2016)	Sensitivity-0.82 Specificity-0.96 Precision-0.88 Accuracy-0.94 F Measure- 0.84	Sensitivity-0.80 Specificity-0.96 Precision-0.84 Accuracy-0.91 F Measure- 0.82
thoughts.rdf	Sensitivity-0.88 Specificity-0.99 Precision-0.90 Accuracy-0.96 F Meassure-0.88	Sensitivity-0.81 Specificity-0.99 Precision-0.86 Accuracy-0.98 F Meassure-0.83

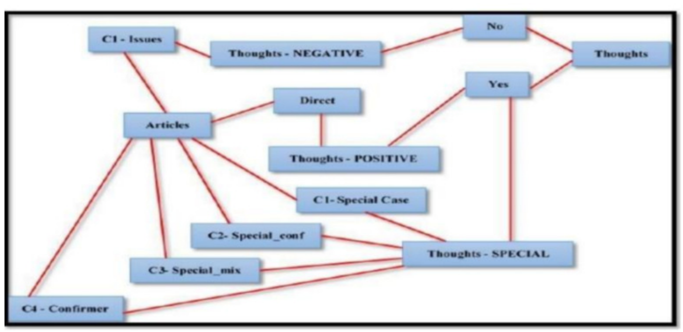

Fig. 13. High-level taxonomic structure of the thoughts ontology.

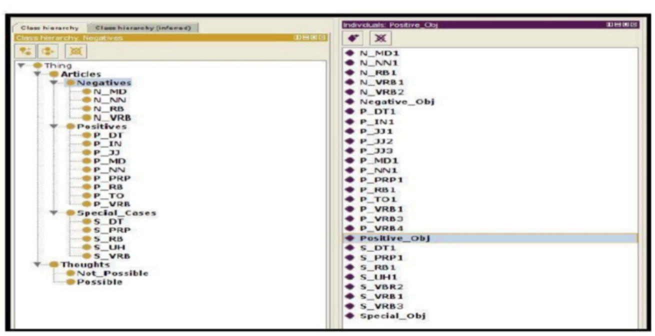

Fig. 14. Thoughts ontology lexicon organization in protégé.

Therefore, in fulfilling this requirement, QALD prototype is executed on the thoughts ontology with multiples of natural language queries, where results are cross-referenced with the stakeholders' knowledge (i.e. consultant psychiatrists and psychologists). After experimenting with the QALD prototype on the thoughts ontology, stakeholders involved in the evaluation process were interviewed. The same version of a controlled interview session is conducted for each stakeholder, based on multiple questions derived according to the CCP framework. CCP framework is a widely used mechanism for opinion segmentation. This framework comprises three main perspectives, which can be mapped with 'Wh' questions to be used in the interviewing process (Stockdale *et al*., 2006). CCP framework's perspectives and questionnaire mapping related to this research are elaborated in Table 2.

