Chapter 30
Semantic Web-Based Knowledge Extraction: Upper Ontology Guided Crime Knowledge Discovery

Kaneeka Vidanage, Noor Maizura Mohamad Noor, Rosmayathi Mohemad and Zuriana Abu Bakar

1 Introduction

Crime is a controversial social hazard, spreading across the globe. As pointed by Ghani in [1], the root cause of crime rate escalation across the world is identified as urbanization. The author further elaborates in the paper, how this actually happens. As claimed by Ghani [1], many rural residents progressively approach to suburban and city areas assuming they would get more business, occupational opportunities and a comfortable life zone. Even though, the majority will not succeed to attain these goals. In fact, they will start to experience, higher cost of living in suburban areas compared to the rural life habitats they are familiar with. Therefore, soon, this becomes unbearable for many, and it will act as the catalysts to incept criminal activities [1, 2]. As elaborated by Badiora and Afon in [3], these crime rates are escalating at an alarming rate, adversely affecting the socio-economic and quality of life. Both Ajaegbu [4] and Katsina [5] argue unemployment and the economic hardships mixed with urbanization are the main triggering points for deadly crimes.

As there is emerging enthusiasm growing in semantic Web-based technologies, it is decided to seek the potentials of using semantic technologies to resolve the problem of crime escalation.

With the introduction of the semantic Web, the whole Internet has become a Web of knowledge. This has become even more fascinating as the knowledge represented in the semantic Web is machine-readable [6]. Even though construction of a domain constrained semantic Web or in other words, an ontology from the scratch is not an easy task [7], as there are no fully automated methods identified up to date. Human intervention is essential for knowledge verification [8].

K. Vidanage (✉) · N. M. M. Noor · R. Mohemad · Z. A. Bakar
School of Informatics and Applied Mathematics, University Malaysia Terengganu (UMT), Kuala Nerus, Malaysia
e-mail: Kaneeka.online@gmail.com

Therefore, as mentioned above, a few facts are apparent. Those would be ontology construction is not an easy task [7, 8]. But if created, the advantages of it would be ample, as of their machine and human readability [6]. Therefore, as much as possible maximum benefits should be derived from created ontologies. Even though, obtaining maximum outcomes from created ontologies have become impossible due to two main causes.

One is the difficulty of understanding the schematic structure of a defined ontology or the knowledge model [9, 10, 11]. Because without knowing the schematic structure, knowledge retrieval from an existing ontology would not be feasible [12, 11]. The second aspect is the requirement of high technical knowledge to understand a schema and necessity of writing queries for knowledge retrieval [11, 12, 13]. Both these facts imposing a great barrier on ontology reusability as well as knowledge comprehension presented in ontology formats [7, 14]. Therefore, the implications of this will affect both technical and non-technical audiences in multiple ways [15]. Additionally, most of the hardly created ontologies will become stagnated soon on the Internet, wasting intellectual and cognitive efforts of the creators [7].

Crime domain as a whole is a very vast discipline with lots of sub-regions as weapons, standard operational procedures (SOPs), evidence gathering, etc. [16]. Ontologies being domain rich conceptualizations [17], what if, a specific sub-region related crime domain knowledge is presented in the form of an ontology? Then, with the introduction of the proposed knowledge discovery framework, despite all the technical barriers conversed above, non-technical criminologists, criminology students also can infer the knowledge represented via the ontology. Because suggested knowledge extraction framework is capable of encapsulating complex schematics and querying barriers related to the ontology and presenting the stored knowledge in natural English. Non-tech crime specialists and students are also language literate, and hence, suggested framework will efficiently disseminate relevant information amidst specialists and students, facilitating investigations and learning requirements. This is one potential use case of the suggested framework and crime domain. But, this framework is not only limited to the crime domain. It is planned to design as a domain and schemata independent framework, to support the rapid growth of use in semantic technologies, both inside and outside the computing domain, facilitating non-technical consultants' usage requirements as well [15, 18].

Therefore, this paper is focusing on proposing an architectural design, leading towards a domain and schema independent, semantic knowledge-based, knowledge extraction framework, to maximize the use of existing knowledge models and at the same time facilitates natural language-based knowledge dissemination among both newly created or existing knowledge models, making them to be the main research contributions.

