

Review Article

Review of Computer-Based Academic Research on the Quran

¹Aliyu Rufai Yauri, ²Rabiah A. Kadir, ³Azreen Azman and ⁴Masrah A.A. Murad

¹Department of Computer Science,

²Institute of Visual Informatics, National University of Malaysia, Selangor,

³Department of Multimedia,

⁴Department of Information Systems, Faculty of Computer Science and Information Technology,
Universiti Putra Malaysia, Serdang, Malaysia

Abstract: This study presents a review on the computer based academic researches related to Quran. The main motivation of the paper is because of the growing interest in the Quran knowledge globally, where today we are witnessing several researches on Quran being the main source of knowledge and law for Muslims. Today, computer based research on the Quran is being conducted globally by both Muslims and non-Muslims from various institutions. The study will give an up to date on computer based researches conducted previously and will assist in furthering up researches in the area.

Keywords: Data mining, information retrieval, knowledge representation, linguistic, Quran, semantic

INTRODUCTION

Over the years, there has been a growing interest in the content of Islamic knowledge by both Muslims and non-Muslims, especially knowledge of the Quran. The Quran is the main source of knowledge, wisdom and law for Muslims. Since the first revelation, the Holy Quran has been among the most influential books that exist (Qurat and Amna, 2011). The first revelation was believed to be from God to Prophet Muhammad was through the Angel Jibreel and continued for a period of 23 years (Sharaf and Atwell, 2009). The Quran is a book that covers a wide range of knowledge which consists of 114 chapters with 6,236 verses covering many themes and concepts that make up the divine knowledge and law. These Quran verses are believed by Muslims to discuss almost every aspect of life which is expected to guide humans through their life. In order to explore this divine knowledge have been engaged in conducting researches that processes the contents of the Quran using computers.

Recently, there has been an increasing amount of literature that focuses on research concerning the computation of Quran knowledge (Zameer and Abdul Siddiqi, 2013). Quite a large number of users from both Muslim and non-Muslim communities are now able to access various computing research, such as linguistic processing, information retrieval, semantic searches, question answering and data mining. We are still witnessing a growth in the computerization of Quranic content and the knowledge it contains.

The purpose of this study is to presents an overview of the previously reported computer-based academic researches on Quran domain. The paper reports previous computing researches conducted by researchers in Quran domain. These previously reported researches on application of computer in Quran domain are further classified according to the different computing field each research belongs. Finally, the paper suggests possible future directions in computer based Quran domain researches. This study will assist computer based researchers on Quran domain to further researches in the domain either based on previous work reported in this study or taking up research from the future direction suggested in this study.

LITERATURE REVIEW ON COMPUTER-BASED RESEARCH ON THE QURAN

This chapter provides a review of the various computer based academic research on the Quran reported previously. It contains the summary of the computer based academic researches and the classification of these researches according to the computing field they belongs is provided. In this study, the computer based academic research is classified into linguistic analysis of the Holy Quran content, data mining of Quran Data, knowledge representation of the Quran Data, information retrieval and semantic search of Quran content. The remaining part of the chapter will provide the detailed classification and explanation of the previously reported computer-based research on Quran domain.

Corresponding Author: Aliyu Rufai Yauri, Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang, Malaysia

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

Linguistic analysis of the content of the Holy Quran:

Linguistic analysis is one of the areas of computing that researchers of Quranic computation have been exploring over the years. Linguistic analysis is an analysis that employs paradigms from the field of linguistics that are designed to study language as a primary focus of inquiry, i.e., it is mainly concerned with using computers to model the natural language content (Howley *et al.*, 2013). Over the years, several works have been reported concerning linguistic processing of the content of the Holy Quran.

Various research (Dukes and Habash, 2010; Dror *et al.*, 2004) has reported different linguistic processing of Quranic content. In Dukes *et al.* (2010), a novel study of the linguistic processing of the Quran was proposed, which contained morphological analysis and part of speech tagging of the Quranic verses. The work presented the Treebank, a syntactic representation of the verses of the Holy Quran. The main objective of the research was to clearly analyze and show the meaning of the Quranic text.

