Indonesian Journal of Electrical Engineering and Computer Science
Vol. 28, No. 2, November 2022, pp. 878~889
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v28.i2.pp878-889

A systematic literature review of automatic ontology construction

Zayanah Zafirah Zulkipli[1], Ruhaila Maskat[1], Noor Hasimah Ibrahim Teo[2]
[1]Information Systems Department, Faculty of Computer and Mathematical Sciences,
Universiti Teknologi Mara (UiTM) Shah Alam, Selangor, Malaysia
[2]Computer Science Department, Faculty of Computer and Mathematical Sciences,
Universiti Teknologi Mara (UiTM) Jasin, Melaka, Malaysia

Article Info

Article history:

Received Mar 17, 2022
Revised Jul 6, 2022
Accepted Aug 16, 2022

Keywords:

Fully-automatic method
Ontology
Ontology construction
Ontology evaluation
Ontology learning
Semi-automatic method

ABSTRACT

Ontologies have gotten a lot of interest as a knowledge representation approach in recent years. However, constructing an ontology manually can be a difficult task. The alternative way is to automate the ontology construction, either by performing a semi or fully-automatic approach. In this paper, we will conduct a systematic literature review that will focus on a comparative analysis of different techniques relating to both semi and fully-automatic ontology construction using several techniques and an automated approach applied. The goal is to identify the distribution, methodology, automated part, evaluation method, main tools, and technologies used to construct the automatic ontology. This paper will review academic documents published in peer-reviewed venues from 2017 to 2021, based on a four-step selection process of identification, screening, eligibility, and inclusion for the selection process. To examine these documents, a systematic review was conducted and five main research questions were answered. The results indicate that automatic ontology construction could give higher complexity, shorter time, and reduce the role of the expert knowledge to evaluate ontology than manual ontology construction. Finally, we summarize the most commonly used methods in automatic ontology construction, which we believe will serve as a foundation for future multidisciplinary research.

Corresponding Author:

Ruhaila Maskat
Faculty of Computer and Mathematical Sciences (FSKM), Universiti Teknologi Mara (UiTM) Shah Alam
40450 Shah Alam, Selangor, Malaysia
Email: ruhaila@fskm.uitm.edu.my

1. INTRODUCTION

The ontology can be referred to as the elements of domain knowledge that will be constructed in a machine-interpretable language [1]. Besides, it is known as a formal and structural way of representing the concepts and relations of a shared conceptualization, whereby the concepts, relations, attributes, and hierarchies present in the domain [2]. It is also called a specification of a conceptualization [3]. Ontologies reflecting domain knowledge were utilized to drive the application's design and provide the system with semantic technology capabilities [4]. Before forming the ontology, the ontology engineer must first identify the main elements, which include concepts, relationships, functions, individuals or instances, and axioms [5]. These five components as a formalization of knowledge in ontologies [6]. Ontologies are essential components of the informatics ecosystem that supports life science research, allowing for the analysis of large datasets, data standardization and integration, search, and discovery [7]. Furthermore, in the era of Big

Data, there is an urgent need to transform the way to model, organize, and refine data, given the vast amount of information and data available today on the internet. Thus, designing ontologies and using them to maximize the benefit of accessing and extracting valuable implicit and explicit knowledge from structured and unstructured data is one method of modelling data efficiently [8].

The ontology development methodology can be referred to as the set of activities that need to be performed when constructing ontologies, to ensure the clarity, coherence, extendibility, reusability, and reliability of the ontology [9]. Unfortunately, there is no standard methodology for constructing the ontology that has been agreed upon [10]. There is no one-size-fits-all approach to developing ontologies. Instead, the idea of combining different methodologies and techniques is supported by practitioners in the field of ontologies [11]. However, the general phases involved in the ontology construction methodology proposed by [12], as shown in Figure 1, can be used as a guideline to construct the ontology. The phases include ontology specification, ontology conceptualization, ontology implementation, ontology validation, and evaluation. In the ontology specification phase, the purpose of ontology building and system development, maintenance, and objects of the ontology application need to be identified, and the domain and field of interest need to be described [12]. Besides, the collection of data also occurs at this phase. Next, in the ontology conceptualization phase, the domain conceptual model is defined, the ontology structure is identified, and ontology relations are mapped. After that, the ontology is constructed in the implementation phase by using the required implementation tool. After ontology construction, the ontology needs to be validated and evaluate either by using domain experts or quantify the precision and recall of the ontology. The ontology can be evaluated in different approaches accordingly.

Ontology construction can be performed manually, semi-automatically, and fully automatically. Based on the knowledge acquisition method, the methodology of ontology can be classified into automating, semi-automated, and manual (from scratch) [13]. The manual method included the interaction between the knowledge analyst and the expert while the automatic ontology learning methods included text mining and knowledge extraction [13]. Because using human intervention in manual ontology construction, is a very complex and tedious task, thus many methods proposed offer automatic or semi-automatic ways for ontology construction [14]. Besides that, the exponential growth of unstructured data on the internet has made the learning of automated ontology from unstructured text a hot topic in various research areas. Thus, with automatic ontology construction, it will significantly reduce the labor cost and time of labor required to build ontologies [15]. Several approaches for automatic ontology construction have been discussed in the literature. For example, information extraction (IE) methods, natural language processing (NLP), and comparison with knowledge references are used to build and populate ontologies automatically and semi-automatically [16].

Ontology learning (OL) is a semi-automated process for creating, maintaining, and transferring various types of information into an ontology with minimal human intervention to ensure better knowledge representation and sharing [17]. OL is the most important step toward lowering the cost of ontology building, which refers to a collection of semi-automated frameworks for creating and maintaining ontologies [18]. Therefore, OL from a text can be referred to as the process of supporting the semi-automated development of ontologies from text [19]. However, in semi-automatic, minimal human intervention in one or more ontology design tasks can affect the quality of generated ontology [20]. Meanwhile, it differs from a fully automated ontology construction where the entire construction is delegated to a software system in an automatic process [20]. In addition, fully automatic ontology construction has a higher level of complexity, takes less time, and reduces the role of expert knowledge in ontology evaluation [21].