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

2   Related Work

Ontologies are domain rich conceptualizations. This is how Spasic et al. have defined ontologies in [17]. Ontologies can be defined to present knowledge associated with any domain. Currently, also there are thousands of already defined ontologies available on the Internet. Few of the locations where you could access these predefined ontologies are on Vocab.org [19], Swoogle [20], LOV [21] and Protégé Wiki [22]. AberOWL [23] and BioPortal [24] are two other important ontology repositories, which contains thousands of biomedical ontologies. Additionally, Covert To RDF [25] provides numerous converters capable of translating comma-separated values (CSV) and spreadsheet-based data into Resource Description Framework (RDF) format. (RDF) or OWL (Ontology Web Language) formats, are the most popular formats for semantic Web knowledge representations [26]. Protégé IDE for the ontology development also promotes a variety of plugin series [27] which are capable of converting a variety of data formats into above-listed RDF or OWL formats. Therefore, as conversed above, it is very apparent that there are plenty of existing ontologies and ontology browsers as well as alternative mechanisms of converting information dispersed across the Web into semantic Web friendlier formats. This is good evidence to show the current enthusiasm towards semantic Web-based technologies. Consequently, it can be determined as there is enormous potential for ontology reusability and ontology creations.

Even though, as already conversed above, complexity in understanding the schemas, ontology querying barriers, inability associated with comprehension of OWL or RDF-based technical representations act as main bottlenecks for both tech and non-tech audiences in hindering the ontology reusability and creation [9, 10, 11, 12, 13] despite the above-discussed potentials.

As we are trying to align the application outcomes of this research towards the criminal domain due to the timely relevance of that region, next it will be investigated on what are the already existing intelligent computational prototypes linked with the criminal domain analysis. As the first result, it is located Masitha et al. had attempted in [28] to come up with a crime ontology with the intention of facilitating relevant officials to react quickly on crime matters. In this research, they have constrained their crime domain to motorcycle thefts only because crime, in general, is a very vast subject discipline [16]. Researchers have initially investigated multiples of criminal case reports and determined a few important elements as vital in extracting crime information.

After this, they have come up with a taxonomized structure representing the interrelations among identified components to assure methodical storing of crimerelated information collected. Eventually, the designed ontology is implemented via the Top-Braid Standard composer tool. Additionally, researchers have developed a case base repository to store information associated with similar types of cases occurred in past. Once the user enters the new information about the reported case, initially it will be reasoned by the implemented crime ontology, and afterwards, crime ontology will talk with the relevant case base and will extract out the most related set of past

crimes occurred, evidence collected, decisions arrived and variety of other necessary information.

As the next finding, another ontology and decision support system [29] working in combination to reason about evidence collected from a crime scene. Investigation analysis of complex and dynamic crimes is not an easy task.

There need to be established and carefully governed procedures [29]. In fulfilling this requirement, having an ontology in place to guide the respective decision support systems as needed for the crime investigation mining will enhance the overall throughput of the entire process.

Likewise, this discussion can be continued as there are a sufficient amount of researches where intelligent systems have been developed for the criminal domain.

The accuracy, reliability and practicality of these knowledge models would be high in value as those are developed and tested by teams of professors, researchers and experts in the respective fields [30]. Even though, the plight is, criminologists and students following criminology pathways are not computer specialists. Hence, they cannot comprehend the valuable pieces of knowledge expressed, in semantic Web forms though they are available free of charge on the Internet [26]. Authors of this paper consider this as an utter cognitive and intellectual waste of the experts and researchers who are involved in developing those knowledge structures as they become soon stagnant on the Web, after serving one specific purpose. From the other hand, computational reusability of those will also become very low due to the difficulties of comprehending complex schematics, as being ignorant about the schema, querying for knowledge retrieval will be infeasible [9, 10, 11, 12, 13]. Therefore, what if there is a framework introduced, which can overcome those technical barriers and hurdles associated with effective knowledge reusability and dissemination?

For instance, as one practical use case, assume the process of training criminology officers, detectives and investigators. They need to gain a comprehensive knowledge of crime types, evidence gathering procedures, crime analysis, etc. [31]. In fulfilling these purposes, there are ample of carefully designed knowledge models available in popular semantic Web formats on the Internet, mostly free of charge [16]. If required formats of knowledge models are not available, these could be created through collaborative efforts of computer scientists and criminology domain specialist. In fact, this could be a one-time effort. Because then the created knowledge models can be used again and again over batches of criminology students and specialists for effective knowledge dissemination assuring knowledge reusability aspects as well.