Similarly, Dukes and Habash (2010) presented a syntactic annotation that showed the grammatical dependency of the Quran verse structures. The paper described an approach to morphological annotation of the Quranic Arabic content, which was initially verified manually and then computationally analyzed to find a morphological representation of the Quranic corpus in order to enable user searches of the verses of the Quran and see a morphological representation of the Quran verse selected.

Dror *et al.* (2004) presented a computational system for morphological analysis and annotation of the Holy Quran, which was mainly for research and teaching purposes. The work processed several queries from the Quran text that made reference to words and linguistic attributes. The system used a finite state toolbox to undertake a morphological analysis of the words in the Quran.

Thabet (2004) proposed a research that developed a new stemming technique for words in the Quran.

The approach presented a new stemming approach based on a light stemming technique that used a transliterated version of the Quran in Western script.

Quran data mining: Data mining is the process of discovering insightful, interesting and novel patterns, as well as descriptive, understandable and predictive models from large-scale data (Zaki and Meira Jr., 2014). Data mining techniques have been applied in various researches on the contents of the Quran.

In the area of data mining, Quranic computation work can be found in Ali (2012). The work proposed an approach that represented the Quran text corpus as a graph and applied frequent sub-path mining algorithm to generate frequent patterns. A frequent pattern is a pattern that frequently appears in a given data collection in order to observe knowledge in the set of data. In this study, they tried to observe useful rules of association and correlation within the Quran content, such as words

that frequently come together in the Quran or identical verses.

Thabet (2004) applied a data mining technique to develop a methodology to discover the thematic structure of the Quran. The research analyzed the text of the Quran and classified the chapters of the Quran based on the subject to identify the relationship between the chapters by abstracting lexical frequency data and then applying hierarchical cluster analysis to the data. The work is useful for various semantic applications.

Quran knowledge representations: Knowledge representation is the method used to encode knowledge in an intelligent system's knowledgebase. The object of knowledge representation is to express knowledge in computer-tractable form, such that it can be used to help intelligent systems perform well (habil.sc.ing. Janis Grundspenkis, sc.ing, Alla Anohina-Naumecca, nd). Quite a number of studies on the Knowledge representation of Quran content have been researched recently.

Knowledge representation research on Quran content can be found in the study of (Sharaf and Atwell, 2009), this research proposed a model for building a Quranic verb corpus using Frame Net frames. The work identified all the verbs from the Quran for building predicates from the Quran content. This enabled the development of Quran ontology. They studied the verbs in their context in the Quran and then compared them with matching frames evoked in an English frame.

Saad *et al.* (2009) presented a methodology that automatically generated ontology instances from the unstructured document of Al-Quran, Hadith and other related Islamic knowledge domains. Their system extracted concepts and built taxonomy of Islamic knowledge. Their main approach was the integration of ontology learning, ontology population and a text-mining framework for the extraction of information from Islamic knowledge sources. Their system mainly involved pattern extraction of various Islamic concepts.

Additionally, QurAna is a computational research into the Quran (Sharaf and Atwell, 2012). In this study, original Quran text was used to develop a large corpus for Quran related knowledge, in which personal pronouns were tagged with their antecedents. The antecedents were composed of a list of ontology concepts from the Holy Quran. The corpus can be used by researchers in several Quran-related applications, such as for training purposes, extracting empirical patterns and in rules for creating new anaphora resolution approaches.

Qursim was presented in Sharaf and Atwell (2012) and comprised system that linked related, semantically similar verses, which formed a large corpus from the Quranic data. The corpus could be used in computational linguistics and machine translation problem solving and other related research.

In Nassourou (2012), a support vector machine was used to categorize chapters of the Quran by their place of revelation, i.e., Mecca and Medina. Each of the chapters was categorized into clusters using a fuzzy-single linkage clustering technique, in order to correspond to the major phases of the life of Prophet Mohammad.

The Semantic Quran was a research project into Quran computation presented by Sherif and Ngomo (2013). They described a Semantic Quran dataset, which was a multilingual RDF representation of a translation of the Holy Quran. The dataset creation was based on integrating data from various semi-structured sources and aligning them to ontology designed to represent multilingual data from sources with a hierarchical structure. The Semantic Quran was designed to be friendly, with a natural-language interchange format containing explicit morpho-syntactic information on the terms used.

