Jurnal Teknologi

Full Paper

EXTRACTING KNOWLEDGE FROM ENGLISH TRANSLATED QURAN USING NLP PATTERN

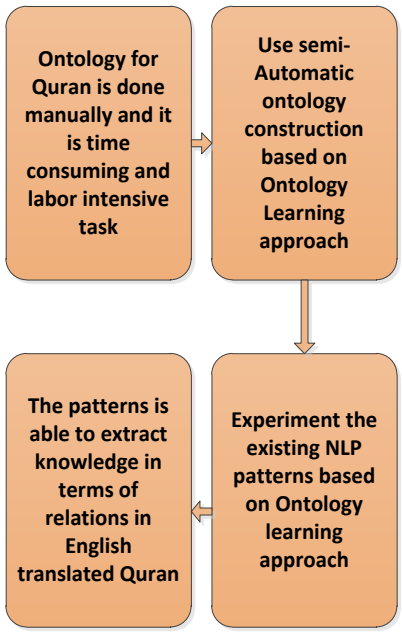
Rohana Ismail[a,b*], Zainab Abu Bakar[a], Nurazzah Abd. Rahman[a]

[a]Faculty of Computer and Mathematical Sciences, Universiti Teknologi Mara, Shah Alam, Selangor, Malaysia
[b]Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Kuala Terengganu, Terengganu, Malaysia

Article history
Received
15 May 2015
Received in revised form
1 July 2015
Accepted
11 August 2015

*Corresponding author
rohana@unisza.edu.my

Graphical abstract



Abstract

Ontology is able to represent knowledge from an abstract view into formal semantics. It is essential for the success of knowledge-based systems because it has been used to share vocabulary, discover new knowledge, flexible access of knowledge and easy integration of knowledge. Currently, Ontology from Quran is not complete and most of the development is done manually. Manual development of ontology is time consuming and labor intensive task. Hence, the automatic or semi-automatic ontology development which is a field of Ontology Learning is needed to efficiently extract knowledge and transform it into Ontology. Current techniques employed in Ontology Learning are based on statistical and Natural Language Processing. This paper provides result from an experiment to extract knowledge using the existing Natural Language Processing (NLP) Pattern based on the Ontology Learning approach. Initial experiment shows that the pattern could be used to extract knowledge in terms of relations that exist in English translated Quran. In addition, NLP could also use to identify new pattern that can be further explored.

Keywords: Ontology, NLP pattern, Quran

1.0 INTRODUCTION

There has been a growing demand to understand the knowledge from Quran since more than a quarter of the world populations adheres to the Islamic faith and this proportion is expected to increase[1]. Many people cannot read Arabic and may use Quran translated version such as English, Urdu and Malay. Translation help readers to understand but reading the translation of Quran chapter by chapter still cannot give clear understanding about certain topics or themes. Users need to relate a topic with few other verses that can be found in different chapters to get a whole picture of the topic [2]. This is due to the scattered knowledge in the Quran [3], [4] which are the main uniqueness of the Quran. Through the availability of Semantic Web technology, the scattered knowledge in Quran can be represented as ontology. Ontology is a way of representing knowledge in an abstract model for a certain domain with a formal semantics [5]. The ontology can be represented as a network of interconnected Resource Description Framework (RDF) triples [6].

At present, few ontologies has been developed for Quran. One of the ontology is the ontology that has been developed based on the Arabic word "time" in the Quran [7]. The purpose of the ontology is to

develop a computational model for representing the semantic meaning of an Arabic word using semantic field theory and componential analysis. The research focus is the componential analysis of the characteristics of the words. A classification of general concepts was generated and the established model was mapped to the structure of the ontology. This ontology contains 18 classes which 7 classes are wide and common to every field of semantics and 11 classes are specific to the semantic field of "Time".