Next, what about the existing frameworks based knowledge extraction from OWL or RDF knowledge models? Subsequently, it is decided to widen up the literature investigation on the assessment of similar frameworks.

Ghorbel et al. proposed a tool by the name "Memo Graph" in [32]. This is an ontology taxonomy visualization tool. It will depict the taxonomic structure of the ontology in visual forms. But, we cannot expect non-technical audiences to look at taxonomy and get an idea to query the ontology. Because still, SPARQL-based querying will be a barrier for them in knowledge retrieval. Another similar type of system located is Semantic Web Portal developed by Ding et al. [33]. Going beyond the graphical visualization of the taxonomical structure of the ontology, this system

is capable of visualizing the triple structure as well. Furthermore, as claimed by Ding et al. in [33], this system can work with any domain. Even though, there is a major bottleneck. That is before applying Semantic Web Portal to a selected domain, portal ontology needs to be created. Portal ontology provides domain associated axioms, facilitating the operation of Semantic Web. Therefore, again, it is not acceptable from a non-technical user to create a portal ontology, to feed the domain knowledge to the Semantic Web Portal. Further, this tool does not have the feature of knowledge extraction and presentation in natural language.

Other than those visualization tools, in [34], team of researchers have implemented a system which is capable of natural language querying of an ontology. However, it is a static, domain and schema-dependent software. In their research, they had created an accommodation ontology and English to SPARQL conversion, and SPARQL queries are statically mapped to the accommodation ontology. Hence, the structure proposed in [34] is not a framework which is capable of the domain and schema-independence analysis as it is statically bound to one specific ontology only.

Therefore, as reviewed so far, the following reflections can be derived. There is great enthusiasm over the use of semantic technologies to overcome, recurring social issues. Even for the considered criminal domain also, it is possible to locate multiples of intelligent systems already deployed. Even though, as previously pointed out in the literature review above, semantic technologies are not doing good in the dimensions of knowledge reusability and dissemination. Further, there are no proper existing frameworks located, which are capable of natural language-based knowledge extraction from popular semantic Web formats with domain-independent knowledge models. Therefore, as the gap of this research, it can be concluded the issues associated with poor performance in reusability and dissemination aspects of the semantic technologies and none existence of domain and schema-independent, natural language-based knowledge extraction framework. Consequently, the remainder of this paper is emphasizing on a proposed architectural design leading towards the framework to address the above-listed gaps, which could be interpreted as the contribution of this research.

3   Methodology

In the process of finalizing an appropriate architecture for this framework, it is identified in [35], a group of researchers have suggested the concept of divide and conquer. As they have pointed out, attempting to fulfil all the tasks from one module will increase the complexity and coupling associated with the module. In [36], they further suggest the definition of a complex problem space via multiple resolution layers will provide the opportunity of attention to detail in knowledge modelling and analysis aspects, and it will further prevent the single module being flooded with lots of information and resulting with complex schematic structures or conditions. Therefore, considering those suggestions, as the first decision point it is concluded that the proposed framework should also comprise with multiple resolution layers,

and hence, the process of knowledge extraction in natural language from RDF and OWL knowledge models is not an easy, straightforward task.

As the next step, arguments presented in ontology designing aspects are considered. Multiple of researchers have justified ontologies would be the strongest conceptual representations associated with elaborating a complex domain, and they would be ideal in process enforcement as well [6, 8]. Here, in this research also, knowledge extraction from RDF/OWL knowledge models is also a complex process, which needs to be enforced carefully. Therefore, recommendations provided in numerous research papers evaluated confirmed the intention of using ontologies for this research as well. Next, question is how to determine the architecture or the structure of the ontology? For this concern, Sowa [37] has provided a good explanation in his research paper. As a claim by Sower [37], there are multiple types of ontologies. Top-level ontologies describe more generalistic concepts associated with a "Thing". When gradually reaching from top to bottom, knowledge represented in ontologies will also vary from meta-concepts to application and from application to the domain and at the bottommost layer, having task ontologies focusing on specified aspects associated with the considered domain. In this research, the process of knowledge extraction from a given RDF/OWL model is domain-independent. Therefore, it confirms that an upper-level ontology needs to be designed as the considered procedure for knowledge extraction is not domain-specific, and it is generalistic.