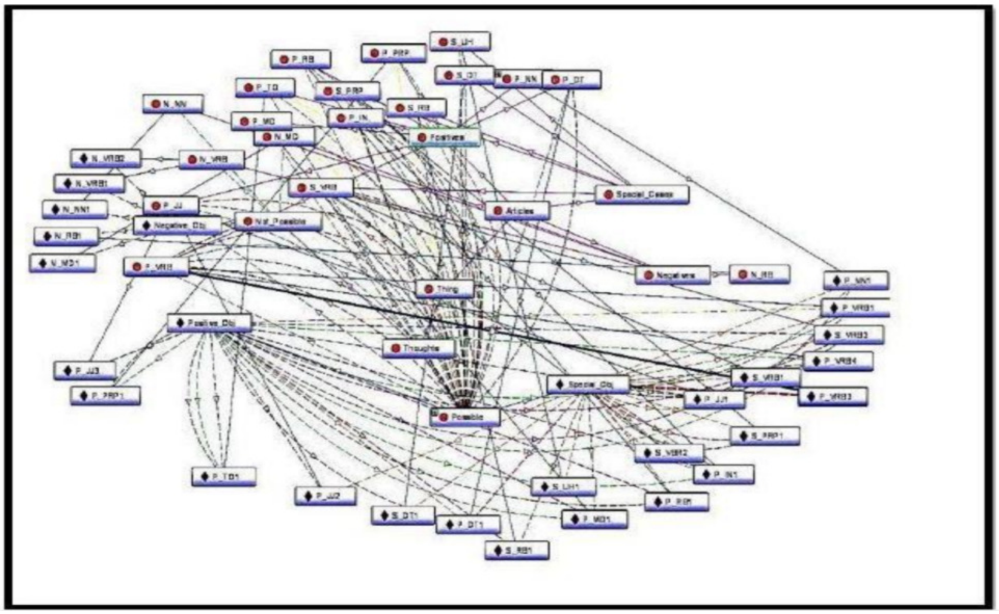

Fig. 15. Completed thoughts ontology (protégé visualization).

Table 2: CCP framework and questionnaire mapping.

CCP Framework Perspective	'Wh' questionnaire terms	Mapping question	Encoded facet	Index Code
Content	{What}	What are the advantages/disadvantages you experienced in this prototype?	Prototype Benefits	PB
		What is your opinion about the answers provided by this prototype?	Prototype Views	PV
Context	{Why}	Why do you think this type of prototype is useful / not useful?	Prototype Uses	PU
		Why do you think this prototype is accurate / not?	Prototype Accuracy	PA
Process	{How}	How do you think to use this prototype for ontology construction/verification?	Prototype Process	PP
		How could this prototype support ontology constructions?	Prototype Support	PS

Aforementioned questions are used to control the sequence of the interview sessions being conducted. This assures uniformity and more focused emphasis on the interviews being conducted. Responses provided by the stakeholders (count of 9 individuals) are recorded and later converted into text form to facilitate the assessment. Afterwards, indexed coded, thematic assessment is conducted for each of the interview transcript available in its text format. Thematic analysis is a very popular and credible qualitative evaluation technique, which is a disciplined neutral mechanism. This is an ideal mechanism to derive a quantitative essence for qualitative opinions, facilitating analytical assessment [37]. All interview transcripts are iteratively assessed for the indexed clusters derived from the CCP framework, as mentioned in Table 2 above. Fig. 16 below depicts the portion of the excel sheet used to organise the calculations associated with thematic analysis. Sections of the figure are made blacked-out to ensurethe professional ethics and adhering to the data protection law and privacy concerns of the textual feedback provided by consultants. Henceforth, that collected counts of expressions belonging to thematic categories are graphically presented in the form of a bar chart as visible in Fig. 17.

no	Response	Consultant code	Criteria Code	Category	Code
1	Sound idea	NF	PB	PrototypeBenefits	PB
2	Results provided are OK	NF	PV	Prototype Views	PV
3	Seems to be fucntioning accurately	NF	PB	Prototype Uses	PU
4	Works with good accuracy	NF	PU	Prototype Accuracy	PA
5	Works well	NF	PP	Prototype Process	PP
6	No big issues located	NF	PS	PrototypeSupport	PS
7	[redacted]	NF	PU		
8	[redacted]	CG	PP		
9	[redacted]	CG	PS		
10	[redacted]	CG	PU		
11	[redacted]	CG	PU	**Consultant**	**Designation**
12	[redacted]	CG	PB	[redacted]	Consultant Psychiatrist
13	[redacted]	CG	PV	[redacted]	Consultant Psychiatrist
14	[redacted]	CG	PV	[redacted]	Consultant Psychiatrist
15	[redacted]	CG	PA	[redacted]	Consultant Psychiatrist
16	[redacted]	CG	PP	[redacted]	Consultant Psychiatrist
17	[redacted]	CG	PP	[redacted]	Consultant Psychiatrist
18	[redacted]	CG	PV	[redacted]	Consultant Psychiatrist
19	[redacted]	DW	PA	[redacted]	Consultant Psychiatrist
20	[redacted]	DW	PS	[redacted]	Clinical Psychologist
21	[redacted]	DW	PU	[redacted]	Clinical Psychologist
22	[redacted]	DW	PU	[redacted]	Senior Councelor

Fig. 16. Organization of the qualitative feedbacks of consultants on the prototype for thematic analysis.