Another Ontology is the Quran Ontology that has been developed by the Arabic Language Computing Research at the University of Leeds [8]. It is a comprehensive ontology for Quran. However the ontology contains insufficient concepts. For example Hajj is just a part of a Calendar event and part part of an Event, but not a part of other concepts that might be related to Hajj. The ontology used Name Entity Recognition (NER) technique as a learning approach to extract concepts and instances in Quran. Named entities in verses, such as the names of historic people and places mentioned in the Quran, are linked to concepts in the ontology as part of named entity tagging. The graph is a network of 300 linked concepts with 350 relations.

The development of an ontology is important as a platform to represent, store, integrate and access this knowledge. Furthermore it could be used to facilitates inferences[9]. However, to extract knowledge and transform the knowledge into ontology presents a challenge in ontology development [10]. Efficient knowledge extraction requires proper handling the preprocessing task, identification of concepts or key entities, identifications of relationship between entities or concepts and performing co-reference resolution to merge information about the same entity [11].

There are few approaches that can be used to extract knowledge from natural language text which is the Ontology Learning approach [12], Information Extraction approach [13], Ontology Based Information Extraction (OBIE) approach [14] and Semantic Annotation [15]. However, different approaches learn different elements to extract and produce different results.

The Ontology Learning approach attempts to extract concepts and relations from text. Five types of output in Ontology Learning are terms, concepts, taxonomy relations, non-taxonomy relations, and axioms [16], [17]. Previous survey revealed that current research in Ontology Learning are either at the stage of enhancing terms extractions or discovering complex discovery [17]. Although the term extraction is more or less stabilized with the performance above 90% in F-score [17] but this does not happen in Quran text translation. So far, research to discover the term extraction from Quran text translation performs 77% recall after enhancing the Natural Language Processing (NLP) pattern and ranking using statistical analysis [3].

This paper addresses the lack of current ontology development for Quran. This paper also discusses the needs for semi-automatic approach in ontology development. Part of this study is to evaluate existing NLP patterns based on Ontology Learning approach. An experiment has been conducted using three existing patterns.

2.0 RELATED WORK

To date, many ontology have been developed for Quran focusing on certain themes or domains [3], [7], [18]. Ontology for Quran as a whole, covering all domains in Quran is still not complete. For example, the ontology for 'solat' only covers 10% from the entire Quran[3]. Ontology for representing semantic lexicons of Arabic about 'time' only focus on verses related to 'time'[7]. Another ontology, covers only verses in *Juz' Amma* [18]. Only ontology from University of Leeds [8] cover the entire Quran. Although the ontology does cover the entire Quran, yet the ontology lacks sufficient description of concepts and has been extended by other authors to suit with their semantic search [19].

The selection of domains or verses to develop the ontology is based on the purposes of a project. Most of the construction were purposely done to support projects such as semantic search[19], [20], knowledge based [3], [18] and knowledge representation [7]. Other reason for selecting the verses are the difficulty exists that in understanding the meaning of each word in Quran. Some verses needs to be interpreted by domain expert such as there are two definitions (direct or indirect) referring to animals [20]. This definition related to concepts that are present in Quran.

In general, most of the ontology development for Quran are done manually starting from the extraction part until the ontology development [7], [18], [20]. For example, the ontology that have been developed for *Juz' Amma* used manual extraction of knowledge [8]. The research done focused on the merging methodology approach, which made the ontology development more effective and intuitive. The authors merged the METHONTOLOGY methodology [21] with Gruninger and Fox's methodology [22]. However, the manual extraction to form ontology remains a labor intensive and time consuming task [23], [12]. A few researchers have construct ontology using at least a semi-automatic approach where the extraction are done automatically and the development are done manually [3], [8]. This process falls under the Ontology Learning approach. The Ontology Learning is a field of research that aims to support the difficulty and time consuming task of knowledge extraction or acquisition by automated techniques [24]. It attempts to extract ontological elements (conceptual knowledge) from input and building ontology from the extracted elements [25], [26]. The approach has been used not only to learn and extract knowledge from Quran but also in other

domains such as terrorism [27], sport [27] and tourism[28].

