

## Towards Context-Sensitive Domain of Islamic Knowledge Ontology Extraction

Saidah Saad  
*Faculty of Info.Sc. and  
Tech  
Universiti Kebangsaan  
Malaysia, UKM Bangi,*

Naomie Salim  
*Faculty of Comp. Science  
and Info. Systems  
Universiti Teknologi  
Malaysia, Skudai, Johor*

Hakim Zainal  
*Faculty of Islamic Study  
Universiti Kebangsaan  
Malaysia,  
UKM Bangi, Selangor*

### Abstract

*Ontology is one of the essential topics in the scope of an important area of current computer science and Semantic Web. Ontologies present well defined, straightforward and standardized form of the repositories (vast and reliable knowledge) where it can be interoperable and machine understandable. There are many possible utilization of ontologies from automatic annotation of web resources to domain representation and reasoning task. Ontology is an effective conceptualism used for the semantic web. However there is none of the research try to construct an ontology from Islamic knowledge which consist of Holy Quran, Hadiths and etc. Therefore as a first stage, in this paper we try to propose a simple methodology in order to extract a concept based on Al-Quran. Finally, we discuss about the experiment that have been conducted.*

### 1. Introduction

Most information resources nowadays on Islamic Knowledge such as translations of Al-Quran and Hadiths articles, that are available on the internet are considered as natural language text documents. However, the classical human readable data resources (like electronic books or web sites) present serious problems for achieving machine interoperability. This is why a structured way for representing information is required and *ontologies* (machine-processable representations that contain the semantic information of a domain), can be very useful.

Every researchers working in the areas of information retrieval, intelligent integration information, education, e-commerce and semantic web seek to find categories of textual resources using various fully automatic methods and, others working mainly in the areas of thesauri [2] and ontologies [7] predefine conceptual structures and assign metadata to the documents that confirm to these conceptual structures.

On one hand, the categorisation of documents is (comparatively) cheap, but the quality of its document categorisation for larger sets of target categories as well as the understandability of its results is often quite low. On the other hand, the quality of manual metadata may be very good, but the cost of building an ontology and adding manual metadata typically are one or several orders of magnitude higher than for automatic approaches. To gain both advantages, while diminishing both their drawbacks at once, the authors of this paper propose an approach of integrated ontology learning, ontology population and text mining framework in order to build the ABox component and used the available Quranic indexing as a TBox (metadata). Therefore, a methodology will be presented to extract information from the Islamic Knowledge sources to build an ontology for a given domain. Moreover the most of the representative documents for the ontology concept were determined by the domain experts. Throughout this research, translations on the meaning of the AL-Quran at DEED collection by [5] have been referred to by focusing on prayer or '*salat*' as a sample.

The aim of the paper is to describe on the Islamic concept and a methodology which is composed of rules and techniques of extracting concept and to build the ontology of Islamic Knowledge. Following that, discussions will be made on the experiments that have been conducted based on the aforementioned rules.

### 2. Islamic Concept

"This Concordance of the Quran satisfies a paramount need of those - and there are millions of them - who have no command of the Arabic language and yet desire to understand the Quran." [4]

Conceived as a reference work based on fourteen centuries of Islamic scholarship, The Islamic Knowledge Ontology will bring to the English-speaking world a unique semantic meaning in order to understand the Islamic Messages as described in the Quran and Hadith . Its potential readers include academic scholars, whether specialising in the field of Quranic studies or working more generally in

other disciplines, as well as Muslims and non-Muslims looking for an authentic source of in-depth and scholarly knowledge on the Quran and its message.

For both Muslims and non-Muslims, the need for a comprehensive meaning of the Quran has been in existence since long before. The need has become an essential for the Muslims because only about 20 percent of them now understand the language of the Quran. The inability to directly access the multi-layered linguistic richness of the original text of the Quran and the number of non-Muslims who have actually opened one of the many available translations of the Quran have increased exponentially in recent years. After reading the first few pages, however, a vast majority of these readers find it impossible to continue, for they encounter a text unlike any they have ever read [11]. In the multi-layered coherence of the Quran, all its themes emerge in short passages, creating an inimitable interplay between its imagery, oaths, parables, chronicles, warnings, and glad tidings. With so many elements of text coalescing, separating, reuniting, and reemphasising each other at numerous levels, the result can be a total incomprehensibility and confusion. However, there is no doubt that the content of the Quran which is of paramount importance in the life of Islam and the Muslims, the language and the vocabulary of revelation, both divinely chosen, are also important. When a Muslim considers the Quran, three elements are of significance to him: what was revealed, the language in which it was revealed, and the vocabulary that was selected to convey the revelation [4] with the semantic meaning.