In this paper, we have identified several methods and methodologies proposed for developing ontologies automatically. For example, to construct semi-automatic ontology, Semi-automatic Sentiment Domain Ontology Building Using Synsets (SASOBUS) methodology was proposed by [22] where term selection phase as its automate part; Sabença framework in the form of modular methodology [23] where constructor, converter, weigher, extractor and exporter modules as automating part; Norms2Onto, a semi-automatic ontology construction method based on machine learning algorithms [17]; Learn2Construct, LDA-based construction learning method [18] to construct fully automatic ontology; WEB2ONTO, an approach for constructing ontology automatically from web pages [24]; ontology learning method and many others. This paper will assist fellow researchers and practitioners in better understanding, and thus implementing, the methods of automatic ontology construction in various fields.

We review the existing literature on the construction of automatic ontologies in this paper. Our main goal is to investigate why an automatic ontology is being built and what methods or techniques are being used to build the ontology automatically. This section also explores an extension of previously elaborated knowledge systematization for ontology learning approaches [14] and the analysis of the chosen approaches allows for the identification of both general trends and missing aspects in ontology learning approaches [25]. As a result, we observe that there is still a gap in the research for planning and designing the methodology of automatic ontology construction.

In this study, we present a systematic review which, to the best of our knowledge, reviews literature on automatic ontology construction covering up to the year 2021. This review is expected to enlighten other researchers interested in this field of study. This review is not exhaustive yet can act as a navigational guide. The contributions of our review are summarized as follows:

- The main research published recently regarding automatic ontology construction is investigated.
- The approach and methodology proposed in automatic ontology construction are identified and discussed.
- The main tools and technologies related to automatic ontology construction are identified.
- A discussion about the validation process and main results of the previous research has been conducted.

This paper is divided as follows: Section 2 presents the conducted systematic review methodology that consists of the definition of research questions, search phase, inclusion and exclusion criteria, and paper eligibility screening. Section 3 presents the result of the systematic review that consists of answers to the research questions, and finally, the conclusion of the systematic review and our thoughts on directions for future work are presented in section 4.

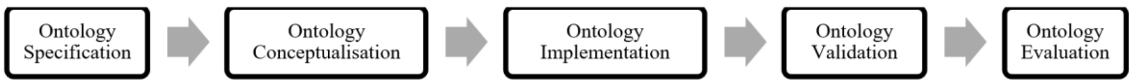


Figure 1. Ontology construction methodology standard phase

2. RESEARCH METHOD

For this paper, the systematic review (SR) was conducted using the preferred reporting items for systematic reviews and meta-analysis (PRISMA) approach as done by [26]. PRISMA is an evidence-based minimum set of items used to guide the development and structure of SRs and other meta-analyses. It was designed to help researchers to do literature reviews systematically and transparently report how the review was done which lead to the findings [27]. Thus, by adapting the PRISMA approach in our paper, the reviewing protocol includes three steps, namely definition of research questions, search phase, and specification of inclusion and exclusion criteria. The specification of these steps for our research study is described in the subsequent sections.

2.1. Definition of research questions

This SR is laid out in a way that covers the scope of research reviewed by categorizing and reviewing existing related publications. The first step consists of defining the research questions to precisely describe the coverage rate of existing works. By studying related works, we can provide several insights that can then help researchers develop new ideas. The research questions used in our SR are described in Table 1.

2.2. Search phase

To conduct our SR, the first step is to define the information sources. As depicted in Table 2, various academic databases, digital libraries, and search engines-both academic and open access-have been searched. The next step consists of defining procedures for exploring the scientific and technical documentation that these searches returned, to find relevant papers to our context. The proposed procedure is based on two main steps: i) determining search terms from the previous research questions to obtain a set of keywords; and ii) determining queries that will be used to find and collect all related results by using Boolean operators AND/OR. The total number of papers identified in the first phase was 223 with search terms in the title that can be the most relevant. Table 3 depicts the search queries used in this paper.

2.3. Inclusion and exclusion criteria

In order to refine search results, we use a set of inclusion criteria (IC) and exclusion criteria (EC) to determine relevant papers (Table 4). Studies that fail to answer EC are ignored; in addition, a screening process is applied to select relevant papers to our context. The screening process is based on three IC steps:

a) Abstract-based step: we discard irrelevant results based on information and keywords found in paper abstracts. Papers whose abstracts satisfied at least 40% of IC were kept for further processing.
b) Full-text-based step: we discard results that did not address or refer to the research terms in Table 3, i.e., papers that only represent minor aspects of the search terms represented in their abstracts.
c) Quality-analysis-based step: we apply a quality analysis to the remaining results by removing those that did not satisfy any of the following criteria:

- C1: The paper discusses a comprehensive solution to automatic ontology construction.

– C2: The paper includes the technical implementation of the proposed solution.
– C3: The paper includes related works.
– C4: The paper presents a discussion of the obtained results.

Table 1. Research questions

Research question	Motivation
RQ1. What is the distribution per year, authors, domain application, and types of publication of published papers related to automatic ontology construction?	The answer to this question allows identifying when, where, and by whom the research studies have been conducted.
RQ2. What is the methodology used to construct automatic ontology and its application?	The answer to this question illustrates the different phases of methodology for constructing automatic ontology.
RQ3. Which phase or part of ontology construction can be automated?	The answer to this question helps to explore the main difficulties that arise when constructing the ontology and to figure out the methods/techniques/algorithms applied that contribute to the automation.
RQ4. What evaluation method was used and the main results have been drawn based on the evaluation method used?	The answer to this question identifies the methods used to evaluate the quality of the constructed ontology and presents the main outcomes of the studied works.
RQ5. What are the main tools and technologies related to automatic ontology construction?	The answer to this question helps identify appropriate tools and techniques adopted by current applications.

Table 2. Search sources

Source	Type	URL
Science Direct-Elsevier	Digital library	http://www.sciencedirect.com/
Scopus	Search engine	http://www.scopus.com/
IEEE Xplore	Digital library	http://ieeexplore.ieee.org/Xplore/home.jsp
ACM Digital library	Digital library	http://dl.acm.org/dl.cfm
Web of science	Search engine	https://www.webofknowledge.com/
Google Scholar	Search engine	https://scholar.google.com/
ResearchGate	Social networking site	https://www.researchgate.net/

Table 3. Search queries

TITLE-ABS-KEY
S1 (automatic ontology AND construction)
S2 (automatic ontology AND development)
S3 (automatic ontology AND building)
S4 (automatic ontology AND framework)
S5 (automatic ontology AND design)

Table 4. List of the IC and EC

Inclusion criteria	Exclusion criteria
– Studies are published during the period between 2017 and 2021. – Studies should meet at least one of the search terms. – Studies should be published/in-press at a journal, conference, or magazine. – Studies should provide answers to the research questions. – The search is performed based on the title, abstract, and full text.	– Studies that are not written in English. – Duplicated papers. – Studies with missing full text. – Papers not directly relevant to automatic ontology construction.