Existing Ontology Learning tools include Text2Onto [28], ASIUM [29] ,HASTI [26], Text-to-Onto [30], [24] and CRCTOL [27], [31]. Most of the systems use different text resources such as Germany and Persian text. Unfortunately these systems cannot be used for the Quran texts which have different formatting style and text structure. In terms of extracted elements, the systems are able to extract different elements such as terms, concepts, taxonomy and non-taxonomy relations and sometimes axioms.

The technique employed in Ontology Learning may vary depending on the tasks and input. The techniques can be categorize into statistics-based, linguistic-based, logic-based or hybrid [17].

In ***statistics-based***, the techniques employed mostly derived from Information Retrieval, Machine Learning and Data Mining. Statistic-based are more established in the early stage in the Ontology Learning because of the lack of semantics relations between the components of text. In some cases, the statistical technique may be skewed due to data sparsity and difficulty to extract suitable semantic information from domain specific such as Quran [32]. Some of the common techniques include clustering, Latent Semantic Analysis, co-occurrence analysis. The clustering techniques find the similarity of terms to be grouped together. This process will discover concepts or construct hierarchy [33]. The process of clustering can either create agglomerative clustering or create divisive clustering. The Latent Semantic Analysis (LSA) has been applied to reduce the dimension of the data sparseness [34]. Another technique is the relevance analyses that use the occurrence of terms in a document and in a corpus. The common relevance measures from an Information Retrieval field is the term frequency-inverse document frequency (TF-IDF) [35] and others based on language analysis [36].

The second technique is based on ***Linguistics***. This technique involved the Natural Language Processing (NLP) tool which relies on Part-Of-Speech (POS) tagging, syntactic structure analysis, sentence parsing, and dependency analysis to analyse the structure of text. The syntactic structures and text dependency can be produced by POS tagging and sentence parsing (parser). Brill Tagger and Minipar are the examples of the POS. The use of semantic lexicons resources is able to help analyzing the text structure. The lexicons can either be general, such as Word Sense Disambiguation [37] and WordNet, or domain specific, such as the Quranic Glossary [3]. Linguistic techniques also depend on patterns that exist in the text. The patterns were created based on the POS tagging. For example, the lexico-syntactic patterns based on Hearst pattern [38] aims to find hyponym and hypernym relations by discovering regular expression patterns in text. The patterns capture the hyponym and hypernym for instance **such as NP, such NP as {NP,}* {(or|and)} NP**. KnowItAll system also used patterns in their extractor [39]. The patterns can extract a class or concept from document warehouse and dedicated web sites. Some syntactic pattern also adapted from Hearst's pattern and others were developed independently. The example of the pattern used are **NP1 {",} "such as" NPList2, NP1 {",} "and other" NP2, NP1 "is a" NP2, "the" NP1 "of" NP2 "is" NP3**. A lot of works used patterns to find relationship between concepts or relations. The pattern is a simple technique yet able to find some hidden information.

In particular, NLP patterns based on Ontology Learning approach that has been employed for 'solat' domain is able to perform the term extraction task for Quran text translation [3]. The patterns achieved 77% performance of recall after enhancing the patterns using statistical technique. The patterns were created based on NLP. It try to map with Quran formatting style and text structure to find ontological elements such as "part of" relations, "synonym" relations, concepts and instances. The existing patterns could be used as a basis to learn some relations for other test collection.

3.0 METHODOLOGY

3.1 Development of Quran Ontology

Development of Quran ontology as a knowledge based needs to consider the Quran resources, expertise and technological elements. The knowledge based must integrate these elements in order to be a successful knowledge based. Figure 1 exhibits the interaction between these elements that should be deliberated in the Quran Knowledge Based Model.

