Domain and Schema Independent Semantic Model Verbalization: A Conceptual Overview

Kaneeka Vidanage[1], Noor Maizura Mohamad Noor[2], Rosmayati Mohemad[3], Zuriana Abu Bakar[4]
School of Informatics and Applied Mathematics, University Malaysia Terengganu (UMT)
Kuala Nerus, Terengganu, Malaysia

***Abstract*—Semantic Web-based technologies have become extremely popular and its a success that has spread across many domains, additional to the computer science domain. Nevertheless, the reusability aspects associated with the created and available semantic knowledge models are very low. The main bottleneck associated with this issue is, the difficulty associated in understanding the complex schema of a knowledge model created and barriers associated with querying the knowledge models using SPARQL or SQWRL query formulations. This research emphasizes on proposing a verbalizer which can go beyond existing Controlled Natural Language (CNL) type verbalizers and to verbalizer knowledge stored in a knowledge model file written in either RDF or OWL format, despite its domain and schematics.**

Keywords—Ontology; OWL; RDF; Verbalize; Schema

I. INTRODUCTION

Ontologies are domain rich conceptualizations [1]. That is the definition given for ontologies by Spasic et.al in [1]. Resource Description Framework (RDF) and the Ontology Web Language (OWL) are the most prominent and World Wide Web Consortium (W3C) accredited standards for creating ontologies [2]. The initial idea of ontologies was elicited from the concept of Semantic Web by Tim Berners-Lee for the first time [3]. However, though the idea initially emerged in 2001, by 2013, almost more than 4 million web domains have incorporated semantic web technologies to their web sites [4]. This clearly depicts the massive growth of semantic web across the entire globe proving its remarkable success. Further to that, as claimed by Feigenbaum in [5] and Kashyap in [12] the noteworthy feature related to semantic web`s utmost success is the potential of both human and machine readability of semantic web`s knowledge representations.

The concept of ontologies that emerged from the initial idea of semantic web can be recognized as a very effective technology among contemporary computer researchers and enthusiasts. For means of justification, it can be easily pointed out, that, already thousands of ontologies developed for a variety of purposes are available in online repositories, almost with free accessibility. To name a few repositories where these predefined ontologies are available will be Vocab.Org [14], Swoogle [15], LOV [16], Protégé Wiki [17], AberOWL [18] and BioPortal [19]. Among them also, it's significant to emphasize both [18] and [19] are ontology repositories solely with human bioscience and diseases related aspects. This is critical evidence to point out that the ontologies as domain rich conceptualizations [1] are doing extremely well in other domains as well, without limiting itself to the computer science domain.

The taxonomic structure, positioning the individuals, assertions between individuals, object and data properties can be effectively visualized through a tool like Protégé [6] or Top Braid composer [7]. Even though, graphical visualization will also not be adequate in most cases, as it`s not only the computer scientists or ontologists who are expecting to seek and gain the benefits of ontologies. Even for computer scientists or ontologists as well, it would be a really challenging task to understand the schema of an ontology developed by another set of researchers [13]. On the other hand, it has been stated that the development of an ontology from scratch is not an easy task as so far, no 100% automated mechanisms are available on ontology construction. Human intervention is essential [20]. Therefore, as claimed by [15][16], methods and mechanisms need to be sought after to enhance the reusability and overcome technological barriers associated, with an understanding of already created ontologies. The effective outcome of this would be, knowledge dissemination associated with existing ontologies will be further improved, enhancing the reusability aspects as well. Additionally, it will prevent a precious piece of information resource being stagnated on the internet after serving, only the one specific purpose it had been created for, which is recognized as an utmost cognitive waste as well by [21].

The ontology-based applications are not only limited to the computing domain. Medical sciences, pharmaceutical sciences [8][9], library sciences [5], law [10], criminology sciences [11] and ample other industries also comprehensively utilize the benefits of ontologies. This brings out the argument, potential capabilities of ontologies are not only sought after within the computer domain. As already conversed, several other disciplines are also very much keen on integrating the capabilities of ontologies to fulfil their discipline-specific requirements as well.

This setting will clearly open up the atmosphere to point out the greatest two bottlenecks associated with semantic web based ontologies, which will be leading to the research question discussed in this paper.

Firstly, understanding of the schema of an ontology written in RDF or OWL is a greatly challenging task even for computer scientists or ontologists as well. Therefore, for non-technical consultants like medical professionals, lawyers, criminologists, it would be a great obstacle [10] [11] [13].

Because without properly understanding the schema of the ontology, queries cannot be written to fulfil appropriate knowledge requirements. Secondly, writing of SPARQL or SQWRL queries to prorogate knowledge retrieval could be mostly an infeasible and unfair task to be expected from a non-technical specialist [22] [23] [24].

Therefore, as already conversed, these two issues will act as critical bottlenecks hindering the effective usage of semantic technologies within and outside of the computing domain. One of the potential solutions to overcome this technical barrier is to introduce ontology verbalizers. Verbalizers are capable of extracting knowledge represented in an OWL or RDF knowledge model and presenting it in a human understandable natural language [24].

But there are several problems associated with existing ontology verbalizers as well, hence most of them are domain and schema dependent, which means, they can work only with one domain as they have been tightly glued to one specific onology`s schema only [25][26]. The other issue is most of the verbalizers cannot work with both RDF and OWL formats and they work with either one of these and not with both, which again increases complexity in finding a suitable verbalizer for a required task [26] [27]. Eventually, most of the existing verbalizers can verbalize the knowledge in an ontology to a Controlled Natural Language (CNL) format only. CNL is a primitive English representation of triple sequences stored in an ontology model, which is not a conversational and readable English output which could be understood by anyone [28].