In order to develop a complete Islamic Ontology, Al-Quran will be the main source because it is divinely chosen and described as a way of life to human beings. But in order to understand, organise or work with Al-Quran, there are prerequisites to be considered which are [6]:

- i. Prioritising on the interpretation of al-Quran by using its own verses: This is because if there is a simple verse of the al-Quran as found in certain chapters, its elaboration alternatively is found on the other chapter of the al-Quran.
- ii. Interpreting verses assisted by the Prophet Muhammad SAW's hadiths: Hadiths from the Prophet act as an elaborator and clarifier of the al-Quran.
- iii. If there is no such explanation in the Hadith, then words from the Prophet's Fellow Acquaintances are referred to: The Fellow Acquaintances were more conversant with the definitions of the al-Quran verses since they witnessed for themselves of the various events and scenarios during the period of when the Quran was revealed to the Prophet. They were also able to perfectly understand those definitions through genuine knowledge and religious practices.
- iv. If there is no such explanation in the al-Quran, Hadith and the Prophets's Acquaintances, then words from the *Tabi'ins* are referred to. There were also several *Tabi'ins* who gathered the interpretations from the Acquaintances while extracting the decrees and debating the evidences based on the correct source intake.
- v. Knowing the Arabic language and its branch of knowledge: al-Quran was delivered by using the Arabic language. Therefore, the understanding of its content relies greatly on the description of the word's meaning and insight based on the contexts. Meanings can differ due to the changes in the case marks (*i'rab*). Those knowledge include *Nahu* (syntax), neurology (morphology), sentence construction knowledge that affects meaning, knowing the clear and implicit definition, as well as *Balaghah* knowledge that is translated through 3 main components; al-Ma'ani, al-Bayan and al-Badi' along with I'jaz's queries.
- vi. Knowing the basic principles on the various knowledge related to al-Quran: Such as Qira'at knowledge (that assists in reciting of the al-Quran), Tauhid knowledge (to ensure that there shall be no interpretation that deviates from the truth especially in verses that involves Allah and His attributes), Usul Fiqh knowledge, basic Tafsir knowledge (with emphasis on things that help in terms of definition clarification, and verse comprehension such as asbab al-nuzul, nasikh mansukh and so on.)
- vii. Having caution in comprehension: This allows interpreters to choose (*tarjih*-ing) one particular definition in comparison with other definitions or extracting a definition that is accurate in accordance with the Islamic facts.

According to the development of the Islamic Knowledge Ontology, the need for a compilation of Ontology in different fields of Islamic Knowledge was certainly obvious. Accordingly, the authors took the first step in this way and started to compile this great project. So far, the compilation of the Ontology of the Islamic Knowledge, the dictionary, glossary and general ontology of Islamic Knowledge are currently taking place. To support the Ontology, some short and useful information have been collected that will prepare the correct methods for the compilation of the lexicon.

### 3. Understanding Al-Qur'an

The Quran certainly contains knowledge of the past and the future, and comprehending it from the Quran. However, this does not contravene with the fact that the Quran has literal meanings that are

understandable by anyone acquainted with the Arabic language and its rules, which one may act upon once he or she becomes clear after due research for supporting indications.

However, in order to understand the language of the Quran, it is therefore a prerequisite to fully grasp its meanings. Hence many Muslims have been taking the initiative to learn this language. Others, who have not done so, make use of translations, which act as an indirect mean of knowing the language, since the meanings of the Quran in the translations have been rendered into their respective mother tongues so that they may familiarize themselves with the message from Allah.

In other way to understand the Quran is to know what roughly actually content of the Quran. The Quranic framework is deriving in Figure 1 as mentioned in [2].

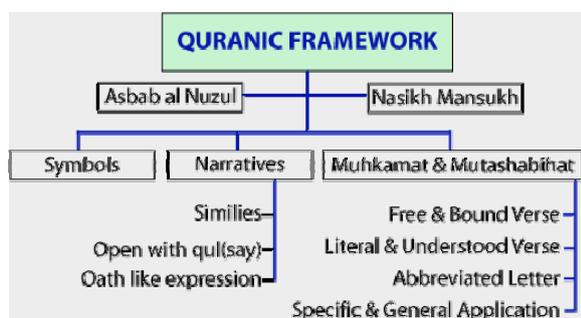


Figure 1. Contents of Quran Representation

The Quran contains many narratives. These narratives illustrate and underline important aspects of the Quranic message which fulfill their functions in a variety of ways. Some of the common patterns are as follows:

- i. Explanation of the general message of Islam.
- ii. General guidance and reminder.
- iii. Strengthening the conviction of the Prophet and the believers.
- iv. Reminder of the earlier prophets and their struggles.
- v. Indication for the continuity and truth of Muhammad's message.
- vi. Providing arguments against some opponents of Islam.

