Research Article
Quranic Verse Extraction base on Concepts using OWL-DL Ontology

Aliyu Rufai Yauri, Rabiah Abdul Kadir, Azreen Azman and Masrah Azrifah Azmi Murad
Faculty of Computer Science and Information Technology, Universiti Putra Malaysia,
43400 UPM Serdang, Selangor, Malaysia

Abstract: In recent years, there has been a global growing demand for Islamic knowledge by both Muslims and non-Muslims. This has brought about a number of automated applications that ease the retrieval of knowledge from the Holy Book, being the major source of Knowledge in Islam. However, the current retrieval methods in the Quranic domain lack adequate semantic search capabilities; they are mostly based on the keywords matching approach. There is a lack of adequate linked data to provide a better description of concepts found in the Holy Quran. In this study we propose an Ontology assisted semantic search system in the Qur'an domain. The system makes use of Quran ontology and various relationships and restrictions. This will enable the user to semantically search for verses related to their query in Al-Quran. The system has improved the search capability of the Holy Quran knowledge to 95 percent accuracy level.

Keywords: Islamic knowledge, keyword matching, ontology, semantic technology

INTRODUCTION

Over the years, there has been growing interest concerning Islamic knowledge by both Muslims and non-Muslims, especially of the Quran. This is because of the belief that the Quran is the main source of knowledge, wisdom and law for Muslims. People in search of what Islam is and what it is not, want easier ways in which they can gain access to answers to their various queries. However, Islamic queries, doubts and worries are best answered by the Holy Quran. Since first revelation, the Holy Quran remains among the most influential books that exist (Qurat ul Ain and Amna, 2011). The first revelation was believed to be revealed to Prophet Muhammad (SAW) through Angel Jibreel directly from God. The Quran is a book that covers wide range of knowledge. It consists of 114 chapters with 6236 verses covering many themes and concepts that make up the divine knowledge and law. In spite of its great influence, very little is known about the Holy Quran. This may be because of the nature in which the Quran is represented. A considerable amount of time is needed before one can go through the Quran and acquire knowledge from its divine wisdom. Researchers, including interpreters, jurisprudents, theologians and Hadith experts have all invested a lifetime of assets of scientific and technical views in contemplating this divine miracle (Qurat ul Ain and Amna, 2011). However, data about the Quran are not computationally linked enough to provide a better description of the knowledge involved therein. This has

made searches of the Quran more of keyword matching and keyword based search without any semantic retrieves of more irrelevant Information. User navigation through a large set of independent sources often leads to users being lost. In addition, large amounts of knowledge are left underutilized (Qurat ul Ain and Amna, 2011). Research has stressed that, there has been no significant effort to facilitate semantic searching and linking of the verses of the Holy Qur'an with the relevant scholarly and authentic literature, to produce a carefully, well-informed interpretation (Sumayya *et al*., 2009). This has been the main motivation of this research.

In this study, we have proposed a Model that makes use of semantic Web technologies (ontology) to model Quran domain knowledge. Web Ontology language (OWL) is the core element of 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. Our system is composed of a Quran concepts semantic search model. We used ontology to link the concepts found in the Holy Quran with various relationships that exist between these concepts. We linked the concepts with various verses of the Holy Quran where they are referred to, discussed or mentioned. We used the existing Ontology from the University of Leeds, UK, which is graphically represented in a network of 300 linked concepts with 350 relations. We used protégée

Corresponding Author: Aliyu Rufai Yauri, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

software to store this Quran ontology. The tool provides a reasoning capability to make inference over our linked data. With this, we can retrieve Concepts and their various linked relationships. We will provide in detail explanations concerning how we created our Model in the later part of the study.

LITERATURE REVIEW

Semantic technology, in other words, a web of linked data, is a mechanism that provides an opportunity for deriving knowledge from a particular domain interest. The main objective of the Semantic Web is to provide machine process able web content. The Semantic Web is described to be an extension of the current Web in which information is given a well-defined meaning, thereby better enabling computers and people to work in cooperation (Orgun *et al*., 2006). One of the core elements of the semantic Web is the Ontology. Ontology conceptualizes a domain into a machine-readable format. Ontology is a mechanism through which knowledge is represented in the form of concepts, nodes that link the relationships between these concepts and restrictions.