All these bottlenecks form the pathway to the research question to be discussed in this research which is to be "How to effectively verbalize both OWL or RDF based ontology, despite its domain and schema?"

The remaining section of the paper will discuss, about related works, methodology, results and discussion, evaluation and conclusion, respectively.

II. Related Works

It`s already conversed in the introduction section as well, there are two critical bottlenecks recognized to be hindering the reusability of existing knowledge models as well as adversely affecting the use of ontologies in other domains as well. To quickly revise, firstly the complexity associated with comprehension of the schematics, as without properly knowing the schematics of the ontology, writing appropriate queries for the knowledge retrieval would be infeasible [10] [11] [13]. Secondly, even after the hurdle of comprehending schema is achieved as the next step, writing of accurate SPARQL or SQWRL queries to achieve the knowledge retrieval demands [22] [23] [24] will become a critical challenge. Users should have a sound knowledge about triple concepts of the ontologies and the relevant syntaxes as well as RDF and OWL axiom related concepts to properly write a query to fulfil knowledge requirements.

It was already stated in the introduction section, it`s not always computer specialists or ontologists only who will be seeking the usage of ontologies [13]. Hence, these challenges would hinder the spreading of benefits of the ontologies to wider audiences within and out of the computer science arena.

One possible potential to overcome this barrier is to use ontology verbalizers. Nevertheless, there are ample issues associated with ontology verbalizers as well, as already conferred in the introduction. The researchers will investigate that aspect more deeply through the assessment of existing verbalizers.

Two such pieces of evidence for verbalizers are [29] and [30], which are acting as domain and schema dependent because both of these verbalizers are statically mapped to DBpedia and accommodation ontology. Hence, those verbalizers are not open to verbalize any other ontology file fed into it. MIKAT [31] is another verbalizer which is specifically defined for the breast cancer domain. This verbalizer acts by providing necessary assistance and guidance to the clinical investigations made by the consultants on their patients, related to breast cancer ailments and diagnosis [31]. In the same way [32] points outs another verbalizer which is specifically defined for the colonoscopy domain. This verbalizer has the capability of annotating video footages of an ongoing colonoscopy. Therefore, none of the verbalizers discussed above is capable of functioning as a generalistic verbalizer.

On the other hand, Noy et.al [33] pointed out, the most popular two formats associated with ontological knowledge representation are RDF and OWL. Even though most of the verbalizers available currently cannot work with both, but only with either OWL or RDF which is another bottleneck to be sorted out, when finding a verbalizer for a verbalization task. For instance, in [34] there is one such verbalizer which can work only with OWL and not in RDF format.

The other issue associated with the verbalizers are, they have still not reached the level of verbalizing the knowledge in the form of conversational English which can be read and understood by everybody. Most verbalizers extract and present the knowledge in CNL formats as already discussed in the introduction section as well. Attempt-to-Control–English (ACE) is one such popular form of CNL output [30] [35]. ACE is again a primitive English representation extracting out the triple arrangement in the knowledge file and it`s not enhanced as conversational English which could be read and understood by everybody.

Therefore, it's very apparent; there is a research gap to be addressed, on verbalizing an ontology despite its RDF or OWL as well as regardless of its domain and schema as well. In other words, the requirement of a generalized verbalizer, which can verbalize knowledge with a more mature level than ACE, is the research gap to be addressed.

III. Methodology

The main emphasis of this research is to come up with a generalized verbalizer, which can extract knowledge from ontologies despite their domain and schema as well as regardless of whether they are written in RDF or OWL formats. In fulfilling these goals, as the initial step, the comprehensive literature analysis was conducted via seeking for latest research and journal articles from credible repositories such as Springer, ACM Science Direct, IEEE etc. Keywords such as "ontology verbalizers, generalistic verbalizers, domain and schema independent verbalizers etc.."

are used to streamline the search results received from the research repositories mentioned above.

Even though lots of valid pieces of information has been collected, however, a proper solution addressing the research gap of generalized verbalizer which can work with both RDF and OWL, despite the domain and schema is not located. Further to that, it is also found, there is a deficiency issue in terms of verbalizing the knowledge as most of the existing verbalizers have not reached to a level beyond ACE as already conferred in detail, in the related works section.

Consequently, all these facts collected created a solid platform to further brainstorm and to continue the research. After completion of multiple brainstorming sessions with field experts and consultants, eventually, the following execution flow is derived as the initial step of extracting the required raw facts from the RDF / OWL knowledge models to commence up with the verbalization process. The proposed flow for the required facts extraction is denoted below, as in Fig. 1.

As illustrated in Fig. 1, the initial step is to check for the format of the knowledge model type. Because depending on its RDF or OWL appropriate extraction procedures needs to be triggered. Unless the user has to be notified with a suitable error message, claiming uploaded file type is not supported etc. Once the format verification stage is passed, the information extraction phase can be continued.

It`s decided to extract information, in sequential order of the individuals one at a time and one after another, rather than extracting information from here and there of the knowledge model, as it would adversely affect to the coherence related with further processing of the extracted information.