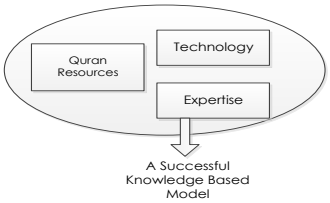


Figure 1 The Quran Knowledge Based Model

First element is the Quran Resources. It should include related Quran documents such as English translated Quran and Quran Glossary Index. Limited chapters or verses related to a certain topics or themes could be used as training documents. It needs to be carefully select to assist understanding about a certain topic of the Quran. Moreover, the Quran is a domain specific text which verse has features of *muhkamat* (verse that have clear meaning) and *mutashabihat* (verse can have many meaning and needs to acquire the correct

guidance) [40]. Also, the use of general thesaurus resources such as WordNet [3] is not suitable for this resources although it is recommended for expanding or specifying information in other ontologies [41].

Second element is the expertise which involves human intervention to make sure the quality of a knowledge based. Two types of the expertise could be included; 1) domain expert of Quran which involve in filtering the relevant terms, concepts, relations and others from Quran resources. The expertise also will evaluate the ontology that has been created. 2) Ontology engineer or ontologist that will develop the ontology from extracted Quran resources.

Last element is the technological elements such as semantic web technology and extraction system. The semantic web technology will be used as a tool that enables the extracted knowledge to be machines-readable and interpretable. In semantic web, the conceptual model of this knowledge can be represented as Resource Description Framework (RDF) and widely recognized as a standard for annotating web documents. RDF is a framework for representing information about resources in a graph form. The RDF provides a simple way to describe resources on the web. It is a directed, labelled graph data format to represent information in the Web. The RDF metadata model is in the form of "subject-predicate-object" expressions, called triples in RDF terminology. The example of RDF metadata model for chapter 73, verse:1 can be represented as in Figure 2.

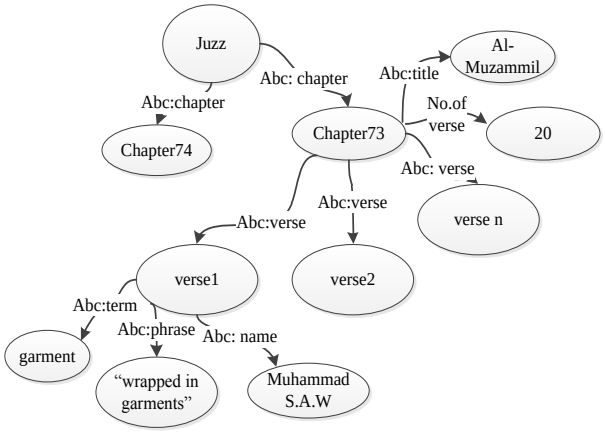


Figure 2 RDF document graph

The editor tools such as Topbraid composer will make the process of developing knowledge based easier. Task such as adding, changing and implementing new concepts and relationships can be just as simple. Moreover the tool will be easily integrated with the high performance knowledge storage like AllegroGraph which can store billions of RDF triples. On the other hand, the extraction system will extract all relevant terms, concepts and relations from the Quran resources. The extracted elements will be used by ontology engineer to form knowledge based.

3.2 The Natural Language Processing (NLP) Patterns

Most of the Ontology Learning processes were extract terms, concepts, relations and axioms. However, this experiment has only carried out to extract relations that exist in the text. As an initial phase, the existing patterns based on [3], [40], [42], [43] has been used to extract relations such as taxonomic relations (Is-A, Part-of relations) and non-taxonomic relations (synonym and definition relations).

In NLP, the main technique relies on Part-Of-Speech (POS) tagging to analyse the pattern exist in the text. This experiment has employed General Architecture of Text Engineering (GATE) tool to prepare the POS tagging. In GATE, the ANNIE (A Nearly-New Information Extraction System) POS tagger has been chosen to perform POS tagging. The POS will assign parts of speech to each word such as noun, verb, adjective, etc. Traditionally, there are eight POS which are noun, verb, pronoun, preposition, adverb, conjunction, adjective and article. This process will generate the annotated text with their POS tagging. Then, the structure of annotated text has been analysed to find patterns. After that this pattern will be matched with the text structure to find relations.

