Using Ontology to Define the Structure of the Holy Quran

Malek Zakarya AL Ksasbeh, Mohd Syazwan Abdullah, Wan Rozaini Sheik Osman, Farihana Elyana
Department of Computer Sciences, College of Arts and Sciences
Universiti Utara Malaysia, 06010 UUM-Sintok, Kedah, Malaysia
Malekzk@hotmail.com, {syazwan, rozai174}@uum.edu.my, farihan_poliperlis@yahoo.com

ABSTRACT
Ontology have gained attention from various fields in the past years, especially in the area of artificial intelligence, a sub-area of computer science. This paper discusses about ontology, ontology languages, ontology tools and how it can be used in organising the knowledge in the Holy Quran. A simple example using ontology for Holy Quran is presented as work-in-progress.

Key Words: Ontology, Ontology languages, Ontology tools, Holy Quran

1. Introduction
In the past 20 years ontologies have emerged as an important research area in Computer Science. Ontology origins, from a philosophical point of view, are found in the ancient Greece. Ontology is a philosophical discipline dealing with the nature and the organization of actuality. In the modern era, ontologies have been created to share and reuse knowledge across domains and tasks. Currently, they are widely used in artificial intelligence, knowledge engineering, and computer science, in applications related to knowledge management, e-commerce, natural language processing, intelligent integration information, information retrieval, database design and integration, etc. One of the goals is to reduce or eliminate the conceptual and terminological confusion among the members of a virtual community of users (humans or computer programs) that need to share electronic documents and information of various kinds. This is achieved by identifying and defining a set of relevant concepts that characterize a given application domain. This paper discusses important issues as regards of using the ontology to define the holy Quran structure. It tries to come up with new ontology system related to Quran structure definition. Moreover, the main reason behind designing this system is to eliminate the Quran terms confusion especially for non-Arabic speakers. The paper organized as follows: section 1 introduces the paper, while section 2 discusses about ontology. And section 3 present different ontology languages. Then, section 4 details the ontology tools widely used. Section 5 briefly presents the Holy Quran and section 6 shows how ontology can be used for organising knowledge in Quran and finally section 7 concludes the paper and future work.

2. Ontology
An ontology is term defined as a formal, explicit specification of a shared conceptualization[1]. Conceptualization refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. Explicit means that the type of concepts used, and the constraints of their use, are explicitly defined. Formal refers to the fact that the ontology should be machine readable. Shared reflects the notion that an ontology captures consensual knowledge, that is, it is not
private of some individual, but accepted by a group.

Another definition focused on the form of an ontology is given; an ontology is term defined the basic terms and relations comprising the vocabulary of atopic area as well as the rules for combining terms and relations to define extensions to the vocabulary [2]. Other approaches have defined ontologies an explicit specification of a conceptualization [3]. Ontology is a body of knowledge describing some domain [4] and the term used to refer to the shared understanding of some domain of interest [5]. Therefore, ontology is the term used to explain and represent an area of knowledge, and includes definitions of basic concepts in the domain and the relationships between them. Ontology can be used in many fields and applications that need to share domain information.

Ontologies have been established to be useful for:

- Retrieving the appropriate information from documents by providing a structure to annotate the contents of a document with semantic information [6].
- Integrating the information from various sources by providing a structure for its organisation and facilitating the exchange of data, knowledge and models [7, 8].
- Ensuring consistency and correctness by formulating constraints on the content of information [8].
- Creating libraries of interchangeable and reusable models [7, 8].
- Supporting inference to derive additional knowledge from a set of facts [9, 10].

The need of an ontology comes from the fact that we are building a semantic web that needs all means of knowledge representation and sharing mechanisms; the Ontologies are the best alternative way to share common understanding of the structure of information among people or software agents, enable the reuse of domain knowledge as models for numerous diverse domains need to represent the notion of time. This kind of representation includes the notions of time intervals, points in time, relative measures of time, and so on. If one group of researchers develops such an ontology in detail, others can simply reuse it for their domains. Furthermore, if a large ontology needs to be produced, several existing ontologies can be integrated describing portions of the large domain.

The creation of ontology is not a goal, but rather a tool to define a set of data/info and their structure for other programs or agents to use. This enables the utilization of knowledge bases for problem-solving methods, domain independent applications, and software agent data exchange among endless uses. This is why the creation is not the goal, instead aiming to create common ontologies. The expectation is that these will be several ontologies for the numerous users/agents over the semantic web urge researchers to look for bridges between those ontologies. While finding a standard is time consuming, a common standard ontology will allow representing entities in a unified method by following a common structure and pattern. This standard will bring the (Rule layer) in the semantic web layers to work properly and operate as expected, which in its turn should allow other processes to be simple rather than tedious and complex.