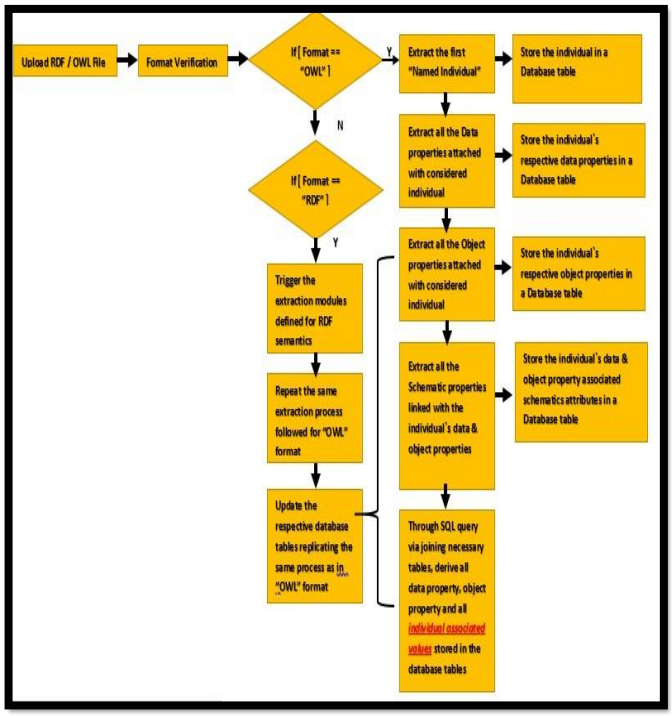

Fig. 1. Information Extraction Process from RDF / OWL Ontology File.

The process would be to select the first individual located in the knowledge model file. Here the individual means the entity of the knowledge instance located in the knowledge file. Then a sequential scan can be conducted throughout the entire document to extract all, subject, object, data properties, object properties, axioms, schematic information etc of the considered individual. Then using separately defined decision analysis and information extraction methodologies, all those information extracted can be again verified and carefully stored in a series of database tables designed according to the schematic structure specified below in Fig. 2. The individual element extraction processors will not be discussed in this paper as it is out of the scope of this paper, whereas the main emphasis of this paper will be on the verbalization process.

Here as denoted in Fig. 2, for each of the individual captured, an autogenerated fact Id can be introduced as a primary key as a mechanism of preserving information consistency associated with the respective individual. Then, using inheritance, properties are derived as data properties and object properties. These properties could not be overlapping, in the sense; a data property cannot be an object property and vice-versa. That`s why the "OR" disjoint constraint is used. Apart from that, another schematics relation is also introduced in the schema to take a track of special RDFS or OWL axioms linked with the individual's expressions. This measurement is taken to overcome the possible information losses associated with RDFS or OWL constraints defined in the knowledge model file when describing the individual's capabilities, domains, ranges etc.

Additionally, another table is defined to keep a track of individual's contexts. Because when it comes to verbalizing, the knowledge stored in an ontology, tracking the context would be very important in expressing the proper meaning out to the end user. It is intended to use discourse representation theory proposed by [36] [37] for the purpose of capturing the individuals' contexts. Fig. 3 mentioned below depicts the overall process flow associated with the discourse representation theory applied for this research. But the entire concept of discourse representation theory is not fully elaborated here, as it is outside the bounds of this paper. Therefore, the interested reader is encouraged to read the article mentioned in [37].

After completion of the process associated with extracting facts from the knowledge model file (depicted in Fig. 1) all extracted facts will be stored in the database schema proposed in Fig. 2. Then, that information can be again accessed in an individual-specific manner, assuring the triple sequence order to perform context assessment as per the discourse representation theory, as illustrated in Fig. 3 above. All extracted pieces of information, associated with the individual can be fed to a hash map and can perform, part of the speech tagging assessment and lexical analysis to trace potential changes causing on contextual differentiation. Then accordingly, context alert flag needs to be updated and it has to be supplied back to the verbalization module, along with the suggested pronouns to be used, which is technically referred as the discourse referent.

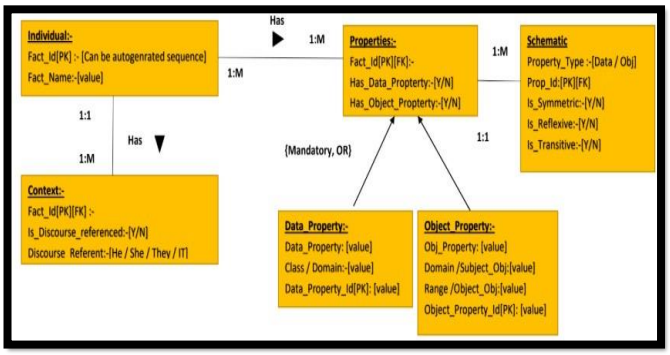

Fig. 2. Database Schema.

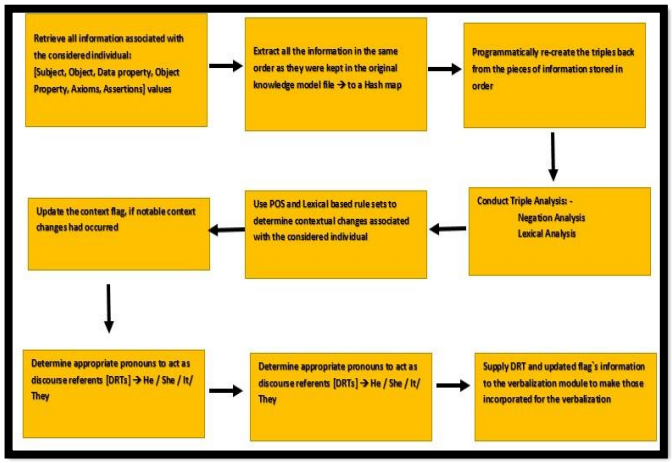

Fig. 3. Discourse Representation Theory Associated Process Flow.

Now all steps are in line to commence the verbalization process. The flow associated with the verbalization mechanism will be discussed in detailed under, results and discussions section, which is the next part of this research article.