The Quran also employs similes that explain certain truths or to drive home important points of the message, by likening it to something well-known or by describing it in a pictorial manner. Example:

‘He sends down water (rain) from the sky, and the valleys flow according to their measure, but the flood bears away the foam that mounts up to the surface, and (also) from that (ore) which they heat in the fire in order to make ornaments or utensils, rises a foam like unto it, thus does Allâh (by parables) show forth truth and falsehood. Then, as for the foam it passes away as scum upon the banks, while that which is for the good of mankind remains in the

earth. Thus Allâh sets forth parables (for the truth and falsehood, i.e. Belief and disbelief).’(13:17).

Another pattern that occurs in Quran is the words begin with ‘Qul’ (say). There are more than 200 passages in the Quran begin with this word, which is a guideline to the Prophet Muhammad to address the words following this introduction to his audience in a particular situation, such as. in replying to any question that has been raised, or as an assertion of a matter of belief, or announcement of a legal ruling, etc.

‘Say: "Nothing shall ever happen to us except what Allâh has ordained for us. He is our Maulâ (Lord, Helper and Protector)." And in Allâh let the believers put their trust.’ (9: 51).

In a number of places the Quran employs oath-like expressions. Their function is to strengthen and support an argument, and to disperse doubts in the mind of the listener. In the Arabic text, these passages often begin with the word ‘wa’ or the phrase ‘la uqsimu’ (indeed I swear). Example:

‘And by the sun and its brightness; And by the moon as it follows it (the sun); And by the day as it shows up (the sun's) brightness; And by the night as it conceals it (the sun); And by the heaven and Him Who built it; And by the earth and Him Who spread it, And by Nafs (Adam or a person or a soul, etc.), and Him Who perfected him in proportion;’(91:1-7)

The Quran also have features of muhkamat and mutashabihat. Example of mutashabihat:

‘The Most Beneficent (Allâh) Istawâ (rose over) the (Mighty) Throne (in a manner that suits His Majesty).’(20: 5).

Where Mutashabihat is described as follows:

- i. Something known to Allah only for example ‘abbreviated letters’ such as Aliff Lam Mim (2:1)
- ii. Something with more than one dimension for example “...to one whom I have created with Both My Hands ...”(38:75). Allah’s Hand being refer as the Powerful of Allah.
- iii. Something that requires further explanation. For example “...Your wives are a tilth for you, so go to your tilth, when or how you will...” (2:223). The Quran doesn’t explain how to approach a wife in detail. But the hadith explain it (tafsir Ibnu Kathir, 2009)

Example of muhkamat:

‘O you who believe! When you contract a debt for a fixed period, write it down. Let a scribe write it down in justice between you. Let not the scribe refuse to write as Allâh has taught him, so let him write. Let him (the debtor) who incurs the liability dictate, and he must fear Allâh, his Lord, and diminish not anything of what he owes. But if the debtor is of poor understanding, or weak, or is unable himself to dictate, then let his guardian dictate in justice. And get two witnesses out of your own men.

And if there are not two men (available), then a man and two women, such as you agree for witnesses, so that if one of them (two women) errs, the other can remind her. And the witnesses should not refuse when they are called on (for evidence). You should not become weary to write it (your contract), whether it be small or big, for its fixed term, that is more just with Allâh; more solid as evidence, and more convenient to prevent doubts among yourselves, save when it is a present trade which you carry out on the spot among yourselves, then there is no sin on you if you do not write it down. But take witnesses whenever you make a commercial contract. Let neither scribe nor witness suffer any harm, but if you do (such harm), it would be wickedness in you. So be afraid of Allâh; and Allâh teaches you. And Allâh is the All-Knower of each and everything...' (2: 282).

Where muhkamat is described as follows:

- i. Something of which knowledge was desired.
- ii. Something with only one dimension.
- iii. Something sufficient in meaning, requiring no further explanation.

Hence in the Quran those ayats dealing with halal and haram, punishments, inheritance, promise and threat, etc. belong to muhkamat, while those concerning the attributes of Allah, the true nature of the resurrection, judgments and life after death etc. belong to the mutashabihat.

Some verses of the Quran are of a very wide, general application, e.g. including all human beings, or all Muslims etc. Other ayats are restricted in their application to certain special circumstances only. Example:

'Let there be no obscenity, nor wickedness nor wrangling in the Hajj' (2:187).

Furthermore one also distinguishes between 'general verses' which remain general, and others which intend to act according to a specific meaning.

Example:

'...And Hajj (pilgrimage to Makkah) to the House (Ka'bah) is a duty that mankind owes to Allâh, those who can afford the expenses ...'. (3: 97).