There are many advantages of ontology, but the main advantage is the ability to organize information into hierarchically ordered taxonomies of concepts, and to define attributes and relationships between these concepts, and describe information according to different contexts [11].

3. Ontology Languages

Several ontology representation languages have been developed in the last few years[12]: RDF(Resource Description Framework), OIL(Ontology Inference Layer), DAML(DARPA Agent Markup Language), DAML+OIL,OWL (Ontology
Web Language )and others . RDF seem to be will positioned to become the standard to represent ontologies in the future [13].

3.1 RDF
RDF (Resource Description Framework) [14] is the first language developed particularly for the Semantic Web. RDF was developed as a language for adding machine-readable metadata to existing data on the web. On top of the XML layer comes the RDF and the RDFS; where RDF is a framework for meta-data that provides the needed interoperability between applications which exchange machine understandable data/information over the web. One of the main important aspects the RDF will add to the semantic web cake layers presented by Berners-Lee is that RDF with digital signatures will be a key in building the “web of trust” for e-commerce, collaboration, and other documents. RDF is a declarative language which uses XML in a certain standard to represent meta-data in the form about properties and relations of items over the WWW, and a specification governing the interoperability of applications in terms of metadata property sets.

RDF Schema [15] extends RDF with some basic (frame-based) ontological modeling primitives. There are such primitives as classes, properties and instances. Also, the instance-of and subclass-of relationships have been introduced.

Building RDF and RDFS on top of XML provides many advantages for the semantic web beside that the XML is widely used. RDF data will become part of the semantic web and it will be easily readable by any system in the near future.

3.2 OIL
OIL (Ontology Inference Layer) [16] is a Web ontology language, which layers on top of RDF(S). OIL was developed by the European IST project On-to-Knowledge to provide simple and neat semantics’ ontology using syntax definitions from XML and RDF to maintain the backward compatibility and still support modeling primitives by description logics and frame-based orientation. The main contribution of OIL is providing the means for describing structured vocabulary with well-defined semantics. Actually, OIL is not completely layered, because there is a part of RDF(S), which is not a part of OIL. Reification and meta-classes in RDFS are not allowed in OIL, which means that not all valid RDF(S) is valid OIL [17].

OIL itself is also layered into four layers:

1. **Core OIL:** This is that part of RDF(S) which has a meaning in OIL. This is RDF(S) without reification and without mixing instances and classes.
2. **Standard OIL:** all necessary mainstream modeling primitives are captured, precisely specifying the semantics and making complete inference viable.
3. **Instance OIL:** Allows the definition of instances in the ontology.
4. **Heavy OIL:** originally reserved for possible future extensions of OIL; OIL is now no longer in used, so heavy OIL is no longer relevant; all current research focuses on OWL.

OIL took in consideration the point of view of three different communities:

1. Formal semantics and efficient reasoning support as provided by description logics[18].
2. Standard proposal for syntactical exchange notations as provided by the Web community[19].
3. Epistemologically rich modeling primitives as provided by the frame-based community[20].

3.3 DAML
DAML (DARPA Agent Markup Language)[21] is an ontology language developed by DARBA; an XML based (semantic) language to connect information on pages to a machine-
readable semantics (ontology). It is also based primarily on RDF to represent semantic relations compatible with current and futuristic web technologies. DAML realized that it must go beyond the embedded semantic agreements inherent in XML and other controlled community-specific languages, and shift to make semantic-entities and markup a primary goal. In order to do so, it allowed designing of meaning and sharing of high-level concepts by communities to extend simple ontologies for their own use, to permit implicit information to be recognized. Mechanisms are provided for explicit representation of business models, services, and processes.

Some of the advantages of the research model of DAML over the other markup approaches were that it will allow semantic interoperability at a level which has syntactic interoperability in XML. Also the goal of DAML which could be an advantage if the model fully implements include creating technology to enable agents to dynamically identify and understand distinct information sources and provide interoperability between those agents semantically. The tasks in their research to achieve this goal was allowing users to provide machine readable semantic annotations for specific communities-of-interest, embedding DAML markup within WebPages and information sources in a transparent manner to be beneficial for users.

