Journal of Engineering and Applied Sciences 11 (3): 629-634, 2016
ISSN: 1816-949X

A Systematic Review Analysis for Quran Verses Retrieval

[1]Muhammad Afifi Mohamad Safee, [1,2]Madihah Mohd Saudi,
[1, 2]Sakinah Ali Pitchay and [1,2]Farida Ridzuan
[1]Faculty of Science and Technology,
[2]Islamic Science Institute (ISI), Universiti Sains Islam Malaysia (USIM),
Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia

Abstract: Now a days, a proper method for presenting information is crucial due to the increase of mass volume and complexity of information. As for Muslim, Quran is referred as one of the main resources for guidance and reference for their daily life. Currently, there are many methods have been used for Quran verses retrieval. However, there is still lack of studies on the scientific way for Quran verses retrieval. As a consequence, it is quite hard to obtain accurate information and might lead to wrong information being retrieved. Therefore, this study investigates and evaluates the strengths and weaknesses of existing works. Furthermore, a case study on a new Quran ontology alignment model by integrating the upper ontology (Quran corpus) and domain ontology (general medical) for Quran verses extraction is discussed in this study. Based on the findings, ontology remodeling is required to fix the inconsistency of the ontology by integrating ontology alignment technique and fuzzy algorithm.

Key words: Complexity, guidance, retrieval, integrating ontology alignment technique, Quran ontology alignment model, domain ontology

INTRODUCTION

By year 2020, 40 Zettabyte of data will be considered as a common amount of data that will be used worldwide and the demand will be increased tremendously. However, there are few challenges will raise up due to the data growth such as data security, data diversity and transfer speed (Khan *et al*., 2014). The characteristics of big data are volume, variety, velocity and value which makes it different from the traditional database systems (Leary, 2013).

Al Quran is referred as one of the primary resource for Muslim. Muslim believes that the Quran contains information related with science, prediction of future, law, and others where those information can be retrieved throughout the. Recently, there are more studies and research work related with Quran and sciences (Arbaoui *et al*., 2013). Therefore, it is crucial to store, manage and process the information efficiently.

Currently, there are two techniques for knowledge representation for Quran which are known as keyword-based and semantic based. Keyword-based technique finds possible word combination that is relevant to the keyword while semantic-based technique finds similarity based on matching the contextual meaning of keyword (Yauri *et al*., 2013). Therefore, a new Quran ontology alignment model by integrating the upper ontology (Quran corpus) and domain ontology (general medical) for Quran verses extraction is proposed.

Literature review: Previously, keyword-based approach was used for information retrieval for Arabic language. That approach can be referred in the existing work done by Jarrah (2006). They design Optical Character Recognition (OCR) for recognition of the Arabic alphabet. It computes image-morphological features on input of word images. Moreover, Chatbot has been created to search and to match based on the keyword to find relevance verses for Arabic Web (Abu Shawar, 2011). The Chatbot matches the important word from the search field and performs simple keyword based searching to find the relevant verses from Arabic website. Semantic search is based on the retrieving documents where it employs the semantic analysis of their contents using natural language processing (Giunchiglia *et al*., 2009). Semantic search differs from Syntactic search by exploiting the meaning of words, thus avoiding many of the well-known problems of syntactic search as discussed in this study. The example

Corresponding Author: Muhammad Afifi Mohamad Safee, Faculty of Science and Technology,
Universiti Sains Islam Malaysia (USIM), Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia

Table 1: Comparison study of Arabic verses retrieval model

Authors	Approach	Method
Yauri *et al.* (2013)	Semantic-based	Combination of natural language and semantic
Shoaib *et al.* (2009)	Semantic-based	Cross Lingual Information System (CLIR) Dictionary-based approach
Ku-mahamud (2014)	Semantic-based	Conceptual graph and semantic network object based knowledge representation
Al-Jarrah (2006)	Keyword-based	Optical Character Recognition (OCR) for recognition of Arabic alphabet
Abu Shawar (2011)	Keyword-based	Pattern matching technique
Al-Tahrawi *et al.* (2015)	Keyword-based	Text classification using polynomial networks