2.4. Paper eligibility screening

The selection of relevant articles for this review was done through the following key steps: identification, screening, eligibility, and inclusion as detailed by the PRISMA flow diagram. Figure 2 illustrates the systematic search-strategy process. Initially, we obtained 223 records. These results were reduced to 188 after excluding duplicates. Then, eligibility criteria based on the title and abstract were applied to these 188 and this elimination round reduced our results to 62; further eligibility criteria based on the full text finally enabled us to obtain 21 relevant papers. These 21 papers were analyzed in-depth to extract the results presented in the next section.

From the 21 papers screened from the review process, 11 papers (52%) constructed ontology by a fully automatic approach, 8 papers (38%) used a semi-automatically way to construct ontology, and 2 papers (10%) presented surveys of automatic ontology construction. Figure 3 shows the percentages in a graphical representation of the papers that we have reviewed.

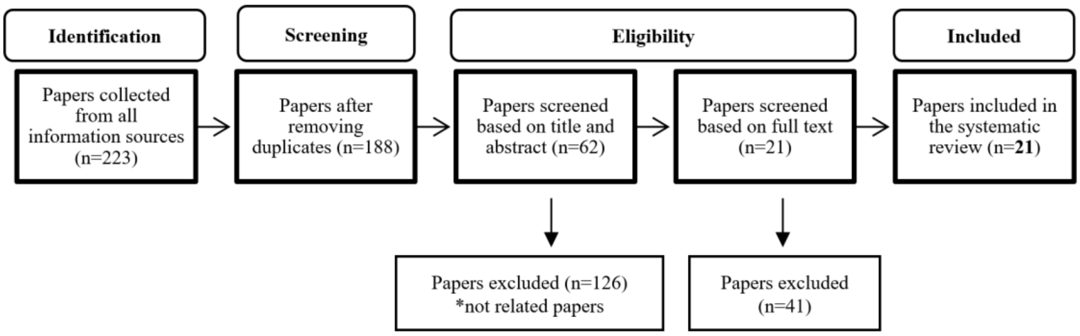


Figure 2. The systematic review process

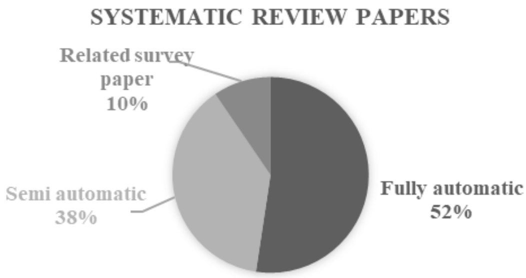


Figure 3. Related papers reviewed

3. RESULTS

In this section, we discussed the result of the review specific to the earlier formulated research questions. This review consists of 21 articles which were systematically selected and fell within the coverage of automatic ontology construction. These answers help researchers to know related recent literature, the methodologies they used, what was automated, the methods used to evaluate the constructed ontology and tools used.

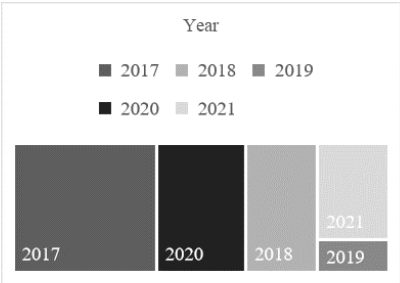
Answer to research question RQ1: **What is the distribution per year, authors, domain application, and types of publication of published papers related to automatic ontology construction?**

The collected papers relating to automatic ontology construction originated from various domains such as online retail, biomedical, public security, information security (IS), Quran, Arabic, Dubai government services, Alzheimer's disease, agriculture, Chinese tax, job portal, sentiment, and ontology learning. The total number of most related publications was 19. Figure 4(a) and Figure 4(b) display the distribution of selected papers by (a) publication year and (b) publisher. The diagram in Figure 4(a) demonstrates that the trend of construction of automatic ontology decreased starting from the year 2018 until 2019. However, this trend increased in the year 2020 and decreased again in the year 2021. From the diagram, we can conclude that the need for automatic ontology construction is still in the field of research across the knowledge engineering community. Most of the selected papers were published in ACM Digital Library, ResearchGate, and ScienceDirect with a total of 10 out of 19 papers followed by Google Scholar, IEEE Xplore, and Scopus altogether with a total of 9 papers as shown in Figure 4(b).

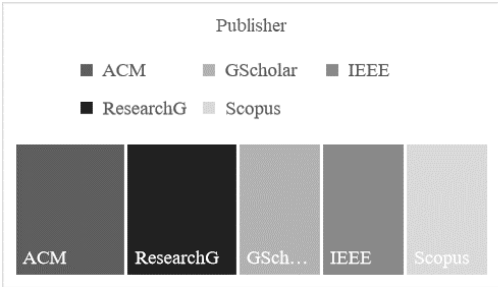
Answer to research question RQ2. **What is the methodology used to construct automatic ontology and its application?**

There were several methodologies with different phases that were implemented in the construction of automatic ontology. Table 5 shows the list of methodology names and methodology phases involved in constructing automatic ontology. After reviewing and analyzing all the papers, we identified that papers 2, 6, 7, 10, 11, 13, 14, 15, 17, 18, and 19 have used fully automatic ontology construction. Fully automatic is sometimes also called automatic ontology construction. Meanwhile, papers 1, 3, 4, 5, 8, 9, 12, and 16 used semi-automatic ontology construction. Both fully and semi-automatic ontology construction have different method and methodology phases involved. Table 5 shows the list of 19 papers with their proposed methodology. To note, we removed papers 20 and 21 from the list considering these two papers served as related survey papers.

From all the papers, we found that only paper 3, 4, 9, 11, and 18 has the specific name of the methodology. For example, these 5 papers have used SOSABUS, Sabenca framework, Norms2Onto, Learn2Construct, and WEB2ONTO methodology respectively. While others have used generic names for their proposed methodology like semi-automatic methodology [12], rule-based approach and FCA/RCA approach [19], bottom-up generic ontology learning approach [28], and fully automatic methodology [29].