Smith proposes [38] mechanics theory, which is enforcing on procedures than declarative aspects associated with the domain. Researchers have recognized mechanics theory to be much appropriate in constructing task ontologies. Helix spindle theory [39] comprises three main stages which are continuing in an incremental manner until a satisfied criterion is met. The first stage is the conceptualizing stage. In this stage, in-depth brainstorming and conceptualizing will take place, pertaining to the considered use case. The output of this phase would be a natural language-based reflective description of the finalized mental image. Then, the second phase would be the elaboration phase, where this natural language description derived from phase one will be graphically represented via a taxonomical structure. Eventually, the final phase would be the definition phase, where the ontology construction will be completed with the required knowledge injections. All these phases are iterative and incremental and should be logically interconnected with each phase as necessary.

Therefore, it is decided to use helix spindle method [39] for the taxonomy derivation of the ontology and mechanics theory [38] for the knowledge injection to the created taxonomic structure. The step-wise progression of the methodology followed is graphically presented in Fig. 1, depicted below.

4   Results and Discussions

Figure 2 illustrates the communicational architecture in between the instructional upper ontology and the relevant decision support systems integrated to the respective endpoints of the upper ontology.
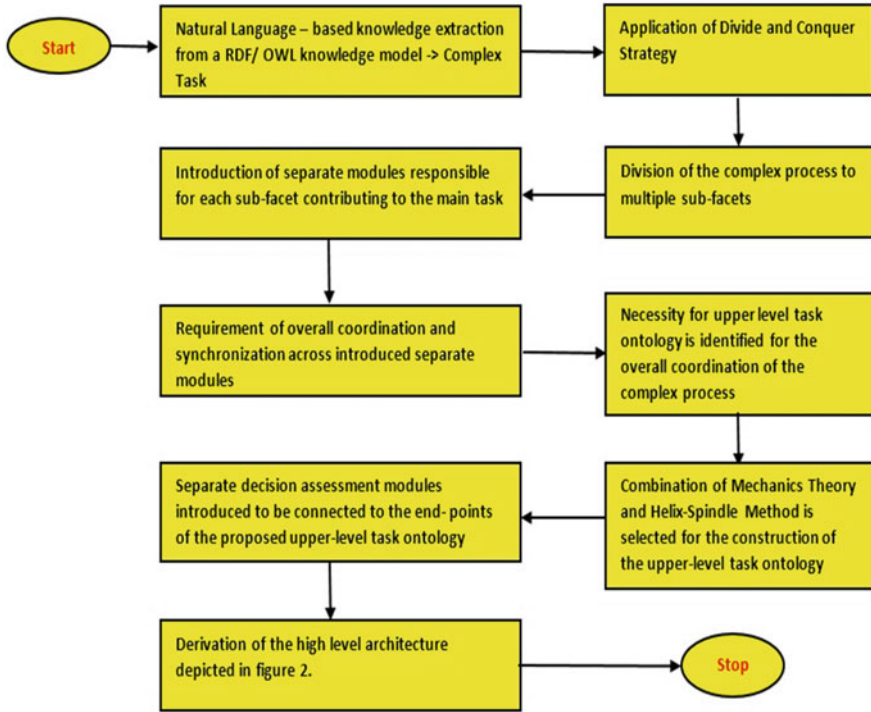
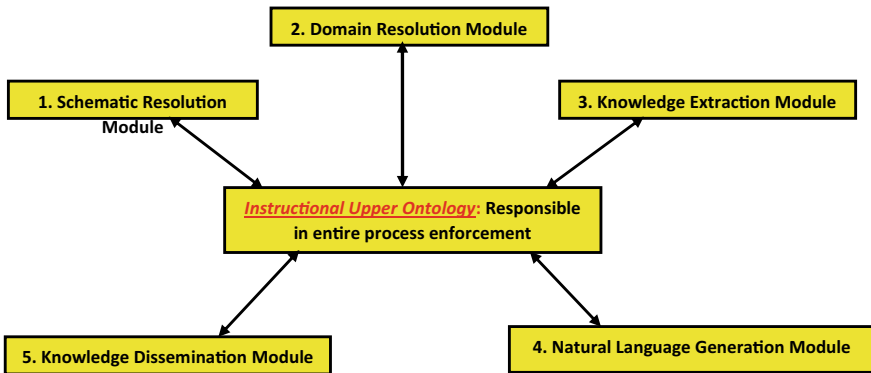**Fig. 1** Overall flow of the methodology

