Journal of Theoretical and Applied Information Technology
15th February 2018. Vol.96. No 3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

QURAN ONTOLOGY: REVIEW ON RECENT DEVELOPMENT AND OPEN RESEARCH ISSUES

[1]NANNA SURYANA, [2,3]FANDY SETYO UTOMO, [4]MOHD SANUSI AZMI

[1,2,4]Faculty of Information and Communication Technology,

Universiti Teknikal Malaysia Melaka, Melaka, Malaysia

[3]Department of Information Systems, STMIK AMIKOM Purwokerto, Purwokerto, Indonesia

E-mail: [1]nsuryana@utem.edu.my, [3]fandy_setyo_utomo@amikompurwokerto.ac.id, [4]sanusi@utem.edu.my

ABSTRACT

Quran is the holy book of Muslims that contains the commandment of words of Allah. Quran provides instructions and guidance to humankind in achieving happiness in life in the world and the hereafter. As a holy book, Quran contains rich knowledge and scientific facts. However, humans have difficulty in understanding the Quran content. It is caused by the fact that the meaning of the searched message content depends on the interpretation. Ontology able to store the knowledge representation of Holy Quran. This paper studies recent ontology on Holy Quran research. We investigate the current trends and technology being applied. This investigation cover on several aspects, such as outcomes of previous studies, language which used on ontology development, coverage area of Quran ontology, datasets, tools to perform ontology development ontology population techniques, approaches used to integrate the knowledge of Quran and other resources into ontology, ontology testing techniques, and limitations on previous research. This review has identified four major issues involved in Quran ontology, i.e. availability of Quran ontology in various translation, ontology resources, automated process of *Meronymy* relationship extraction, and Instances Classification. The review of existing studies will allow future researchers to have a broad and useful background knowledge on primary and essential aspects of this research field.

Keywords: *Information Retrieval, Ontology, Quran Ontology, Ontology Extraction, Ontology Population.*

1. INTRODUCTION

Quran is the holy book of Muslims that contains the commandment of words of Allah. Quran provides instructions and guidance to humankind in achieving happiness in life in the world and the hereafter. As a holy book, Quran contains rich knowledge and scientific facts. Many scientific facts have been discovered and proved by modern science discovery. However, there are still a lot knowledge and scientific facts that have not been discovered by science. Humans have difficulty in understanding the Quran content. It is caused by the fact that the meaning of the searched message content depends on the interpretation. Furthermore, Quran is rich of synonyms, polysemy, homonyms, and semantics of words.

Quran knowledge can be categorized into *muhkamat* (i.e., semantic verse sentence is clear) and *mutashabihat* (i.e., semantic verse sentence is not clear) that require a high level of complex explanations. *Mutashabihat* verses require elaboration from different sources, such as Hadith, *Tafsir*, and *Asbab Al Nuzul* to complete the knowledge that involves the verses. Based on these categories, there are two approaches which able to comprehend the Islamic values contained in the Quran, namely the literal or textual, and contextual approach. Literal approach is a method to interpret textual meaning of Quran without reference to another source, while the contextual approach is a method to interpret Quran by observing the semantic or contextual information related to Quran, such as the information obtained from Hadith, Tafsir, and Asbab Al Nuzul.

A search for information contained in the Quran can be done using the Information Retrieval (IR). Semantic ontology technique is one of IR method. Ontology able to store the knowledge representation of Holy Quran. While for providing relevant information to the users, semantic search approach able to extract the knowledge of Holy Quran from an ontology. Ontologies are used for making explicit conceptual knowledge models, providing the semantic vocabulary, and allowing this domain knowledge available to be shared by

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

different applications in a great variety of domains [1–3]. Domain Knowledge consists of classes (concepts), properties (attributes), class instances (class members), and property instances (relationships) [4–6]. While, property instances can be divided into two types, i.e. Object property and Data property [7–9]. Semantic search engines able to describe the information precisely based on the contextual meaning of terms in the corpus and the query [10–12].

This paper studies recent ontology on Holy Quran research. We try to investigate the current trends and technology being applied. This investigation cover on several criteria, such as outcomes of previous studies, language which used on ontology development, coverage area of Quran ontology, Datasets, ontology population techniques, approaches used to integrate the knowledge of Quran and other resources into ontology, ontology testing techniques, and limitations on previous research. This study focuses on research works trying to analyze and understand knowledge in Holy Quran by applying semantic-ontology approaches. Purpose of our study is to identify the potential issues on Quran Ontology and to find the feasibility of helpful features, which could be applied for ontology. The review of existing studies will allow future researchers to have a broad and useful background knowledge on the primary and essential aspects of this research field.

This paper is organized as follows. Section 2 describes materials and methods to collect the literature. Section 3 presents the review of recent studies on Quran ontology. Section 4 discusses open research issues. Finally, Section 5 concludes the critical points in this paper.

2. MATERIALS AND METHODS

The research questions (RQ) are specified to keep the review focused. The research questions addressed by this literature review are:

- **RQ1:** What kind of languages is the most used for ontology on Holy Quran?
- **RQ2:** What kind of datasets is the most used for ontology on Holy Quran?
- **RQ3:** What kind of tools able to perform ontology development?
- **RQ4:** How do the existing studies perform Quran ontology development?
- **RQ5:** How do the existing ontologies integrate the knowledge of Quran and other resources like Hadith and Tafsir?
- **RQ6:** How to evaluate the existing Quran ontologies?

The search process for literature consists of several activities, such as selecting digital library, defining the search string, executing the search string, and retrieving a list of primary studies from digital libraries. Digital libraries are used in the search process, i.e. ACM digital library, Emerald, IEEE, Science Direct, Springer Link, Scopus, and Google Scholar. The inclusion and exclusion criteria were used for selecting the primary studies. These criteria are shown in Table 1.

Table 1: Inclusion and Exclusion Criteria

Inclusion	Studies discuss developing and testing the ontology performance.
	For studies that have both the conference and journal versions, only the journal version will be included.
	Literature was limited to the year of publication: Jan. 2013 - April 2017.
Exclusion	Studies not written in English.

3. RECENT STUDIES ON ONTOLOGY DEVELOPMENT FOR HOLY QURAN

This Sub-Section is organized as follows. Sub-Section 3.1 describes significant article publications between January 2013 and April 2017. Sub-Section 3.2 presents outcomes of previous studies. Sub-Section 3.3 shows languages commonly used in Quran ontology. Sub-Section 3.4 shows datasets commonly used in Quran ontology development. Sub-Section 3.5 shows tools often used in ontology development. Sub-Section 3.6 describes existing approaches to perform ontology development. Sub-Section 3.7 presents existing ontologies integrate the knowledge of Quran and other resources. Sub-Section 3.8 shows evaluation techniques to test the ontology. Finally, Sub-Section 3.9 discusses limitations on previous research.