An ontology-based information extraction system is seen as a system that processes unstructured or semi-structured natural language text through a mechanism guided by ontologies to extract certain types of information and presents the output using ontology (Daya and Dejing, 2010). Ontology creation may be done either manually, automatically or semi-automatically depending on the researcher's choice and research condition. The vision of ontology learning includes a number of complementary disciplines, such as machine learning, natural language processing, data mining, adaptive information extraction and so on that feed on different types of unstructured, semi structured and fully structured data and support semi automatic and also cooperative ontology engineering (Ontology Research Team, UITM, 2011). However, the majority of the constructed ontologies today are done manually or semi-automatically because of the complexity of some concepts and the maintenance may need human Intervention (Rizwan, 2010; Hui and Jamie, 2008). Other ontology based information extraction work exists in other domains like business, tourism, education, health and so on. Business ontology based information extraction can be found in the work of Horacio *et al.* (2007). In this research, an ontology-based Information extraction system for business Intelligence was developed. This intelligent system is used to gather information pertaining to international companies and countries/regions. The technology extracts relevant semantic information, which is expressed in the ontology and is used in business intelligence processes, such as risk management, IT operational risk management as well as internationalization.

A system that uses an automatic extraction method to acquire ontology from the Quran and Hadith domain text was developed (Ontology Research Team, UITM, 2011). They developed a technique that mines ontologies form the Quran and Hadith. Their ontology constitutes 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 where they 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-Yahy *et al*., 2010). In this study the entire noun concept found in Al-Quran were used for the creation of their ontology. Dataquest is a framework for modelling and retrieving knowledge from distributed knowledge sources primarily related to the Holy Qur'an and related scholarly texts, with the use of Semantic Web, Information Extraction and Natural Language Processing techniques (Ontology, 2011). The documents are annotated using the domain ontology and then a semantic based intelligent search. In this research, they collected all sorts of documents found on the Web that are related to the Holy Quran. The system will allow you to invoke a keyword search of information related to Al-Quran. In this case, the research does not focus on Al-Quran alone but also related documents. Therefore, in this case not all the concepts and verses may be represented in the search systems.

Another related study is the "Quran Search for a Concept "Tool and Website (Noorhan, 2009). This project developed a bilingual (English/Arabic) comprehensive search tool for the Holy Quran. There study was mainly a keyword search for concrete concepts as well as for abstract concepts that are found in the Holy Quran. (Horacio *et al*., 2007) presented a methodology that will automatically generate ontology instances form the unstructured document of Al-Quran, hadith and other related Islamic knowledge domains. Their system extracts concepts and builds the 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 was mainly pattern extraction of various Islamic concepts.

Sufism (Islamic mysticism) domain ontology was developed in order to extract data relating to Sufism (Al-Yahy *et al*., 2010). The system enables the user to extract various information that is related to Islamic mysticism Sufism. Leveraging semantic Web technology for standardized knowledge modelling and retrieval from the holy Quran and religious text is another related study done in this context (Sumayya *et al*., 2009). The system was designed for improving Quran knowledge sharing, storing, modelling reasoning and retrieval from diverse Islamic domain sources.

Most of the Quran search systems works are based on key word search, in which the user needs to have the correct keywords in order to search for desired information. This has left large amount of Quran information underutilized.

SYSTEM MODEL

To improve the search capability of the current Quran knowledge search systems, we propose a semantic search system, which goes beyond the traditional keyword search. This system enables users to semantically search for verses relating to concepts found in the Quran and their corresponding relationships. The System includes the creation of a Quran ontology model, which is composed of important Quran concepts found in the Holy Quran and the annotation of these concepts with various properties and restrictions. The semantic search model enables the user to semantically search for the desired knowledge from the Holy Quran. Figure 1 shows the graphical representation of the designed model.

As shown in Fig. 1 signifies a typical semantic search framework that takes the user query as input using the Manchester owl query language. The user query is then matched against the knowledge base comprising Quran ontology annotated concepts by the query model. The query model uses an inference engine to make an inference based on the user query using annotated ontology properties to come up with an answer to the user's query. We will explain in detail in the remaining part of the study.