Fig. 2 Communicational architecture in between instructional upper ontology and decision support systems the relevant

Natural language-based knowledge extraction from RDF or OWL semantic formats is not an easy task. Hence, to methodically perform the task according to the enforced process by the instructional upper ontology, multiples of DSS systems are introduced. Knowledge extraction from RDF or OWL-based semantic formats is executed through multiple resolution layers, synchronized under the control instructional upper ontology. Figure 3 mentioned will elaborate on the step-wise execution of the knowledge extraction process associated with the uploaded RDF/OWL-based knowledge model file. Each cell in Fig. 3 clearly elaborates the steps to be executed in one after another sequentially, from start to end. This entire process can be governed by the communicational architectural structure proposed in Fig. 2.
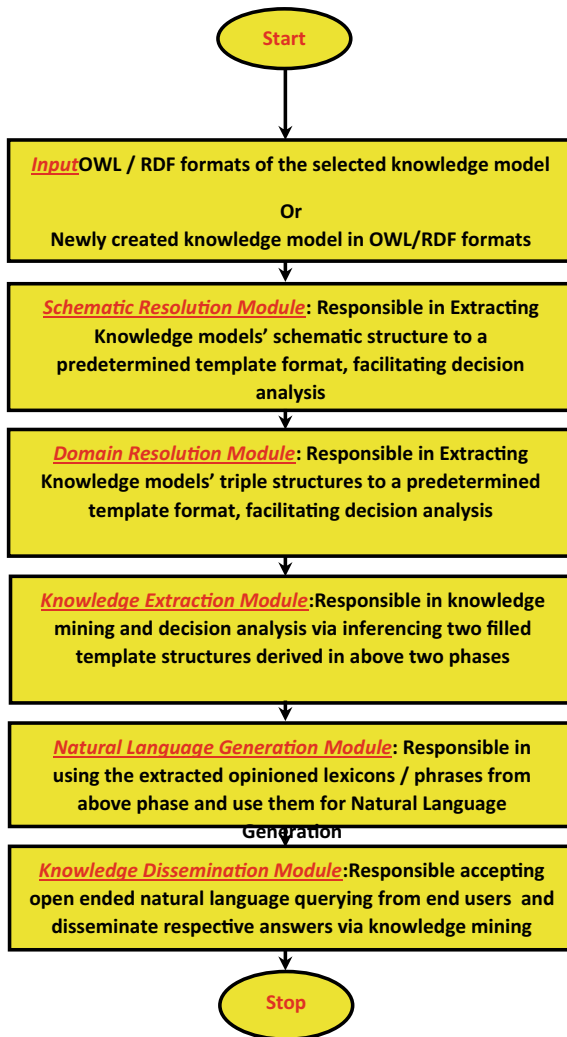

Fig. 3 Overall execution flow of the proposed framework

5   Evaluation

An evaluation strategy can be planned as mentioned in below Fig. 4, to assess the efficacy of the proposed framework for knowledge extraction from semantic Web-based knowledge models.