Of the 'special meanings', there are several varieties. Usually some kind of condition or limitation is specified. Example:

'They ask you concerning menstruation. Say: that is an Adha (a harmful thing for a husband to have a sexual intercourse with his wife while she is having her menses), therefore keep away from women during menses and go not unto them till they have purified (from menses and have taken a bath). And when they have purified themselves, then go in unto them as Allâh has ordained for you (go in unto them in any manner as long as it is in their vagina). Truly, Allâh loves those who turn unto Him in repentance and loves those who purify themselves (by taking a

bath and cleaning and washing thoroughly their private parts, bodies, for their prayers, etc.)'. (2: 222).

'Free' and 'Bound' verses in Quran are those where some of the ahkam verses are valid, 'free' (mutlaq) from any conditions or circumstances, while others are 'bound' (muqayyad) to special conditions or situations, and apply only therein. Examples:

'But whosoever cannot afford (that), and then he should fast for three days. That is the expiation for the oaths when you have sworn...' (5:89)

It is therefore free, which is left to one's discretion whether to fast three days consecutively or with interruptions. Example :

'..and you find no water, then perform Tayammum with clean earth and rub therewith your faces and hands..' (5: 6).

Quran also has few 'Literal' and 'Understood' meanings. The meaning of certain ayats is derived from the literal wording while the others are derived from what is understood by them. There are several kinds of literal understanding. The first concerns a clear text, without ambiguity. Example:

'.. but if he cannot afford it, he should observe Saum (fasts) three days during the Hajj and seven days after his return (to his home), making ten days in all..' (2: 196).

In other cases, the text may be somewhat ambiguous in its expression but it is rather obvious as far as the meaning is concerned. Example:

'..keep away from women during menses and go not unto them till they have purified (from menses and have taken a bath)..'. (2: 222).

According to the above text, some actually refer to the end of a woman's menstrual period, or the completion of the bath after the menstruation; and by this, the second is seen as more obvious.

The so-called 'abbreviated letters' are an important section of the mutashabihat insofar as their meanings are not known. In technical language the word is used for certain letters found at the beginning of several surahs of the Quran, called 'the abbreviated letters'.

The Quran also introduced new ideas and new concepts, and for their expression it provided new terms and continuously used metaphors. Similarly the Quran also introduced the usage of symbols. In the Quran metaphors assumed the status of symbols on account of their frequency of usage. It is one of the miracles of the Quran that become aware in the situation of our after life during our lifetime only through the means of symbols. Symbols have always a possibility of alteration or deviation in their meaning. The Quran being an eternal message the symbols in it are not subject to revision. The meanings assigned to a symbol must last forever. It is one of the miracles of the Quran that it has successfully expressed feelings which cannot be

expressed in normal words. The Quran's figurative style creates an atmosphere in which concepts and meanings appear in the form of symbols. Example:

'They said: "O Shu'aib! Does your Salat (prayer) (i.e. the prayers which you offer has spoiled your mind, so you) command that we leave off what our fathers used to worship, or that we leave off doing what we like with our property? Verily, you are the forbearer, right-minded!" (They said this sarcastically).' (11:87)

In the above ayah the word salat is used as a symbol for 'din'. In Surah 111, 'Abu Lahab' is used as a symbol for those who oppose the 'Divine Message'. In the Quran 'the wind' is used as a symbol for intellectual progress and hope.

The other thing that has to be considered in order to understand the contents of the Quran is called 'asbab al-nuzul'. It is where the Quran has been revealed for guidance, for all times and situations to come. However, various ayats were revealed at a particular time in history and in particular circumstances. The Arabic word 'sabab' (the plural is l- asbab) means reason or cause, and 'marifa asbab al-nuzul' is the knowledge about the reasons of the revelations, i.e. the knowledge about the particular events and circumstances in history that is related to the revelation of particular passages from the Quran.

The knowledge of asbab al-nuzul helps to distinguish between these two by:

- i. Clarifying the events and circumstances, which are connected with the revelation of certain ayats.
- ii. Illustrating the application of such ayats by referring to certain situations, when the Companions of the Prophet found them to be proper and applicable.

In the modern computer age on more content of the Quran was discovered. For example discovery done by Dr. Rashid Khalifah (2009) of which the earlier Muslim scholars were not aware of.

Dr. Khalifa based his theory on exact science i.e., mathematics for mathematics is never partial. To prove to the satisfaction of every atheist and every agnostic that the Quran is the word of Allah he took the aid of computer. By counting every letter and every word of the Quran Dr. Rashad khalifa established that the Quran has its own mathematical code. The computer data revealed that there was close relationship between the letters as well as the words of the Quran and number 19. Thus a mathematical proof was provided for the unique nature of the Quran. Dr. Khalifa based his theory on the following ayah of the Quran.