3.1 Significant Article Publications

Based on the searching results, there are twenty literatures which are relevant to the search condition. There are seven studies in 2013, two studies in 2014, four studies in 2015, and seven studies in 2016. There is no publication further about ontology extraction on Holy Quran between January 2017 and April 2017 in Digital Libraries. Summary about number of studies on each digital libraries is described in Figure 1. Based on Figure 1, we collected literature from several sources, i.e. Conference, Journal, and Thesis.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

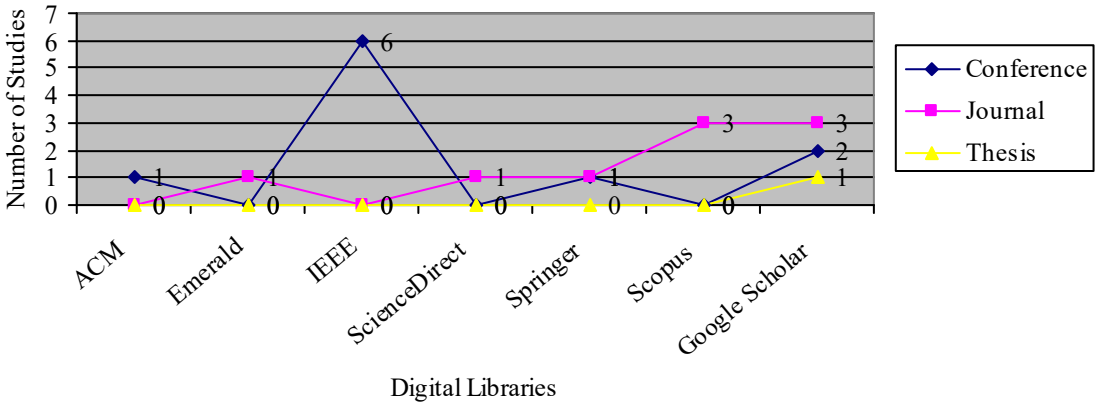

Figure 1: Search Results on Digital Libraries

3.2 Outcomes of Previous Studies

Based on our investigation toward literature, there are several types of outcome from previous studies, namely, develop patterns for ontology population, develop the existing ontologies, develop an ontology in particular domain, develop ontology model, develop a new ontology, patterns testing to extract knowledge, and develop a statistical parser. Figure 2 describes distributions of outcome types from previous research. According to Figure 2, we concluded most researchers from the previous studies focus on developing ontologies in particular domain, such as ontology-related to *Juz' Amma*, prayer, and stories of the prophets.

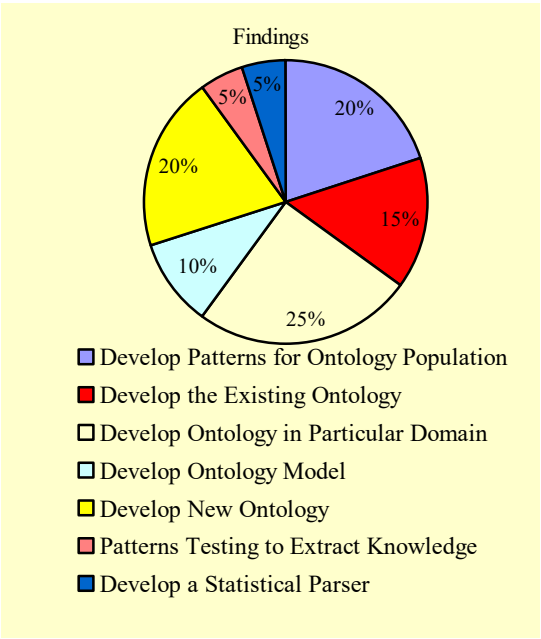

Figure 2: Distributions of Outcome Types from Previous Research

Based on Fig. 2, the coverage area of ontology can be divided into two types, i.e. ontology which covers entire the Quran and ontology which includes the particular domain. Studies conducted by [13–23] covered entire the Quran, while some studies covered the particular Quran domain. Research conducted by [24, 25] developed the ontology-related to faith and deed main themes. Whereas studies by [26, 27] built the ontology-related to *Juz' Amma*. Furthermore, study by [28] developed the ontology-related to the stories of the prophets. Research conducted by [29] created the ontology-related to *salat*. Next, study by [30] developed the ontology-related to living creatures. Then, research conducted by [31] built the ontology- related to Place Nouns, and study by [32] created the ontology-related to Nature domain.

There are several elements to provide a better understanding of the verses, i.e. Tafsir, Hadith, Quran translation, and verses classification according to topics and themes of each verse. Intended meaning of verses could only be acquired by connecting the sentences or words with the central theme [17, 25, 33, 34]. Based on our observation against literature, there are some studies which applied themes classification on their ontologies. Studies conducted by [18, 24, 25, 27] applied themes classification. They adjusted themes classification in their ontologies manually by insert one by one theme which related to each verse into ontologies. Study by [27], they classified themes based on Quran experts. Besides using Quran experts to validate themes on each verse, studies by [24, 25] also used Syammil Al-Quran Miracle the Reference as a source to perform themes classification. Finally, research by [18] used only Tafsir books as a source to perform classification. Table 2 describes findings previous research on

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Quran ontology. Then, Section 3.9 explain deficiencies on previous studies.

Table 2: Findings on Previous Research

Cite	Findings
[28]	Ontology construction using pattern-based schemes and associations rules
[29]	The rules and pattern approach using NLP for extracting the Quranic knowledge
[13]	Statistical dependency-based parser (Hybrid Statistical Parser)
[14]	They have built 650 additional descriptions of the properties into [13] ontology
[24]	Quran ontology based on faith and deed main themes
[30]	Quran ontology-related to living creatures
[26, 27]	Quran ontology for Juz' Amma
[15]	Enhanced version of Hearst's Algorithm [35] to an Arabic corpus
[25]	A model for Quran ontology based on Quran themes and concepts
[16]	Quran Vocabulary (QVOC) ontology
[17]	Quran ontology based on extract relations from four main existing patterns
[31]	Ontology model for Arabic language vocabulary associated with "Place Nouns"
[18]	quranontology.com
[19]	New ontologies based on aligned the four main existing ontologies
[20]	Question Answering framework on Holy Quran
[21]	An ontology learning model based on a hybrid method which combines lexico-syntactic patterns and association rules for an English Quranic translation
[22]	Graph database of English Quran translation
[23]	Indonesian Quranic ontology, which contains 222 concepts
[32]	Quran ontology model related to "Nature" domain

3.3 Most Language Which Used on Ontology Development

Based on our observation against literature, ontology of Holy Quran consists of mono language or multi-languages Quran translation. Figure 3 shows the distribution of monolingual and multilingual Quran translation ontologies in previous studies.