A system that used an automatic extraction method to acquire ontology from the Quran and Hadith domain text was developed (Saad and Salim, 2008). The technique mined ontologies from the Quran and Hadith. These ontologies constitute a specific vocabulary used to describe a particular model of the Islamic world, plus a set of explicit assumptions regarding the intended meaning of the words in the vocabulary that mainly focused on concepts relating to solat.

A computational model for representing the Arabic lexicon using ontology was developed based on the field theory of semantics for the linguistic domain, using data from Al-Quran (Al-Yahya *et al.*, 2010). In this study, the entire noun concept found in Al-Quran was used for the creation of their ontology.

Information retrieval: Information retrieval is another popular computing field that has witnessed several studies aimed at the easy and effective retrieval of Quranic content. Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers) (Christopher and Hinrich, 2009). Today there is a growing number of information retrieval systems through which the user can access Quran data both online and offline.

A Quran information retrieval system based on the use of formal methods was presented by Al-Gharaibeh *et al.* (2011). This research described the use of formal methods for Quran natural language processing search systems and used Z notation to express the formal specifications of the text-based search technique, synonym based search systems and stem based search systems used in the Quran search systems. The system was based on a keyword search in which the user is allowed to search using keywords to retrieve relevant verses from the Quran.

Noordin and Othman (2006) offered a system for retrieving Quran content and important knowledge derived from the Quran. The system surveyed and retrieved Quranic texts and knowledge from various websites that represent Quranic texts. The focus of the system design was on translation, texts, recitation, exegesis, Hadith and historical concepts or objects mentioned in the Quran. The system helps the user identify the meaning of verses in the Quran and their proper citation and knowledge generated from these verses.

Nassourou (2011) reported another system that supports the retrieval of Quranic content. In this system, the verses were clustered by chapter and a weight assigned to each cluster based on the number of verses it contained, so that users could easily identify the most relevant areas and identify places of revelation in the verses. Users could see complete results and select a section to zoom in on and click on an indicator to view a table containing verses with their corresponding English translation.

Semantic search of the Quran: With the introduction of semantic technology, research on Semantic Searches of Quran content is becoming increasingly popular. A semantic search is a data retrieval mechanism that integrates the capabilities of the Semantic Web and search engines in order to get more precise results than the current search engine. A semantic search enables computers to think, reason, manipulate data and provide humans with the information they need in the way that they need it (Kassim and Rahmany, 2009). It uses Semantic Web technology to manipulate and interpret a user's natural language queries and match them against information in the knowledgebase in order to extract semantic knowledge (Bettina *et al.*, 2010). This enables computers to accept complex queries use semantically annotated documents, reason and make inferences and finally, present good results to the user.

Khan *et al.* (2013) described a Quranic-based semantic search system that used ontology to search for important knowledge from the Holy Quran. This system used Quran ontology to enable users to search for living creation concepts that are mentioned in the Quran. Users can make queries about things mentioned in the Quran, such as animals, using SPARQL query language. The system requires users to be familiar with the complex syntax of SPARQL query language. The system is good for users who are familiar with the syntax of the SPARQL query language; however, users may find it difficult to use this system in order to retrieve the desired information.

Abbas (2009) improved on using structured syntax by enabling the use of natural language for querying. This project developed a bilingual (English/Arabic) comprehensive search tool for the Holy Quran. Their

work mainly involved a keyword search for concrete and abstract concepts found in the Holy Quran.

Shahzadi and Shaheen (2011) proposed a semantic network of religious repositories. This system helps in creating a customized semantic network, which enables semantic searches against any word or concept, a parser and a customized story builder. In this system, users query in order to get relevant information related to Islamic religious entities from the holy Quran.

Dataquest is a work (Qurat and Amna, 2011) based on a traditional keyword search and a pre-defined facet-based search system. Dataquest is a framework for modeling and retrieving knowledge from distributed knowledge sources primarily related to the Holy Quran and related scholarly texts, with the use of the Semantic Web, information extraction and natural language processing techniques. The documents are annotated using the domain ontology and then a semantic-based intelligent search. In this research, they collected all sorts of documents related to the Holy Quran that they

found on the Web. The system allows one to invoke a concept search of information related to Al-Quran.