of the semantic search can be found in the existing work done by Al-Tahrawi and coauthors which created the Arabic Text Classification Using Polynomial Networks. The algorithm was used in text pre-processing step to remove noise and reduce the number of terms in dataset. Work by Yauri *et al.* (2013) applied the ontology semantic based approach to represent the Quran domain for searching. They used dataset from Leeds University for evaluation. The ontology is made up from important noun concept found in Quran. The system displays a complete expression of statement that requested using query. Cross-Language Information Retrieval (CLIR) method is used by (Yahya *et al.*, 2013). The Quran concept is applied for disambiguating translation of the query. It aims to improve dictionary-based query translation. Lastly, work done by Ku-Mahamud and coauthors used the conceptual graph and semantic network to present extracted information where this research only focused for female terms extraction. Table 1 summarizes the discussed existing works.

Based on the evaluation of the previous researchers, one of the main challenges for keyword retrieval model is the loss of keyword semantics (contextual meanings) which gave many irrelevant result, which is called as the Syntactic search. The search procedure is essentially based on the syntactic matching of document and query representation. Therefore, semantic search is known as one of the promising way to overcome the problems.

Ontology is the semantic search backbone, and the success of any semantic application depends on the design and development of the Ontology. In general, there are 2 main kinds of ontology, the domain specific ontology and upper ontology. The domain specific ontology represents the particular meaning of the term as they apply to that particular domain. For example, the word "bank" has many different meaning. An ontology about the domain of nature would model the "river bank" meaning of the word, while an ontology about the domain of economy would model the "bank balance sheet" and "bank notes" while 'upper Ontology' which is a representation of the common concepts that are applicable via a wide range of domain ontologies (Jarrar, 2006).

Onto select Ontology library shows that there is a lack of Arabic Ontologies library and about 49% of Ontologies are created for Latin character set. Since, each language has its own linguistic environment and cultured context, so each language needs its own Ontology. As most developed Ontologies are in English, there is an urgent need for the development of Arabic Ontologies to be used as the base of the Arabic retrieval applications. Several researches concerned the building and population of Arabic Ontologies. The most common ontologies for English are Upper Model, Wordnet, Sumo and OpenCyc, while most common work is Arabic Wordnet.

Until 2015, several ontologies have been developed for Quran focusing on certain themes or domains (Saad, 2011). Ontology for Quran as a whole, covering all domains in Quran is still not complete. For example, the ontology for pray only covers pray perspective from the entire Quran (Saad *et al.*, 2011). While work by Al-Yahya and Al-Khalifa developed semantic lexicons of Arabic that only focus on verses related to time (Al-Yahya and Al-Khalifa 2010). Another ontology, covers only verses in Juz Amma (Khan *et al.*, 2013) developed ontology for living creatures including animals and birds mentioned in Holy Quran. Only ontology from University of Leeds covers the entire Quran. However, the ontology lack of sufficient description of concept and has been extended by other authors to suit with their semantic search (Yauri *et al.*, 2013). Another work has been done by Zamil and Al-Radaideh (2014). using lexicon-syntactic patterns where it gives poor performance on ambiguous text for Arabic text. Table 2 summarizes the previous works of development of ontology for Holy Quran.

Based on Table 2 review, not much works have been done for domain independent ontology or known as upper ontology. Most domain independent ontology is not able to retrieve the relevance result when tackle with the ambiguous meaning queries. Most ontology built for the Quran is incomplete and focused in a specific domain. These approaches failed to retrieve relevance results for cross domain representation. Hence, the results produced lower precision and recall of Quran verses. Some of researchers suggested ontology alignment approach to tackle this problem (Shvaiko and Euzenat, 2007).