3.4 DAML+OIL
DAML+OIL [22] is an ontology language developed by DAML (Darpa Agent Markup Language) and the developers of OIL. DAML is US based and the developers of OIL are mainly EU based. The American and European ontology communities joined efforts to create a common ontology for semantic markup; and merged those two ontology languages within agreements for a better concept acceptance by the users’ society, while depending heavily on the OIL specifications. The advances made for DAML+OIL over the original languages is conceived through appending ontological primitive of Object-Oriented (OO) and frame-based systems, and a formal rigor of expressive (DL). The implementation of the ontology through the OO approach builds on a structure of domain(s) described in terms of classes, properties, and axioms which state their characteristics.

DAML+OIL is defined by a standard model-theoretic semantics based on interpretations [19]. The interpretation itself consists of a discourse and interpretation function. The research communities tend to accept DAML+OIL as an ontology language due to what it includes from basic structures from the original languages and its own additional view for the common standard and the scopes of different communities producing sufficient expressiveness gained wide acceptance for the implementation perspectives, hoping to be able to manifest new technologies and services using this language. Yet it is still evolving and being developed and open for enhancement toward future trends; joining the base two languages works well as a delivery platform for ontology, and a language for ontology representation, how ever, it is still not fully efficient as an exchange, sharing, and modeling format. It needs to be adopted very carefully especially with domains without enough ontologies to cover or new domains.

3.5 OWL
OWL (ontology web language) [23] is an ontology language for the semantic web not as a radical creation but a build-up on OIL and DAML+OIL implementing actual layering on RDF(S) while considering anticipations of industrial and academic researchers involved and anxious for the active semantic web, where the main features of it are very comparable to those languages.

Though layering OWL on top of RDF added high compatibility with the web sources, but, however, we can not regard OWL model theory as extension of the
RDF model theory; where RDF(S) can present conclusions through syntactic definition which OWL cant due to constraints and restrictions expressed in OWL and not in RDF.
The aim of developing OWL was to enhance ontology evolution and interoperability, and reach internationalization to reach everyone through the ease of use and by XML Syntax compatible with the current sources, and by a balance between expressiveness and scalability, and through adding the inconsistency detection, sharing ontologies can be a method for better capabilities enabled through this language.
Main components of OWL are ontologies, Axioms, and Facts. Ontologies are a sequence of axioms and facts referred to by URIs, in addition to inclusion references for other ontologies. Axioms associate partial or complete specifications of characteristics and other logical information for class and property IDs. Facts state information about particular class’s individuals and their properties and values.

OWL subsidiaries are based on the level of interaction the ontology’s developer wants to have; each sublanguage presents a different level of expressiveness to meet the required goals[24]:

- **OWL-Lite:** Roughly consists of RDFS plus equality and 0/1-cardinality. Layered and easy-going language for tool builders. Developed to capture many of the commonly used features of DAML+OIL. It attempts to provide more functionality than RDFS, which is important in order to support web applications.
- **OWL DL:** Contains the whole OWL vocabulary, interpreted under a number of simple constraints. Primary among these can be found the type separation. Class identifiers cannot simultaneously be properties or individuals. Similarly, properties cannot be individuals.
- **OWL Full:** Composed of the complete vocabulary but interpreted more broadly than in OWL DL. A class can be treated simultaneously as a collection of individuals (the class extension) and as an individual in its own right (the class intension).

DAML+OIL and OWL add two more pieces of the final puzzle, one simple and one complicated:
The simple part: OWL adds the ability to indicate when two classes or properties are identical; which allows bodies of data in different schemas be linked together.
The complex part: OWL declarations provide additional information to let rule-checking and theorem-proving systems work with RDF data.

4. Ontology Tools
Several ontology tools have been developed in the last few years: Apollo, OILEd, Ontolingua Server, WebOnto, protégé and others. Protégé seems to be well positioned to become the standard to create ontologies nowadays [25].

4.1 Protégé
Protégé [25] is an open-source ontology development environment with functionality for editing classes, slots (properties), and instances. Protégé makes it not only possible to extend the metamodel but also to customize the user interface freely. Protégé’s user interface consists of several screens, called tabs, each of which displays a different aspect of the ontology in a specialized view. Each of the tabs can include arbitrary Java components. Most of the existing tabs provide an explorer-style view of the model, with a tree on the left hand side and details of the selected node on the right hand side. The details of the selected object are typically displayed by means of forms. The forms consist of configurable components, called widgets. Typically, each widget displays one property of the selected object [26].
There are standard widgets for the most common property types, but ontology developers are free to replace the default widgets with specialized components. Widgets, tabs, and back-ends are called plugins. Protégé’s architecture makes it possible to add and activate plugins dynamically; so that the default system’s appearance and behavior can be completely adapted to a project’s needs [26].