Figure 4 suggests creating a suitable ontology, covering a sub-discipline related to the crime domain, unless if no appropriate existing ontology is located. The ideal mechanism will be to create a suitable ontology covering a specialized sub-discipline associated with the crime domain. This can be done via brainstorming with a few crime specialists on a selected crime domain.
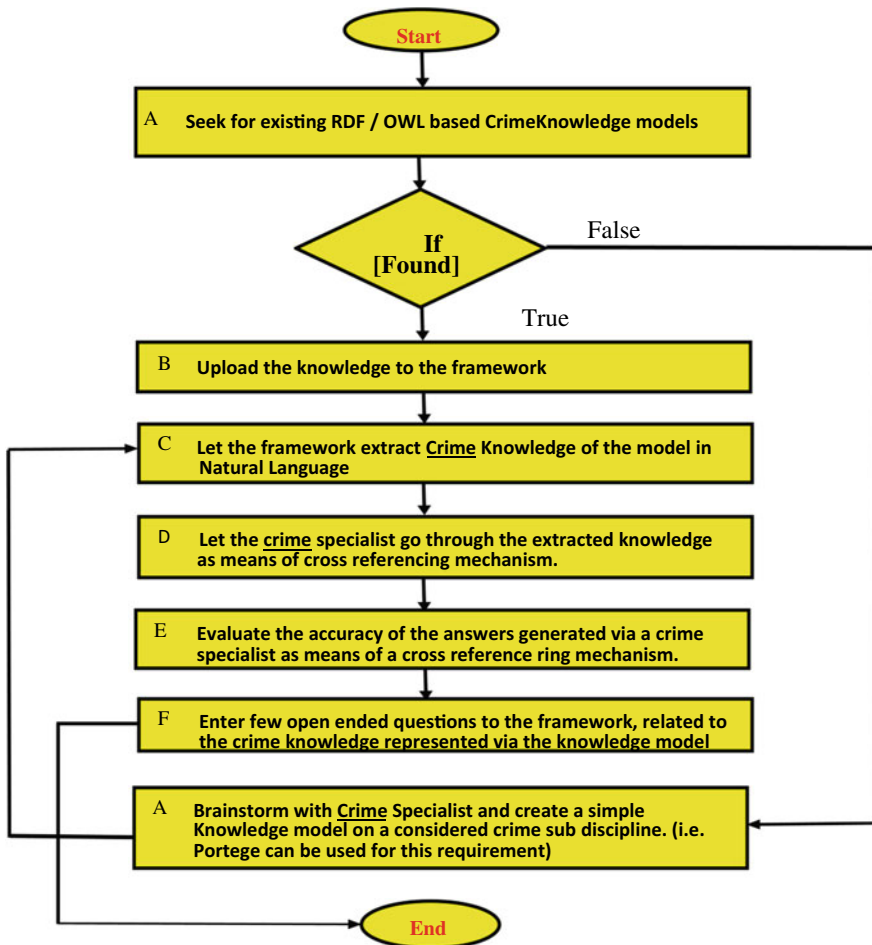

Fig. 4   Suggested evaluation strategy

Table 1 Test statistics of the proposed architecture

Measurement	Accomplishment
Sensitivity	0.78
Precision	0.90
Accuracy	0.86
F-measure	0.80

Crafted ontology can be presented to the suggested framework either as an OWL document or an RDF document. Then, let the framework extract the axioms stored in the ontology and verbalize it in natural English. Henceforth, the verbalized contents can be cross-referenced with the crime specialists who have involved in the ontology creation phase. Then, through their instincts on verbalized output, it can be verified is there any information or important aspects have been missed/lost. Using it as the main evaluation platform, a confusion matrix can be derived to determine the true positives, false positives, false natives and true negatives associated with the verbalized output and ontology contents.

Via the usage of those parameters, evaluation matrices such as recall, precision and *F*-measure can be derived, to numerically asses the throughput of the verbalization process.

Table 1, depicts the test statistics derived, after exercising the above framework on a crafted crime ontology. Verbalized results are cross-referenced with crime specialists, and Fig. 5 demonstrates the mechanisms associated with deriving of true positives, true negatives, false positives and false negatives leading towards the test statistics calculations presented in Table 1.
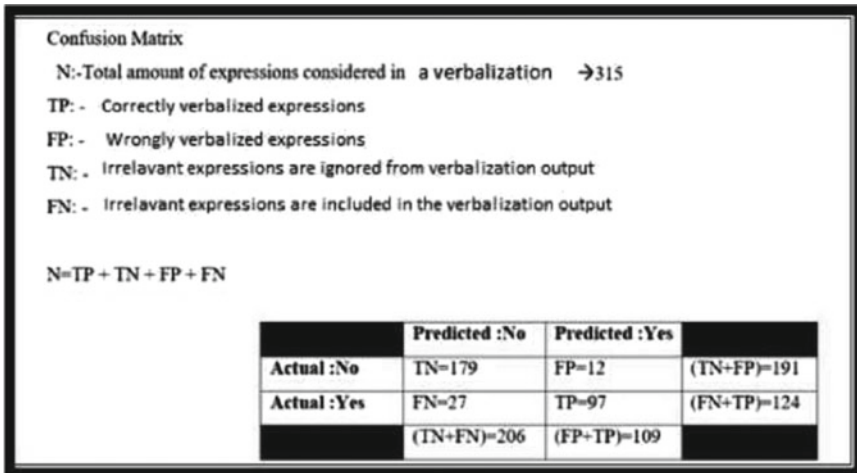