IV. Results and Discussion

Up to now, the process associated with extracting important pieces of information from the RDF or OWL knowledge files, storing them in the database schema, application of discourse representation theory to ensure context sensing is, already discussed and illustrated in Fig. 1, 2 and 3, respectively.

The next important aspect is to discuss the verbalization process in detail. Upon the completion of the information extraction phase illustrated in Fig. 1, all individual-specific information is stored in the database schema presented in Fig. 2. Further, as in Fig. 3, discourse representation theory is applied to ensure context sensing information is recognized and proper discourse referents are introduced and context flags are updated as required. Then all these processed information, stored back in the database can be retrieved again to a hash map. It`s very important to keep on track, only the specific individual associated information is required to be extracted from the database. This will be feasible, as the information stored in the database is also governed and linked via entity and referential integrity constraints defined in the database schema.

The subsequent step would be to apply the Rapid Automatic Keyword Extraction (RAKE) algorithm [38] to all individual-specific lexical information derived as of one pool. RAKE algorithm will compute correlations amidst all the other individual-specific lexicons, within the pool and will derive a context relevancy value. Then all these lexicons' context relevancy values, lexicons, additionally added identification index values, need to be carefully stored in a temporary table which will be used later as a look-up grid for appropriate lexicon extraction. This process is graphically visualized in Fig. 4.

Once all these individual specific, lexicon`s context relevancy values, are extracted to the temporary table, the next step, intended is to apply, K-Means clustering algorithm [39]. K-Means is an unsupervised machine learning clustering algorithm, which could be used to cluster heterogeneous pieces of information into mostly feasible homogenous sets of clusters. Hence the intended goal is on generic verbalization, it has been decided to have 04 defined clusters representing, "introduction related—C1", "elaboration related–C2", "analysis related–C3" and eventually "conclusions related—C4" as those would be the ideal coverage which could be expected in terms of generic verbalization. Therefore, in terms of the K-Means clustering algorithm, it has been decided to specify K value as K=4, representing the four clusters from C1 to C4 derived above.

The ultimate expectation would be, once the K-Means clustering algorithm is applied to the consolidated individual, lexicon`s context relevancy values, need to be segmented into a mostly optimal 04 clusters, where within the cluster, information has to be mostly homogeneous. K-Means clustering algorithm works on the underlying concept of Euclidian distance. Therefore, the RAKE algorithms` context relevancy values would be much useful to segment the lexicons, considering their context relevancy values and grouping them into 04 homogeneous clusters. The entire process associated with the application of the K-Means algorithm is graphically illustrated in Fig. 5.

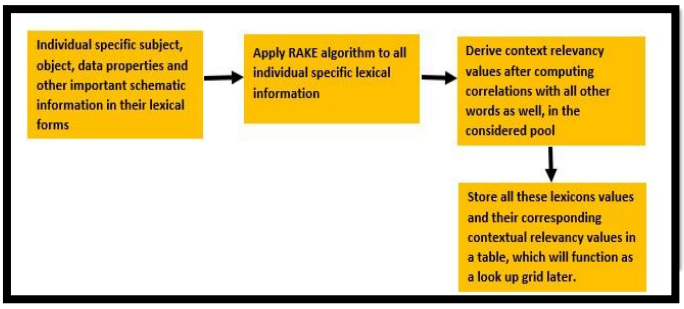

Fig. 4. Applying RAKE Algorithm to Get Context Relevancy Values.

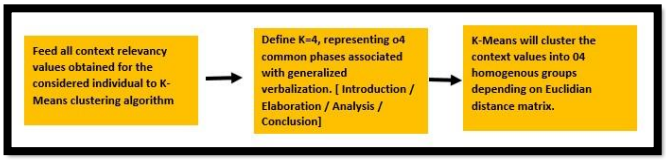

Fig. 5. Applying the K-Means Algorithm.

Afterwards, context-sensitive values residing in each cluster can be converted up back to the individual`s lexicons via referring them back with the look-up grid maintained earlier. This will result in four clusters of homogenous lexical groups belonging to the same individual.

The next step is to define four specified templates to cover up the verbalization scope of "introduction", "elaboration", "analysis" and "conclusion" which is synchronized up with the four clusters derived from K-Means clustering. But the issue is hence the K-Means clustering is an unsupervised clustering algorithm, and it`s not practical to directly mention which cluster can be mapped as "introduction" or "elaboration" or "analysis" or "conclusion" or vice-versa.

To overcome that issue, first –order-logic based Prologue rules can be introduced inside each of the four templates proposed, denoting their phased specific execution level, as "introduction" or "elaboration" or "analysis" or "conclusion". Then at the time of inferencing, these rules will execute and fill the templates with appropriate data elements derived from the knowledge model file. Then a thematic mapping algorithm like Latent Semantic Indexing (LSI) [40] can be used to determine, which template matches with which cluster. LSI is a very intelligent algorithm, which does a lot more than simple keyword comparison. The Single Value Decomposition (SVD) mechanism used in LSI functions as a dimensionality reduction mechanism via integrating all related dimensions to one specific theme. This SVD strategy is a key contributor resulting in the refined intelligent behaviour of the LSI algorithm [41]. Here the template does contain only a gist of information and using it as the triggering point, with the help of LSI algorithm, the appropriate cluster can be identified. The process of clusters and template matching is clearly noted in Fig. 6 mentioned below.

Eventually, via referring to the context-sensitive values of each of the lexicons belonging to a specified cluster, all lexicons within the cluster can be sorted from max to min of its context sensitive values. This will allow locating the nucleus and the satellites as per the Rhetoric Structure Theory (RST) which plays a vital role in micro-planning of sentences [42]. This will improve the readability of the text generated from the verbalizer. According to RST, the nucleus will postulate the most important fact associated within the context and the satellites will become the associative facts which will be elaborating the nucleus.