Quran ontology model: We use existing Quran ontology from Leeds University United Kingdom. Leeds Quran ontology uses knowledge representation to define the key concepts in the Quran and shows the relationships between these concepts using predicate logic. We adapted the ontology reuse method for our system. Ontology reuse can be defined as the process in which available ontological knowledge is used as input to generate new ontologies (Elena *et al*., 2003). Ontology reuse gives the opportunity of improving the capabilities and knowledge of the existing ontology. Our ontology was built from the already existing Quran ontology built at the University of Leeds United Kingdom. As mentioned earlier, Leeds Ontology comprises 300 concepts and about 350 relationships linking the concepts. These 300 concepts include noun concepts mentioned in al-Qur'an. We used the 300 noun concepts identified by Leeds ontology as the scope of our research. Relationships in other word properties give more description to the concepts and the restrictions show how the properties will be used. These properties and restrictions will be used by the search model to infer over a given query. However the existing ontology only has 350 relationships, which lack sufficient description of concepts for our semantic search. Therefore, we have contributed by designing more relationships and restrictions using sources from the Quran, hadith, Islamic websites and other Islamic related resources. We will see the breakdown of how our ontology is represented below.

We have added a few concepts that enabled us to incorporate the complete Quran into our Quran
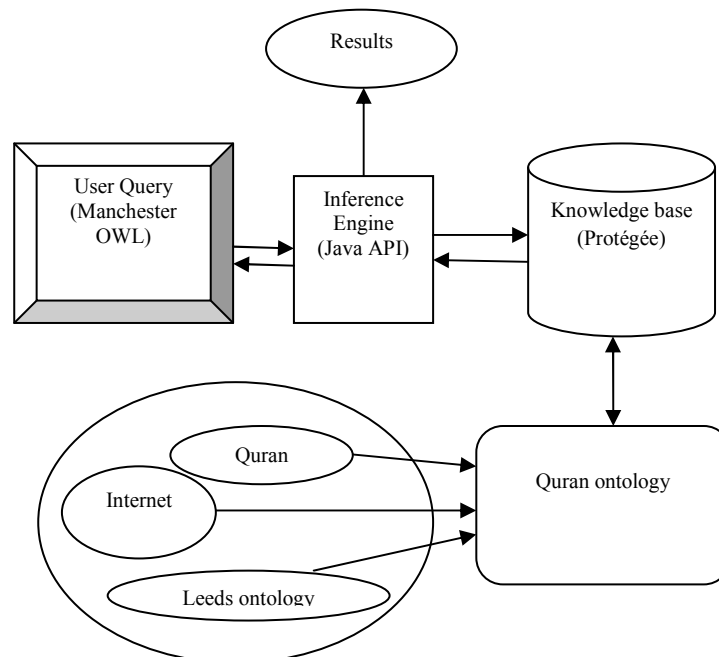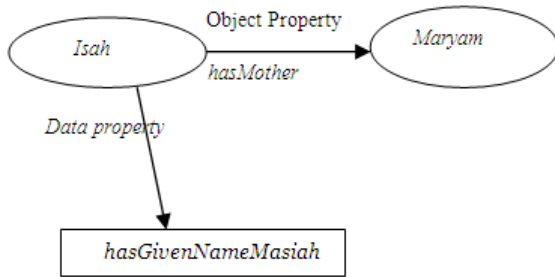

Fig. 1: Research framework
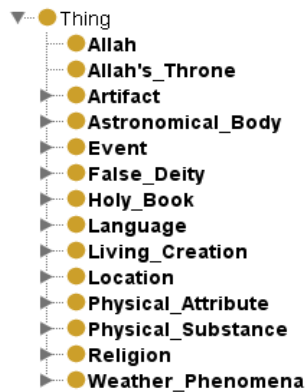
Fig. 2: Properties demonstration

Fig. 3: Quran ontology concepts

ontology. We already have the Quran as a concept and added sub-concepts to the Quran concept as we can see below.

Concept: Quran:
Sub-concept: (juzz, chapters, verses): This will support us by including Quran verses in out generated triple. Likewise, the 350 relationships are not adequate to handle the many possible queries by the users. As mentioned earlier, in order for our system to cope with more user queries, we introduced more relationships between concepts and imposed more restrictions in the ontology. Building more relationships and posing more restrictions will help our system cope with many more possible user queries. We have built 650 additional relationships. These additional concepts and relationships are used by our search system Model to make inference over query. When the ontology is well created, the assertion capabilities help to semantically search for the desired knowledge. We used protégée ontology editor to store our ontology and linked the concept through relationships. Protégé also allows us to insert some restrictions over the concept. The information below will give us more insight into how our Ontology was developed including the concepts, relationships and retractions.