Confusion Matrix

N:-Total amount of expressions considered in a verbalization →315

TP: - Correctly verbalized expressions

FP: - Wrongly verbalized expressions

TN: - Irrelavant expressions are ignored from verbalization output

FN: - Irrelavant expressions are included in the verbalization output

N=TP + TN + FP + FN

	Predicted :No	Predicted :Yes	
Actual :No	TN=179	FP=12	(TN+FP)=191
Actual :Yes	FN=27	TP=97	(FN+TP)=124
	(TN+FN)=206	(FP+TP)=109	

Fig. 5 Verbalized results assessment

For the experimented scenario of crime knowledge ontology, proposed architecture-based verbalizer has presented a reasonable performance with an approximate overall accuracy above 80%. Even though, it is suggested to verify the performance of this architecture, via exercising it on multiple more knowledge models obtained from a variety of domains.

6   Conclusion

As a futuristic resolution, authors of this paper have proposed an architectural structure for a potential framework which could resolve the technical barriers associated with the effective use of new and existing semantic Web-based knowledge models. Further, as a functional outcome of the proposed framework, it will facilitate natural language-based information dissemination, allowing criminologists, detectives, police officers and students to experience the benefits of the semantic Web, though they are not ontologists or computer specialists. Finally, as the main contribution of this research, the practical applications of this suggested framework design will widen the horizons of semantic Web-based knowledge comprehension and dissemination as its applications are not limited to the crime domain only, making a progressive step for the betterment of mankind.

References

1. Ghani ZA (2017) A comparative study of urban crime between Malaysia and Nigeria. J Urban Manage 6(1):19–29. https://doi.org/10.1016/j.jum.2017.03.001
2. Soh MB (2012) Crime and urbanization: revisited malaysian case. Procedia Soc Behav Sci 42:291–299. https://doi.org/10.1016/j.sbspro.2012.04.193
3. Badiora AI, Afon AO (2013) The spatial pattern of crime in Nigerian traditional city: the Ile-Ife experience. Int J Criminol Sociol Theory 6(3):15–28
4. Ajaegbu OO (2012) Rising youth unemployment and violent crime in Nigeria. Am J Soc Issues Humanit 2(5):315–321
5. Katsina AM (2013) Trend analysis of poverty and urban crime in Nigeria since 1999. Int J Arts Commer 1(2)
6. Kashyap V (2008) Ontologies and schemas. In: The Semantic Web, pp 79–135. https://doi.org/10.1007/978-3-540764526_5
7. Trokanas N, Cecelja F (2016) Ontology evaluation for reuse in the domain of process systems engineering. Comput Chem Eng 85:177–187. https://doi.org/10.1016/j.compchemeng.2015.12.003
8. Noy N, McGuiness D (2001) Ontology development 101: a guide to creating your first ontology. Stanford University, Stanford
9. Chergui W, Zidat S, Marir F (2018) An approach to the acquisition of tacit knowledge based on an ontological model. J King Saud Univ Comput Inf Sci. https://doi.org/10.1016/j.jksuci.2018.09.0129
10. Alavi M, Leidner DE (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 JR (1983) The architecture of cognition. Harvard University Press, Cambridge, MA

11. Kotis K, Lanzenberger M (2008) Ontology matching: current status, dilemmas and future challenges. In: 2008 international conference on complex, intelligent and software intensive systems. https://doi.org/10.1109/cisis.2008.28

12. Gutierrez-Basulto V, Ibanez-Garcia Y, Kontchakov R, Kostylev EV (2015) Queries with negation and inequalities over lightweight ontologies. SSRN Electron J. https://doi.org/10.2139/ssrn.3199213

13. Shvaiko P, Euzenat J (2005) A survey of schema-based matching approaches. J Data Semant IV

14. Tartir S, Arpinar IB, Sheth AP (2010) Ontological evaluation and validation. Theory Appl Ontol Comput Appl 115–130. https://doi.org/10.1007/978-90-481-8847-5_5

15. Berners-Lee T (2001) The semantic web (PDF). Scientific American

16. Usman U, Yakubu M, Bello AZ (2012) An investigation on the rate of crime in Sokoto state using principal component analysis

17. Spasic I, Ananiadou S, McNaught J, Kumar A (2005) Text mining and ontologies in biomedicine: Making sense of raw text. Brief Bioinform 6(3):239–251. https://doi.org/10.1093/bib/6.3.239

