Verifying Ontology Increments through Domain and Schema Independent Verbalization

Kaneeka Vidanage, Noor Maizura Mohamad Noor, Rosmayati Mohemad and Zuriana Aby Bakar

Faculty of Ocean Engineering Technology and Informatics, University Malaysia Terengganu, Malaysia

Summary
Collaborative ontology construction is the latest trend in developing ontologies. In this technique domain specialists and ontologists need to work together. Because of the complexity associated with ontology construction, it`s done in an iterative and incremental fashion. After each iteration, an ontology increment will be produced. Current ontology increment is always an enhanced version of the previous increment. Each ontology increment has to be verified for its accuracy. Domain specialists' contribution is very significant in accomplishing this necessity. Unfortunately, non-computing domain specialists (i.e. medical doctors, bankers, lawyers) are illiterate on semantic concepts. Therefore, validating the accuracy of the ontology increment is a complex hurdle for them. This research proposes verbalization approach to address this complexity.

Key words:
Ontology-increment, Ontologists, Schema, Verbalization

1. Introduction

An ontology is a domain specific conceptualization, which is machine and human readable. It`s impossible to develop an ontology at once. Therefore, an iterative and incremental fashion is followed. Each iteration produces an improved version of an ontology increment. Before proceeding to the next ontology increment, accuracy of the current increment has to be verified [1-3]. Domain specialists` input is very significant in accomplishing this requirement. But non-computing domain specialists are not experts in ontology engineering. They don`t have the skillset to comprehend an ontology increment available in form of a (Resource Description Framework) RDF or (Ontology Web Language) OWL file. Further, their understanding is inadequate to comprehend the visualized schemata of an ontology increment also. Accuracy of the embedded knowledge into an ontology cannot be verified, as they have no knowledge in understanding the schematic structure and writing appropriate SPARQL or SQWRL queries. These challenges make the domain specialists' knowledge verification role impossible [4-5].

Verbalization is the technique of converting complex technical semantics into understandable natural language. Therefore, verbalization is proposed as an ideal strategy to comprehend the knowledge embedded in RDF / OWL files, without the need for writing SPARQL queries and resolving the technological bottlenecks [10-11]. Even though, existing verbalization techniques have multiple limitations

2. Related Works

Majority of the existing verbalizers can verbalize up to the level of control natural language (CNL). This is not colloquial general English, where a non-technical layman can understand. Attempt to control English (ACE) is a popular form of CNL, which is similar to ASSEMBLY. But this form of verbalization does not resolve the technical barrier faced by the non-computing domain specialists. Therefore, the verbalizers` produce output in CNL form is not a solution for the problem investigated [9-14].

Next is the issue of extensive configuration requirements. If a verbalizer to be used in verbalizing a knowledge model associated for a specific domain, first it has to be configured. Because the verbalizer has to understand the schemata and the domain to assure a productive verbalization output. Therefore, the same verbalizer cannot be used for multiple domains without going through the extensive configuration phase. Hence, it can be presented as majority of the existing verbalizers are domain and schema dependent. For an example MIKAT (Medical Imaging and Advanced Knowledge Technologies) verbalizer is limited only to the breast cancer domain. It cannot be used with any other domains [18-20].

Majority of the current verbalizers are depending on the Natural Language Generation (NLG) pipeline. Principals of NLG is relying on the Rhetoric Structure Theory (RST). According to RST, two main concepts remain as nucleus and satellite. Nucleus represent the main concept and satellite represents the related concepts linked with the main concept. Nucleus and satellite will be changing from domain to domain. If these two concepts are not properly recognized in relation to the domain, accurate verbalizations will not happen. Hence, a technically intensive manual phase has to be applied to perform the lexical mappings via assigning proper weightages. This is another huge limitations associated with existing verbalizers [15-17].

NaturalOWL [15] is a popular verbalizer with the deficiency of extensive manual mapping phase. This phase

Manuscript received January 5, 2021
Manuscript revised January 20, 2021
https://doi.org/10.22937/IJCSNS.2021.21.1.6

is technically referred as portal configuration phase. Live OWL Documentation Environment (LODE) is another very popular verbalizer [21]. This is a web service which accepts the ontology increment file to be verbalized. The biggest restriction of LODE is, it`s designed to function with de-facto standard meta models like Dublin-Core and FOAF. Because LODE uses XSLT technology to interact with those de-facto standard meta models end-points. Therefore, it is a must to have semantic annotations related with the de-facto standard meta models, in the ontology increment to be verbalized. The next limitation is, LODE accepts ontology increments with a Cool-URI. The pseudo-URL structure expected by the LODE web service is https://w3id.org/lode/optionalparameters/published_ontology_url. Therefore, the respective ontology increment to be verbalized has to be published in the web, which is another restriction of LODE. Verbalizers as OnToology [22] and Widoco [23] are also having similar restrictions as those also have been developed on top of LODE.