Concept: Concepts, also known as classes, are a special kind of resource representation where resources that share common characteristics or are similar in some

way are grouped together. Concepts are described using formal mathematical descriptions that state precisely the requirements for membership of the class. For instance the concept "Prophet" is classed with members (Muhammad, Isa, Ibrahim, Moses among others). We have about 300 concepts including the sub-concepts in our Ontology as seen in Fig. 3. A concept may also be a sub-concept of a particular concept. For example, the concept "Prophet" is a sub-concept of "Human" in the Quran ontology.

Sub concept: Individual members of a class or concept are referred to as sub concepts of the class. Example in the class above, (Muhammad, Isa, Ibrahim, Moses) is sub concepts of a class "Prophet". A sub concept may also be a concept or class of another sub concept.

Equivalents: Equivalents are used to map two concepts as being one entity. For example we may say "Jesus" is-equivalent-to "Isa"; this signifies that, whenever the user uses either Jesus or Isah, it should be represented as one entity. So irrespective of whether the user used Jesus or Isa the system will retrieve the same information. Therefore, although the user may not necessarily search for the exact keyword, the equivalent class may help the user find what he needs without using the exact keyword. This helps us deal with the ambiguity in natural language.

Property: Properties are used to describe resources; they give more description to a class/concept or subclass. They are used to serve as annotators to concepts. This provides a better description about data in ontology. They provide the relationship between a concept and the data about the concept. Properties provide Concepts and concept inheritance relationship provides the semantics. Concept inheritance relationships are linked through the properties of the concepts. Properties are object property and data property. An object property describes the relationship between concept and concepts. While data property is used for relationship between concept and its literal. A description of the properties can be seen in the Fig. 2

In Fig. 2 shows the graphical representation of object and data property. From the graph, both Isah and Maryam are concepts in the Quran ontology. The has Mother relation is an Object property, which associate Isah and Maryam concepts. The property has Given Name Masiah is a data property that associates the concept Isah with its literal. We have built many of these relationships and apply many restrictions in our Ontology creation. We have manually used protégée to create our ontology. We used a top down ontology development process, the development process starts with the definition of the most general concepts in the domain and, subsequently, their various sub concepts, their necessary relationships and we related them to various verses in which they are mentioned in the Quran.

Inverse: In every object property there may be a corresponding inverse property. For instance, we have seen the object property has Mother before, between Isah and Maryam relationship; therefore, there will be a corresponding inverse property has Child. So, if Isah has Mother Maryam, then because of the inverse property we can infer that Maryam has Child Isah.

From Fig. 3 shows how Quran ontology concepts are represented in protégée ontology editor. Thing is the main concept or class, which is the standard starting point of protégée. Our Quran ontology is composed of 14 concepts/classes. These concepts are represented in protégée as a sub concept of Thing. Each of the 14 concepts serves as a class to other abstract sub concepts. Protégé provides a good environment for the creation or storing of concepts, concept inheritance relationships and concept instances. All of these concepts are linked together with various properties. This provides a better description for the Quranic concepts for our semantic search model.

Semantic query model: Ontology in protégée is represented in a RDF (Resource Description Frame Work) graph. RDF is a formal language for describing structured information. It enables applications to exchange information on the web while still preserving their original meaning. RDF enables us to represent our ontology using a set of nodes that are linked by a directed edge. These nodes, which are linked edges, are represented as triple base representation to represent our Quran ontology. Each triple in this case can be assigned a web address. The nodes represent various concepts found in the Quran and the edges are the relationship or properties of the concepts as in Fig. 2. As mentioned earlier, concept inheritance relationship provides semantically annotated information, which gives a better description of the concepts. These annotated concepts enable us to semantically retrieve important knowledge from the Quran. The Query model is an ongoing research, but for the purpose of this study we will use protégée built in reasoner which is used to access an external DIG compliant reasoner, thereby enabling inferences to be made about the classes and individuals in the ontology. The DL Query tab in protégée provides a good feature for searching a classified ontology. It is a standard Protégé plug-in. The query language that is supported by the plug-in is based on the Manchester OWL syntax.

The Manchester OWL Syntax is a new syntax that has been designed for writing OWL class expressions (Ontology of Quran Concept). It is influenced by both the OWL Abstract Syntax and the DL style syntax, which uses description logic symbols, such as the universal quantifier ($\forall$) or the existential quantifier ($\exists$). A quantifier is "an operator that limits the variables of a proposition".