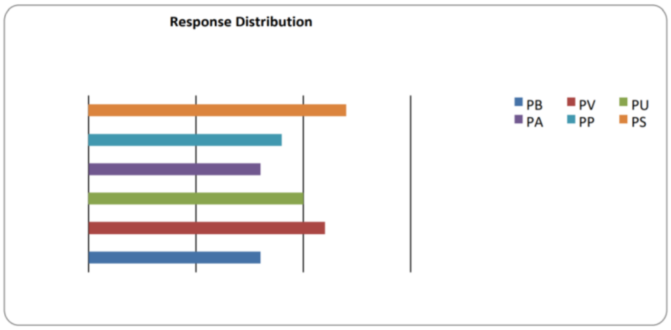

Fig. 17. Stakeholder`s response distribution.

Eventually, as defined in the research methodology (figure 1), the iterative framework is used to verify the overall efficacy of the research prototype, against the research objectives accomplishment. Table 3 mentioned below contains each step of the iterative framework and, its stepwise accomplishment justifications, extracted from the research. The iterative framework is a well-renowned qualitative result interpretation framework proposed by [38].

Table 3: Application of Iterative Framework.

Steps in Iterative Framework	Justification Elaborations
01 What are the data telling me?	Quantitative Metrics: - As depicted in table 1, above, multiple quantitative matrices are utilized to validate the effectiveness of the constructed QALD prototype and its operational effectiveness. Qualitative Assessment: - With the involvement of the stakeholders contributed for the thoughts ontology construction, empirical assessment of QALD prototype is done, in terms of the results returned, accuracy, usability, technical aid provided & etc. Henceforth, as visible in figure 17 above, the stakeholders' opinion distribution is plotted in a bar chart. As the collective reconciliation, it can be concluded both the quantitative and qualitative experimental phases have yielded with satisfactory results.
02 What do I want to know?	The overall efficacy of the QALD prototype developed in terms of domain/schema independence and minimized human work burden on configuration requirements.
03 Is there a dialectical relationship between step 01 & 02?	In the first quantitative phase of evaluation, the QALD prototype is experimented against five different ontology schemas, under two phases as the direct phase and reasoning phase. In both those phases quantitative matrices are calculated to determine the overall efficacy of the QALD prototype and it`s apparent the overall operation has yielded successful results. In the second phase of evaluation, a qualitative assessment is conducted on the QALD operation and the results output by the prototype. Stakeholders' opinions are thematically assessed and outcomes are graphically depicted in the form of a bar chart. Both quantitative and qualitative evaluation phases conducted on the criteria of the domain/schema independence and minimized human work burden on configuration requirements, have collectively yielded successful outcomes. Therefore, as the overall final reflection, as per the iterative framework rationale, it can be concluded as, there is a positive and satisfactory link between step-01 and step-02, which reflects the overall efficacy of the QALD prototype resulted from this research.

VI. CONCLUSION

In collaborative ontology construction, domain specialists and ontologists should have proper glueing, in order to result in an error-free applied ontology. In accomplishing this goal, ontology increment reviewing in each of its iterations is an extremely important, yet practically infeasible task.

The main barrier for this is the ill-literacy on semantic concepts and SPARQL querying, experienced by non-computing domain specialists.

Basic concepts associated with semantic web, such as the data-properties, No. object-properties, notion of triples can be shared with non-computing experts in the form of few voCamp sessions as already discussed above. Even though transferring the skillset of SPARQL querying and query formation is not an easy task, due to its requirement of prolonged experience and expertise in the computing domain since query formation is linked with schema comprehension as well. This hinders the active engagement of the non-computing domain specialists (i.e. medical doctors, lawyers, business professionals) in taking part in the knowledge reviewing stages associated with the ongoing development of the ontology.

