Aligning and Merging Ontology in Al-Quran Domain

Mohammad Alqahtani[1] and Eric Atwell[2]
[1] University of Leeds, scmmal@leeds.ac.uk
[2] University of Leeds, E.S.Atwell@leeds.ac.uk

Al-Quran is the main sacred text of Muslims and the primary resource for the Islamic sciences and Arabic language (Ali 1951). Al-Quran consists of less than 80,000 words forming 114 chapters (Atwell et al. 2011). A chapter (Surah) contains a varying number of verses (Ayat).

Many researchers have built different Quranic ontologies to facilitate the retrieval of knowledge from Al-Quran, with a minority of these ontologies covering all Al-Quran chapters (Alrehaili & Atwell 2014). The term "ontology" is defined as an explicit specification of concepts, attributes, and relations in a domain (Gruber n.d.). Common components of ontologies include classes (concepts), attributes, relations, function terms, restrictions, and axioms. The concepts are entities of interest in a particular domain. These concepts are structured into taxonomy tree or un-taxonomy tree. Each tree node represents a concept that is a specialization of its ancestor. The concept is related to a set of instances. Additionally, the concept has a set of attributes. Relations are ways in which concepts and instances can be linked to each other.

The three common ontologies covering Al-Quran are: Quranic topics (Abbas 2009), Arabic Quran Corpus (Dukes 2013), and QuranA (Muhammad 2012). The Quranic-topics ontology contains nearly 1,100 Quranic concrete and abstract concepts linked to all verses of Al-Quran. (Abbas 2009) used existing Quranic concepts from the Islamic scholarly book *Mushaf Al Tajweed*. These concepts in the index have an aggregation relationship; the hierarchy of concepts is non-reflexive, non-symmetric, and transitive.

Secondly, (Dukes 2013) extracted 300 concepts and 350 relations from Al-Quran. The relationship types connecting concepts using predicate logic are Part-of and IS-A. The ontology is based on a famous book describing Al-Quran called '*Tafsir Ibn Kathir'* (Abdul-Rahman 2009).

Thirdly, (Muhammad 2012) developed an ontology for Al Quran in the scope of pronoun antecedents. This ontology consists of 1,050 concepts and more than 2,700 relations. Additionally, the relationship types connecting concepts are has-antecedent, has-concept and has–a-segment.

These Quranic ontologies have different scopes and format, and one ontology does not cover all concepts of Al-Quran (Alrehaili & Atwell 2014). Therefore, these ontologies need an alignment and normalization. Ontology alignment is a process of finding one-to-one correspondence via both ontology entities. The primary goal of ontology alignment is to integrate different ontologies of the same domain (Zaeri & Nematbakhsh n.d.).

The main objective of merging Quranic ontologies is to pioneer a research enriching the raw Arabic Quran text with Islamic ontology. Additionally, this combined ontology will help in understanding Al-Quran. Moreover, aligning the Quranic ontologies will increase the coverage of the domain of Al-Quran in various scopes. Furthermore, the alignment will enhance the knowledge extraction from Al-Quran.

Three modules are used to align Quranic ontologies: normalization, terminological approach, and structural approach (Euzenat & Shvaiko n.d.).In the normalization process, all ontologies are reformatted to have the same file format.

Terminological techniques are divided into string-based and language-based approaches. The string-based approach matches entities based on the similarity between letters of the two words, such as "author" and "authority" being more similar than "author" and "writer." However, the language-based technique aligns two entities that share the same meaning, such as "newspaper" and "article."

On the other hand, the structural approaches detect correspondences between entities depending on an internal structure of the entity and how it is connected to other entities. In other words, structural method matches entities based on the ontology graph. Most of the existing alignment tools exploit terminological techniques as the initial step, then use the structural techniques to improve the outcomes ontology.

REFERENCES

Abbas, N., 2009. Quran'search for a Concept'Tool and Website. *Unpublished Dissertation, Leed*.

Abdul-Rahman, M.S., 2009. *Tafsir Ibn Kathir Juz' 1 (Part 1): Al-Fatihah 1 to Al-Baqarah 141 2nd Edition*, MSA Publication Limited. Available at: https://books.google.co.uk/books?id=sIryWTYotyYC.

Ali, A.Y., 1951. The Holy Quran. *Lahore, Pakistan: Ahmadiyyah Anjuman Ishaat Islam*.

Alrehaili, S.M. & Atwell, E., 2014. Computational ontologies for semantic tagging of the Quran: A survey of past approaches. *LREC 2014 Proceedings*.

Atwell, E. et al., 2011. An Artificial Intelligence approach to Arabic and Islamic content on the internet. In *Proceedings of NITS 3rd National Information Technology Symposium*.

Dukes, K., 2013. *Statistical Parsing by Machine Learning from a Classical Arabic Treebank*. School of Computing, University of Leeds. Available at: http://corpus.quran.com/.

Euzenat, J. & Shvaiko, P., Ontology matching. *2007*. Available at: http://link.springer.com/content/pdf/10.1007/978-3-642-38721-0.pdf [Accessed August 19, 2015].

Gruber, T., Ontology (Computer Science) - definition in Encyclopedia of Database Systems. Available at: http://tomgruber.org/writing/ontology-definition-2007.htm [Accessed August 20, 2015].

Muhammad, A.B., 2012. *Annotation of Conceptual Co-reference and Text Mining the Qur'an*,

Zaeri, A. & Nematbakhsh, M., A Semantic Search Algorithm for Ontology Matching. *semantic-web-journal.net*. Available at: http://semantic-web-journal.net/sites/default/files/swj254.pdf [Accessed August 19, 2015].