"...Over it are nineteen." (Quran 74 : 30)

According to the earlier commentators the pronoun 'it' refers to the hell and they explained that there are 19 angels guarding the doors of hell. According to Dr. Rashad Khalifa the pronoun 'it' referred to the Quran and that number 19 is assigned

to the Quran. There are several more intricate facts have also been mentioned by Dr. Rashad Khalifa in support of his contention.

Nevertheless, all of the contents of Quran can be classified into fifteen categories or subjects according to AlKabi(2005) and Saidah & Naomie (2007).

According to Tuncer(2006), there are two important aspects of understanding Quran based on Fethullah Gulen's Methodology of Interpreting Quran. First, everything either wet or dry exists in Quran al-Kerim, and the other is to keep a balance in explaining all kinds of matters with Quran verses. In the light of these two principles, He interprets Quran under 4 main headings.

- i. Quran as a miracle and Quran's telling many things with less words.
- ii. Contextuality among verses where Quran has context within all surahs and verses.
- iii. Interrelation between modern sciences and Quran verses
- iv. New and authentic views in the interpretation of Quran. Example where Gulen's point of view on 'dabbetul ard' and AIDS and others.

Islamic Knowledge is much more complicated by the fact that Arabic is a language and its structure gives different meaning within 'context' resulting in various interpretations. There will be other characteristic or features would be raise by time. These are based on the full understanding and research towards the Quran.

#### 4. Islamic Knowledge Ontology vs. Other Ontologies

The most evident distinction in developing Islamic knowledge ontology in relation to other ontologies, is the domain itself since the documents applied have different characteristics from other ordinary documents.

In addition to that, translations into the particular mother tongue languages by the interpreters are different from one person to another based on their respective interpretation knowledge and how they interpret the texts according to their understanding.

Islamic knowledge is also different from other ontologies as it has a range of knowledge sequences that must be complied as emphasised in [10], which involve the rules in creating knowledge on Islam.

Islamic knowledge that will be developed is deep-rooted from the Qur'an, in which Qur'an has the *muhkamat* and *mutashabihat* verses as explained earlier. In the meantime, *mutashabihat* verses require elaboration from different sources (as mentioned in [10]) in order to complete the ontologies that involved the verses. By looking at the previous researches conducted by the ontology researchers,

information from the applied documents is direct and easily comprehended without having to rely on other documents such as data on tourism [5], data related to biography of the extracted artist [2] and so forth. This is unlike the use of other ontologies in order to expand or specify the current available information, for instance WordNet.

Islamic knowledge also requires *Asbab al-nusul* for the purpose of understanding contents of the Qur'an as knowledge on it helps one to understand:

- i. The direct and immediate meaning and implication of an *ayat* (verse), as it can be seen within its original context.
- ii. The imminent reason underlying a legal ruling.
- iii. The original intent of the *ayat*.
- iv. Whether the meaning of an *ayat* is specific or of general application, and if so, under which circumstances it is to be applied.
- v. The historical situation at the time of the Prophet and the development of the early Muslim community.

Without knowing the *sebab* (reason), one might easily conclude that this revelation permits the Muslims to face any direction when performing prayer, while it is well-known that facing the *qiblat* is one of the conditions without which prayer becomes invalid. The circumstances in which this revelation occurred explain its implications. On the other hand, other ontologies may not require the written *sebab* as everything is clearly stated in the document.

Based on [10], the pattern being extracted covers all verses as each verse is equally important. The extracted verse is a phrase that later undergoes the pruning process in order to obtain a fully optimised pattern. Meanwhile for other ontologies, the developed pattern is based on the Hearst pattern [5], syntactic dependency that involves subject, object and prepositional phrase (PP) that are related to verb [4], or the one that is based on the probabilistic method [9].

Another thing to be considered is in the multi-layered coherence of the Qur'an; together with all of its themes that emerge in short passages, creating an inimitable interplay between its imageries, oaths, parables, chronicles, warnings, and glad tidings. With so many elements of text coalescing, separating, reuniting, and reemphasising each other at numerous levels, the result can be a total incomprehensibility and confusion. This has consequently caused for a proper and comprehensive ontology to be generated in order to connect the information contained in the Qur'an and later elaborated by other supporting documents.

The following Table 1 summarises the differences between the Islamic Knowledge Ontology in comparison with the other ontologies.