18. Guha RV (2013) Light at the end of the tunnel. In: International semantic web conference 2013 keynote

19. Davis I (2014) vocab.org—a URI space for vocabularies. Retrieved 16 Feb 2019 from http://vocab.org/

20. Yu L (2007) Swoogle. In: Introduction to the semantic web and semantic web services, pp 145–157. https://doi.org/10.1201/9781584889342.pt3

21. Linked Open Vocabularies (2013) Linked open vocabularies (LOV). Retrieved 6 Dec 6 2018 from https://lov.linkeddata.es/dataset/lov/

22. Musen MA, The Protégé Team (2013) Protégé ontology editor. Encycl Syst Biol:1763–1765. https://doi.org/10.1007/978-1-4419-9863-7_1104

23. Slater L, Gkoutos GV, Schofield PN, Hoehndorf R (2016) Using AberOWL for fast and scalable reasoning over BioPortal ontologies. J Biomed Semant 7(1). https://doi.org/10.1186/s13326-016-0090-0

24. Faria D, Jiménez-Ruiz E, Pesquita C, Santos E, Couto FM (2014) Towards annotating potential incoherences in bioportal mappings. Semant Web ISWC: 17–32. https://doi.org/10.1007/978-3-31911915-1_2

25. W3C (2018) ConverterToRdf W3C Wiki. Retrieved 6 Dec 2018 from https://www.w3.org/wiki/ConverterToRdf#CSV_.28Comma-Separated_Values.29

26. Zenuni X, Raufi B, Ismaili F, Ajdari J (2015) State of the art of semantic web for healthcare. Procedia Soc Behav Sci 195:1990–1998. https://doi.org/10.1016/j.sbspro.2015.06.213

27. Protege—DataMaster (2014) DataMaster—Protege Wiki. Retrieved 6 Dec 2018 from https://protegewiki.stanford.edu/wiki/DataMaster

28. Abdul Jalil M, Ling CP, Noor NMM, Mohd F (2017) Knowledge representation model for crime analysis. Procedia Comput Sci 116:484–491. https://doi.org/10.1016/j.procs.2017.10.067

29. Dzemydiene D, Kazemikaitiene E (2005) Ontology-based decision support system for crime investigation processes. In: Vasilecas O, Wojtkowski W, Zupancic J, Caplinskas A, Wojtkowski WG, Wrycza S (eds) Springer, United States

30. Caldarola EG, Rinaldi AM (2016) An approach to ontology integration for ontology reuse. In: 2016 IEEE 17th international conference on information reuse and integration (IRI). https://doi.org/10.1109/iri.2016.58

31. Tremblay G (2015) Cultural industries, creative economy and the information society. In: Power, media, culture. https://doi.org/10.1057/9781137540089.0011

32. Ghorbel F, Ellouze N, Métais E, Hamdi F, Gargouri F, Herradi N (2016) MEMO GRAPH: an ontology visualization tool for everyone. Procedia Comput Sci 96:265–274. https://doi.org/10.1016/j.procs.2016.08.139

33. Ding Y, Sun Y, Chen B, Borner K, Ding L, Wild D, Wu M, DiFranzo D, Fuenzalida AG, Li D, Milojevic S Toma I (2010) Semantic web portal: a platform for better browsing and visualizing semantic data. In: Active media technology, pp 448–460. https://doi.org/10.1007/978-3-642154706_46
34. Habernal I, Konopík M (2013) SWSNL: semantic web search using natural language. Expert Syst Appl 40(9):3649–3664. https://doi.org/10.1016/j.eswa.2012.12.070
35. Yang P, Tang K, Yao X (2018) Turning high-dimensional optimization into computationally expensive optimization 22
36. Abeysiriwardana PC, Kodituwakku SR (2012) Ontology-based information extraction for disease intelligence. Int J Res Comput Sci 2(6):7–19. https://doi.org/10.7815/ijorcs.26.2012.051
37. Sowa JF (1999) Knowledge representation: logical, philosophical, and computational foundations. Brooks/Cole, Pacific Grove, CA
38. Smith B (2003) Ontology and information systems. SUNY at Buffalo, Buffalo, NY
39. Kishore R, Zang H, Ramesh R (2004) Computational ontologies and information systems foundations. In: Communications of the association for information systems 14. https://doi.org/10.17705/1cais.01408