Proposition in logic is viewed as a statement that is either true or false. Propositional Logic is a static discipline of statements that lack semantic content.

Example:

$P \rightarrow$ Muhammed is a Messenger
$Q \rightarrow$ The list of Prophets includes Muhammed
$R \rightarrow$ Abu $-$ Bakr is a prophet

From the above statements, both P and Q statements are true. While Statement R is false, Abu-Bakr is a companion of a Prophet not a Prophet.

The universal quantifier, which is represented by an upside-down A: $\forall$ means "for all":

- Let $P(x) = x + 1 > x$
- $\forall x\, P(x)$

For all values of x, $P(x)$ is true"
For all values in Manchester Owl syntax is represented as "only" all Values From. For example "Maryam has Child only Isa". In order for us to infer that this statement is true, it must be true for all cases. So In order to prove that a universal quantification is true, it must be shown for ALL cases. In addition, in order to prove that a universal quantification is false, it must be shown to be false for only ONE case.

The existential quantifier ($\exists$) is represented by a backwards E represented as $\exists$ meaning "there exists".

- We can state the following:
- $\exists x\, P(x)$

There exists (a value of) x such that $P(x)$ is true"
The syntax uses existential quantifiers to represent that one of the statements or conditions is true. Existential quantifiers are represented as $\exists$ some values from. Muhammad has Children some Male. So every male child of Muhammad can be retrieved. In order to show an existential quantification is true, you only have to find ONE value. In addition, in order to show an existential quantification is false, you have to show it's false for ALL values.

Propositional logic is the study of how simple propositions can come together to make more compound propositions. However, form compound propositions by using logical connectives (logical operators) to form propositional "molecules".

In this query expression language Manchester owl syntax, these logical connectors in Table 1 such as $\wedge$, $\neg$ $\vee$, are replaced by keywords, such as "and", "not" and "or". In addition, the quantifier symbols $\forall$ and $\exists$ are replaced by keywords, such as "some" and "only". This helps the inference engine make inferences over the query receive to retrieve the corresponding answer to such query.

For example, (Human and (is-a prophet) and (is-a Messenger), this refers to those humans who are both Prophets and Messengers of Allah. (Human and (is-a prophet) or (is-a Messenger), in this case either Prophet

Table 1: Logical connectives

Operator	Symbol	Usage
Negation	¬	Not
Conjunction	∧	And
Disjunction	∨	Or
Exclusive or	⊕	Xor
Conditional	→	If, then
Biconditional	↔	Iff

Table 2: Sample queries set

Sample queries
Where do Muslims face during prayers?
Where does Quran name Prophet Muhammad as Ahmed?
Who is the last of the prophets?
Who is the angel that communicates between Prophet Muhammad and Allah?
Why do Muslims marry more than one wife?
Is the consumption of alcohol prohibited?
Why should a person only follow Islam?
Can a Muslim man marry a Christian or a Jew?
Does Islam promote violence and bloodshed?
What are the halal foods mentioned in Quran?
What are the non halal foods mentioned in the Quran?
Is Christianity mentioned in the Quran?
Will Christians enter paradise?
What are the historical lands mentioned in the Quran?

Table 3: Results of recall, precision for the experiments

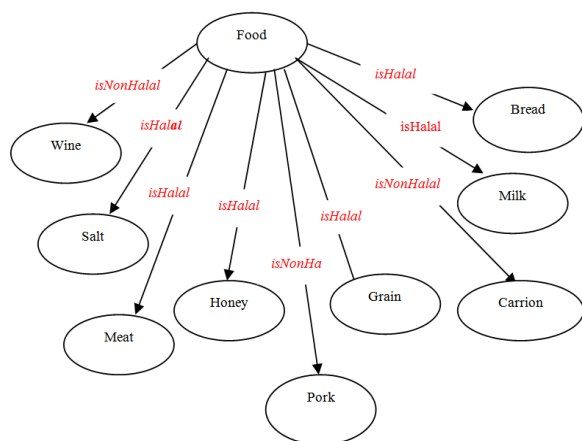| Queries	Average precision	Average recall
40 queries	0.94	0.96

Fig. 4: Query: is Halal some Food

or Messenger satisfied the condition. We will see more examples of how we use these logical connectives for our search.

EXPERIMENTS

In our model, we have categorised the queries we used into simple and complex queries. For our experiment, we used 40 queries asked by ordinary people.