TABLE 1: DIFFERENCES BETWEEN ISLAMIC ONTOLOGY AND OTHERS ONTOLOGY

Criteria	Ontologies	Islamic Ontology Knowledge
Content Phrase	Direct	Narratives, <i>Muhkamat</i> and <i>mutashabihat</i>
Sources Reference	Other document, other ontology	Document based on sequence priority : Quran, Hadith, Tabiin, ijmak Ulama, Qias Asbab al-nuzul
Pattern extraction	NP, PP S O V PP	Phrase, compound noun, case letter.
Relevancy	Statistical techniques, clustering	Every sentence is important and relevant to certain subject
Content Document	One paragraph for one message.	Multi-layered coherence of the Quran, all its themes emerge in short passages, creating an inimitable interplay between its imageries, oaths, parables, chronicles, warnings, and glad tidings

## 5. Methodology

According to [1], the general subjects that are found in Quran to be relevant to Muslim scholars' classifications are Islamic basic (pillars of Islam), faith, general and political relations, science and art, holy Quran, organising of financial relations, human and social relations, Al-Jihad, religions, judicial relations, working, stories and history, human and ethical relations, trade, agriculture and industry and call for Allah (*Dakwah*). Therefore, the terminology or phrase extracted will be based on the general subjects as its guidance.

In the Islamic Pillar, it consists of faith or belief in the Oneness of God and the finality of the prophethood of Muhammad; establishment of the daily prayers; concern for and almsgiving to the needy; self-purification through fasting; and the pilgrimage to Makkah for those who are able. For this paper, the focus will be placed on the daily prayer.

In developing an Islamic Ontology, two principal components of ontology are distinguished. The first component is the metadata component that is often called terminological component which is composed by taxonomy of concepts (TBox components). The second one is the assertion component which contains concrete representations of these concepts that are often called instances (ABox components).

In order to develop an Islamic Ontology mentioned before, these are informal questions that the ontology must be able to answer and will be used to check that the ontology is fit for purpose. The following are the examples of possible competency questions in domain solat :

- What is solat (prayer)
- What type of solat
- When can I perform solat
- What characteristics should I have if I want to perform solat or Prerequisites of the Prayer
- Where can I perform solat
- What is purification
- What is ablution (wudu')
- What is tayammum, the dry ablution
- What is menstruation
- What is azhan, call to prayer
- What is obligatory acts of prayer
- What is sunnah acts of prayer

### 5.1 Pre-Processing The Document

In generating this system, there are rules to be understood and considered such as below:

- i. Interpretation of meaning is only generated on one particular layer only, which is what is portrayed in the verse without looking at the hidden meaning.
- ii. In the al-Quran, a complete verse may consist of between 1-8 phrases. Connectors are used to join one verse to another, while the verse is separated from one point to another by comma signs.
- iii. For an achieved verse of the al-Quran that contains the word 'prayer' or 'solah' for instance, therefore that verse must not only be considered, but also the previous and following verse.
- iv. A phrase that needs to be parsed must undergo a definition's temporary replacement process to simplify the ongoing parsing process. For example: *Those who believe* or *You who believe* is replaced with *Mukminun*. This is obtained from the list of predetermined REPLACEMENT.

### 5.2 Defining the domain concepts in the Islamic ontology

In order to identify the key concepts that exist in the domain, their properties and the relationships that hold between them, these are the steps to be followed:

- i. term extraction, where all the terms exist in the Quran are important,
- ii. synonym extraction,
- iii. extracting concepts (It was terms that indicate a concept within the domain population. The concept formation should ideally provide an intension definition of concepts, its extension and the lexical that is being referred to [6] and its legislation [9]), and
- iv. extracting relation (taxonomic (IS-A and part-of relation) and non-taxonomic relation (refers to any relation between concepts except the IS-A relations)) [10].

In order to develop an ontology, understandings must be obtained on:

- i. formatting of the Al-Quran and
- ii. text pattern texts sources. Each document gives different text formatting and different pattern because it depends on the interpretation of the translators.

In Al-Quran translation to the meaning document, there are texts formatting that can help in developing concepts.

### 5.3 FORMATTING OF THE AL-QURAN

Two types of brackets as employed in the text document are discovered: rounded ( ) and squared [ ]. This may be explained as follows:-

- The use of the rounded ( ) brackets: if NP( ), it gives a synonym and/or a meaning to the Arabic term that is being used in the documents and if ~NP( ), it gives a complete word that does not appear in that passage - for example

*Say (O Muhammad SAW): "Verily, my **Salat** (**prayer**), .....*

After run the POS tagging, the result is“(NP (NN Salat) (-LRB- -LRB-) (NN prayer) (-RRB- -RRB-))”

This indicates that  
Noun Salat ( Noun prayer)

This can be formalize it by this statement,  
if NP<sub>0</sub>(NP<sub>i</sub>\*)

**NP<sub>0</sub>•synonym = NP<sub>1</sub> ... (FORMAT 1)**

- The use of the squared [ ] brackets : It is being used to make the sentence portion of the text more complete or to elaborate more detail about the concept, for example