The problem of the separated upper ontology and domain-dependent ontology due to lack complete for of data mining domain. Historically, the needs for ontology

Table 2: Related works on ontology for Holy Quran

Author	Type of	Weakness	Strength
Saad *et a l.* (2011)	Domain Specific	Only covers pray perspective	Automatic extraction from Quran in order to build ontology.
Iqbal *et al.* (2013)	Domain Specific	Only covers Juzu Amma (30th) from Quran	Methodology used that it possesses a high level degree of formality.
Ullah Khan *et al.* (2013)	Domain Specific	Only covers living creatures mentioned in Holy Quran	Quranic Wordnet contains definition of concepts as defined in Holy Quran is introduced.
Yunus *et al.* (2010)	Upper Ontology	Not able to distinguish ambiguous meanings of certain formal representation in Malay, English and Arabic.	Using semantic cross language (CLIR) for information extraction from Quran
Yahya *et al.* (2013)	Domain Specific	Not able to retrieve relevant document at concept matching part after query translation/	Cross-Language Information Retrieval (CLIR) for disambiguating translation of the query and to improve the dictionary-based query translation
Muhammad (2012)	Domain Specific	This research restricts the domain to the Qur'anic text	Qur'anic pronoun tagging, a domain concept ontology of the Qur'an emanating from pronoun referents
Yunus *et al.* (2010)	Upper Ontology	Failed to retrieve complex ambiguous queries	Enhancement of Leeds Ontology by adding acts (Solat, Zakat, Zina, Honesty)
Muhammad (2012)	Domain Specific	The ontology only covers two themes: "Iman" (faith) and "Akhlaq" (deed)	Ontology-based representation which consist of Juz, Chapter and Vers, differs from traditional method

Fig. 1: Pseudo code of Fuzzy logic

alignment rise up because of the need to integrate heterogeneous databases that independently within each other, hence having their own data vocabulary. However, there are many different classifications to types of heterogeneity (Shvaiko and Euzenat, 2007). Stated that there are three most obvious types of heterogeneity. First one is known as syntactic heterogeneity, where the two ontologies are not expressed in the same ontology language. This kind of mismatch is generally tackled at the theoretical level when one constructs similarities different languages. Next is the terminological heterogeneity where variation in name when referring to the same entities in different ontologies. This heterogeneity caused implementation of different natural languages. Lastly, conceptual heterogeneity also called semantic heterogeneity because of differences in modelling the same domain of interest. This heterogeneity occurs because of using different axioms for defining concepts or due to the use of totally different concepts. Therefore, ontology alignment/matching are required to tackle this heterogeneity issue. It aims to support semantic interoperability between different information systems (Vennesland, 2015).

Shvaiko and Euzenat (2013) and Cerdeira *et al.* (2015) suggested context-based ontology matching uses external resources in order to help establish contextualization between the ontologies. The external resources can be formal or informal. Formal resources are typically ontological structures using a formal language such as OWL or RDF and less formal resources such as Word Net (Jean-Mary *et al.*, 2009) and use of informal resources such as web sites to identify correspondences (Paulheim and Hertling 2013).

Fuzzy logic is a form of many-valued logic in which the truth values of variables may be any real number between 0 and 1 that can handle the concept of partial truth where the truth value may range between completely true and completely false (Zadeh, 1988). Figure 1 describes the pseudo code of Fuzzy logic that will be implemented as a matcher evaluator for ontology alignment process (Zadeh, 1988). The selection of matching can be determined on threshold level of the Fuzzy logic. The thresholds level in fuzzy logic can be used to constrain the limits of the acceptable similarity value.

MATERIALS AND METHODS

The overall research processes or known as the experimental framework used in this research as illustrated in Fig. 2. The framework is inspired from ontological framework called as Agreement Maker Light by Faria *et al.*, 2013). The architecture can be altered and customize to assemble a new architecture based on user needs. Agreement Maker Light is one of the best systems that has been evaluated at Ontology Alignment