In addition, English translated Quran from Hillali Khan has been used in this experiment. It provides more explanation for a certain verse rather than literal translation. Besides, the translation summarizes the information according to tafsir of AlTabari, Al-Qurtubi and Ibn kathir with comments based on Sahih Al-Bukhari. Besides, the existing patterns has been created and tested only for this translation. Different translation may produce different patterns.

In terms of verses, the moderately long chapters (the chapters that contain an average of 55 verses) are being chosen for testing. There are five moderately long chapters has been used in this experiment which is chapter *Al-Maarij*, *Al-Jinn, Nuh, Al-Muzammil* and *Al-Muddathir*. The chapters contain 20 verses to 56 verses.

The selection of test collection is done because executing patterns requires a deep analysis on the text structure and it is tedious task to test the entire Quran for 114 chapters with 6236 verses. Also, the selection is based on chapters and not by a certain theme or domain. This is because in order to get a complete sentence based on some verse, we have to consider other previous verses. On the other hand, previous experiment used only used 73 verses of the Quran based on verses related to prayer or "solat"[40]. Meanwhile this experiment used 176 verses which consists of 3043 words.

Based on existing NLP patterns, There are three patterns have been chosen for testing. The patterns

are based on 1) the formatting of Quran text structure and 2) the patterns of Quran text. The following NLP pattern has been used to test the performance of the extraction.

Pattern1

NP_0 [| (i.e. P_1, P_2,.....P_n) |] -> P_1, P_2.... P_n part of NP_0

If after "open Square bracket "[" or round bracket "(" i.e. then If $NP_0[P_1]$* then P_1, P_2.... P_n part of NP_0

Pattern1 use to extract PartOf relations between concepts

Example of the text and the POS tagging based on the new test collection as follows.

"...the Mujrim, (criminal, sinner, disbeliever, etc.)..."

the/DT Mujrim/NNP ,/, (/(criminal/JJ ,/, sinner/NN ,/, disbeliever/NN ,/, etc/FW ./.)/)

Pattern2

NP_0 {:|!} (P_1, and)* ....., and P_n

After NP then the symbol ':' or '!' exist, P* part of NP_0

Pattern2 use to extract PartOf relations between concepts

Example of the text and the POS tagging based on the new test collection as follows.

"...Nawafil non-obligatory acts of worship: prayers, charity, fasting, Hajj and 'Umrah, etc.)...."

Nawafil/NNP non-obligatory/JJ acts/NNS of/IN worship/NN :/: prayers/NNS ,/, charity/NN, fasting/VBG ,/, Hajj/NNP and 'Umrah/NNP ,/, etc/FW ./.

Pattern3

if $NP_0(NP_1*)$ and word.lenght <= 5 then
NP_0•synonym = NP_1
else
p $NP_0(P_1)$•Definition = P_1

Pattern 3 use to extract synonym and definition relations

Example of the text and the POS tagging based on on the new test collection as follows.

"...Salat (prayers)..."

Salat/NNP (/(prayers/NNS)/)

The NP (Noun Phrase) and P (Phrase) represent concepts or sub concepts that exist in Quran. The NP is based on any noun, combination of two words with a noun as a head, pronoun that exists in the text or can be a combination of verb with adjective. P is a phrase based on NP or VP (verb phrase).

The evaluation of extracted elements is done based on the match of the three patterns with the annotated POS tagging for Quran. Formula to calculate the Performance of Extraction (PoE) is as follows:

$$PoE = \frac{\text{No. of Correct Match}}{\text{Total of Extraction}} \quad (1)$$

Where:
No. of Correct Match = Number of correct matches from all chapters for each pattern

Total of Extraction = Total of correct extraction from all chapters for each pattern.