(a)



(b)

Figure 4. Distribution of selected papers by publication year and publisher (a) publication year and (b) publisher

Answer to research question RQ3. **Which phase or part of ontology construction can be automated?**

After analyzing all 19 papers, we found that there were two types of ontology construction involved; fully automatic and semi-automatic. Both fully and semi-automatic has different approach and techniques used concerning their methodology phases. Based on [20] fully automatic process involves the complete construction of the ontology that is delegated to a software system. Meanwhile, the semi-automatic process requires human or manual intervention in one or more ontology design tasks which can affect the quality of generated ontology [20]. Thus we consider that the fully automatic ontology construction is automated in each phase of the methodology. Therefore, the automated phase or part has been identified only in semi-automatic ontology construction; papers 1, 3, 4, 5, 8, 9, 12, and 16. Table 6 shows the list of automated phases in the construction of semi-automatic ontology in which the involved phases follow the standard methodology as in Figure 1.

To construct an automated ontology in a semi-automated ontology paper, we have identified the automated phases and activities involved in each paper. From the review, most of the mentioned papers have identified their automation in the ontology specification phase which is the first phase of the standard methodology as shown in Figure 1. For example [17], [22], [30] and [31] has done their automatic construction in ontology specification phase. Meanwhile, [19] has done automation in the ontology conceptualization phase, the second phase of the methodology, and [12] in the implementation phase which is the third phase of the methodology. However, [11] and [23] have done automatic construction in more than one phase whereas [11] has done automatic construction in the ontology specification and evaluation phase, while [23] in the ontology specification, conceptualization, and implementation phase. As can be seen from Table 6, most of the papers have done automatic construction in the ontology specification phase by doing term extraction activity. Sharef *et al.* [11] has used the engine based on POS tagging coupled with pattern-based extraction techniques to automatically extracts the candidate terms from the competency questions, in [17] has used TF-IDF measure to extract relevant concepts from the corpus, in [22] used relevance score based on DP and DC to extract useful terms, and [23] performs the activity of constructor and converter for pre-processing and use TF-IDF for term weighing. Besides, [30] has applied NLP to extract important information from academic papers by using Python 2.7 and NLTK, meanwhile [31] has used a term scoring algorithm for term and phrases extraction from textual corpus to generate the domain terminology. In the ontology conceptualization phase, [19] has generated Arabic compound structures through a set of POS patterns using the platform NooJ and Xerox Morphology System for transliteration, and [23] use Hearst's (1992) method to extract the taxonomic relations. To implement the ontology, [12] has added the Celfie plugin to extract data from an Excel sheet to create the ontology, the individuals, data properties, and object properties, and apply the Fact++ reasoning technique to the ontology to gather logical consequences from a set of asserted facts or axioms, and [23] use OWL language to build an ontology. Lastly, [11] also done automation in ontology evaluation by using DL as formal query language and executing using DL-Query Tab to check ontology correctness, using concept map as a reference frame to compare with implemented ontology, and OntoGraf Plugin to capture different snapshots of implemented ontology.

Table 5. List of method

Reference	Methodology name	Methodology phases	Degree of automation
Paper 1 [12]	Semi-automatic methodology	Specification, conceptualization, implementation, validation, evaluation	Semi-automatic
Paper 2 [8]	No specific methodology name mentioned	NLP, entity discovery, semantic entity enrichment, RDF triple extraction, syntactic patterns, ontology factory	Fully automatic
Paper 3 [22]	SASOBUS	Ontology structure, skeletal ontology, term selection, hierarchical relations	Semi-automatic
Paper 4 [23]	Sabença framework	Constructor, parser, converter, tagger, weigher, extractor, exporter, replacing modules in the framework	Semi-automatic
Paper 5 [11]	New proposed approach by F. Dalia, A. Safia, A. Mostafa (2017), motivation from Brusa et al., 2008	Elaborating the motivation scenarios and competency questions, automatic extraction of potential terms from the competency questions, construction of concept map, transforming the concept map to ontology, evaluation	Semi-automatic
Paper 6 [3]	Arabic ontology framework	Extraction, XML schema parsing, ontology generation, refinement, and evaluation	Fully automatic
Paper 7 [32]	No specific methodology name mentioned	Data extraction and validation, data processing and NLP tasks, mapping rules process, ontology construction, and generation	Fully automatic
Paper 8 [19]	Rule-based approach and FCA/RCA approach	General relations deciphering, General objectProperty relations specification	Semi-automatic
Paper 9 [17]	Norms2Onto	Pre-processing, learning, visualization	Semi-automatic
Paper 10 [28]	A bottom-up generic ontology learning approach	Pre-processor, concept extractor, concept to domain mapper, concept pair extractor, taxonomic relation extractor, non-taxonomic relation extractor	Fully automatic
Paper 11 [18]	Learn2Construct, LDA-based	Pre-processing, terms extraction, topic modeling, concepts & relations extraction, ontology visualization	Fully automatic
Paper 12 [30]	Semi-automatic ontology learning method	Information exploration, relationship construction, ontology verification	Semi-automatic
Paper 13 [15]	The ontology generation life cycle was inspired by I. Bedini and B. Nguyen (2007)	Generation, refinement, mapping	Fully automatic
Paper 14 [21]	No specific methodology name mentioned	Term and relation extraction, matching with Alzheimer glossary, matching the ontology design patterns, score computation similarity term and relations with ODPs, ontology building, evaluation	Fully automatic
Paper 15 [33]	Rule-based reasoning algorithm RelExOnt	Identification of equivalent terms: has_synonym relation, identification of hierarchical relation: is_type_of, identification of instances: is a relation, identification of intercrops: is_intercrop relation	Fully automatic
Paper 16 [31]	Middle out approach	Pre-processing, term extraction, relationship extraction	Semi-automatic
Paper 17 [20]	No specific methodology name mention	Web scrapping, concepts classification, pre-processing, archetypes identification, ontology design	Fully automatic
Paper 18 [24]	WEB2ONTO	Remove HTML tags, images and fetch paragraphs only to work on them, to determine the co-reference in paragraphs and replace them with their source mention, to extract triples from sentences (subject, action, object), and get the sources from the verb in action (word stem), to determine the type of subject and object, to get synonyms for action using WordNet, to check if the subject and object already exist in ontology to avoid duplication	Fully automatic
Paper 19 [29]	Fully automatic methodology	Key, Value-based Web data extraction, a combiner and partitioner-based integration of log files, automatic ontology construction based on Flat File Parsing	Fully automatic