The Protégé main advantage today’s is the combined power of the plug-in components and the ability to tailor the environment by developing custom components [27].

4.2 Apollo

Apollo[28] is a user friendly ontology development application. The design was motivated by researcher’s experiences working with industrial partners who wished to use knowledge modelling techniques, but required an easy to use and understand syntax and environment. Apollo supports all the basic primitives of knowledge modelling: ontologies, classes, instances, functions and relations. Full consistency checking is done while editing, for example, detecting the use of undefined classes. Apollo has its own internal language for storing the ontologies, but can also export the ontology into different representation languages, as required by the user. Apollo is implemented in Java.

4.3 OILEd

OILEd [29] is a graphical ontology editor developed to allow the user to build ontologies using DAML+OIL by University of Manchester. The knowledge model of OILEd is based on that of DAML+OIL, though this is extended by the use of a frame-like presentation for modelling, therefore OILEd offers a familiar frame-like paradigm for modeling while still supporting the rich expressiveness of DAML+OIL where required. Classes are defined in terms of their superclasses and property restrictions, with additional axioms capturing further relationships such as disjointness. The expressive knowledge model allows the use of complex composite descriptions as role fillers. This is in contrast to many existing frame-based editors, where such anonymous frames must be named before they can be used as models.

The main task that OILEd is targeted at is that of editing ontologies or schemas, as opposed to knowledge acquisition or the construction of large knowledge bases of instances.

4.4 Ontolingua Server

The Ontolingua Server [30] is a set of tools and services that support the building of shared ontologies between distributed groups, and have been developed by the Knowledge Systems Laboratory (KSL) at Stanford University. The ontology server architecture provides access to a library of ontologies, translators to languages (Prolog, CORBA IDL, CLIPS, Loom, etc.) and an editor to create and browse ontologies. Remote editors can browse and edit ontologies, and remote or local applications can access any of the ontologies in the ontology library using the OKBC (Open Knowledge Based Connectivity) protocol.

4.5 WebOnto

WebOnto [31] is a tool developed by the Knowledge Media Institute (KMi) of the Open University (England). It supports the collaborative browsing, creation and editing of ontologies, which are represented in the knowledge modelling language OCML (Options Configuration Modeling Language). Its main features are: management of ontologies using a graphical interface; the automatic generation of instance editing forms from class definitions, support for PSMs and tasks modelling; inspection of elements, taking into account the inheritance of properties and consistency checking; a full tell and ask interface, and support for collaborative work, by means of broadcast/ receive and making annotations (using Tadzebao).
5. The Holy Quran

The Quran [32] is the central religious text of Islam. Muslims believe the Qur’an to be the book of divine guidance and direction for mankind, and consider the original Arabic text to be the final revelation of Allah (God).

The Quran speaks of the best rules relating to social life, commerce, marriage, inheritance, penal law, international law, and so on. But the Quran is not a book in the ordinary sense - it is a collection of the Words of Allah, were revealed, during the course of twenty-three years, to His messenger sent among human beings. The Quran employs graphically the word "king" for Allah, and "slave" for man. When a king desires to communicate a message to His slave, He sends a messenger, and gives His instructions to his envoy. Therefore, there are certain things understood and implied; there are repetitions, and even changes of the forms of expression. Thus, Allah speaks sometimes in the first person and sometimes in the third. He says "I" as well as "We" and "He", but never "They". It is a collection of revelations sent occasion by occasion (and this fact must be recalled to the beginner) and one should therefore read it again and again in order to be able to grasp the meaning better. It has directions for every person, every place and for all time.

The beauty of Quran lies with its eloquence and purity of its impressive style of language. The diction and style of the Quran are magnificent and appropriate to its Divine quality. Its recitation moves the spirit even for those who only listen to it without understanding it. The words of Quran are clear and balance and explain important concepts in a precious but simple manner. The speech of the Quran is harmonious balance and consistent and every word is irreplaceable. In spite of claims about other books, Quran is the most read and the only completely memorized book throughout its history. It is also the most studied book in the world. It has stimulated development of entire disciplines of knowledge dealing with its reading, writing, and interpretation. Hundreds of Millions of copies have been printed or handwritten in nearly every part of the world, without any variant in the text.