Ultimately, prologue governed phase specific templates can instruct the SimpleNLG framework to carry out the domain and schema independent verbalization process. The final step associated with domain and schema independent verbalization is as mentioned in Fig. 7.

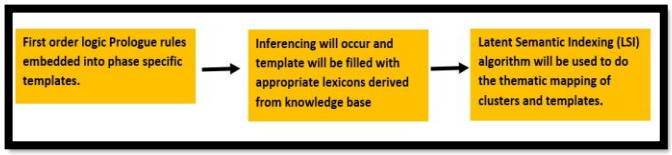

Fig. 6. Cluster and Template Matching.

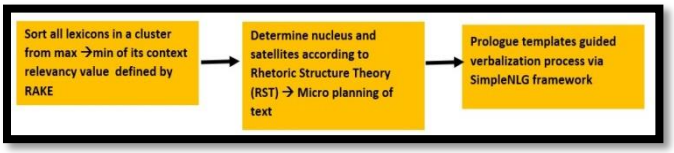

Fig. 7. Final Steps of the Verbalization Process.

V. EVALUATION

For the evaluation purposes of the domain and schema independent verbalizer, the process depicted in Fig. 8 below is utilized. The crime domain is selected as the application domain to test the verbalizer. The reason for that is, the research team already works with the government police and crime officers for several other crime analysis types of research, ongoing.

Initially, an existing knowledge model on the domain of crime knowledge is sought after. Then, later on, it's decided, in order to more accurately perform the evaluation process, to create a simple knowledge model associated with the sub-discipline of evidence handling related to a crime scene. Few of the crime officers (i.e. -6) are interviewed and a potential knowledge model on the domain of evidence handling was created via the use of Protégé in RDF. Then through Protégé Integrated Development Environment (IDE), same RDF knowledge model is converted to it`s OWL counterpart.

Subsequently, the RDF version of the knowledge model is uploaded to the verbalizer and the contents are verbalized in English. Then, the generated output is provided to a few of the crime experts in the department and they are asked to verify the effectiveness in the facets of understandability, information loss and reliability aspects. Henceforth, the same process is followed for the OWL version of the knowledge model as well. Interviewed feedbacks obtained from the crime specialists are thematically and statistically assessed via thematic analysis evaluation methodology.

As specified in the literature, thematic analysis is a very effective mechanism, which can be applied to any discipline in evaluating qualitative data [43]. Use of thematic analysis in computer science discipline is also very prominent. For instance, in [44] Porter et al. have stated the effectiveness in the use of thematic analysis for designing a proper UI/UX for e-personas leading into an e-government identification project. Likewise as suggested by multiple pieces of evidence [44-45], in developing information systems which make close interactions with humans, critical emphasis should be given to the interaction experiences [45] Without utilizing proper subjectivist approaches like thematic analysis, unforeseen negative interaction experiences cannot be extracted, which would ultimately lead to a failed deployment of the information system [43-45].

The suggested verbalizer in this research is also going to be a system, which closely interacts with the end user. Because the content verbalized by the verbalizer should be understood by the end user with almost no ambiguity. In order to check that dimension, users' personal interaction experience with the

system is very important. Therefore, these arrangements clearly point out the suitability of the thematic analysis evaluation technique to be used for this research as well. In the process of thematic analysis, the first step would be, detailed and repetitive reading of the interview feedbacks, documented through the interview process.

Then, rationales emerged out are carefully analyzed. Ideologies which have a close coherence are aggregated into groups and organized as facets. When interviewing more and more end users/subject specialists (i.e. crime officers), coherent feedbacks obtained are caused to accumulate the facet counters defined for each facet group, representing a numerical/statistical overview on the insights collected. The concept of Evaluation Onion proposed by [46] is utilized in mapping the facts emerged out from repetitive interview feedback reading, with facet criteria presented as facet groups.

Evaluation onion, which is also referred to as CCP framework [46] is recommended by Eslami et al. (2017) in [47] and many other researchers as well, in determining evaluation criteria's be used in the assessment of Information Systems which closely interacts with the ends users. Fig. 9 denoted below will present the overview of evaluation onion concept which is also referred to as CCP framework.

Table I will clearly illustrate how the recognized facts from interview feedbacks can be mapped within the CCP framework. Bold "Wh " question criteria's will clearly demonstrate how the CCP framework aspects are interlinked with interview comments facets.

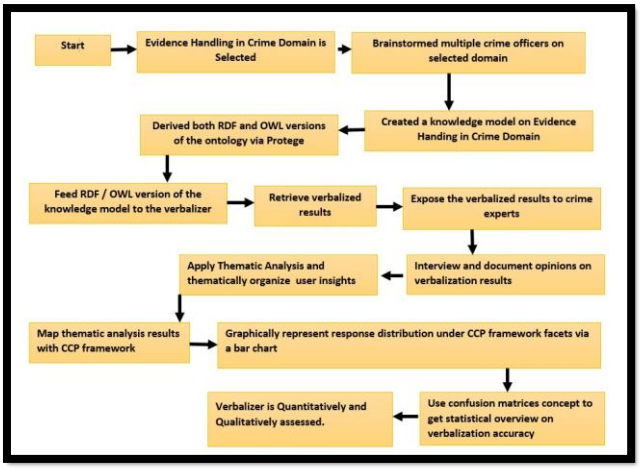

Fig. 8. Complete Evaluation Process.