3. Results and Discussion

As there are several limitations associated with existing verbalization approaches, new domain and schema independent verbalization technique, free from manual configurations is proposed. The proposed technique comprises of three main phases as depicted in figure 1.

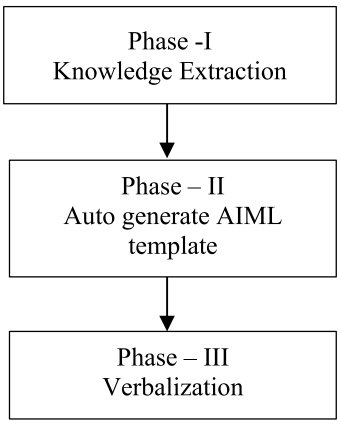

Fig.1 Execution phases of the proposed verbalization approach

3.1 Knowledge Extraction Phase

Workflow associated with the knowledge extraction phase is depicted in figure 2.

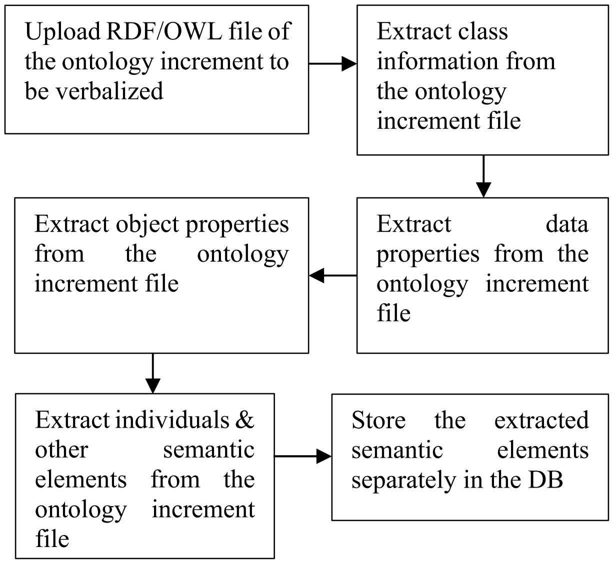

Fig. 2 Workflow of the knowledge extraction phase

Semantic elements have to be extracted one by one as depicted in figure 2 above and store them in the database structure defined in figure 3. Conditional verification has to be enforced, as all semantic elements could not be receding in an ontology increment.

In phase –II as depicted in figure 4, ontology increment specific AIML file will be auto generated. This is a very important step, in accomplishing domain and schema independent verbalization. The main reason for selecting the AIML technology is, unlike technologies like DialogFlow [25], RAZA [24], IBM Watson [26], AIML [27-29] does not require a domain specific training set to train the verbalizing model. Further it`s free and lot of external integration support is available.