4.0 RESULTS AND DISCUSSION

A total of 176 verses (3043 words) from five chapters have been tested. Based on Table 1, it is found that Pattern1 and Pattern2 are able to extract *PartOf* relations between concepts. Pattern1 has higher correct match of relations compared to the Pattern2 with the performance of 23.91% correct match. Meanwhile, Pattern2 only perform 2.17% of correct match. False match of extracting the *PartOf* relation exist in Pattern2 when the POS tagging annotated the first and the second terms as NNP even though the terms are just an upper initial for the sentence. For example, the extracted term "*Verily*" is not *PartOf* the term "*Nay*" even though both were extracted as NNP for a sentence.

Table 1 Result Based on Patern1, Patern2 and Pattern3

Pattern	Correct Match	False Match	Performance of Extraction (%)
Pattern 1	11	0	23.91
Pattern 2	1	3	2.17
Pattern 3	34	32	73.91
Total	46	35	100

Most of the exclamation marks in Pattern2 were used at the end of a sentence and not in the middle of a sentence. Meanwhile, in Pattern1, the false extraction relations exist when it used bracket to explain or elaborate sentences. The extracted NP is not *PartOf* relations for other NP. For example, the term "*who have submitted to Allah*" is not PartOf the term "Muslim" but it is more to elaborate or explain the term "Muslim". Pattern3 has the highest correct matched compared to Pattern2 and Pattern1 with the performance of 73.91%. The pattern is able to extract *synonym* and *definition* relations that exist in the test collection.

The details of the extraction of synonyms and definition relations for Pattern 3 are shown in Table 2. Based on the table, the performance of correct matches for synonym relations is 85.29% which is higher than Definition relations. Only 51.52% of the extracted *synonym* and *definitions* relations are correct while 48.48% are wrong. Meanwhile, the extracted *synonym* relations are higher compared to extracted *definition* relation with the performance of 38.46%. It also found that the false extracted *synonym* relations are due to the annotated POS tagging. For example the POS tagging for a sentence is *garments/NNS (Prophet/NNP Muhammad/NNP SAW/NNP)*. In this sentence, the term "*Prophet Muhammad SAW*" which is NP is not a synonym to the term "*garments*". The brackets used in that sentence were to explain other NP which is the term "*you*" that exist in the sentence. As previously discussed, the use of bracket is to explain or elaborate about terms in the sentence.

Table 2 Results of Extracted Synonyms and Definition Relations based on Patern3

Relations	Correct Match	False Match	Performance of Extraction %
Synonym	29	24	85.29%
Definition	5	8	14.71%
Total	34	32	
Performance of correct extracted Synonym & Definition	51.52%	48.48%	

This experiment also found that the use of rounded brackets '( )' and square brackets '[]' are not consistent. Sometimes the brackets can be *synonym* or can be *definition* relations. Mostly the square brackets were used if there are rounded bracket in the sentence such as the term "*Messenger[Musa (Moses)]*. Based on the above example, it can be used as a basis to extract "*isA*" relation. It indicates that the term "*Musa*" or "*Moses*" "isA" "*Messenger*". Therefore, the use of both square and round brackets simultaneously can be formalized as a new pattern as follows:

If $NP_0[NP_1(NP_2)]$
NP_0 isA= NP_1, NP_2

The new pattern needs to be tested to retrieve more relevant results for constructing the ontology. Apart from that, the three patterns needs to be modified especially pattern3 to obtain higher match of synonyms and definition relations.

5.0 CONCLUSION

The existing NLP patterns are able to extract relations such as PartOf, synonym and definition. The extracted output will be used for the next phase which is the ontology development. However, with the inconsistent used of brackets and high dependency on POS tagging, the NLP patterns needs to be modified to extract more correct relations. On the other hand, the patterns conform to the formatting style of Quran, therefore it becomes simple and promising technique to create a new pattern based on finding relations in the Quran.