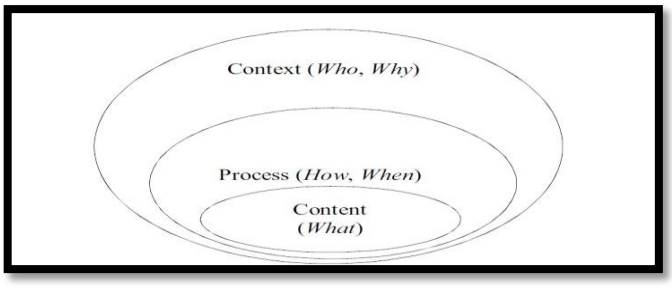

Fig. 9. CCP Framework by Farby and others (Farbey, Land, & Targett, 1993).

TABLE I. FACETS MAPPING INTO CCP FRAMEWORK

Facets	Indexed Code	CCP framework mapping question
Verbalizer Accuracy.	**PAC**	**What** is the level of accuracy experienced in this system?
Verbalizer Applications.	**PAP**	**How** this system could be useful for end users?
Verbalizer Assistance.	**PAS**	**Who** would be benefitted by the use of this system?
Verbalizer Importance	**PAI**	**Why** this research is important to end-users / experts?

Fig. 10 depicts the distribution of frequencies associated with each facet mentioned above.

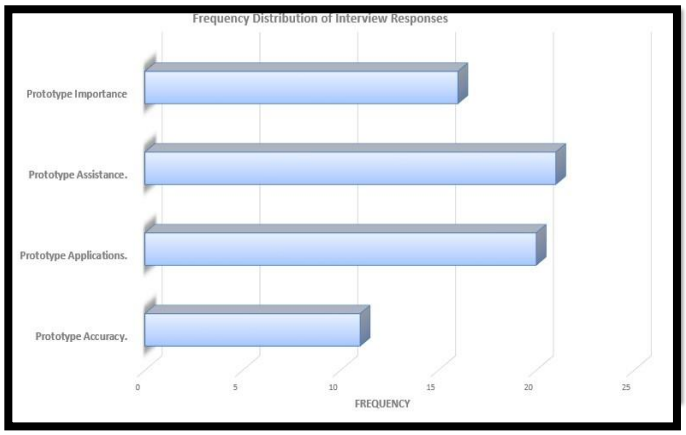

Fig. 10. End users Response Distribution Against Defined Facets.

As the second phase of evaluation, a statistical assessment is also conducted with the use of evaluation metrics. True Positives (TP), True Negatives (TN), False Positives (FP), False Negatives (FN), associated with the verbalization process is verified. The crime officers involved in the interview process, for the output evaluation, also involved in the creation of the knowledge model. Hence, they have the proper ideology in verifying the accuracy aspects associated with the verbalization, to recognize, are there any information losses, misinterpretations etc.

Subsequently, typical test measurements such as recall, precision and F-measures are derived, on the True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) confusion matrix element values derived. Table II depicts the statistics derived in one specific verbalization instance associated with 315 text expressions. Fig. 11 illustrates the calculation process initiated from the confusion matrices concept.

Here TP denotes accurate verbalization of expressions extracted from the knowledge model and FPs denotes incorrect verbalizations of the existing expressions in the knowledge model. Likewise, FN denotes misinterpretations resulted in verbalization process and eventually, TN denotes exclusion of less important axioms which occurred during the verbalization process. This will mainly occur via the functionality of the RAKE algorithm discussed above.

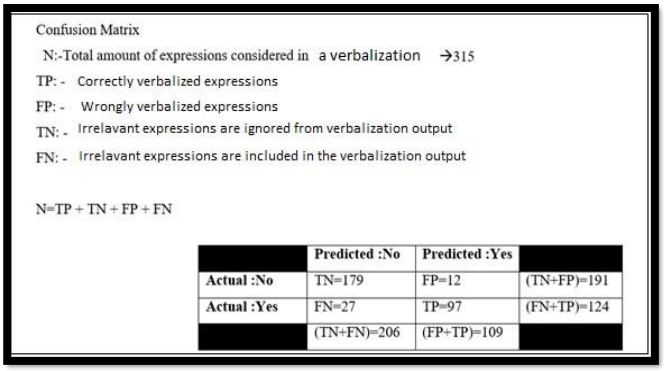

Fig. 11. Formulation of Test Statistics for Verbalization.

Quantitative test statistics derived for the verbalization took place for a specified instance, is logged in Table II.

TABLE II. TEST STATISTICS FOR A VERBALIZATION INSTANCE

Measurement	Accomplishment
Sensitivity	0.78
Precision	0.90
Accuracy	0.86
F-Measure	0.8

The verbalizer proposed in this research is qualitatively (via thematic analysis and CCP framework) and quantitatively assessed (via test statistics) as conversed above.

VI. CONCLUSION

The research gap attempted to address in this research article is to derive a potential resolution on the issue of domain and schema independent ontology verbalization, despite RDF or OWL formats. As already conversed in the paper, even there are few existing verbalizers located; most of them have multiples of issues, which are already discussed in related works section etc. Therefore, as a means of overcoming those deficiencies, this new conceptual arrangement of the verbalizer and its internal algorithmic functionalities are reviewed and evaluated in this paper. It is assumed, these findings will further contribute in making use of semantic technologies more applicable and addressable across a vast range of domains, despite the technical bottlenecks. However, the verbalizer proposed needs to be tested on several multiple domains, to further enhance and stably justify its accuracy.

VII. LIMITATIONS AND FUTURE WORK.

The most challenging aspect of the domain and schema independent verbalization is inability to use a large dataset as the training corpus to train the verbalizer to effectively perform the verbalization process, on any given domain. Because, at the stance, used a specified training dataset to train the verbalizer it will not further be a domain-independent verbalizer. Therefore, to rationally handle this requirement, a combination of algorithms and techniques such as RAKE algorithm on key phrase extraction, K-Means unsupervised learning algorithm, Prologue enabled phased specific templates and Rhetoric Structure Theory and SimpleNLG framework has been utilized as a pipeline of technologies (see Fig. 12).