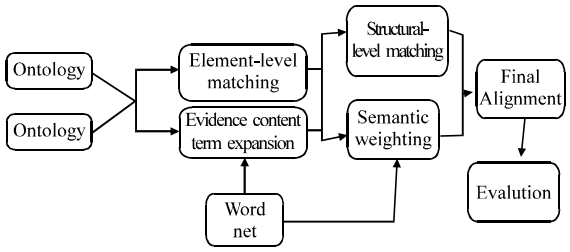

Fig. 2: Proposed Framework for ontology matching

Evaluation Initiative (OAEI) Paulheim and Hertling (2013). Quran Ontology from QuranOntology.com (Hakkoum, 2014) and Ontology for General Medical terminology ontology Xiang will be used as input ontologies for the new model. Arabic Wordnet will be used as external resources for lexical matching between ontologies. For the lab environment, open source applications were used as displayed in Table 3. Here, fuzzy algorithm will be implemented as the matcher for ontology mapping. It will measure the numerical value that reflects the semantic similarity between terms.

The proposed framework consists of two modules: the ontology loading module and the ontology matching module. The ontology loading module will load the input ontology files and construct ontology objects. On the other hand, ontology matching module will align ontology objects by combining one or more matching algorithms.

The proposed framework consists of two modules: the ontology loading module and the ontology matching module. The ontology loading module will load the input ontology files and construct ontology objects. Furthermore, ontology matching module will align ontology objects by combining one or more matching algorithms. The ontology matching module allows embedding of any matching algorithm where the Fuzzy Logic Algorithm will be embedded. A fuzzy ontology O_F is defined as a 5-tuple (Abu Shawar, 2011):

$$OF = \{C, P_F, R_F, A_F, I\}$$

Where, C is the set of classes which represents various entities in some domain being modeled. The classes are name by one or more natural language terms and are normally referenced within the ontology by a unique identifier. P_f is a set of properties. Any $p_F \in P_F$ is defined as a 3- tuple:

$$p_F = \{c, v, q\}$$

c = It is a class
v = It is value of p_F.
q = It is a qualifier

Table 3: Software Used in the development and testing

Software	Function
Protege	Open source visual ontology editor for ontologies modelling
Agreement Maker Light	Automated and efficient ontology matching system
Arabic WordNet	Sources of contextual information
Eclipse	Java Editing tools to develop the system
Java SE Runtime	Requirement to run Agreement Maker Light
Environment 7	ontology matching system
OntoLT	Tool to mapping between linguistic entities in text and class/slot candidates in Protege

An instance is "This watch is very expensive". Price is explained as property p_F watch, expensive and very is explained as c, v and q. The R_f is the set of relationships between classes. Any $r_F \in R_F$ is defined as a 4-tuple:

$$r_F = \{c_1, c_2, t, f\}$$

Where:

c_1 and c_2 = Two classes
t = Type of relationship between c_1 and c_2
t = One of synonym of, kind of, part of, instance of, property of
f = A membership degree (0, 1) of t

In this research, only kind of relationship is considered. A_f is a set of fuzzy axioms, usually formalized into some logic language. These axioms specify additional constraints on the ontology and can be used in ontology consistency checking and for inferring new knowledge from the ontology through some inference mechanism. I is set of all instances.

RESULTS AND DISCUSSION

Preliminary study: An ontology matching test have been developed using ontology from Quran Ontologies (Hakkoum, 2014) and Ontobee human general diseases. The objective of the matching test is to find correspondences between semantically related entities of ontologies. These correspondences later on can be used for various tasks such as ontology merging and data translation (Shvaiko and Euzenat, 2013) (Fig.3).

Agreement Maker Light tool has been used for matching process. The stand alone tool was executed using Java Virtual Machine environment. The tool manages to detect 45 classes, 67 properties and 1449 individuals in Quran ontology and 99 classes on human disease ontology as displayed on Fig. 4. From the automated matching process, it is found that output