*And seek help in patience and As-Salat (the prayer) and truly it is extremely heavy and hard except for **Al-Khashi'un** [i.e. the true believers in Allah - those who obey Allah with full submission, fear much from His Punishment, and believe in His Promise (Paradise, etc.) and in His Warnings (Hell, etc.)].*

This indicates that  
If after “open squared bracket [“ i.e then  
if NP<sub>0</sub>[ NP<sub>1</sub>, NP<sub>2</sub>.....NP<sub>n</sub>] then

**NP<sub>1</sub>, NP<sub>2</sub>...NP<sub>n</sub> PartOf NP<sub>0</sub> (FORMAT 2)**

### 5.4 Letter Case

In Al-Quran translation, all of the uppercases indicate the concept (except the starting word of sentences) for example *the Book*

### 5.5 Pattern

The three patterns have been used to identify part-of relationships between the concepts that are referred by the two terms in the text. In order to identify phrases representing concepts, henceforth  $P_0$ , and phrases representing subconcept, henceforth  $P_n$  (where  $n = 1 \dots n$ ), that phrase can be VP or NP.  $NP_0$  representing a superconcept. A shallow parsing technique (Stanford Parser) is used based on matching regular expressions over part-of-speech tags to identify the phrases as described above.

**PATTERN 1:  $NP_0^* \{is,are\} ((P_n)^* and^*)^*$**

*The believers, men and women, are Auliya..*

if NP (suffix with er) + gender (men/women) must represent as a two group for example Auliya is part-of “the believers men” and “the believers women”

**PATTERN 2:  $NP_0 \{:,!\} P_1 \text{ and } P_2 \dots \text{ and } P_n$**

*.... a covenant from the Children of Israel, (saying): Worship none but Allah (Alone) and be dutiful and good to parents, and to kindred, and to orphans and Al-Masakîn(the poor), and speak good to people and perform As-Salat (Iaamat-as-Salat), and give Zakat...*

Where  $P^*$  PartOf NP

### 5.6 Compound Noun.

The following pattern makes two or more nouns/adjective combined to form a single noun. Where it can be extract in three patterns:

- i. The  $N_1N_0$  /The  $N_0$  of  $N_1$
- ii.  $NP(JJ, NP_0) / NP(NP_0(PP NP))$
- iii. head of Noun  
where  $N_0$  is part-of  $N_1N_0$  /The  $N_0$  of  $N_1$

For example “funeral prayer” is part of “pray”

### 5.7 Verb As A Relations

### RELATION 1: $NP_0 V NP_1$

*We explain the Ayat (proofs, .....)*

*My, We, Us, Lord* when used in the middle of the verse with capital letter, it will change to Allah. So from above verse, we get the pattern **explain (Allah, Ayah)**

### 5.8 Apposition

A noun, noun phrase, or series of nouns used to identify or rename another noun, noun phrase, or pronoun. The following pattern makes use of the fact that certain entities appearing in a text are further described in terms of an apposition as in

*And when We made the House (Kaaba) a place of return, a sanctuary for mankind: They took the place where Ibrahim stood (to pray) as a place of prayer.....*

The pattern capturing this intuition looks as follows: **APPOSITION:  $NP_0$  a / as a  $NP_1$**

This indicates that

$NP_1$  is-a  $NP_0$

### 5.9 Copula

Copula is a word that is used to link the subject of a sentence with a predicate where the predicates are an adjective and a noun, so the linking verb "is". It is probably the most explicit way of expressing that a certain entity is an instance of a certain concept is by the verb ‘to be’, where the general is **COPULA:  $NP_0$  verb-to-be  $NP_1$**  for example

*Allah is Ever OftPardoning, OftForgiving*

This indicates that

$NP_1$  is-a  $NP_0$

## 6. Results & Discussion

Table 1. Result for format and pattern extraction

TECHNIQUES	HITs	%
Format 1	124	21.83
Format 2	12	2.11
Letter Case	331	58.27
Pattern 1	6	1.06
Pattern 2	2	0.35
Compound Noun	54	9.51
Relation 1	32	5.63
Apposition	2	0.35
Copula	5	0.88
<b>ALL</b>	568	100.00

In 73 verses of the Al-Quran regarding solah or prayer which consists of 3378 words, an analysis was held in order to obtain information for the generation of the ontology. The analysis process was carried out based on the information generated by the domain experts. From there, rules generation was performed and the results are as discussed in the above 4.2. Based on the format and rules that have been created, then the extraction process takes place. The results are as explained in the following Table 1.