Baqai *et al.* (2009) improved the facet search by assisting the user in formulating the user queries in order to retrieve information from the Holy Quran. The system was designed for improving Quran knowledge-sharing, storing, modeling, reasoning and retrieval from diverse Islamic domain sources. This study is based on small fragment queries and, thus, queries with multiple sentences cannot be answered by the system. This system has the limitation of not catering for ambiguity, i.e., there is no provision for disambiguating the user queries in case a different vocabulary from that in the knowledgebase is used.

Aliyu *et al.* (2013) used a Semantic Web technology application to present a framework that identifies historical concepts from the Holy Quran based on user natural language query. Applying Semantic Web technology to various heterogeneous

Table 1: Summary of the computer-based Quran research

| Research | Computing field | Quran computing methods | Output |
|--|--------------------------|--|--|
| Dukes and Habash (2010) | Linguistic analysis | Morphological annotation of Quran Arabic | Part of speech tagging of Quran words |
| Dror <i>et al.</i> (2004) | Linguistic analysis | Morphological analysis and annotation of the Quran | Quran morphological analyzer |
| Thabet (2004) | Linguistic analysis | Stemming of Quran transliteration words | Quran stemmer |
| Ali (2012) | Data mining | Sub-path mining algorithm for frequent pattern identification | Identify frequent verses come together in Quran |
| Thabet (2004) | Data mining | Classification of relatedness of chapters in the Quran | Cluster of related Quran chapters |
| Sharaf and Atwell (2009) | Knowledge representation | Building of lexical database for Quranic verb contents | Corpus of Quran verb content |
| Saad <i>et al.</i> (2009) | Knowledge representation | Extraction of information from Islamic knowledge sources | Quran ontology |
| Sharaf and Atwell (2012) | Knowledge representation | Annotation of Quranic pronouns | Quran corpus linked related, semantically similar verses |
| Sharaf and Atwell (2012) | Knowledge representation | Linking of related, semantically similar verse | Quran corpus |
| Nassourou (2012) | Knowledge representation | Categorizing of Quranic chapters by their place of revelation | Quran cluster of related chapters |
| Sherif and Ngomo (2013) | Knowledge representation | Integration and annotation of Quranic data from various sources | Semantic Quran dataset |
| Saad and Salim (2008) | Knowledge representation | Ontology extraction based on fuzzy logic | Quranic ontology |
| Al-Yahya and Al-Khalifa (2010) | Knowledge representation | Development of Quranic ontology | Quranic ontology |
| Al-Gharaibeh <i>et al.</i> (2011) | Information retrieval | Formal methods for Quranic natural language processing search systems | Quranic retrieval system |
| Noordin and Othman (2006) | Information retrieval | Retrieving Quranic content and important knowledge derived from the Quran | Quranic retrieval system |
| Nassourou (2011) | Information retrieval | Verses are clustered by chapter | Quranic retrieval system |
| Khan <i>et al.</i> (2013) | Semantic search | Search for living creation concepts that are mentioned in the Quran | Quranic concepts |
| Abbas (2009) | Semantic search | Development of a bilingual (English/Arabic) comprehensive search tool for the Holy Quran | Quranic concepts |
| Shahzadi and Shaheen (2011) | Semantic search | Customized semantic network | Quranic concepts |
| Dataquest is a work in Khan <i>et al.</i> (2013) | Semantic search | Modelling and retrieving knowledge from distributed knowledge sources | Quranic concepts |
| Aliyu <i>et al.</i> (2013) | Semantic search | Semantic web technology application to present a framework that identifies historical concepts from the holy Quran | Historical Quranic concepts |
| Aliyu <i>et al.</i> (2013) | Semantic search | Modelled Quranic ontology for Concept search | Quranic concepts |

Islamic information sources facilitates access to Islamic data with higher precision.