Since effective ontology construction is always an iterative and incremental task, ideally the knowledge verification should be incorporated with every iterative increment.

The QALD prototype proposed in this research, eliminates the necessity of being literate on the semantic concepts and SPARQL querying skill-set. Hence with the opportunity provided to query any ontology in natural language, despite its underlying schemata and awareness of SPARQL, even a non-computing domain specialist can effectively engage in the knowledge verification process. Therefore, this opens the avenue to verify an ongoing ontology development project's knowledge verification iterations to be operated smoothly even by non-computing domain specialists in natural language.

However as a limitation of this mechanism, it can be highlighted that this QALD prototype will work only for lexical based knowledge models and it will not work for mathematical or formulae oriented ontologies as the natural language processing plays a vital part for the accurate functionality of this system.

Ultimately, this tool vastly contributes to the construction of error-free applied ontologies by providing a convenient platform for iterative and incremental knowledge verification needs to be fulfilled by non-computing domain specialists.

VII. FUTURE SCOPE

As the future expands on this research, it needs to be further and rigorously tested on numerous domains along with more complex reasoning based natural language queries. Additionally, frequently asked questions (FAQ) module to be integrated in order to disseminate the basic information associated with the ontology is under investigation. Since this will create an even more comfortable platform for the stakeholders to be aware of the organization of the underlying knowledge model, it will facilitate in inquiring more complex and relevant questions from the prototype to experiment and validate the ontology of concern.

Conflict of Interest.

REFERENCES

[1]. Spasic, I., Ananiadou, S., McNaught, J., & Kumar, A. (2005). Text mining and ontologies in biomedicine: Making sense of raw text. *Briefings in Bioinformatics, 6*(3), 239-251.

[2]. Kashyap V (2008). Ontologies and Schemas. The Semantic Web, 79-135. doi:10.1007/9783-540 764526_5

[3]. Abeysiriwardana, P. C., & amp; Kodituwakku, S. R. (2012). Ontology-Based Information Extraction for Disease Intelligence. *International Journal of Research in Computer Science, 2(*6), 7-19. doi:10.7815/ijorcs.26.2012.051

[4]. Katifori, A., Vassilakis, C., & Dix, A. (2010). Ontologies and the brain: Using spreading activation through ontologies to support personal interaction. *Cognitive Systems Research, 11*(1), 25-41. doi:10.1016/j.cogsys.2009.02.001

[5]. Yin, J. Hou, (2016). Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes. *Neurocomputing 174*, 643–650.

[6]. Ramanathan V. Guha (2013). Light at the End of the Tunnel". International Semantic Web Conference 2013 Keynote.

[7]. Alobaid, A., Garijo, D., Poveda-Villalón, M., Santana-Perez, I., Fernández-Izquierdo, A., & Corcho, O. (2019). Automating ontology engineering support activities with OnToology. *Journal of Web Semantics, 57,* 100472. doi:10.1016/j.websem.2018.09.003

[8]. Simperl, E., & Luczak-Rösch, M. (2013). Collaborative ontology engineering: a survey. The Knowledge Engineering *Review, 29*(1), 101-131. doi:10.1017/s0269888913000192

[9]. Saha, S., Usman, Z., Li, W., Jones, S., & Shah, N. (2019). Core domain ontology for joining processes to consolidate welding standards. *Robotics and Computer-Integrated Manufacturing, 59,* 417-430. doi:10.1016/j.rcim.2019.05.010

[10]. Elgammal, A., & Turetken, O. (2015). Lifecycle Business Process Compliance Management: A Semantically-Enabled Framework. *2015 International Conference on Cloud Computing (ICCC*). doi:10.1109/cloudcomp.2015.7149646

[11]. De Nicola, A., & Missikoff, M. (2016). A lightweight methodology for rapid ontology engineering. *Communications of the ACM, 59*(3), 79-86. doi:10.1145/2818359doi:10.1093/bib/6.3.239