The table shows that the most prominent technique in order to extract concept and generate ontology based on Al-Quran is an uppercase letter and in translation Al-Quran, the translator try give more meaning and explanations on certain word especially on Arabic word. Another reason is the Islamic concept (usually state or refer in Arabic terms) has special characteristics that distinguish it from all other concepts, giving it an independent personality and a particular temperament. This concept cannot be intermixed with any other nor can anything be added to it from other concepts (Sayyid Qutb, 1992).

But, there are verses of extracted Al-Quran that consist only concepts without relation. Therefore, it cannot be relate with other concepts or relations in order to form an ontology. This because, in certain verses there is more than one verse have to be consider in order to get a complete sentence. For example :

2:43 And perform As-Salat, and give Zakat, and Irka along with ArRakiun

In order to get a complete sentence based on verse above (surah Baqarah, verse 43), we have to consider another three verses before.

From the analysis based on subject salat or prayer, we found that there are twelve verses have the same problem as mention.

Beside that, there are sixteen verses cannot be relate with any format, rules or pattern describe before. Example :

Recite (O Muhammad SAW) what has been revealed to you of the Book (the Quran), and perform As-Salat (IqamatasSalat).

From the observation shows that the simple format and pattern that being described can be used in order to form ontology. But it can be upgrade by using complex NLP techniques and also using machine learning techniques.

## 7. Conclusion and Future Work

This work presents an approach to the automatic generation of ontology instances from a collection of unstructured documents known as the Al-Quran. The presented approach is stimulated based on the combination of natural language processing techniques, Information Extraction (IE) and Text Mining techniques. Based on the traditional IE systems, the authors have applied and defined a grammar and extraction rule to obtain the ontology instances. The system tried to form the correct partial instances by taking the words and entities appearing in the texts, and combining them to form correct and complete instances.

In the future, the authors intend to extract the whole Al-Quran based on the pattern described in order to evaluate the accuracy and completeness of the rules.

In the multi-layered coherence of the Quran, all of its themes emerge in short passages, creating an inimitable interplay between its imagery, oaths, parables, chronicles, warnings, and glad tidings. With so many elements of text coalescing, separating, reuniting, and reemphasising each other at numerous levels, the result can be a total incomprehensibility and confusion. Therefore, hadiths are required in order to create an ontology that is considered to be more complete, in which hadiths act as definer of the definitions in Al-Quran. For that matter, the next agenda is to combine Al-Quran and Hadiths based on the similar subject, which is Solah or prayer.

## 8. Acknowledgement

The work reported in this paper has been supported by the Ontology Extraction Tool for Quran and Hadith Project (01-01-06-SF0539), funded by Ministry of Science, Technology and Innovation, Malaysia.

## 9. References

- [1] AL-Kabi M.N. Kanaan G. And Al-Shalabi R. 2005. Statistical Classifier of the Holy Quran Verses (Fatiha and Yaseen Chapters). *Journal of Applied Sciences* 5(3): 580-583,
- [2] B. J. Wielinga. A. Th. Schreiber. J. Wielemaker. J. A. C. "From thesaurus to ontology". *International Conference On Knowledge Capture*, Columbia, Canada. 2001, Pages: 194 - 201 .
- [3] Cimiano. P. 2006. *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications* Springer. November 2006.
- [4] Hanna E. Kassis. *A Concordance of the Quran*. Published by University of California Press, 1983.

[5] Hasanuddin Ahmed. 2004. *Introducing the Quran : How to Study and Understand the Quran*. Goodword Books.

[6] M.T.Al-Hilali, M. M. Khan. *Complete Interpretation of the Meaning of The Noble Quran In the English Language*. <http://www.iiu.edu.my/deed/quran/nobelquran/index.html>. 2009.

[7] Manna' Al-Qattan. 1978. *Mabahith Fi 'Ulum Al-Quran*. Beirut: Muassasah Ar-Risalah.

[8] S. Staab and R. Studer, editors. *Handbook on Ontologies*. Springer, 2004.

[9] Saidah Saad, Naomie Salim, 2007, *Terminological Extraction from Text to Build an Ontology in Islamic Knowledge*. Postgraduate Annual Research Seminar 2008, UTM, Skudai.

[10] Saidah Saad, Naomie Salim, 2008, *Methodology of Ontology Extraction for Islamic Knowledge Text*. Postgraduate Annual Research Seminar 2008, UTM, Skudai.

[11] Sayyid Qutb. 1992. *The Islamic Concept and Its Characteristics*. American Trust Publications.

[12] Shamsfard, M and Barforoush, AA, 2002, "An introduction to HASTI: an ontology learning system" *Proceedings of 6th Conference on Artificial Intelligence and Soft Computing (ASC'2002)*.

[13] *The Integrated Encyclopedia of the Quran (IEQ)* . <http://www.iequran.com/ieq-br.pdf>. 2009.