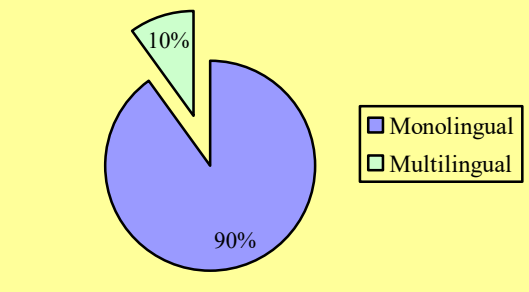

Figure 3: Distribution of Monolingual and Multilingual on Quran Ontologies

Based on Figure 3, research conducted by [16, 26] applied multi-language of Quran translation on their ontologies. Study by [16] applied 43 languages of Quran translation, while study by [26] applied 2 languages on their ontology, i.e. English and Malay. Figure 4 describes the distribution of publications in monolingual Quran translation which has been applied to previous research. Based on Figure 4, we draw a conclusion, Arabic and English Quran translation are the most applied to previous research.

Studies by [13, 15, 18–20, 28, 31] applied the Arabic language on their ontologies, while studies by [14, 17, 21, 22, 29, 30, 32] applied the English Quran translation on their ontologies. Furthermore, research conducted by [24, 25, 27] applied the Malay Quran translation. Finally, study by [23] applied the Indonesian Quran translation on their ontology.

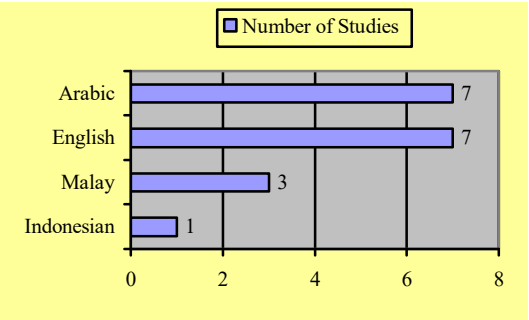

Figure 4: Distribution of Languages Quran Translation

3.4 Datasets Used for Ontology Development

Investigation results of our study towards datasets which has been used to perform ontology development, previous researchers could use two types of datasets to applied, i.e. Private and Public datasets. There are 15 studies used public dataset, and one study used private and hybrid (public and private) dataset.

Based on our observation towards literature, there are two types of data format which are used in

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

previous research, i.e. structured data and unstructured data format [36–38]. Structured data include ontologies, HTML and back-end Deep Web databases, XML and CSV (comma separated value). Whereas unstructured data include free text, web pages, emails, blogs, reports, and news articles. Based on previous research, there are ten studies applied structured data format and seven studies used unstructured data format. Table 3 describes datasets that be used in previous research. Based on data from Table 3, we can draw conclusion Quran ontology from Dukes [13] is the most widely used by researchers to develop their Quran ontology. However, Dukes ontology is not supported by contextual information related to the verses, but more focuses on connecting different concepts or keywords with related verses [17, 26, 27, 31].

Table 3: Datasets on Previous Research

Cite	Dataset	Available
[27–29]	Quran	-
[13]	Quran from http://tanzil.net	Public
[14, 22, 23, 31]	Quran ontology from [13]	Public
[24, 25]	Syammil Al-Quran Miracle the Reference	Public
[30]	English Quranic translation text by Pickthall	Public
[26]	The Qur'an Corpus for *Juz' Amma* from [39]	Private
[15]	Quran, 1000 documents from Arabic newspapers and blogs	Public
[16]	Quran ontology from [13] and 42 manual translations from http://tanzil.net	Public
[17]	English translated Quran from Hilali and Khan	Public
[18]	Quran ontology from [13] and QVOC ontology [16]	Hybrid
[19]	They aligned the four existing ontologies from [13, 16, 34, 40]	Public
[20]	Quran Vocabulary (QVOC) ontology [16]	Public
[21]	English translation text by Yusuf Ali; and Hilali and Khan	Public
[32]	English translated Quran from Sahih International	Public

3.5 Tools to Perform Ontology Development

Investigation results of our study towards tools which has been used to perform ontology development, there are two types of tools which have been used by the previous researcher, i.e. Natural Language Processing (NLP) tools and ontology development tools. NLP tools are used to perform tasks like parse, morphology analysis, and part of speech (POS) tagging. While ontology development tools are used to develop an ontology.

Studies conducted by [14, 18, 20, 22–27, 30–32] used Protege to develop an ontology. Besides using Protege as a tool to build the ontology, some researcher like [16, 18, 20] also used Apache JENA to develop the ontology. Based on our observation towards literature, NLP tools which have been used in previous research, i.e. KP-Miner [28], Shereen Khoja stemmer ([15, 28]), Standford Parser ([15, 29]), Buckwalter Arabic Morphological Analyzer [13], General Architecture for Text Engineering (GATE) [17], Stanford NLP Segmenter and Arabic Toolkit Service [19], Stanford CoreNLP [21]. Table 4 describes tools name and their functions which have been used in previous research.

Table 4: NLP and Ontology Development Tools

Tools Name	Function
KP-Miner	Arabic Natural language processing (NLP) tools
Shereen Khoja stemmer	Arabic stemmer
Stanford Parser	POS tagging
Buckwalter Arabic Morphological Analyzer	Morphology Analysis
Protege	Tools to develop ontology
Apache JENA	Java framework for building Semantic Web and Linked Data applications
General Architecture for Text Engineering	Tools for diverse language processing tasks, e.g. parsers, morphology, tagging, Information Retrieval tools, Information Extraction components for various languages, and many others
Stanford NLP Segmenter	Token pre-processing
Arabic Toolkit Service	Arabic Stemmer
Stanford CoreNLP	Stanford Core NLP integrates many of Stanford's NLP tools, including the part-of-speech (POS) tagger, the named entity recognizer (NER), the parser, the co-reference resolution system, sentiment analysis, bootstrapped pattern learning, and the open information extraction tools

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.6 Approaches Used for Ontology Development on Holy Quran

Based on our observation against literature, two methods used in previous research to perform ontology development are manual (non-automated process) and automated process. This automated process is well-known as ontology population. Ontology population is the process of learning the extensions for concepts, relations, and instances from the Natural Language text, then inserting them into the existing ontology [41–44]. Based on previous research, there are twelve studies applied automated process and eight studies used non-automated process.

Ontology population from Holy Quran by automated process can be done with four approaches, i.e. Rule-based, Natural Language Processing (NLP), statistical, and Hybrid approaches [17, 21, 45–49]. Figure 5 describes a total distribution of utilization four approaches automated process on previous research.

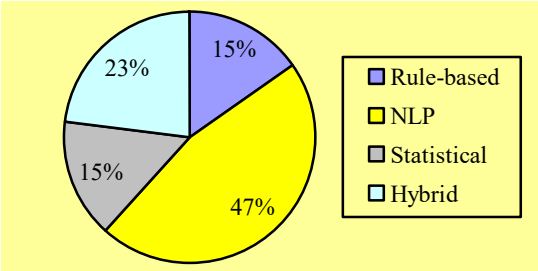

Figure 5: Total Distribution of Automated Process Approaches

Based on Figure 5, there are two studies used Rule-based and statistical approaches, six studies used NLP approaches, and three studies applied hybrid approaches.