[12]. Ingram, J., & Gaskell, P. (2019). Searching for meaning: Co-constructing ontologies with stakeholders for smarter search engines in agriculture. *NJAS - Wageningen Journal of Life Sciences,* 100300. doi:10.1016/j.njas.2019.04.006\

[13]. Strohmaier, M., Walk, S., P Pschko, J., Lamprecht, D., Tudorache, T., Nyulas, C. Noy, N. F. (2013). How Ontologies are Made: Studying the Hidden Social Dynamics Behind Collaborative Ontology Engineering Projects. *SSRN Electronic Journal*. doi:10.2139/ssrn.3199036

[14]. Ramesh, P., Vivekavardhan, J., & Bharathi, K. (2015). Metadata Diversity, Interoperability and Resource Discovery Issues and Challenges.

[15]. McDaniel, M., & Storey, V. C. (2019). Evaluating Domain Ontologies. *ACM Computing Surveys, 52*(4), 1-44. doi:10.1145/3329124

[16]. Façanha, R. L., Cavalcanti, M. C., & Campos, M. L. (2019). A Systematic Approach to Review Legacy Schemas Based on Ontological Analysis. *Metadata and Semantic Research,* 63-75. doi:10.1007/978-3-030-14401-2_6
[17]. Trokanas, N., & Cecelja, F. (2016). Ontology evaluation for reuse in the domain of Process Systems Engineering. *Computers & Chemical Engineering, 85,* 177-187. doi:10.1016/j.compchemeng.2015.12.003
[18]. Munir, K., & Anjum, M. S. (2018). The use of ontologies for effective knowledge modelling and information retrieval. *Appl Comput Inform., 14*(2): 116–126, ISSN 2210-8327.
[19]. Elve, A. T., & Preisig, H. A. (2019). From ontology to executable program code. *Computers & Chemical Engineering*, *122*, 383-394.
[20]. El Ghosh, M., Naja, H., Abdulrab, H., & Khalil, M. (2017). Ontology Learning Process as a Bottom-up Strategy for Building Domain-specific Ontology from Legal Texts. Proceedings of the *9th International Conference on Agents and Artificial Intelligence*. doi:10.5220/0006188004730480
[21]. John, S., Shah, N., & Smalov, L. (2016). Incremental and Iterative Agile Methodology (IIAM): Hybrid Approach for Ontology Design towards Semantic Web-Based Educational Systems Development. *International Journal of Knowledge Engineering, 2*(1), 13-19. doi:10.18178/ijke.2016.2.1.044
[22]. Fathalla, S., Lange, C., & Auer, S. (2019). A Human-Friendly Query Generation Frontend for a Scientific Events Knowledge Graph. TPDL.
[23]. Steinmetz, N., Arning, A., & Sattler, K. (2019). From Natural Language Questions to SPARQL Queries: A Pattern-based Approach. BTW.
[24]. Soru, T., Marx, E., Moussallem, D., Publio, G., Valdestilhas, A., Esteves, D., & Baron, C. (2017). SPARQL as a Foreign Language. ArXiv, abs/1708.07624.
[25]. Pradel, C., Haemmerlé, O., & Hernandez, N. (2013). Natural language query interpretation into SPARQL using patterns.
[26]. Soru, T., Marx, E., Valdestilhas, A., Esteves, D., Moussallem, D., & Publio, G. (2018). Neural Machine Translation for Query Construction and Composition. ArXiv, abs/1806.10478.
[27]. Hamon, T., Grabar, N., & Mougin, F. (2017). Querying biomedical Linked Data with natural language questions. *Semantic Web, 8,* 581-599.
[28]. Rencis, E. (2018). On Keyword-Based Ad-Hoc Querying of Hospital Data Stored in Semistar Data Ontologies. *Procedia Computer Science, 138*, 27-32. doi:10.1016/j.procs.2018.10.005
[29]. Beirade, F., Azzoune, H., & Zegour, D. E. (2019). Semantic query for Quranic ontology. *Journal of King Saud University - Computer and Information Sciences.* doi:10.1016/j.jksuci.2019.04.005