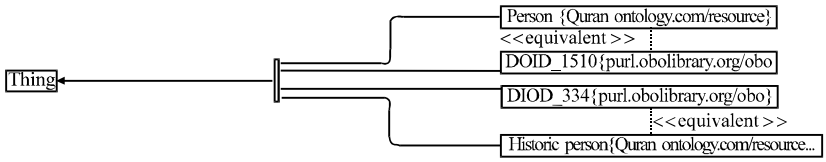

Fig. 3: Ontology matching visualization

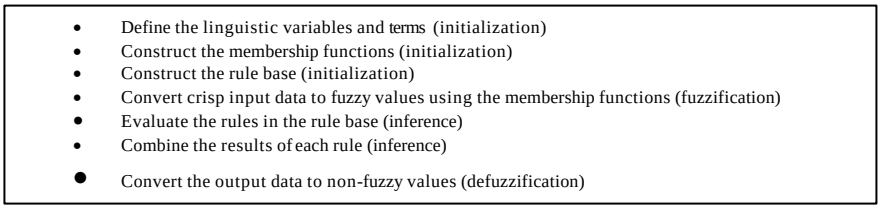

Fig. 4: Ontology matching process

mappings found two similar class from these two ontologies. First, the matcher match person class from Quran ontology and a personality disorder from human general disease ontology. Second, the matcher match historic person class from Quran ontology and histrionic personality from human general diseases ontology. Figure 3 shows the visualization of the ontology mapping.

The output mapping is less accurate in term of contextual meaning and more relevance mapping of other similar class are expected. The result showed an enhancement need to be done in part of consistency of ontology. It is also found that the integrated ontology is inconsistent. The inconsistency of ontology leads to less accurate results of matching. The input ontology needs to be remodeled and fixed to get more accurate match. This is despite the fact that justifications are crucial for repairing inconsistent ontologies and can be used as a basis for paraconsistent reasoning (Horridge *et al*., 2009). There are numerous tools such as Cons VISor and Oil Ed to fix the inconsistency of ontology (Baclawski *et al*., 2002). These tools are a consistency checker for formal ontologies, including both traditional data modeling languages and the more recent ontology languages.

CONCLUSION

This study presented a systematic analysis for Quran verses retrieval techniques where gaps in the existing works have been identified for further improvement. Based on the findings for further improvement of the Qur'anic ontology matching approach, it can be concluded that there is inconsistency on ontology that lead to less accurate results of matching process. Therefore, further investigation needs to be carried out to fix the inconsistency of ontology. However, this research scope and dataset is limited for medical domain. It is expected, a new Quranic data repository for medical domain will be developed based on the proposed model with a better accuracy retrieval rate. This study is a part of larger project to design Qur'anic intelligent search.

ACKNOWLEDGMENTS

The researchers would like to express their gratitude to Universiti Sains Islam Malaysia (USIM) and Institute Science Islam (ISI) for the support and facilities provided. This research study is supported by Ministry of Education grants [PPP/UCG-0114/ FQS/ 30/11714], RAGS grants [USIM/RAGS/FST/36/50814], [USIM/ RAGS/ FST/ 36/50914] and [USIM/NRGS-P4/FST/8404/52113].

REFERENCES

Abu Shawar, B., 2011. A Chatbot as a natural web interface to Arabic web QA. Int. J. Emerg. Technol. Educ., 6: 37-43.

Al-Yahy, M., H. Al-Khalifa, H. Al-Khalifa, A. Bahanshal, I. Al-Odah and N. Al-Helwah, 2010. An ontological model for representing semantic lexicons: An application on time nouns in the Holy Quran. Arabian J. Sci. Eng., 35: 21-35.

Arbaoui, A., Y.M. Alginahi and M. Menacer, 2013. Strategies for collecting electronic resources on the quranic researches. Strategies, 3: 57-78.

Baclawski, K., M.M. Kokar, R. Waldinger and P.A. Kogut, 2002. Consistency Checking of Semantic Web Ontologies. In: The Semantic Web-ISWC 2002. Horrocks, I. and J. Hendler (Eds.). Springer Berlin Heidelberg, Heidelberg, Germany, ISBN: 978-3-540-43760-4, pp: 454-459.

Cerdeira, L.O, F.J.R. Martinez and G.A. Rodriguez, 2015. Ontology matching: A literature review. Expert Syst. Appl., 42: 949-971.