Studies conducted by [21, 28] extracted concepts and semantic relations with different approaches. For concepts extraction, [21] extracted noun phrases to obtain the candidate terms. Then, ranked the terms by computing linear combination to get high value which indicates probable concepts. Furthermore, to extract taxonomic relations, they applied patterns from [35, 41, 50]. While, to extract non-taxonomic relation, they used the association rules. A study by [28] extracted the concepts by using *Shereen Khoja* Stemmer. While to obtain relations between concepts, they used *Apriori* algorithm.

For semantic relationships extraction, a study by [14] manually added more relations and restrictions to [13] ontology. While studies by [15, 17, 29] used an automated process. To extract "is-a" relations, they modified patterns in [51] and improved the models for copula by modified copula rule from [41] by adding "are/was/were" keywords. Then, to extract "part-of" relations, they formulated three rule-based patterns to obtain it. Furthermore, a study by [15] performed semantic relations extraction with several steps. First, they performed lexical-syntactic patterns. Second, expanded the pattern by using Arabic *WordNet* to check the synonym and hyponym on the pattern. Finally, they performed pattern filtering and aggregation by applied the coverage metric. The last one, research conducted by [17] used the patterns from [50, 52–54] to extract taxonomic and non-taxonomic relations.

To extract concepts, semantic relations, properties, and terms, studies conducted by [13, 19] used different approaches. To extract it, [13] used Hybrid Statistical Parser (HSP). During training HSP used Support Vector Machine (SVM) and LIBSVM algorithm. Meanwhile, a study by [19] obtained concepts, semantic relations, properties, and terms extraction by integrated four main existing ontologies.

Research conducted by [24, 25, 27] performed concepts, properties, semantic relations, Named Entity extraction manually. Meanwhile, studies conducted by [20, 30] extracted the concepts and properties. They [30] performed extraction by manually added it into their ontology. Furthermore, a study by [20] performed concepts and properties extraction into their ontology by several steps. They applied normalization, stemming, synonyms generation, synonyms validation, and input the text into Gazetteer using Apache JENA.

To extract concepts, properties, semantic relations, studies by [26, 32] added it manually into their ontologies. While research by [22] added concepts, properties, semantic relations by an automated process. To determine the concepts and properties, they used POS tagging. Whereas to extract semantic relations, they applied Graph-based of English Quran translation.

To extract terms, concepts, semantic relations, properties, and Named Entity, a study by [18] added it manually into their ontologies. They performed extraction by using Apache JENA. Meanwhile, studies conducted by [16, 23, 31] focused on single extraction by using automated process. A study by [16] performed properties extraction. They used Apache JENA to extract it into the ontology. While [31] focused on terms (nouns) extraction. Finally, research by [23] focused on concept extraction. They determined concepts by applied TF-IDF for term weighting,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

vector space model and the cosine similarity measure for relevance score calculation.

3.7 Integration the Knowledge of Quran and Other resources into Ontology

The Relationships in the ontology could be classified into two categories, i.e. Taxonomic and non-Taxonomic [51, 55]. Taxonomic relations consists of "is-a" (class inclusion) and "part-of" (subset hierarchy) relationship [51, 56]. While a non-taxonomic focus on a relationship between concepts (related verbs) and label relationships [57, 58]. Table 5 describes common linguistic and semantic relations in an ontology [59–62].

Table 5: Typical Linguistic and Semantic Relations

Linguistic Relation	Example	Correspondence Type
Synonymy	Beautiful, Pretty	Equal, means
Antonymy	Cold, Hot	Mismatch, unlike, disjoint
Hyponymy	Banana, Fruit	is-a, is-a-kind-of, kind-of
Hypernymy	Vehicle, Car	inverse-is-a
Meronymy	Tire, Car	part-of, is-a-part-of
Holonymy	Tree, Leaf	has-a

Based on Table 5, Hyponymy, Hypernymy, and Meronymy are part of Taxonomic relations. While Synonymy, Antonymy, and Holonymy are part of non-taxonomic relationships. "is-a" and "inverse is-a" are relations between classes or instances to classes. While "part-of" is relation between classes. "has-a" is relation between class with the values in properties of that class. Finally, "equal" and "disjoint" are relations which used to relate class to class, class to property, property to another property, and term to another term.

Table 6 shows the results of our observation towards on previous research, number of concepts and relations, and relationship types in the ontology.

Table 6: Ontologies on Previous Research

Cite	Concepts	Relations	Relations Type
[28]	N/A	N/A	Part-of, equal, is-a
[29]	N/A	N/A	Is-a, part-of
[13]	300	350	Is-a, part-of, equal
[14]	300	650	Is-a, part-of, equal
[24]	11	26	Is-a, part-of
[30]	N/A	N/A	N/A
[26]	5	11	Is-a, has-a, part-of
[27]	36	N/A	Is-a, part-of
[15]	N/A	N/A	Is-a, inverse-is-a, has-a, part-of
[25]	11	26	Is-a, part-of
[16]	4	6	Part-of
[17]	N/A	N/A	Equal, part-of
[31]	9	58	Is-a, has-a, part-of, equal
[18]	14	N/A	Has-a, part-of, equal
[19]	N/A	N/A	Is-a, equal, part-of, has-a
[20]	45	N/A	Part-of, is-a
[21]	N/A	N/A	Part-of, subject- verb-object relation
[22]	300	350	Is-a, part-of, equal
[23]	222	N/A	Is-a, part-of, equal
[32]	N/A	N/A	Part-of, has-a

Based on our observation, studies by [14, 26, 27], they associate the verses with Tafsir by using data property and object property. "Has-a" is a relation which is used to relate a verse with the Tafsir as a value in properties. Figure 6 describes an example of relationships on Quran ontology. As shown in Figure 6, there are some example relationships on Quran ontology, i.e. "is-a" (*Hyponymy*), "part-of" (*Meronymy*), "equal" (*Synonymy*), and "has-a" (*Holonymy*) relations.