Aliyu *et al.* (2013) proposed an Ontology assisted semantic search system in the Quran domain. The system makes use of Quran ontology and various relationships and restrictions. This enables the user to semantically search for verses related to their query in Al-Quran. The system makes use of semantic Web technologies (ontology) to model Quran domain knowledge. Web Ontology Language (OWL) is the core element of the Semantic Web, which consists of statements that defined concepts, relationship and constraints. Ontologies are used to capture knowledge about some domains of interest by describing concepts in the domains and relationships that are held between those concepts.

Summary on the classification of computer-based Quran research: In this section, a comprehensive overview of the computational tools for the Quran reported previously is presented. Table 1 shows summary of the research by classifying each research based on the computing field to which it belongs.

Table 1 presents a summary of the various computer related research on the Quran. The table also shows the classification of the research according to the different computing fields to which they belong. Various fields of computing have been applied to the computation of Quran content, such as linguistic processing, data mining, information retrieval and semantic search.

DISCUSSION

With the growing interest in exploration of Quranic content, future direction on the Quran will be more in terms of knowledge representation and semantic search systems. There is a need for building a richer Quran knowledgebase that will support a variety of Quranic applications and facilitate the retrieval of Quranic knowledge. Another future direction we intend to take is to incorporate the Hadith ontology by merging the Quranic and Hadith ontology into the semantic search system, as Islamic related queries may not be answered by the contents in the Quran alone. Therefore Hadith, being the second source of Islamic law, will improve the ontology to a much richer capacity that can handle more queries than using the Quran alone. By doing so, a semantic search will be able to retrieve the corresponding Quran verses as well as related Hadith.

Developing more intelligent search systems will be another major future aim for research on the Quran. Most of the existing systems are based on concept search that do not support deep semantic analyses of the query. For example, most users simply want a yes or no answer, such as is a Muslim allowed to marry a Christian? The user may simply need to know yes or no. In fact even some of the queries that fail or have no answer could simply be answered by a yes/no answer

without failing to return anything. Therefore, part of the future research challenge that is not addressed in this thesis is Boolean queries. The system in this thesis does not cover yes/ no or true/false questions which are quite popular in Islamic related queries. Future work should incorporate support for yes/no queries so that users can ask queries to which the target answer is yes or no, or true or false.

CONCLUSION

The paper presents the review paper on the previously reported computer base researches on Quran. The paper reviewed previous reported researches works conducted by researchers on the Quran domain. These reviewed researches are then classified according to the various computing field each research belongs. Furthermore, the paper suggested possible future directions on the computer-based researches on Quran domain. We believe the paper will assist computer based researchers in Quran domain to carry on further researches based on the previous researches reported in this study or based on the future directions as suggested in the study.

REFERENCES

- Abbas, N.H., 2009. Quran 'search for a concept' tool and website. M.Sc. Thesis, The University of Leeds, UK.
- Al-Gharaibeh, A., A. Al-Taani and I. Alsmadi, 2011. The usage of formal methods in Quran search system. Proceeding of the ICICS 2011. Retrieved from: http://www.icics.info/icics/Proceeding/ICICS_paper/61.pdf.
- Ali, I., 2012. Application of a mining algorithm to finding frequent patterns in a text corpus: A case study of the Arabic. *Int. J. Softw. Eng. Appl.*, 6(2012): 127-134.
- Aliyu, R.Y., A.K. Rabiah, Azreen Azman and A.A.M. Masrah, 2013. Quranic verse extraction base on concepts using OWL-DL ontology. *Res. J. Appl. Sci. Eng. Technol.*, 6(23): 4492-4498.
- Aliyu, R.Y., A.K. Rabiah, A. Azreen and A.A.M. Masrah, 2014. Semantic web application for historical concepts search in Al-Quran. *Int. J. Islam. Appl. Comp. Sci. Technol.*, 2(2): 1-7.
- Al-Yahya, M. and 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(2): 21.
- Baqai, S., A. Basharat, H. Khalid, A. Hassan and S. Zafar, 2009. Leveraging semantic web technologies for standardized knowledge modelling and retrieval from the Holy Quran and religious texts. In: ACM Special Interest Group on Artificial Intelligence. Proceedings of the 7th International Conference on Frontiers of Information Technology, New York.