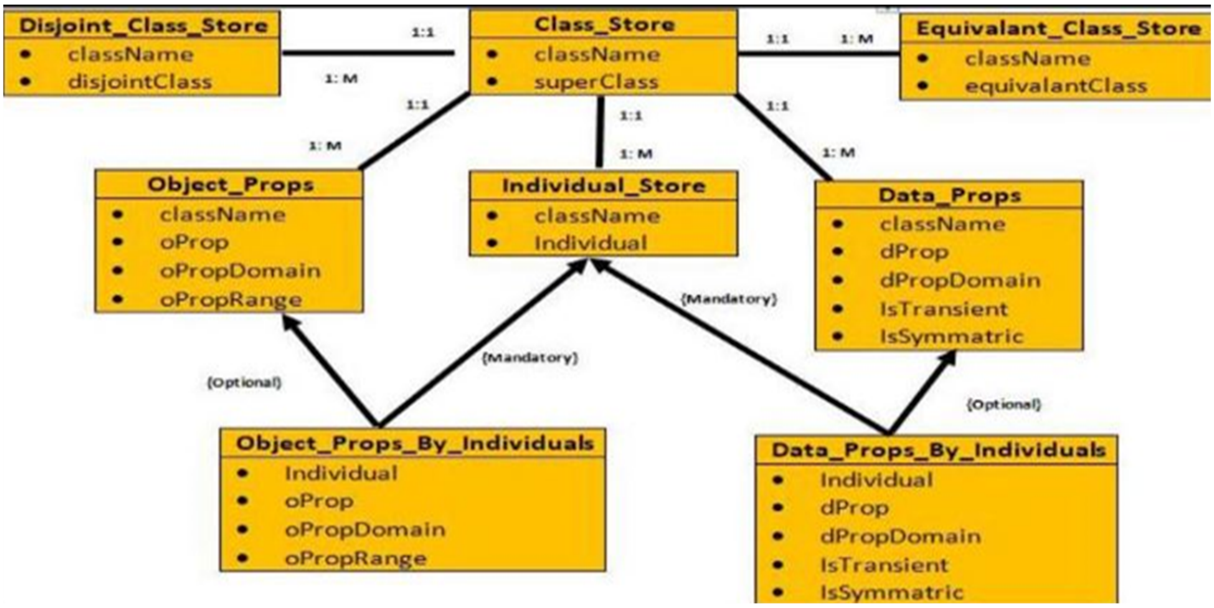

Fig.3 Database Schema

3.2 Auto Generate AIML Template

The workflow associated with phase-II is as depicted in figure 4 below.

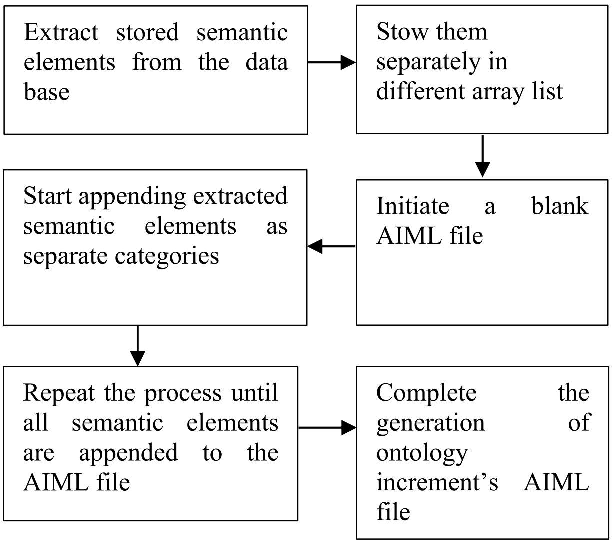

Fig. 4 Workflow for AIML auto generation.

AliceBot [28] is a chatbot engine which derives knowledge from AIML files. Here the as depicted in figure 5, AliceBot accepts sementic element inputs from the database and it will query and locate the respective categories from the auto generated AIML file. Henceforth, ontology specific values extracted from the database will be injected to the placeholders avilable inside the specific category selected in the AIML file.

Eventually, all the processed categories will be fed into a seperately defined arraylist. Once the verbalize request handling is over, the array list will be filled with the contents to be verbalized. By initiating an iterater to traverse through the verbalized contents in the arraylist, item by item can be fed to the iText PDF plugging to accomplish the verbalized report generation. A verbalized report generated using this approach is depicted in figure 6

3.3 Verbalization Phase

The workflow associated with phase-III is as depicted in figure 5 below.

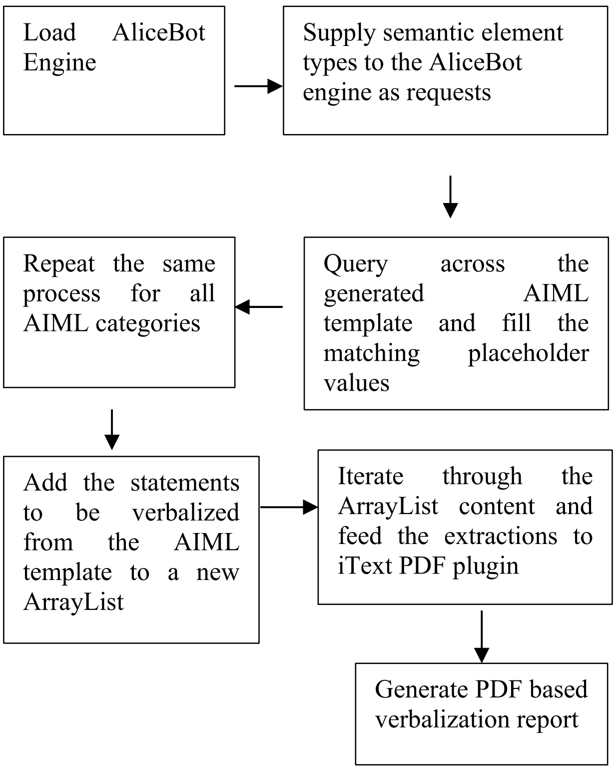

Fig.5 Workflow of the verbalization phase

The importance of this algorithm is, it will ensure domain and schema independent verbalization, free of any configuration requirements. Proposed algorithm is capable enough to generate the ontology increment specific AIML file with the required knowledge embeddings. Therefore, with no manual configurations this algoritm can verbalize any ontology increment despite the domain and schema restrictions. Another notable contribution is, this algorithm has alterted the fucntionality of the AliceBot engine, from a chatbot to a verbalizer.