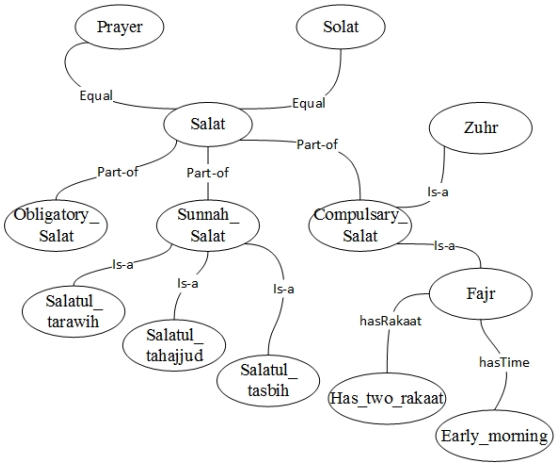

Figure 6: An example of relationships on Quran ontology

3.8 Evaluation Techniques to Test the Ontology

Based on our observation against literature, there are four focus areas to test the performance related to Quran ontology which is used in previous research, i.e. overall ontology performance using precision and recall ([14, 20]) and Descriptive Logic Queries ([18, 26, 27, 30–32]); ontology consistency using *HermiT* 1.3.8 *Reasoner* tool by *Protege* ([25]); themes validation by Quran experts ([24, 25]); pattern extraction evaluation using precision, recall, and F-measure ([15, 21]), Performance of Extraction Metrics ([17]), and True positive, false positive, false negative ([29]).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The evaluation which conducted by [29] used 130 taxonomic relations as a data test. Results from their assessment show that *Cimiano++* has a better result than the proposed approach (*QPattern*) to perform a taxonomic extraction. However, *QPattern* has a better result in False Positive and False Negative value rather than Hearst [35], Cimiano [41], and *Cimiano++* methods. The overall score for *QPattern* is True positive 13.8%; False-positive 1.5%, and False negative 2.3%. Similar to [29], research conducted by [17, 21] also developed patterns to extract semantic relations from English Quran translation. To perform an evaluation, [17] used chapter *Al-Maarij*, *Al-Jinn*, *Nuh*, *Al-Muzammil* and *Al-Muddathir*. These chapters contain 176 verses (3043 words). They used Performance of Extraction (PoE) metrics to measure patterns performance. To extract "part-of" relationships, their pattern has PoE score 23.91%, while for synonym extraction, their pattern has PoE score 73.91%. Meanwhile to perform an evaluation, [21] used two datasets, i.e. The English Quranic translation text by Hilali-Khan and The English Quranic translation text by Yusuf Ali. One hundred Quran verses are taken on each dataset. They used Precision, Recall, and F-measure metrics to quantify patterns performance. Evaluation results on concepts extraction by TF-IDF, TIM, and Linear Combination (LC) shows that LC has the highest precision score with 90% on English Quranic translation text by Hilali-Khan and 93% on English Quranic translation text by Yusuf Ali. Then, the performance of taxonomic and non-taxonomic relations extraction in each dataset has different results. The precision score for taxonomic relations on English Quranic translation text by Yusuf Ali is 85.8%, while non-taxonomic is 84.34%. A decrease in precision score occurs in English Quranic translation text by Hilali-Khan dataset. The precision score for taxonomic relations is 83.3%, while non-taxonomic is 81.34%.

Patterns extraction evaluation on Arabic Quranic was conducted by [15]. They used Classical Arabic (Quran), Modern Standard Arabic (newspapers), and unstructured Arabic texts (blogs) as a dataset. Precision, recall, and F-measure are used to measure extraction performance. Evaluation results show that extraction performance for newspaper dataset has the highest Precision with 89.77%, while on Quran dataset has precision 76.28%. The lack of ambiguity on the newspapers is a factor that causes newspapers has the highest accuracy rather than Quran.

3.9 Limitations on Previous research

There are two approaches which able to comprehend the Islamic values contained in the Quran, namely the literal or textual, and contextual approach. Figure 7 describes a total distribution of literal and contextual approaches which is used in previous studies. According to Figure 7, we can conclude literal approach on Quran ontology development is the most widely used in previous studies.

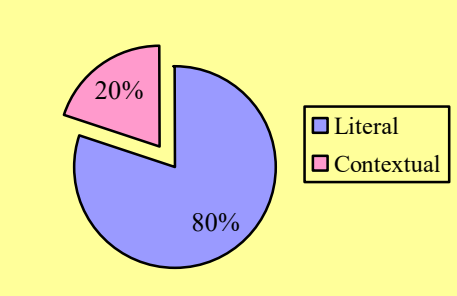

Figure 7: Total Distribution of literal and contextual approaches

Based on Figure 7, there are 20% studies which used the contextual approach in previous research. These studies used several languages, i.e. Arabic, English, and Malay. Each of them has one number of studies. Meanwhile, one study used Multilingual Quran translation, i.e. Malay and English. Ontology on Holy Quran should save all verses and their contextual meaning to assist human being to get better insight and better understanding content of Quran and to avoid misinterpretation.

According to [17, 21, 29] results of performance evaluation on pattern extraction for English Quran translation, we can conclude each rule and pattern could be applied only to specific language. Since every language has different written form, grammar, vocabulary, and syntax [63–65]. Furthermore, each rule has different evaluation results in a diverse dataset with similar language.

Ontology development for Indonesian Quran translation has conducted by [23]. They applied [13] ontology to build Indonesian Translation of Quran (ITQ). To convert English into Indonesian Quran translation from [13] ontology, they used BING Application Programming Interface (API). Refinement towards results of translation from English to Indonesian has done against incorrect translation result terms due to lexical difference and single quote missing. However, improvement against grammatical and structural sentences has not been made yet. Quran translation text should be in a proper format to provide precise information and can be understood by the users.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. OPEN RESEARCH ISSUES

There are open research issues that can be highlighted for Quran ontology. These issues are availability of Quran ontology in various translation, ontology resources, automated process of *meronymy* relationship extraction, and Instances Classification.

A. Availability of Quran Ontology in Various Translation

Based on our observation against literature, there are four languages of Quran translation on previous studies, namely Indonesian, Malay, English, and Arabic. To provide better insight and better understanding content of Quran, ontology on other Quran translation for other *muslim* countries should be built.

B. Ontology Resources

There are several factors which influence information quality from ontology to provide best answers to the users. One of them is the collection of ontology resources. This collection could be *Tafsir*, Hadith, and revelation reasons. Based on our observation, there are 3 studies stored the contextual meaning into an ontology. These studies used several languages, i.e. Arabic, English, and Malay. To assist human being to get better insight and better understanding content of Quran and to avoid misinterpretation, the ontology in other Quran translation should keep the contextual meaning.

C. Automated Process of Meronymy Relationship Extraction

Meronymy or "part-of" is a relation between classes on ontology. Studies conducted by [17, 29] proposed patterns to perform "part-of" extraction on English Quran translation. Based on the evaluation against [29] patterns, their models has True Positive with 13.8%. Meanwhile, [17] proposed models have PoE with 23.91%. These results have led to an issue to develop or improve patterns for extract "part-of" relation. However, rule and pattern could be applied only to specific language. Since every language has different written form, grammar, vocabulary, and syntax.