Faria, D., C. Pesquita, E. Santos, M. Palmonari and I.F. Cruz *et al*., 2013. The Agreement Maker Light Ontology Matching System. In: On the Move to Meaningful Internet Systems: OTM 2013 Conferences, Meersman, R., H. Panetto, D. Tharam, Z. Bellahsene and N. Ritter *et al.* (Eds.). Springer Berlin Heidelberg, Heidelberg, Germany, pp: 527-541.

Giunchiglia, F., U. Kharkevich and I. Zaihrayeu, 2009. Concept search. Semant. Web Res. Appl., 5554: 429-444.

Hakkoum, A. and S. Raghay, 2015. Advanced search in the Quran using semantic modeling. Proceedings of the 2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA), November 17-20, 2015, IEEE, Marrakech, Morocco, pp: 1-4.

Horridge, M., B. Parsia and U. Sattler, 2009. Explaining Inconsistencies in OWL Ontologies. In: International Conference on Scalable Uncertainty Management. Godo, L. and A. Pugliese (Eds.). Springer Berlin Heidelberg, Heidelberg, Germany, ISBN: 978-3-642-04387-1, pp: 124-137.

Jarrah, O.A., S.A. Kiswany, B.A. Gharaibeh, M. Fraiwan and H. Khasawneh, 2006. A new algorithm for arabic optical character recognition. WSEAS. Transac. Inf. Sci. Appl., 3: 832-845.

Jean-Mary, Y.R., E.P. Shironoshita and M.R. Kabuka, 2009. Ontology matching with semantic verification. Web Semant., 7: 235-251.

Khan, H.U., S.M. Saqlain, M. Shoaib and M. Sher, 2013. Ontology based semantic search in Holy Quran. Int. J. Future Comput. Commun., 2: 570-575.

Khan, I.N. I. Yaqoob, T. Abaker, Z. Hashem and W. Inayat *et al*., 2014. Big data: Survey, technologies, opportunities and challenges. Sci. World J., 2014: 1-18.

Leary, D.E.O., 2013. Artificial intelligence and big data. IEEE. Intell. Syst., 2: 96-99.

Paulheim, H. and S. Hertling, 2013. Wesee-match results for OAEI 2013. Proceedings of the 8th International Conference on Ontology Matching-Volume 1111, October 21-21, 2013, CEUR-WS, Aachen, Germany, pp: 197-202.

Saad, S., N. Salim, H. Zainal and Z. Muda, 2011. A process for building domain ontology: An experience in developing Solat ontology. Proceedings of the International Conference on Electrical Engineering and Informatics, July 17-19, 2011, IEEE, Bandung, Indonesia, ISBN: 978-1-4577-0753-7, pp: 1-5.

Saad, S., N. Salim, H. Zainal and Z. Muda, 2011. A process for building domain ontology: An experience in developing Solat ontology. Proceedings of the International Conference on Electrical Engineering and Informatics, July 17-19, 2011, IEEE, Bandung, Indonesia, ISBN: 978-1-4577-0753-7, pp: 1-5.

Shvaiko, P. and J. Euzenat, 2013. Ontology matching: State of the art and future challenges. IEEE Trans. Knowl. Data Eng., 25: 158-176.

Yahya, Z., M.T. Abdullah, A. Azman and R.A. Kadir, 2013. Query translation using concepts similarity based on Quran ontology for cross-language information retrieval. J. Comput. Sci., 9: 889-897.

Yauri, A.R., R.A. Kadir, A. Azman and M.A.A. Murad, 2013. Ontology semantic approach to extraction of knowledge from Holy Quran. Proceedings of the 5th International Conference on Computer Science and Information Technology, March 27-28, 2013, IEEE, Amman, Jordan, pp: 19-23.

Yauri, A.R., R.A. Kadir, A. Azman and M.A.A. Murad, 2013. Quranic verses extraction base on concepts using OWL-DL ontology. Res. J. Appl. Sci. Eng. Technol., 6: 4492-4498.

Zadeh, L.A., 1988. Fuzzy logic. Comput., 21: 83-93.

Zamil, M.G.A. and Q.A. Radaideh, 2014. Automatic extraction of ontological relations from Arabic text. J. King Saud Univ. Comput. Inf. Sci., 26: 462-472.