The proposed technique is utilized in 03 different ontology construction projects. Generated verbalized reports were collaboratively assesed by the domain specialists involved for those projects for three separate iterations. Henceforth, depending on the verification results, confusion matirces were derived as specified in table 1 in next page.

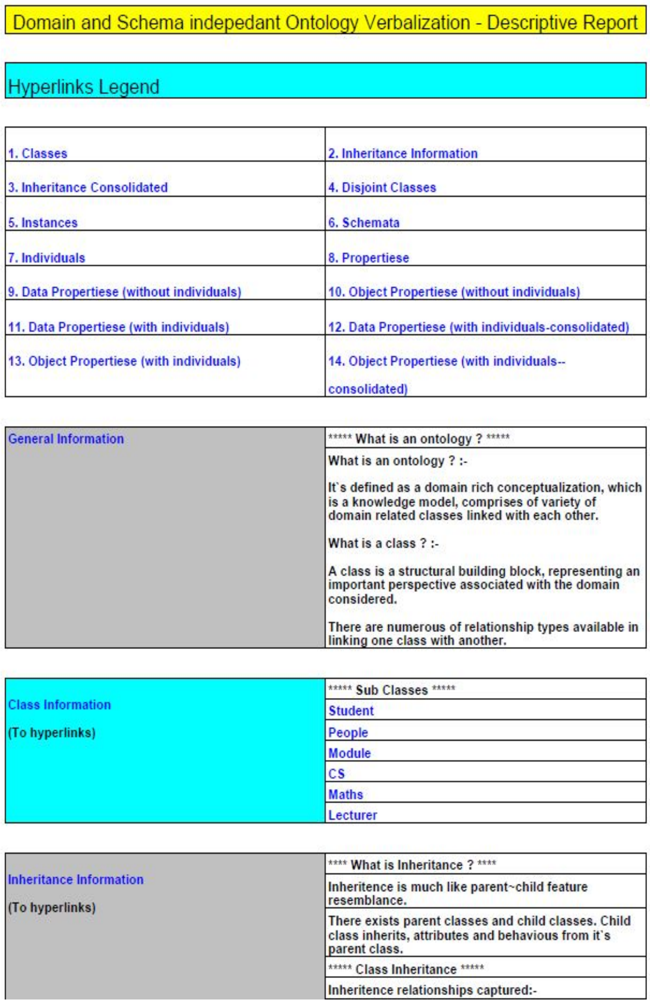

Fig 6 Fraction of a Verbalized Report

Domain	Iteration – 01		Iteration – 02		Iteration – 03		Averaged	
Psychotherapy	Precision	0.84	Precision	0.88	Precision	0.93	Precision	0.88
	Recall	0.78	Recall	0.80	Recall	0.85	Recall	0.81
	Accuracy	0.85	Accuracy	0.88	Accuracy	0.91	Accuracy	0.88
	F-Measure	0.81	F-Measure	0.83	F-Measure	0.88	F-Measure	0.84
Law	Precision	0.77	Precision	0.81	Precision	0.85	Precision	0.81
	Recall	0.71	Recall	0.78	Recall	0.83	Recall	0.77
	Accuracy	0.78	Accuracy	0.80	Accuracy	0.82	Accuracy	0.80
	F-Measure	0.73	F-Measure	0.79	F-Measure	0.84	F-Measure	0.79
Marine Biology	Precision	0.80	Precision	0.82	Precision	0.85	Precision	0.82
	Recall	0.79	Recall	0.83	Recall	0.89	Recall	0.84
	Accuracy	0.83	Accuracy	0.85	Accuracy	0.88	Accuracy	0.85
	F-Measure	0.79	F-Measure	0.82	F-Measure	0.87	F-Measure	0.83

Table 1 Evaluation Results

4. Conclusion

Collaborative ontology engineering is the current trend of developing ontologies. In this mechanism, ontologists and domain specialists need to play a collaborative role with mutual understanding. Domain specialists' contribution is very significant throughout the ontology development cycle for the verification of the accuracy of the conceptualized knowledge embeddings. But domain specialists' like bankers, medical doctors, lawyers being non-technical specialists face lots of hurdles in effectively contributing as domain specialists, due to the technical complexities.