Future work includes the extraction of a larger test collection and improving the NLP patterns based on Ontology Learning approach. The identified new pattern also needs to be experimented further so as to evaluate the performance of extraction of "isA" relation.

References

[1] H. Kettani. 2010. World Muslim Population. *Proc. 8th Hawaii Int. Conf. Arts Humanit. Honolulu, Hawaii, 2010.*

[2] N. Abbas. 2009. Quran'Search for a Concept'Tool and Website.

[3] S. Saad. 2013. Ontology Learning and Population Techniques for English Extended Quranic Translation Text.

[4] S. Baqai, A. Basharat, H. Khalid, A. Hassan, and S. Zafar. 2009. Leveraging Semantic Web Technologies for Standardized Knowledge Modeling and Retrieval from the Holy Qur'an and Religious Texts.

[5] P. Cimiano, J. Völker, and R. Studer. 2006. Ontologies on Demand?-A Description of the State-of-the-Art, Applications, Challenges and Trends for Ontology Learning from Text. *Information, Wiss. und Prax*. 57: 315-320,

[6] K. Hassanzadeh, M. Reformat, W. Pedrycz, I. Jamal, and J. Berezowski. 2013. T2R: System for Converting Textual Documents into RDF Triples. *IEEE/WIC/ACM Int. Jt. Conf. Web Intell. Intell. Agent Technol.* Nov. 2013. 221-228.

[7] M. Al-yahya and H. Al-khalifa. 2010. An Ontology Model for Representing Semantic Lexicons: An Application on Time Nouns in the Holy Quran. 35(2): 21-35.

[8] K. Dukes. 2014. Ontology of Quran Concepts. [Online]. Available: http://corpus.quran.com/ontology.jsp. [Accessed: 12-Jun-2014].

[9] D. Allemang and J. Hendler. 2007. *Semantic Web for the Working Ontologiest.*

[10] K. Byrne and E. Klein. 2010. Automatic Extraction of Archaeological Events from Text. *Proc. 37th Int. Conf. Williamsburg, Virginia, United States Am*. 1-16.

[11] J. Fan and A. Kalyanpur. 2012. Automatic Knowledge Extraction From Documents. 56(3): 1-10.

[12] A. Maedche and S. Staab. 2001. Ontology Learning For The Semantic Web. *IEEE Intell. Syst.* 16: 72-79.

[13] O. Etzioni, A. Fader, J. Christensen, and S. Soderland. 2011. Open Information Extraction: The Second Generation. 3-10.

[14] D. C. Wimalasuriya and D. Dou. 2010. Ontology-Based Information Extraction: An Introduction and a Survey of Current Approaches. *J. Inf. Sci.* 36(3): 306-323.

[15] V. Uren, P. Cimiano, J. Iria, S. Handschuh, M. Vagasvera, E. Motta, and F. Ciravegna. 2006. Semantic Annotation For Knowledge Management: Requirements and A Survey Of The State Of The Art. *Web Semantics: Science, Services and Agents on the World Wide Web.* 4: 14-28.