D. Instances Classification

"is-a" or "Instance-of" is relation between class and their instances. Based on our observation against literature, *instances classification* on previous research is used for verses classification. Previous researchers performed verses classification based on themes or thematic topics in Holy Quran. Verses classification based on thematic topics purpose is to connect the sentences (verses) with the central theme to get a whole picture of the topic and for providing a better comprehension to the users [17, 25, 33]. Furthermore, the primary goal of text (verses) classification is to reduce the searching space by identifying the passages of information that are relevant to the particular topic [66-68]. A study by [18] used seven classes for verses classification on Holy Quran. These classes based on *Tafsir* books. While a study by [27] used 36 classes based on Quran experts. Research conducted by [24, 25] used Syammil Al-Quran Miracle the Reference and Quran experts to perform verses classification. To perform verses classification, they conducted with the non-automated process.

A different approach has conducted by [66] to perform verses classification. They used Surah *Al-Baqarah* from English Quran translation by Abdullah Yusuf Ali as a dataset. This research has used the WEKA toolkit [69] to implement the classification of Surah *Al-Baqarah* based on Back-propagation Neural Network (BPNN) classifier. They classified Surah *Al-Baqarah* into three classes, Fasting, Pilgrimage, and None class that are not related to any of these mentioned classes. Their training set consists of 150 instances of verses. Where 80% of them is used for the data set, while 20% of the remaining data will be utilized for the testing set. Each class has 50 examples as an instances. Before data is processed by BPNN. Firstly, verses is pre-processed by several tasks, such as stop word removal, stemming and Term Frequency (TF) transformation. Subsequently, BPNN compares the weights of all feature sets that are extracted from the Quranic verses with the weights of predefined classes based on the training set to determine each verse to its class. Based on the evaluation, BPNN classifier has F-Score value with 90%.

Verses classification by thematic topics is a part of text classification study. Term Weighting is the basis issue on text classification which affects the level of accuracy of the classification results [70-73]. According to [70, 74, 75], TF-IDF is not an efficient algorithm for text classification, due to TF-IDF ignores the class labeling on training document. Furthermore, Supervised learning like Support Vector Machine (SVM) and Artificial Neural Network (ANN), usually requires a large training corpus to learn a classifier that performs well [76]. Shortcomings from Supervised learning,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

if the dataset sized is small, then the accuracy of the classifier may decline [8, 75, 76]. Holy Quran has a small data set. It has 6236 verses. Research result by [77] showed that SVM classifier accuracy is weak in a small dataset. Challenge in Instances Identification is what technique could be used in a large themes class (class labeling) with small training corpus for high classification accuracy.

5. CONCLUSION

Different from reviewing Quran ontology development by [33] based on parameters such as the outcomes of previous studies, dataset, tools, and Quran ontology development approaches. In this study, we comprehensively discussing about the outcomes of previous studies, languages commonly used in Quran ontology, tools which often used in ontology development, datasets which are used for ontology, the existing studies perform ontology development on Holy Quran, the existing ontologies integrate the knowledge of Quran and other resources, the evaluation techniques for the existing ontologies, limitations on previous research, and the research gaps in the current state of existing ontology on Holy Quran. As a result, many research opportunities are still available along this line and further investigations for Quran ontology in availability of Quran ontology in various translation, ontology resources, automated process of *meronymy* relationship extraction, and Instances Classification. Considering the research issues discussed in this paper, researchers could find the direction for Quran ontology designing methodology, the best possible approaches for meronymy relationship extraction and instances classification that have not been discussed in this paper.

ACKNOWLEDGEMENT

The authors would like to thank the financial support from STMIK AMIKOM Purwokerto and Universiti Teknikal Malaysia Melaka (UTEM) for their assistance in this research.

REFERENCES

[1] D. H. Ngo and Z. Bellahsene, "Overview of YAM++-(not) Yet Another Matcher for ontology alignment task," *Journal of Web Semantics*, vol. 41, 2016, pp. 30–49.

[2] M. C. Suárez-Figueroa, A. Gómez-Pérez, and M. Fernández-López, "Scheduling ontology development projects," *Data and Knowledge Engineering*, vol. 102, 2016, pp. 1–21.

[3] L. Yue, W. Zuo, T. Peng, Y. Wang, and X. Han, "A fuzzy document clustering approach based on domain-specified ontology," *Data and Knowledge Engineering*, vol. 100, 2015, pp. 148–166.

[4] M. Gulić, B. Vrdoljak, and M. Banek, "CroMatcher: An ontology matching system based on automated weighted aggregation and iterative final alignment," *Journal of Web Semantics*, vol. 41, 2016, pp. 50–71.

[5] C. Diamantini, A. Freddi, S. Longhi, D. Potena, and E. Storti, "A goal-oriented, ontology-based methodology to support the design of AAL environments," *Expert Systems with Applications*, vol. 64, 2016, pp. 117–131.

[6] M. del M. R. García, J. García-Nieto, and J. F. Aldana-Montes, "An ontology-based data integration approach for web analytics in e-commerce," *Expert Systems with Applications*, vol. 63, 2016, pp. 20–34.

[7] A. R. Yauri, R. A. Kadir, A. Azman, M. Azrifah, and A. Murad, "Ontology Semantic Approach to Extraction of knowledge from Holy Quran," in *2013 5th International Conference on Computer Science and Information Technology*, 2013, pp. 19–23.

[8] A. H. Asiaee, P. Doshi, T. Minning, and R. L. Tarleton, "A framework for ontology-based question answering with application to parasite immunology," *Journal of biomedical semantics*, vol. 6, 2015, pp. 1–25.

[9] A. Sayed and A. Al Muqrishi, "An efficient and scalable Arabic semantic search engine based on a domain specific ontology and question answering," *International Journal of Web Information Systems*, vol. 12, no. 2, 2012, pp. 242–262.

[10] H. Raviv, O. Kurland, and D. Carmel, "Document Retrieval Using Entity-Based Language Models," in *SIGIR '16 Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, 2016, pp. 65–74.

[11] F. Lashkari, F. Ensan, E. Bagheri, and A. A. Ghorbani, "Efficient Indexing for Semantic Search," *Expert Systems With Applications*, vol. 73, 2017, pp. 92–114.

[12] A. Hinze, C. Taube-Schock, D. Bainbridge, R. Matamua, and J. S. Downie, "Improving Access to Large-scale Digital Libraries

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Through Semantic-enhanced Search and Disambiguation," in *Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries Pages*, 2015, pp. 147–156.

[13] K. Dukes, "Statistical Parsing by Machine Learning from a Classical Arabic Treebank," Ph.D, University of Leeds, Leeds, 2013.

[14] A. R. Yauri, R. A. Kadir, A. Azman, and M. A. A. Murad, "Quranic verse extraction base on concepts using OWL-DL ontology," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 6, no. 23, 2013, pp. 4492–4498.

[15] M. G. H. Al Zamil and Q. Al-Radaideh, "Automatic extraction of ontological relations from Arabic text," *Journal of King Saud University-Computer and Information Sciences*, vol. 26, no. 4, 2014, pp. 462–472.