Answer to research question RQ4. **What evaluation method was used and the main results have been drawn based on automatic ontology construction (fully automatic and semi-automatic)?**

Evaluation of the ontology occurs after the ontology is generated. Generally, the purpose of ontology evaluation is being conducted to ensure the quality and the accuracy of ontology elements (concepts, relations, and keywords) [32]. Besides that, it is important to evaluate the ontology's main aspects to guarantee that representation is the most real according to the domain [5]. Therefore, in this paper, we have identified the evaluation methods used in these papers. Table 7 (see in Appendix) depicts the list of evaluation methods and its result.

Answer to research question RQ5. **What are the main tools and technologies related to automatic ontology construction?**

There were several tools used to develop the ontologies automatically. We divided the tools according to the standard phase of ontology development methodologies such as pre-processing, keyword extraction, implementation, and evaluation as shown in Table 8 (see in Appendix). Most of the ontology methodology has used automatic pre-processing pipelines like NLTK. Overall, for implementation, they use the Protégé software tool and OWL language.

Table 6. List of semi-automatic methodology phase

Reference	Automate phase	Automate activity description
Paper 1 [12]	Phase: Ontology implementation	**Activity: Structuration of ontology** – Add Celfie plugin to extract data from Excel sheet - to create the ontology, the individuals, data properties, and object properties. – Apply Fact++ reasoning technique to the online retail old ontology and evolved online retail ontology - to gather logical consequences from a set of asserted facts or axioms.
Paper 3 [22]	Phase: Ontology specification	**Activity: Term selection** – Use relevance score based on domain pertinence (DP) and domain consensus (DC) to extract useful terms. – Use of synsets in term extraction, concept formation, and concept subsumption. – Implement a Simplified Lesk algorithm to identify a sense of each word for both domain corpus and contrastive corpora for ontology learning.
Paper 4 [23]	Phase: Ontology specification	**Activity: Constructor for pre-processing** – Perform functions: controls communication flow between modules, receives calls from external applications, controls used directory mapping, controls parameterization and customization, and controls all document importation.
	Phase: Ontology specification	**Activity: Converter for pre-processing** – Use Apache POI 3.10 to convert .doc and .docx document type and iText 5.5.2 to convert the .pdf document type.
	Phase: Ontology specification	**Activity: Term weigher for terms extraction** – Use method Term Frequency-Inverse Document Frequency (TF-IDF) for term weighing.
	Phase: Ontology conceptualization	**Activity: Identification of concept and relation extraction** – Use Hearst's (1992) method to extract the taxonomic relations.
	Phase: Ontology implementation	**Activity: Structuration of ontology** – Builds ontological structures in OWL language. – Generate ontological structure containing simple terms, compound terms, and Hearst pattern-based relations.
Paper 5 [11]	Phase: Ontology specification	**Activity: Term extraction** – The engine based on part-of-speech (POS) tagging coupled with pattern-based extraction techniques-automatically extracts the candidate terms (concepts, instances, relationships) from the competency questions.
	Phase: Ontology evaluation	**Activity: To check the completeness and correctness of ontology** – Use the robust procedure to evaluate ontologies at both formal and graphical levels. – Formal level-use competency questions as a reference frame. Use description logic (DL) as a formal query language, then execute using DL-Query Tab available in Protégé IDE to check ontology correctness. – Graphical level-use concept map as a reference frame to compare with implemented ontology. Use OntoGraf Plugin to capture different snapshots of implemented ontology. It also generates diverse combinations of terms in the ontology.
Paper 8 [19]	Phase: Ontology conceptualization	**Activity: General relations deciphering** – Generate Arabic compound structures through a set of POS patterns using the platform NooJ. – Use Xerox Morphology System for transliteration.
	Phase: Ontology conceptualization	**Activity: General object property relations specification** – Adopt a data-driven strategy to find the preposition that can most likely be used to specify the “object property (X, Y)” expression.
Paper 9 [17]	Phase: Ontology specification	**Activity: Term extraction** – Extract relevant concepts from the corpus based on TF-IDF measure to extract concepts from their frequency in the corpus. – Build a learning model to predict the definition of each relevant concept and its subcategory.
Paper 12 [30]	Phase: Ontology specification	**Activity: Information extraction** – Apply NLP to extract important information related to intrusion detection from academic papers by using Python 2.7 and Natural Language Toolkit programming language.
Paper 16 [31]	Phase: Ontology specification	**Activity: Term extraction** – Use term scoring algorithm for term and phrases extraction from textual corpus to generate the domain terminology.

4. CONCLUSION

This paper summarizes the distribution of automatic ontology construction papers per year, authors and domain application, their automatic construction methodology along with evaluation measures, and highlights the automation activity and tools involved in the construction process. A systematic review was conducted using the PRISMA approach and its selection process of identification, screening, eligibility, and inclusion was reported in detail. A total of 19 works were selected from the 223 initially extracted, based on their relevance to the five main research questions we developed. A fully automatic process involves the complete construction of the ontology that is delegated to a software system, while the semi-automatic process requires human or manual intervention in one or more ontology design tasks. From the review, we discovered that the automation has majorly occurred in the ontology specification phase of the methodology by performing term extraction activity. We also looked at various evaluation techniques for automatic ontology and discovered that domain expert evaluation is the best since it gives more accurate results. Our survey of these papers has led us to conclude that automatic ontology construction would help ontologies evolve and save on the cost and time of ontology creation and maintenance. Besides fully automatic construction could speed up and reduce the human's role as an expert to evaluate ontology rather than building ontology manually. However, most of these contributions are newly developed and require further optimization, providing timely future work opportunities for researchers interested in this interdisciplinary field.

ACKNOWLEDGEMENTS

We are extremely grateful to the reviewers who took the time to provide constructive feedback and useful suggestions for improving this article. The Ministry of Higher Education, Malaysia is funding this research through the Fundamental Research Grant Scheme (FRGS) at Universiti Teknologi MARA (UiTM) Shah Alam, Malaysia (FRGS/1/2019/SS05/UITM/02/5). The authors would like to thank the Universiti Teknologi MARA Shah Alam, Malaysia for the support acknowledgments.