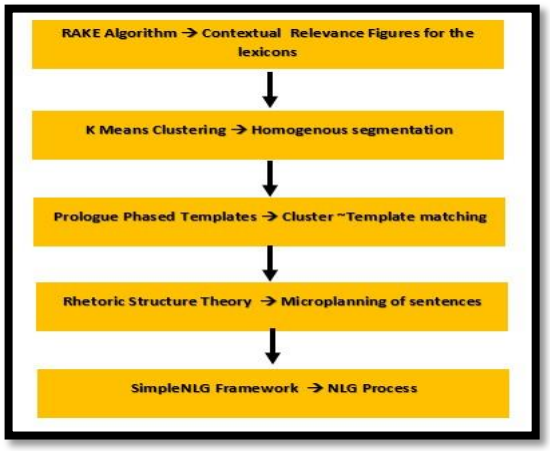

Fig. 12. Pipeline of Technologies.

At the moment, the verbalizer is only evaluated on the crime domain and it has to be further tested on other domains, such as medicine, law, management etc. Then test statistics such as recall, precision, F-measures can be derived for the verbalizer assessing the functionality on several other domains and, it will yield to derive a more normalized and stable outcome on the verbalizer performance.

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.doi:10.1093/bib/6.3.239

[2] Caldarola, E. G., & Rinaldi, A. M. (2016). An Approach to Ontology Integration for Ontology Reuse. 2016 IEEE 17th International Conference on Information Reuse and Integration (IRI). doi:10.1109/iri.2016.58

[3] Berners-Lee, Tim (May 17, 2001). "The Semantic Web" (PDF). Scientific American.

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

[5] Lee Feigenbaum (May 1, 2007). "The Semantic Web in Action". Scientific American

[6] Musen, M. A., & The Protégé Team. (2013). Protégé Ontology Editor. Encyclopedia of Systems Biology, 1763-1765. doi:10.1007/978-1-4419-9863-7_1104

[7] Topbraid Enterprise Data Governance (2019) Retrieved March 6, 2019, from, https://www.topquadrant.com/products/topbraid-enterprise-data-governance/

[8] Bontcheva, K., & Wilks, Y. (2004). Automatic Report Generation from Ontologies: The MIAKT Approach. Natural Language Processing and Information Systems, 324-335. doi:10.1007/978-3540-27779-8_28

[9] Bao, J., Cao, Y., Tavanapong, W., & Honavar, V. (2004). Integration of Domain-Specific and DomainIndependent Ontologies for Colonoscopy Video Database Annotation. Artificial Intelligence Research Laboratory-Iowa State University.

[10] Ku, C., & Leroy, G. (2014). A decision support system: Automated crime report analysis and classification for e-government. Government Information Quarterly, 31(4), 534-544. doi:10.1016/j.giq.2014.08.003

[11] Pinheiro, V., Furtado, V., Pequeno, T., & Nogueira, D. (2010). Natural Language Processing based on Semantic inferentialism for extracting crime information from text. 2010 IEEE International Conference on Intelligence and Security Informatics.

[12] Kashyap V (2008) Ontologies and Schemas. The Semantic Web, 79-135. doi:10.1007/978-3-540 764526_5

[13] Rusu, D., Dali, L., Fortuna, B., Grobelnik, M., & Mladnnic, D. Triple Extraction from sentences. Paper presented at Technical University of Cluj-Napoca, Romania.

[14] Davis, I. (2014). vocab.org - A URI space for vocabularies. Retrieved February 16, 2019, from http://vocab.org/

[15] Yu, L. (2007). Swoogle. Introduction to the Semantic Web and Semantic Web Services,145-157. doi:10.1201/9781584889342.pt3

[16] Vandenbussche, P., Atemezing, G. A., Poveda-Villalón, M., &Vatant, B. (2016). Linked Open Vocabularies (LOV): A gateway to reusable semantic vocabularies on the Web. Semantic Web, 8(3), 437-452. doi:10.3233/sw-160213

[17] Protege. (2018). Protege Ontology Library - Protege Wiki. Retrieved February 16, 2019, fromhttps://protegewiki.stanford.edu/wiki /Protege_Ontology_Library

[18] Slater, L., Gkoutos, G. V., Schofield, P. N., &Hoehndorf, R. (2016). Using AberOWL for fast and scalable reasoning over BioPortal ontologies. Journal of Biomedical Semantics, 7(1). doi:10.1186/s13326-016-0090-0

[19] Faria, D., Jiménez-Ruiz, E., Pesquita, C., Santos, E., &Couto, F. M. (2014). Towards Annotating Potential Incoherences in BioPortal Mappings. The Semantic Web – ISWC 2014, 17-32. doi:10.1007/978-3-319-11915-1_2

[20] 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

[21] Zenuni, X., Raufi, B., Ismaili, F., & Ajdari, J. (2015). State of the Art of Semantic Web for Healthcare. Procedia - Social and Behavioral Sciences, 195, 1990-1998. doi:10.1016/j.sbspro.2015.06.213

[22] Chergui, W., Zidat, S., & Marir, F. (2018). An approach to the acquisition of tacit knowledge based on an ontological model. Journal of King Saud University - Computer and Information Sciences. doi:10.1016/j.jksuci.2018.09.012 9.

[23] Alavi, M., Leidner, D.E., 2001. Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. Manag. Inf. Syst. Q. 25, 107–136. https://doi.org/10.2307/3250961. Anderson, J.R., 1983. The Architecture of Cognition. Harvard University Press, Cambridge, MA.