The proposed verbalization approach has yielded successful results, hence it can verbalize beyond the level of CNL and its configuration free along with domain and schema independence. The effectiveness of this approach is reveled thorough the experimental analysis conducted for three different domains iteratively. Overall results of the accuracy of this approach is close to 82%.

In future, it`s intended to improve the human computer interaction of this approach even further with proper interface and tool tip support.

References

[1] 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

[2] 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

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

[4] 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\

[5] Strohmaier, M., Walk, S., PPschko, 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.319903

[6] Bouayad-Agha, N., Casamayor, G., &Wanner, L. (2014). Natural Language Generation in the context of the Semantic Web. Semantic Web, 5, 493-513.

[7] Dai, Y., Zhang, S., Chen, J., Chen, T., and Zhang, W. Semantic Network Language Generation Based on a Semantic Networks Serialization Grammar. World Wide Web 13, 3(2010), 307–341.

[8] Demir, S., Carberry, S., and Mccoy, K. A Discourseaware Graph-based Content Selection Framework. In Proceedings of the 6th International Natural Language Generation Conference (INLG) (2010), pp. 17–27. \

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

[10] Keet, C. M., &Khumalo, L. (2016). On the verbalization patterns of part-whole relations in isiZulu. Proceedings of the 9th International Natural Language Generation conference. doi:10.18653/v1/w16-6629

[11] KaarelKaljurand 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
[12] Guimarães, M. A., Zisman, R. P., & Renzi, A. B. (2019). Pharmaceutical Online Store Project: Usability, Affordances and Expectations. Advances in Intelligent Systems and Computing, 523-534. doi:10.1007/978-3-030-20040-4_47
[13] Sun, Mi, Olsson, Paulsson, &Harrie. (2019). Utilizing BIM and GIS for Representation and Visualization of 3D Cadastre. ISPRS International Journal of Geo-Information, 8(11), 503. doi:10.3390/ijgi8110503
[14] Dorothy, B., & S. Ramesh Kumar, B. (2018). DORBRI: An Architecture for the DoD Security Breaches Through Quantum IoT. International Conference on Computer Networks and Communication Technologies, 491-496. doi:10.1007/978-981-10-8681-6_44
[15] Androutsopoulos, I., Lampouras, G., &Galanis, D. (2013). Generating Natural Language Descriptions from OWL Ontologies: the NaturalOWL System. Journal of Artificial Intelligence Research, 48, 671-715. doi:10.1613/jair.4017
[16] Mann, W.C., & Thompson, S.A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. Text & Talk, 8, 243 - 281.
[17] Sportelli, F., &Franconi, E. (2016). Formalisation of ORM Derivation Rules and Their Mapping into OWL. On the Move to Meaningful Internet Systems: OTM 2016 Conferences, 827-843. doi:10.1007/978-3-319-48472-3_52
[18] 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
[19] Bao, J., Cao, Y., Tavanapong, W., & Honavar, V. (2004). Integration of DomainSpecific and DomainIndependent Ontologies for Colonoscopy Video Database Annotation. Artificial Intelligence Research LaboratoryIowa State University
[20] Poulovassilis, A., Selmer, P., & Wood, P. T. (2016). Approximation and Relaxation of Semantic Web Path Queries. SSRN ElectronicJournal. doi:10.2139/ssrn.3199265
[21] Motta, E., Mulholland, P., Peroni, S., d'Aquin, M., Gomez-Perez, J.M., Mendez, V.,
[22] Zablith, F.: A novel approach to visualizing and navigating ontologies. In: Aroyo, L., Welty, C., Alani, H., Taylor, J., Bernstein, A., Kagal, L., Noy, N., Blomqvist, E. (eds.) The Semantic Web { ISWC 2011: 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011
[23] 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
[24] Garijo, D. (2017). WIDOCO: A Wizard for Documenting Ontologies. Lecture Notes in Computer Science, 94-102. doi:10.1007/978-3-319-68204-4_9
[25] Singh, A., Ramasubramanian, K., & Shivam, S. (2019). Introduction to Microsoft Bot, RASA, and Google Dialogflow.
[26] Janarthanam, S. (2017). Hands-On Chatbots and Conversational UI Development: Build chatbots and voice user interfaces with Chatfuel, Dialogflow, Microsoft Bot Framework, Twilio, and Alexa Skills.
[27] Ralston, K. (2019). Personalizing Chatbot Conversations with IBM Watson.
[28] Wallace, The Elements of AIML Style, ALICE A. I. Foundation, Inc. March 28, 2003
[29] Wallace, R. S. (2014). AIML 2.0 Working Draft, https://docs.google.com/document/d/1wNT25hJRyup cG5 1aO89UcQEiG - HkXRXusukADpFnDs4/pub