[16] M. A. Sherif and A. C. Ngonga Ngomo, "Semantic Quran A Multilingual Resource for Natural - Language Processing," *Semantic Web*, vol. 6, no. 4, 2015, pp. 339–345.

[17] R. Ismail, Z. Abu Bakar, and N. Abd Rahman, "Extracting knowledge from english translated quran using NLP pattern," *Jurnal Teknologi*, vol. 77, no. 19, 2015, pp. 67–73.

[18] A. Hakkoum and S. Raghay, "Advanced Search in the Qur'an using Semantic modeling," in *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, 2015, pp. 1–4.

[19] M. Alqahtani and E. Atwell, "Arabic Quranic Search Tool Based on Ontology," in *21st International Conference on Applications of Natural Language to Information Systems, NLDB*, 2016, vol. 9612, pp. 478–485.

[20] A. Hakkoum and S. Raghay, "Semantic Q&A System on the Qur'an," *Arabian Journal for Science and Engineering*, vol. 41, no. 12, 2016, pp. 5205–5214.

[21] T. Weaam and S. Saad, "Ontology Population from Quranic Translation Texts Based on a Combination of Linguistic Patterns and Association Rules," *Journal of Theoretical and Applied Information Technology*, vol. 86, no. 2, 2016, pp. 250–257.

[22] M. . Noordin, Sembok, T. M. Tengku, R. Othmana, and R. . Gusmita, "Constructing An Ontology-Based and Graph-Based Knowledge Representation of English Quran," *Jurnal Teknologi*, vol. 78, no. 8–2, 2016, pp. 35–41.

[23] S. J. Putra, K. Hulliyah, N. Hakiem, R. P. Iswara, and A. F. Firmansyah, "Generating Weighted Vector for Concepts in Indonesian Translation of Quran," in *iiWAS '16 Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services*, 2016, pp. 293–297.

[24] A. Ta'a, S. Z. Abidin, M. S. Abdullah, A. B. B. Ali, and M. Ahmad, "Al-Quran Themes Classification Using Ontology," in *Proceedings of the 4th International Conference on Computing and Informatics, ICOCI 2013*, 2013, no. 74, pp. 383–389.

[25] A. Ta'A, M. S. Abdullah, A. B. M. Ali, and M. Ahmad, "Themes-based classification for Al-Quran knowledge ontology," in *International Conference on ICT Convergence*, 2014, pp. 89–94.

[26] R. Iqbal, A. Mustapha, and Z. M. Yusoff, "An experience of developing Quran ontology with contextual information support," *Multicultural Education & Technology Journal*, vol. 7, no. 4, 2013, pp. 333–343.

[27] N. S. H. A. Periamalai, A. Mustapha, and A. Alqurneh, "An Ontology for Juz'Amma based on Expert Knowledge," in *7th International Conference on Computer Science and Information Technology (CSIT)*, 2016, pp. 1–5.

[28] F. Harrag, A. Al-nasser, A. Al-musnad, R. Al-shaya, and A. S. Al-salman, "Using association rules for ontology extraction from a Quran corpus," in *5th International Conference on Arabic Language Processing*, 2014.

[29] S. Saad, S. A. M. Noah, N. Salim, and H. Zainal, "Rules and Natural Language Pattern in Extracting Quranic Knowledge," in *Proceedings - 2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences, NOORIC 2013*, 2013, pp. 381–386.

[30] H. U. Khan, S. M. Saqlain, M. Shoaib, and M. Sher, "Ontology Based Semantic Search in Holy Quran," *International Journal of Future Computer and Communication*, vol. 2, no. 6, 2013, pp. 570–575.

[31] W. Alromima, I. F. Moawad, R. Elgohary,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and M. Aref, "Ontology-based model for Arabic lexicons: An application of the Place Nouns in the Holy Quran," in *11th International Computer Engineering Conference (ICENCO)*, 2015, pp. 137–143.

[32] A. B. M. S. Sadi *et al.*, "Applying Ontological Modeling on Quranic «nature» Domain," in *7th International Conference on Information and Communication Systems (ICICS)*, 2016, pp. 151–155.

[33] N. K. Farooqui and M. F. Noordin, "Knowledge exploration: Selected works on Quran ontology development," *Journal of Theoretical and Applied Information Technology*, vol. 72, no. 3, 2015, pp. 385–393.

[34] N. Abbas, "Quran'search for a Concept'Tool and Website," MSc, The University of Leeds, 2009.

[35] M. A. Hearst, "Automatic Acquisition of Hyponyms from Large Text Corpora," in *International Conference on Computational Linguistics*, 1992, pp. 539–545.

[36] A. Fong, A. Z. Hettinger, and R. M. Ratwani, "Exploring methods for identifying related patient safety events using structured and unstructured data," *Journal of Biomedical Informatics*, vol. 58, 2015, pp. 89–95.

[37] A. McCallum, "Information Extraction: Distilling Structured Data from Unstructured Text," *Queue-Social Computing*, vol. 3, no. 9, 2005, pp. 48–57.

[38] M. J. Cafarella, J. Madhavan, and A. Halevy, "Web-Scale Extraction of Structured Data," *ACM SIGMOD Record*, vol. 37, no. 4, 2008, pp. 55–61.

[39] A. Mustapha, Z. M. Yusoff, and R.-J.-R. Yusof, "The Qur'an Corpus for Juz' Amma," in *LREC'2012 Workshop: LRE-Rel Language Resources and Evaluation for Religious Texts*, 2012, no. May, pp. 54–57.

[40] M. Sharaf and E. S. Atwell, "QurAna: Corpus of the Quran annotated with Pronominal Anaphora.," in *8th International Conference on Language Resources and Evaluation (LREC'12)*, 2012, pp. 130–137.

[41] P. Cimiano, *Ontology Learning and Population from Text*, 1st ed. New York, NY: Springer US, 2006.

[42] R. Witte, R. Krestel, T. Kappler, and P. C. Lockemann, "Converting a Historical Architecture Encyclopedia into a Semantic Knowledge Base," *IEEE Intelligent Systems*, vol. 25, no. 1, 2010, pp. 58–67.

[43] J. A. Reyes and A. Montes, "Learning Discourse Relations from News Reports: An Event-driven Approach," *IEEE Latin America Transactions*, vol. 14, no. 1, 2016, pp. 356–363.

[44] J. A. Reyes-ortiz, M. Bravo, and H. Pablo, "Web Services Ontology Population through Text Classification," in *2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2016, vol. 8, pp. 491–495.

[45] M. Gaeta, F. Orciuoli, S. Paolozzi, and S. Salerno, "Ontology Extraction for Knowledge Reuse: The e-Learning Perspective," *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, vol. 41, no. 4, 2011, pp. 798–809.

[46] S. Mirhosseini, G. Zuccon, B. Koopman, A. Nguyen, and M. Lawley, "Medical Free-Text to Concept Mapping as an Information Retrieval Problem," in *Proceedings of the 2014 Australasian Document Computing Symposium*, 2014, pp. 93–96.