[16] P. Buitelaar, P. Cimiano, and B. Magnini. 2004. Ontology Learning from Text: An Overview. 1-10.
[17] W. Wong, W. Liu, and M. Bennamoun. 2012. Ontology Learning from Text. *ACM Comput. Surv.* 44(4): 1-36.
[18] R. Iqbal and A. Mustapha. 2013. An Experience Of Developing Quran Ontology With Contextual Information Support. 7(4): 333-343.
[19] A. R. Yauri, R. A. Kadir, A. Azman, M. Azrifah, and A. Murad. 2013. Quranic Verse Extraction base on Concepts using OWL-DL Ontology. 6(23): 4492-4498.
[20] H. U. Khan, S. M. Saqlain, M. Shoaib, and M. Sher. 2013. Ontology Based Semantic Search in Holy Quran. 2(6).
[21] M. Fernandez, A. Gomez-Perez, and N. Juristo. 1997. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. 33-40.
[22] M. Gruninger and M. S. Fox. 1995. Methodology for the Design and Evaluation of Ontologies.
[23] A. Gómez-pérez and D. Manzano-macho. 2003. Deliverable 1. 5: A Survey of Ontology Learning Methods aAnd Techniques Ontoweb Consortium.
[24] A. Maedche and S. Staab. 2001. Ontology Learning for the Semantic Web. *IEEE Intell. Syst.*
[25] M. Shamsfrad and A. A. Barforoush. 2003. The State of the Art in Ontology Learning: A Framework for Comparison. *Knowl. Eng. Rev*. 18: 293-316.
[26] M. Shamsfard and A. A. Barforoush. 2004. Learning Ontologies from Natural Language Texts. *Int. J. Hum. Comput. Stud.* 60(1): 17-63.
[27] X. Jiang and A. H. Tan. 2010. CRCTOL: A Semantic-Based Domain Ontology Learning System. *J. Am. Soc. Inf. Sci. Technol.* 61: 150-168.
[28] J. Cimiano, Philipp and Völker. 2005. A Framework for Ontology Learning and Data-Driven Change Discovery. In *Natural Language Processing and Information Systems*. 227-238.
[29] D. Faure and C. Nedellec. 1999. Knowledge Acquisition Of Predicate Argument Structures From Technical Texts Using Machine Learning: The System ASIUM. *Knowl. Acquis. Model. Manag.*
[30] A. Maedche and S. Staab. 2000. The Text-To-Onto Ontology Learning Environment. Softw. Demonstr. *ICCS-2000, Eight Int. Conf. Concept. Struct.*
[31] X. Jiang and A. Tan. 2005. Mining Ontological Knowledge from Domain-Specific Text Documents. *Fifth IEEE Int. Conf. Data Min.* 665-668.
[32] D. Maynard, Y. Li, and W. Peters. 2008. Nlp Techniques For Term Extraction and Ontology Population. *Proc. 2008 Conf. Ontol. Learn. Popul. Bridg. Gap between Text Knowl.* 107-127.
[33] S. Mokarizadeh, P. Küngas, and M. Matskin. 2013. Ontology Acquisition From Web Service Descriptions. *Proc. 28th Annu. ACM Symp. Appl. Comput.-SAC '13*. 325.
[34] S. Deerwester, S. T. Dumais, G. W. Furnas, and T. K. Landauer. 1990. Indexing by Latent Semantic Analysis. *J. Am. Soc. Inf. Sci*. 41(6): 391-407.
[35] G. Salton and C. Buckley. 1988. Term-weighting Approaches in Automatic Text Retrieval. *Inf. Process. Manag.*
[36] J. Ponte and W. Croft. 1998. A Language Modeling Approach To Information Retrieval. In *Proceedings Of The 21st Annual International ACM SIGIR Conference On Research And Development In Information Retrieval.* 275-281.
[37] R. Navigli. 2009. Word Sense Disambiguation: A Survey. 41(2).
[38] M. A. Hearst, "Automatic Acquisition of Hyponyms from Large Text Corpora," pp. 23–28, 1992.
[39] O. Etzioni, A. Popescu, D. S. Weld, D. Downey, and A. Yates. 2004. Web-Scale Information Extraction in KnowItAll (Preliminary Results). 100-110.
[40] S. Saad, N. Salim, and H. Zainal. 2010. Towards Context-Sensitive Domain of Islamic Knowledge Ontology Extraction. 3(1): 197-206.
[41] M. Fernandez, Z. Zhang, V. Lopez, V. Uren, and E. Motta, 2011. Ontology Augmentation: Combining Semantic Web and Text Resources. 9-16.
[42] S. Saad, N. Salim, and H. Zainal. 2009. Islamic Knowledge Ontology Creation.
[43] S. Saad, N. Salim, and H. Zainal. 2009. Pattern Extraction For Islamic Concept. 333-337.