APPENDIX

Table 7. List of evaluation methods of automatic ontology construction and its result

Reference	Evaluation method	Results
Paper 1 [12]	– Use domain experts by measuring the precision and recall of ontology. – Implementation of the quality features dimension by measuring using metrics Cohesion and Conceptualisation (richness of semantic, attribute, and inheritance).	– Precision: 97% – Recall: 72% – Cohesion: 1524.67 – Semantic richness: 0.04 – Attribute richness: 0.17 – Inheritance richness: 36
Paper 2 [8]	– Use task-based evaluation, the gold standard, and OntoGain tools	– Evaluation concept extraction: F-measure of 58.12% for CDR corpus, and 81.68% for SemMedDB – Biomedical taxonomic relation extraction: F-measure of 65.26% using dataset CDR, and 77.44% using dataset SemMedDB – Biomedical non-taxonomic relation extraction: F-measure of 52.78% using CDR corpus, 58.12% using SemMedDB – The comparison with manually constructed baseline Alzheimer ontology: F-measure of 72.48% in terms of concepts detection, 76.27% in relation extraction, and 83.28% in property extraction
Paper 3 [22]	– Use hybrid approach Ont + LCR-Rot-hop (a combination model of ontology with LCR-Rot-hop separated Neural Network)	– Its accuracy was slightly lower (about 2%) than the accuracy of the manual ontology but the user time is significantly lower (about halved). – sOnt + LCR-Rot-hop accuracy: out of sample: 84.49%; in sample: 86.07%; cross-validation: 79.73%
Paper 4 [23]	– Manual validation by domain experts in a public security area	– 283.064 terms were found in the morphosyntactic labeling phase – 5.934 total non-relevant terms for the domain found in term identification – Terms were weighed using the TF-IDF method, which resulted in 20.076 extracted terms
Paper 5 [11]	– The robust evaluation procedure consists of formal (competency questions) and graphical validation (concept map) as a reference frame – Use domain experts to check the effectiveness of the proposed ontology	– Automatic extract the concept and relationships in 50 competency questions which resulted: - response time 0.305999994278 sec for detected concepts and instances; - response time 0.308000087738 sec for detected relationships

Paper 6 [3]	– Use a data-driven evaluation method and tree-based mining	– The output of the ontology is compared with the source document – Apply the similarity measure with the best-obtained result was 65%.
Paper 7 [32]	– Evaluation metrics: Precision, Recall, F1-measure to ensure the quality and the accuracy of ontology elements (concepts, relations, and keywords) in the Protégé tool. – The confusion matrix-to evaluate the concept/relations components in the extracted ontology	– The precision with 87% on average, and recall with 97% on average
Paper 8 [19]	– Validation by using domain experts to verify the accuracy of chosen relations and reliability of proposed rules – Compare the derived ontology to a human-modeled ontology-to evaluate the approach	– The approach to achieving high precision and recall scores: Precisions: 84% and 92%, Recall: 68% and 73%
Paper 9 [17]	– Task-based evaluation method by using conventional measures in information retrieval such as recall (R), precision (P), F-measure (F-score)	– Norms2Onto + Linear SVC: 0.87 P; 0.86 R; 0.86 F-score. – Norms2Onto +Logistic Regression: 0.81 P; 0.5 R; 0.56 F-score. – Norms2Onto +Random Forest: 0.93 P; 0.91 R; 0.92 F-score. – Norms2Onto + Multinominal Naïve Bayes: 0.48 P; 0.23 R; 0.26 F-score. – Result: proposed approach gives better performance in comparison with the random forest algorithm
Paper 10 [28]	– Use domain experts - to evaluate the quality of generated ontology – Precision measure-to judge how far the extracted information is correct for the major phases of the ontology learner	– The experimental result exhibits: 78.75% of precision in candidate term extraction, 79.59% of precision in taxonomy induction, and 55.00% of precision in specific semantic relation extraction for a morphologically complex Amharic language with a limited size corpus.
Paper 11 [18]	– Use criteria-based and task-based evaluation	– Learn2Construct has a precision of 81%, a recall of 79% and an F-score of 79.98% using the corpus. – Learn2Construct has a precision of 95%, a recall of 90%, and an F-score of 92.43% using a larger corpus which is constituted with corpus 2.
Paper 12 [30]	– Use multiple domain experts to verify intrusion detection ontology: 5 Ph.D. students who had a basic knowledge of network security and one of whom was majoring in intrusion detection	– Based on statistics and graphic results – Statistic result: 86 papers out of 168 presented intrusion type and intrusion detection solutions in the title, abstract, introduction, and conclusion part of papers. – Graphic result: construct ontology based on the statistical results using OWLGrEd Software to present the graphic solution-oriented intrusion detection knowledge mapping.
Paper 13 [15]	– Use reference ontology; YAGO and DOLCE to evaluate and verify the generated KGs	– For each concept, c in the temporary ontology OKG, the number of domain inconsistencies $\varepsilon\ (c)$ is calculated as the sum of the differences between the properties in the target ontology, On, and properties in generated ontology OKG
Paper 14 [21]	– To evaluate in terms of complexity, time, effort	– The evaluation compared with the result by Drame et al that construct a semi-automatic ontology – The result of the accuracy value of fully automatic ontology construction is 72%.
Paper 15 [33]	– Domain expert evaluation - expert opinion is used to judge whether the relation extracted between two particular terms does hold in real life	– mOIE performed well for the identification of synonym relations with a precision of 67% and recall of 72% on 200 pages of agricultural data – RelExOnt performs well with an average precision of 86.89%.
Paper 16 [31]	– Use the domain expert evaluation method, 25 experts from the tax bureau and tax firm	– The combined method shows significant improvement both in terms of precision and recall. – The combination of lexical similarity and semantic similarity for taxonomic relation extraction also performs better than their respective benchmarks.
Paper 17 [20]	– Use domain expert to evaluate the domain concept of ontology	– The ontology contains 2,518 concepts and attributes regrouped into 1,166 trivial and non-trivial concerns.
Paper 18 [24]	– WEB2ONTO is compared to the FRED approach since it is the only fully automatic tool – Evaluating information retrieval systems: calculating precision and recall	– FRED: Average total precision: 0.64, recall: 0.64 – WEB2ONTO ontology: Average total precision: 0.84, recall: 0.79. – WEB2ONTO is less than FRED in recovering all information but the correct extracted triples in WEB2ONTO are more than FRED.
Paper 19 [29]	– To evaluate completeness, validity, functional suitability, usability, and adequacy	– The ontology is shared with the ontology engineers to get the feedback for the mentioned parameters. – The constructed ontology got: 80%-in case of validity of intent behind the creation 86.67%-in case of ease of use 86.67%-the usability of the ontology response 86.67%-the purpose of survival.