The early Muslims have observed that many important words have been repeated the exact number of a time which is relevant to the ideas [33]. For example the phrase the "seven heavens" is repeated 7 times the word "day" is repeated 365 times while the word "moon" is repeated 12 times many cases the certain related words have been repeated equal number to its opposite idea. For example the words, "world" and "hereafter" are each repeated 115 times. The word "faith" (iman) (without genitive) is repeated 25 times throughout the Quran, as is also the word "infidelity" or the phrase "covering over the truth." (kufr). When we count the word "say", we come up with the result of 332. We arrive at the same figure when we count the number of times the phrase: "they said" is used. The word "shaytan" is used 88 times. The word "angel" is also repeated 88 times. As explain there many miracle attribute of Holy Quran numerical system may been one of them [33].

The Quran is often hard to understand, being somewhat cryptic in it's message and verses. To fully understand the Quran, one has to refer to the tafsir. The tafsir shed light on each verse of the Quran and give explanations for why each verse was revealed and what were the circumstances surrounding the revelation [32].

5.1 Quran Challenge

Many challenges in the holy Quran have been mentioned as following [33]:

- **Alif-Lam-Ra.** These letters are one of the miracles of the Quran, and nonebut Allah (Alone) knows their meanings. 15:1 Al-Hijr
- **Say:** "If the mankind and the jinns were together to produce the like of this Quran, they could not produce the like thereof, even if they helped one another." 17:88 Al-Isra’.
• And if you (Arab pagans, Jews, and Christians) are in doubt concerning
• that which We have sent down (i.e. the Quran) to Our slave (Muhammad Peace be upon him),
then produce a Surah of the like thereof and call your witnesses
(supporters and helpers) besides Allah, if you are truthful. 2:23 Al-Baqarah.

6. Using Ontology In The Holy Quran
The very important question raised in this research is what are the most important elements of using the ontology in the holy Quran? What are the reasons behind not using simple techniques of a controlled vocabulary comparing with ontologies technique? Many researchers have extensively discussed the advantages of using ontologies. In this paper, we will show how ontology can be used to define the Quran structure in a simple example. Initially, we have discussed the concept of ontology and the holy Quran as scientific resources for Muslims’ community. The researcher contribution is to combine the ontology concepts into the holy Quran context in order to ease and assist non-Arabic speakers Muslims to understand comprehensively the Quran terms denotation. The literature scholars in holy Quran area will be used in this research in order to identify the Quran structures. Then, the research will design and create ontology system based on the related information from the literature review. Eventually, the proposed design will be evaluated and tested in order to determinate the achieved objectives. The general subjects that they found in Quran relevant to Muslim scholars’ classifications are Islamic basic (Islam Pillars), Faith, General and political relations, Science and art, Organizing financial relations, Human and social relations, Al-Jehad, Religions, Judicial relations, Working, Stories and history, Human and ethical relations, Trade, agriculture and industry and Call for Allah (Dawah) [34]. These classifications are most important issues in Islam and should be known well for Arabic and non-Arabic Muslims’ community. Thus, the confusion of these issues for non-Arabic speaker was identified. Therefore, this research tries to reduce the knowledge gap between these classifications and non-Arabic speaker Muslims’ community.

Example of ontology structure in Holy Quran: The Quran consists of vast knowledge and it is organized into Juz, Surah and Verse. Juz has juz number, Surah consist of chapter number and chapter name, while Verse is a sentence where it has the detailed explanation on Quran contents. Explanation in verses covers all aspects of human life and the universe. In this study, we have focused on the prayer topic. Information about the prayer topic was obtained from reading materials about prayer interviewing pious people whom are well versed in Quran well-versed for understand relation among Juz, Surah and Verse. Figure 1 show the organization of the Quran as an ontology structure in Protégé, which shows the relationship among Juz, chapter and verse by general. Figure 2 shows how query can be answered using Protégé.

Figure 1. Organisation of Quran in Figure using Protégé

Figure 2. Query for find verse in Juz 11
7. Conclusion & Future Work
This paper has discussed about ontology, the ontology languages and ontology tools. It shows how ontology can be used in organising knowledge in the Quran, with an example of the prayer. The future work here is in having the whole structure of ontology defined in ontology that would enable users to find information/knowledge about Quran faster.

References


[33] Siddique, "Introduction to Holy Quran."