[30]. Dou, J., Qin, J., Jin, Z., & Li, Z. (2018). Knowledge graph based on domain ontology and natural language processing technology for Chinese intangible cultural heritage. *Journal of Visual Languages & Computing, 48*, 19-28. doi:10.1016/j.jvlc.2018.06.005
[31]. Cimiano, P., Haase, P., Heizmann, J. (2007). Porting natural language interfaces between domains: an experimental user study with the orakel system. In: IUI '07: Proceedings of the *12th international conference on Intelligent user interfaces,* 180–189. ACM, New York, NY, USA.
[32]. Damljanovic, D., Tablan, V., Bontcheva, K. (2008). A text-based query interface to owl ontologies. In: 6th Language Resources and Evaluation Conference (LREC). ELRA, Marrakech, Morocco.
[33]. McCarthy, J. (1980). "Circumscription—A Form of Non-Monotonic Reasoning." *Artificial Intelligence 13*(1–2): 27–39.
[34]. Whittemore R. (2005). Combining evidence in nursing research: methods and implications. *Nursing Research 54*(1): 56-62
[35]. Renzi, A.B., Chammas, A., Agner, L., Greenshpan, J.: Startup Rio: user experience and startups. In: Design, User Experience, and Usability: Design Discourse, Los Angeles, CA, 339–347. Spring (2015)
[36]. Arnold, R.D., & Wade, J. (2015). A Definition of Systems Thinking: A Systems Approach.
[37] Boyatzis, R. (1998). Transforming Qualitative Information: Thematic Analysis and Code Development.
[38]. Srivastava, P., & Hopwood, N. (2009). A Practical Iterative Framework for Qualitative Data Analysis.
[39]. W3.org. (2020). [online] Available at: https://www.w3.org/TR/owl-guide/wine.rdf [Accessed 24 Feb. 2020].
[40]. Chiba, J. (2017). Shakespeare's Ontology.
[41]. Protege.stanford.edu. (2020). [online] Available at: https://protege.stanford.edu/ontologies/pizza/pizza.owl [Accessed 24 Feb. 2020].
[42]. El-Sappagh, S., Ali, F. DDO: a diabetes mellitus diagnosis ontology. *Appl Inform 3*, 5 (2016). https://doi.org/10.1186/s40535-016-0021-2
[43]. Stockdale, R., & Standing, C. (2006). An interpretive approach to evaluating information systems: A content, context, process framework. *European Journal of Operational Research, 173*, 1090-1102.
[44]. Westerinen, A., & Tauber, R. (2017). Ontology development by domain experts (without using the "O"word). *Applied Ontology, 12*(3-4), 299-311. doi:10.3233/ao-170183
[45]. Braun, S., Schmidt, A., Walter, A., Nagypal, G. & Zacharias, V. (2007). Ontology maturing: a collaborative Web 2.0 approach to ontology engineering. In Proceedings of the Workshop on Social and Collaborative Construction of Structured Knowledge at the 16th International World Wide Web Conference (WWW 2007), Edinburgh
[46]. VoCamp Wiki. (2019). Retrieved January 29, 2020, from http://vocamp.org/wiki/Main_Page
[47]. Narock, T., & Fox, P. (2015). The Semantic Web in Earth and Space Science. Current Status and Future Directions. Amsterdam, Netherlands: IOS Press.
[48]. Guizzardi, G. (2019). Ontology, Ontologies and the "I" of FAIR. Data Intelligence, 181191. doi:10.1162/dint_a_00040
[49]. Sanya, I., & Shehab, E. (2014). An ontology framework for developing platform ndependent knowledge-based engineering systems in the aerospace industry. *International Journal of Production Research, 52*(20), 6192-6215. doi:10.1080/00207543.2014.919422

How to cite this article: Vidanage, K., Noor, N. M. M., Mohemad, R. and Bakar, Z. A. (2020). Domain and Schema Independent Question - Answering on Linked Data (QALD): First Order Logic-Based (FOL), Non – SPARQL Knowledge Discovery on Lexical Ontologies. *International Journal on Emerging Technologies, 11*(4): 474–485.