Table 8. Ontology development tools classification

Pre-processing	– Python code for data cleaning [15] – Stanford CoreNLP 3.8.03 toolkit for performing NLP pipeline on contrastive corpora (English book) [22] – Apache OpenNLP 1.5.3 and its models to perform morphosyntactic labeling on the Portuguese language [23]
Keyword extraction	– Plugin (celfie)-extract data from Excel sheet [12] – Python 2.7 for information extraction [30] – Automatic Extraction Dataset System (AEDS) tool to automatically extract all services information from the webpages [32]
Implementation	– X20WL tool to build OWL ontology from XML data source [3] – Cmap Tools (2013) to construct a concept map [11] – Protégé 5.0.0 software / Protégé IDE [11], [12], [19], [23], [28], [29], [32], [33] – Apache Jena API to write ontology in RDF file [8] [24] – Jambalaya plug-in to supports the visualization of Arabic letters in ontology visualization [19] – Neo4j or Watson to generate KG from text [15] – Java API for WordNet Searching (JAWS) library to obtain synsets from WordNet [22]
Evaluation	– YAGO and DOLCE as reference ontologies [15] – HAABSA framework in Python to evaluate created ontology [22] – OntoGain tools [8]

REFERENCES

[1] M. Hamiz, H. Haron, M. Bakri, and N. L. Mohd Lazim, “Ontology model for intake suggestion and preparation for malay confinement dietary recipes,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 17, no. 1, p. 481, Jan. 2020, doi: 10.11591/ijeecs.v17.i1.pp481-488.

[2] M. N. Asim, M. Wasim, M. U. G. Khan, W. Mahmood, and H. M. Abbasi, “A survey of ontology learning techniques and applications,” *Database*, vol. 2018, pp. 1–24, Jan. 2018, doi: 10.1093/database/bay101.

[3] D. Fadl, S. Abbas, and M. Aref, “Automatic Arabic ontology construction framework: an insectivore’s animal case study,” in *International Conference on Advanced Intelligent Systems and Informatics*, Cham: Springer, 2018, pp. 458–466.

[4] J. Elhassouni, A. El Qadi, M. Bazzi, and M. El Haziti, “Modeling with ontologies design patterns: credit scorecard as a case study,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 17, no. 1, pp. 429–439, Jan. 2020, doi: 10.11591/ijeecs.v17.i1.pp429-439.

[5] C. Reyes-Peña and M. Tovar-Vidal, “Ontology: components and evaluation, a review,” *Research in Computing Science*, vol. 148, no. 3, pp. 257–265, Dec. 2019, doi: 10.13053/rcs-148-3-21.

[6] T. Slimani, “A Study Investigating Typical Concepts and Guidelines for Ontology Building,” *Journal of Emerging Trends in Computing and Information Sciences*, vol. 5, no. 12, pp. 886–893, 2015, doi: https://doi.org/10.48550/arXiv.1509.05434.

[7] R. C. Jackson, J. P. Balhoff, E. Douglass, N. L. Harris, C. J. Mungall, and J. A. Overton, “ROBOT: a tool for automating ontology workflows,” *BMC Bioinformatics*, vol. 20, no. 1, p. 407, Dec. 2019, doi: 10.1186/s12859-019-3002-3.

[8] M. Alobaidi, K. M. Malik, and S. Sabra, “Linked open data-based framework for automatic biomedical ontology generation,” *BMC Bioinformatics*, vol. 19, no. 1, p. 319, Dec. 2018, doi: 10.1186/s12859-018-2339-3.

[9] A. Abdelghany, N. Darwish, and H. Hefni, “An Agile Methodology for Ontology Development,” *International Journal of Intelligent Engineering and Systems*, vol. 12, no. 2, pp. 170–181, Apr. 2019, doi: 10.22266/ijies2019.0430.17.

[10] H. A. A. Ghanim and L. Kovács, “Development of ontology-based model to support learning process in LMS,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 1, pp. 507–518, Oct. 2021, doi: 10.11591/ijeecs.v24.i1.pp507-518.

[11] N. M. Sharef, A. Mustapha, M. A. A. Murad, and R. Iqbal, “An ontology development approach using concept maps driven by automatic term extraction,” *International Journal of Information and Communication Technology*, vol. 10, no. 1, p. 51, 2017, doi: 10.1504/IJICT.2017.10001311.

[12] R. Alaa, M. Gawish, and M. Fernández-Veiga, “Improving recommendations for online retail markets based on ontology evolution,” *Electronics*, vol. 10, no. 14, p. 1650, Jul. 2021, doi: 10.3390/electronics10141650.

[13] T. A. Gavrilova and I. A. Leshcheva, “Ontology design and individual cognitive peculiarities: a pilot study,” *Expert Systems with Applications*, vol. 42, no. 8, pp. 3883–3892, May 2015, doi: 10.1016/j.eswa.2015.01.008.

[14] J. Wątróbski, “Ontology learning methods from text - an extensive knowledge-based approach,” *Procedia Computer Science*, vol. 176, pp. 3356–3368, 2020, doi: 10.1016/j.procs.2020.09.061.

[15] S. Elnagar, V. Yoon, and M. Thomas, “An automatic ontology generation framework with an organizational perspective,” in *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2020, vol. 2020-Janua, pp. 4860–4869, doi: 10.24251/HICSS.2020.597.

[16] T. Louge, M. H. Karray, and B. Archimede, “Investigating a Method for Automatic Construction and Population of Ontologies for Services: Performances and Limitations,” in *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*, Oct. 2018, pp. 1–6, doi: 10.1109/AICCSA.2018.8612844.