The example in Fig. 4 is a simple query; the user wants to retrieve all the halal (non-prohibited food) foods mentioned in the Quran. Query: "is Halal some Food", food mentioned in Quran that is halal (prohibited food). We therefore show the verses.

Query1: Is halal some food: The Fig. 4 represents how our knowledge base is represented in a RDF graph. The annotated concepts involving the relationships and retractions generated many triples of subject, predicated and object. The system makes use of the property or restriction that a concept is annotated with to make inference of all those concepts that are food and is halal. So, from the graph, we could see that we have (bread, salt, meat, honey, Grain, Milk, Bread) as halal food that are mentioned in the Quran as requested by the user. Therefore we are able to retrieve the verses through the identified concepts.

We can ask more complex queries like:
Query 2: Angel and (communicate some Muhammad and Allah): Here, the user query is

"which angel Allah use to communicate with Prophet Muhammad". The system follows the same logical concept to retrieve "Jibreel" and corresponding verse where Jibreel is discussed in the Quran (Table 2).

Evaluation: For evaluation of the effectiveness of our system, we used a popular precision and recall technique:

$$\text{Precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

$$\text{Recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{total relevant documents}\}|}$$

Precision and recall methods were used to measure the effectiveness of the search system. The recall measures of how many of the relevant documents were retrieved; while precision measures of how many of the retrieved documents were relevant (Table 3).

From our experiment based on the 40 queries used for the experiment we found that 95 answers given by the system are satisfactory. The proposed approach retrieved 94% correct answers and 6% of the retrieved information is not relevant. While 96% of the expected answers for the queries were retrieved missing only 4%. This is very good result in the retrieval of Holy Quran content compared to tradition keyword search.

CONCLUSION

This study presented a semantic search system for Quranic knowledge using ontology assertion capability. Muslims and non-Muslims can semantically retrieve Verses in relation to their queries relating to the Holy Quran. The system has shown a significant level of effectiveness for the retrieval of Quranic Knowledge. Our System has proven to have a 95% level of accuracy. Our feature study combines both the Quran

Ontology and Hadith ontology in order to build a system capable of handling more possible user Queries.

REFERENCES

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(2C).

Daya, C.W. and D. Dejing, 2010. Ontology-based information extraction: An introduction and a survey of current approaches. J. Inform. Sci., 36: 306-323.

Elena, P.B., M. Malgorzata and T. Robert, 2003. Case Studies on Ontology Reuse. Freie Universitat Berlin, Germany. Electronic Publication, Digital Object Identifiers.

Horacio, S., F. Adam, M. Diana and B. Kalina, 2007. Ontology-based information extraction for business intelligent. Proceeding of the 6th International and 2nd Asian Semantic Web Conference. Springer, Berlin.

Hui, Y. and C. Jamie, 2008. Ontology generation for large email collections. Proceeding of the 9th Annual International Digital Government Research Conference. Georgia, USA.

Noorhan, H.A., 2009. Quran 'search for a concept' tool and website. M.Sc. Thesis, The University of Leeds School of Computing.

Ontology, 2011. Retrieved from: http://en.wikipedia.org/wiki/Ontology (Accessed on: September 27, 2011).

Ontology of Quran Concept, Leeds University, UK. Retrieved from: http://corpus.quran.com/ontology.jsp.

Ontology Research Team, UITM, 2011. Ontology Extraction Tool for Quran and Hadith Using Fuzzy-Swarm Algorithms. Retrieved from: http://webs. cs.utm.my/ontologyGroup/index.php/proposal1 (Accessed on: September 25, 2011).

Orgun, B., M. Dras and A. Nayak, 2006. Approaches for semantic interoperability between domain ontologies. Proceeding of the Australasian Ontology Workshop (AOW). Hobart, Australia.

Qurat ul Ain and B. Amna, 2011. Ontology driven Information Extraction from the Holy Qur'an related Documents. Proceeding of the 26th IEEEP Students' Seminar.

Rizwan, I., 2010. Ontology development for "Sufism" domain. Proceeding of the International Conference on Software and Computing Technology (ICSCT).

Sumayya, B., B. Amna, K. Hira, H. Amna and Z. Shehneela, 2009. Leveraging semantic web technologies for standardized knowledge modeling and retrieval from the Holy Quran and religious texts. Proceeding of the 7th International Conference on Frontiers of Information Technology. New York.