- Bettina, F., G. Giorgio, G. Georg and L. Thomas, 2010. Semantic web search based on ontological conjunctive queries. *Lect. Notes Comput. Sc.*, 5956: 153-172.
- Christopher, D. and P. Hinrich, 2009. *Introduction to Information Retrieval*. Cambridge University Press, New York.
- Dror, J., S. Dudu, T. Rafi and W. Shuly, 2004. Morphological analysis of the Qur'an. *Lit. Linguist. Comput.*, 19(4): 431-452.
- Dukes, K. and N. Habash, 2010. Morphological annotation of Quranic arabic. *Proceeding of the Language Resources and Evaluation Conference (LREC)*, Valletta, Malta.
- Howley, I., E. Mayfield and C.P. Rosé, 2013. Linguistic Analysis Methods for Studying Small Groups. In: Hmelo-Silver, C., A. O'Donnell, C. Chan and C. Chin (Eds.), *International Handbook of Collaborative Learning*. Taylor and Francis Inc, NY, pp: 184-202.
- Kassim, J.M. and M. Rahmany, 2009. Introduction to semantic search engine. *Proceeding of the International Conference on Electrical Engineering and Informatics, Selangor, Malaysia*, pp: 380-386.
- Khan, H., S. Saqlain, M. Shoaib and M. Sher, 2013. Ontology based semantic search in Holy Quran. *Int. J. Future Comp. Commun.*, V2: 229.
- Noordin, M. and R. Othman, 2006. An information retrieval system for Quranic texts: A proposed system design. *Proceedings of the 2nd Information and Communication Technologies (ICTTA'06)*, Damascus, 1: 1704-1709.
- Nassourou, M., 2011. Assisting Analysis and Understanding of Quran Search Results with Interactive Scatter Plots and Tables. Retrieved from: <http://opus.bibliothek.uni-wuerzburg.de/volltexte/2011/5584/>.
- Nassourou, M., 2012. Using machine learning algorithms for categorizing Quranic chapters by major phases of Prophet Mohammad's messengership. *Int. J. Info. Commun. Technol. Res.*, 2(11): 863-871.
- Qurat, A. and B. Amna, 2011. Ontology driven information extraction from the Holy Qur'an related documents. *Proceedings of the 26th IEEE Students' Seminar of Pakistan Navy Engineering College National University of Sciences and Technology, Pakistan*.
- Saad, S. and N. Salim, 2008. Methodology of ontology extraction for islamic knowledge text. *Postgraduate Annual Research Seminar*. Retrieved from: <http://comp.utm.my/pars/files/2013/04/Methodology-of-Ontology-Extraction-for-Islamic-Knowledge-Text.pdf>.
- Saad, S., N. Salim and H. Zainal, 2009. Pattern extraction for Islamic. *Proceeding of the International Conference on Electrical Engineering and Informatics, Selangor, Malaysia*
- Shahzadi, N. and A. Shaheen, 2011. Semantic network based semantic search of religious repository. *Int. J. Comput. Appl.*, 36(9): 975-8887.
- Sharaf, A. and E. Atwell, 2009. Knowledge representation of the Quran through frame semantics: A corpus-based approach. *Proceedings of the 5th Corpus Linguistics Conference, University of Liverpool*, pp: 12.
- Sharaf, A.M. and E.S. Atwell, 2011. QurAna: Corpus of the Quran annotated with pronominal anaphora. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*, European Language Resources Association (ELRA).
- Sharaf, A.M. and E.S. Atwell, 2012. QurSim: A corpus for evaluation of relatedness in short texts. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*, European Language Resources Association (ELRA).
- Sherif, M.A. and A.N. Ngomo, 2013. Semantic Quran: A multilingual resource for natural-language processing. *Semantic Web J.*, 6(4): 339-345.
- Zaki, M.J. and W. Meira Jr., 2014. *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press, Cambridge, pp: 12.
- Zameer, A.A. and A.S. Abdul, 2013. A programming approach for the digital Quran applications. *Int. J. Eng. Comput. Sci.*, 13(5).