[17] A. Drissi, A. Khemiri, S. Sassi, and R. Chbeir, “A new automatic ontology construction method based on machine learning techniques: application on financial corpus,” in *Proceedings of the 13th International Conference on Management of Digital EcoSystems*, Nov. 2021, pp. 57–61, doi: 10.1145/3444757.3485111.

[18] A. Khemiri, A. Drissi, A. Tissaoui, S. Sassi, and R. Chbeir, “Learn2Construct: an automatic ontology construction based on LDA from texual data,” in *Proceedings of the 13th International Conference on Management of Digital EcoSystems*, Nov. 2021, pp. 49–56, doi: 10.1145/3444757.3485110.

[19] I. B. Mezghanni and F. Gargouri, “Deriving ontological semantic relations between Arabic compound nouns concepts,” *Journal of King Saud University - Computer and Information Sciences*, vol. 29, no. 2, pp. 212–228, Apr. 2017, doi: 10.1016/j.jksuci.2017.03.001.

[20] I. Meriah, L. B. A. Rabai, and R. Khedri, “Towards an automatic approach to the design of a generic ontology for information security,” in *2021 Reconciling Data Analytics, Automation, Privacy, and Security: A Big Data Challenge (RDAAPS)*, May 2021, pp. 1–8, doi: 10.1109/RDAAPS48126.2021.9452006.

[21] D. E. Cahyani and I. Wasito, "Automatic ontology construction using text corpora and ontology design patterns (ODPs) in alzheimer's disease," *Jurnal Ilmu Komputer dan Informasi*, vol. 10, no. 2, p. 59, Jun. 2017, doi: 10.21609/jiki.v10i2.374.

[22] E. Dera, F. Frasincar, K. Schouten, and L. Zhuang, "SASOBUS: semi-automatic sentiment domain ontology building using synsets," in *The Semantic Web*, Cham: Springer International Publishing, 2020, pp. 105–120.

[23] N. C. Guimarães and C. L. De Carvalho, "A modular framework for ontology learning from text in Portuguese," *Multi-Science Journal*, vol. 3, no. 3, pp. 37–42, Oct. 2020, doi: 10.33837/msj.v3i3.899.

[24] N. Elmesalmy, M. Hadhoud, and M. Fayeka, "WEB2ONTO: automatic ontology construction approach from web pages," in *2019 15th International Computer Engineering Conference (ICENCO)*, Dec. 2019, pp. 175–182, doi: 10.1109/ICENCO48310.2019.9027405.

[25] A. Konys, "Knowledge systematization for ontology learning methods," *Procedia Computer Science*, vol. 126, pp. 2194–2207, 2018, doi: 10.1016/j.procs.2018.07.229.

[26] Y. Hajjaji, W. Boulila, I. R. Farah, I. Romdhani, and A. Hussain, "Big data and IoT-based applications in smart environments: A systematic review," *Computer Science Review*, vol. 39, p. 100318, Feb. 2021, doi: 10.1016/j.cosrev.2020.100318.

[27] M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. n71, Mar. 2021, doi: 10.1136/bmj.n71.

[28] F. Getahun and K. Woldemariyam, "Integrated ontology learner: towards generic semantic annotation framework," in *Proceedings of the 9th International Conference on Management of Digital EcoSystems*, Nov. 2017, pp. 142–149, doi: 10.1145/3167020.3167042.

[29] U. Kethavarapu and S. Saraswathi, "A frame work to construct automatic ontology with combine and partitioner logics," in *2017 IEEE International Conference on Current Trends in Advanced Computing (ICCTAC)*, Mar. 2017, vol. 2018-Janua, pp. 1–8, doi: 10.1109/ICCTAC.2017.8249989.

[30] Y. Zhang, M. Saberi, and E. Chang, "Semantic-based lightweight ontology learning framework: a case study of intrusion detection ontology," in *Proceedings of the International Conference on Web Intelligence*, Aug. 2017, pp. 1171–1177, doi: 10.1145/3106426.3109053.

[31] Y. Qiu, L. Cheng, and D. Alghazzawi, "Towards a semi-automatic method for building Chinese tax domain ontology," in *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, Jul. 2017, pp. 2530–2539, doi: 10.1109/FSKD.2017.8393173.

[32] A. Albarghothi, W. Saber, and K. Shaalan, "Automatic construction of e-government services ontology from Arabic webpages," *Procedia Computer Science*, vol. 142, pp. 104–113, 2018, doi: 10.1016/j.procs.2018.10.465.

[33] N. Kaushik and N. Chatterjee, "Automatic relationship extraction from agricultural text for ontology construction," *Information Processing in Agriculture*, vol. 5, no. 1, pp. 60–73, Mar. 2018, doi: 10.1016/j.inpa.2017.11.003.

BIOGRAPHIES OF AUTHORS

Zayanah Zafirah Zulkipli is a Ph.D. candidate in Computer Science at the Universiti Teknologi MARA (UiTM) Shah Alam, Malaysia. In 2019, she completed her M.Sc. in Computer Science from the Faculty of Computer and Mathematical Sciences, at the same university. She previously received a B.Sc. in Computer Science from the Faculty of Computer and Mathematical Sciences, UiTM Melaka, in 2018. Her research interests are concentrated in the field of information retrieval and natural language processing. She can be contacted at: zayanahzafirahaz@gmail.com.

Dr. Ruhaila Maskat is a senior lecturer at the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Shah Alam, Malaysia. In 2016, she was awarded a Ph.D. in Computer Science from the University of Manchester, United Kingdom. Her research interest then was in Pay-As-You-Go dataspaces which later evolved to Data Science where she is now an EMC Dell Data Associate as well as holding four other professional certifications from RapidMiner in the areas of machine learning and data engineering. Recently, she was awarded with the Kaggle BIPOC grant. Her current research grant with the Malaysian government involves conducting analytics on social media text to detect mental illness. She can be contacted at: ruhaila@fskm.uitm.edu.my.

Dr. Noor Hasimah Ibrahim Teo is a senior lecturer at the Universiti Teknologi Teknologi MARA Melaka, Malaysia. She received her Ph.D. in Computer Science from the University of Warwick, United Kingdom in the year 2019. Her research interests are concentrated in the field of artificial intelligence and NaturalLanguae Processing. Her most recent research work has focused on processing text documents to generate questions using BERT model classification automatically. She can be contacted at: shimateo@uitm.edu.my.