[24] Gutierrez-Basulto, V., Ibanez-Garcia, Y., Kontchakov, R., & Kostylev, E. V. (2015). Queries with Negation and Inequalities over Lightweight Ontologies. SSRN Electronic Journal. doi:10.2139/ssrn.3199213

[25] Williams, S., Third, A., & Power, R. (2011). Levels of organisation in ontology verbalization. ENLG. Retrieved from https://www.semanticscholar.org/paper/Levels-of-organisation-in-ontologyverbalisation-Williams-Third/08c6a058f5f78cf497 01d2534bf9c6a f3683f9e9

[26] Habernal, I., & Konopík, M. (2013). SWSNL: Semantic Web Search Using Natural Language. Expert Systems with Applications, 40(9), 3649-3664. doi:10.1016/j.eswa.2012.12.070

[27] Poulovassilis, A., Selmer, P., & Wood, P. T. (2016). Approximation and Relaxation of Semantic Web Path Queries. SSRN ElectronicJournal. doi:10.2139/ssrn.3199265

[28] Kaarel Kaljurand and Norbert E. Fuchs. 2007. Verbalizing owl in attempt to controlled English. In Proceedings of Third International Workshop on OWL: Experiences and Directions, Innsbruck, Austria (6th–7th June 2007), volume 258

[29] Poulovassilis, A., Selmer, P., & Wood, P. T. (2016). Approximation and Relaxation of Semantic Web Path Queries. SSRN ElectronicJournal. doi:10.2139/ssrn.3199265

[30] Kaarel Kaljurand and Norbert E. Fuchs. 2007. Verbalizing owl in attemp to controlled English. In Proceedings of Third International Workshop on OWL: Experiences and Directions, Innsbruck, Austria (6th–7th June 2007), volume 258

[31] Bontcheva, K., & Wilks, Y. (2004). Automatic Report Generation from Ontologies: The MIAKT Approach. Natural Language Processing and Information Systems, 324-335. doi:10.1007/978-3540-27779-8_28

[32] Bao, J., Cao, Y., Tavanapong, W., & Honavar, V. (2004). Integration of Domain-Specific and DomainIndependent Ontologies for Colonoscopy Video Database Annotation. Artificial Intelligence Research Laboratory-Iowa State University.

[33] Noy, N., & McGuiness, D. (2001). Ontology Development 101: A Guide to Creating Your First Ontology. Stanford University, Stanford

[34] Bontcheva, K., & Wilks, Y. (2004). Automatic Report Generation from Ontologies: The MIAKT Approach. Natural Language Processing and Information Systems, 324-335. doi:10.1007/978-3540-27779-8_28

[35] Bojars, U., Liepins, R., Gruzitis, N., Cerans, K., & Celms, E. (2016). Extending OWL Ontology Visualizations with Interactive Contextual Verbalization. VOILA@ISWC.

[36] Lascarides, A., & Asher, N. (2008). Segmented Discourse Representation Theory: Dynamic Semantics With Discourse Structure. Computing Meaning, 87-124. doi:10.1007/978-1-4020-5958-2_5 .

[37] Van Eijck, J. (2006). Discourse Representation Theory. Encyclopedia of Language & Linguistics, 660668. doi:10.1016/b0-08-044854-2/01090-7

[38] Gupta, S., Mittal, N., & Kumar, A. (2016). Rake-Pmi Automated Keyphrase Extraction. Proceedings of the International Conference on Informatics and Analytics - ICIA-16. doi:10.1145/2980258.2980463

[39] Wei, S., Yonglin, O., Qingcai, Z., Jiaqiang, H., & Yaying, S. (2018). Unsupervised Machine Learning: Kmeans Clustering Velocity Semblance Auto-Picking. 80th EAGE Conference and Exhibition 2018. doi:10.3997/2214-4609.201800919

[40] Al-Anzi, F. S., & AbuZeina, D. (2018). Enhanced Search for Arabic Language Using Latent Semantic Indexing (LSI). 2018 International Conference on Intelligent and Innovative Computing Applications (ICONIC). doi:10.1109/iconic.2018.8601096

[41] Amini, B., Ibrahim, R., Othman, M. S., & Nematbakhsh, M. A. (2015). A reference ontology for profiling scholar's background knowledge in recommender systems. Expert Systems with Applications, 42(2), 913-928. doi:10.1016/j.eswa.2014.08.031

[42] MANN, W. C., & THOMPSON, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. Text - Interdisciplinary Journal for the Study of Discourse, 8(3). doi:10.1515/text.1.1988.8.3.243

[43] Lapadat, J. (2010). Encyclopedia of Case Study Research. In Encyclopedia of Case Study Research. SAGE Publications

[44] Porter, C., Sasse, M., & Letier, E. (2013). Giving a voice to personas in the design of egovernment identity processes. Research to Design: Challenges of Qualitative Data Representation and Interpretation in HCI-in BCS HCI.

[45] Adams, A., Lunt, P., & Cairns, P. (2008.). A qualitative approach to HCI research. Research Methods for Human–Computer Interaction, 138-157. doi:10.1017/cbo9780511814570.008

[46] Farbey, B. Land and Targett, D (1993). How to assess your IT investment. A Study of Methods and Practice. Butterworth Heinemann, Oxford

[47] Eslami Andargoli, A., Scheepers, H., Rajendran, D., & Sohal, A. (2017). Health information systems evaluation frameworks: A systematic review. International Journal of Medical Informatics, 97, 195-209. doi:10.1016/j.ijmedinf.2016.10.008.