[47] W. Wong, W. Liu, and M. Bennamoun, "Ontology learning from text: A look back and into the future," *ACM Computing Surveys*, vol. 44, no. 4, 2012, pp. 1–36.

[48] R. Shah and S. Jain, "Ontology-based Information Extraction: An Overview and a Study of different Approaches," *International Journal of Computer Applications*, vol. 87, no. 4, 2014, pp. 6–8.

[49] O. Ahmad *et al.*, "A Survey of Searching and Information Extraction on a Classical Text Using Ontology-based semantics modeling: A Case of Quran," *Life Science Journal*, vol. 10, no. 4, 2013, pp. 1370–1377.

[50] S. Saad, "Ontology learning and population techniques for English extended quranic translation text," Ph.D, Universiti Teknologi Malaysia, 2014.

[51] C. Y. Yong, R. Sudirman, K. M. Chew, and N. Salim, "Comparison of Ontology Learning Techniques for Qur'anic Text," in *Proceedings - 2011 International Conference on Future Computer Sciences and Application, ICFCSA 2011*, 2011, pp. 192–196.

[52] S. Saad, N. Salim, and H. Zainal, "Towards Context-Sensitive Domain of Islamic Knowledge Ontology Extraction," *Int. J.*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Infonomics, vol. 3, no. 1, 2010, pp. 197–206.

[53] S. Saad, N. Salim, and H. Zainal, "Pattern Extraction For Islamic Concept," in *2009 International Conference on Electrical Engineering and Informatics*, 2009, no. August, pp. 333–337.

[54] S. Saad, N. Salim, and H. Zainal, "Islamic Knowledge Ontology Creation," in *2009 International Conference for Internet Technology and Secured Transactions, (ICITST)*, 2009, pp. 1–6.

[55] A. Doan, J. Madhavan, R. Dhamankar, P. Domingos, and A. Halevy, "Learning to match ontologies on the Semantic Web," *VLDB J. — The International Journal on Very Large Data Bases*, vol. 12, no. 4, 2003, pp. 303–319.

[56] C. Garcia-Alvarado and C. Ordonez, "Query processing on cubes mapped from ontologies to dimension hierarchies," in *Proceedings of the fifteenth international workshop on Data warehousing and OLAP*, 2012, pp. 57–64.

[57] C. Welty and N. Guarino, "Supporting ontological analysis of taxonomic relationships," *Data and Knowledge Engineering*, vol. 39, no. 1, 2001, pp. 51–74.

[58] M. Kavalec, A. Maedche, and V. Svátek, "Discovery of Lexical Entries for Non-taxonomic Relations in Ontology Learning," in *International Conference on Current Trends in Theory and Practice of Computer Science*, 2004, pp. 249–256.

[59] P. Arnold and E. Rahm, "Enriching ontology mappings with semantic relations," *Data and Knowledge Engineering*, vol. 93, 2014, pp. 1–18.

[60] A. Benabdallah, M. A. Abderrahim, and M. E.-A. Abderrahim, "Extraction of terms and semantic relationships from Arabic texts for automatic construction of an ontology," *International Journal of Speech Technology*, 2017, pp. 1–8.

[61] R. Henriksson, T. Kauppinen, and E. Hyvönen, "Core Geographical Concepts: Case Finnish Geo-Ontology," in *Proceedings of the first international workshop on Location and the web*, 2008, pp. 57–60.

[62] A. Kleshchev and E. Shalfeyeva, "Methodology of organizing the catalogue of ontology properties," *Automatic Documentation and Mathematical Linguistics*, vol. 41, no. 3, 2007, pp. 114–123.

[63] B. Andi-pallawa, "A Comparative Analysis between English and Indonesian Phonological Systems," *International Journal of English Language Education*, vol. 1, no. 3, 2013, pp. 103–129.

[64] B. R. Chiswick and P. W. Miller, "Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages," *Journal of Multilingual and Multicultural Development*, vol. 26, 2005, pp. 1–18.

[65] A. U. Rahayu, "Differences on Language Structure between English and Indonesian," *International Journal of Languages, Literature and Linguistics*, vol. 1, no. 4, 2015, pp. 257–260.

[66] S. K. Hamed and M. J. A. Aziz, "A question answering system on Holy Quran translation based on question expansion technique and Neural Network classification," *Journal of Computer Science*, vol. 12, no. 3, 2016, pp. 169–177.

[67] H. J. Oh, S. H. Myaeng, and M. G. Jang, "Semantic passage segmentation based on sentence topics for question answering," *Information Sciences*, vol. 177, no. 18, 2007, pp. 3696–3717.

[68] B. Baharudin, L. H. Lee, K. Khan, and A. Khan, "A Review of Machine Learning Algorithms for Text-Documents Classification," *Journal of Advances in Information Technology*, vol. 1, no. 1, 2010, pp. 4–20.

[69] G. Holmes, A. Donkin, and I. H. Witten, "WEKA: a machine learning workbench," in *Proceedings of ANZIIS '94 - Australian New Zealnd Intelligent Information Systems Conference*, 1994, pp. 357–361.

[70] K. Chen, Z. Zhang, J. Long, and H. Zhang, "Turning from TF-IDF to TF-IGM for term weighting in text classification," *Expert Systems with Applications*, vol. 66, 2016, pp. 245–260.

[71] T. Sabbah, A. Selamat, M. H. Selamat, R. Ibrahim, and H. Fujita, "Hybridized term-weighting method for Dark Web classification," *Neurocomputing*, vol. 173, 2016, pp. 1908–1926.

[72] M. A. Fattah, "New term weighting schemes with combination of multiple classifiers for sentiment analysis," *Neurocomputing*, vol. 167, 2015, pp. 434–442.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[73] H. J. Escalante *et al.*, "Term-weighting learning via genetic programming for text classification," *Knowledge-Based Systems*, vol. 83, no. 1, 2015, pp. 176–189.

[74] T. Peng, L. Liu, and W. Zuo, "PU text classification enhanced by term frequency–inverse document frequency-improved weighting," *Concurrency Computation Practice and Experience*, vol. 26, no. 3, 2013, pp. 728–741.

[75] Z. H. Deng, K. H. Luo, and H. L. Yu, "A study of supervised term weighting scheme for sentiment analysis," *Expert Systems with Applications*, vol. 41, no. 7, 2014, pp. 3506–3513.

[76] K. Zhang and J. Zhao, "A Chinese question-answering system with question classification and answer clustering," in *Proceedings - 2010 7th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 2010, vol. 6, pp. 2692–2696.

[77] D. Tomás and J. L. Vicedo, "Minimally supervised question classification on fine-grained taxonomies," *Knowledge and Information Systems*, vol. 36, no. 2, 2013, pp. 303–334.