Using Ontology for Associating Web Multimedia Resources With the Holy Quran

Conference Paper · December 2013
DOI: 10.1109/NOORIC.2013.57

3 authors:

- **Mostafa Mahmoud**
  Taibah University
  
  12 PUBLICATIONS  31 CITATIONS
  
  SEE PROFILE

- **Yasser Abdelhamid**
  University of Tabuk
  
  12 PUBLICATIONS  18 CITATIONS
  
  SEE PROFILE

- **Tarek El-Sakka**
  University of Sharjah
  
  5 PUBLICATIONS  0 CITATIONS
  
  SEE PROFILE

All content following this page was uploaded by Tarek El-Sakka on 05 March 2016.

The user has requested enhancement of the downloaded file. All in-text references underlined in blue are added to the original document and are linked to publications on ResearchGate, letting you access and read them immediately.
Abstract—There is a huge wealth of multimedia web resources related to the sciences of the Holy Quran, including "Tafseer" of the Holy Quran, teaching the provisions of recitation, the stories of the Holy Quran, and many other categories of resources. There is no doubt that these resources would help in understanding the verses of the Holy Quran, clarifying its meaning, and specifying the right way of applying provisions of its recitation, especially for children where the audio visual effect of the multimedia resources would have a great impact on them. Ontology is the explicit formal specification of the terms in a domain and the relations between them. It has been used extensively for knowledge sharing, application integration, and the most for providing the capability of machines to understand and exchange web content. This paper addresses the possibility of using Ontology for linking the verses of the Holy Quran with the multimedia resources found on the web. The proposed system will provide the capability of dynamically searching the web for resources related to the verses of the Holy Quran based on common ontological terms.

Keywords— Multimedia Annotation, Ontology-based Application Integration, Ontology-based annotation.

I. INTRODUCTION

The Holy Quran is universally accepted by Muslims to be the infallible Word of God as first revealed by the Angel Gabriel more than fourteen hundred years ago. It provides the rules of conduct that remain fundamental to the Muslim. God put it aware of everything, Provisions and laws, Proverbs and wisdom, Homilies and history, the stories and the system of the universe.

These rules, provisions, laws, proverbs, wisdom, homilies, history, system of the universe, and others, constitute the main concepts and themes of the HQ. Each chapter and each verse of the HQ considers one or more of these concepts, sometimes they are clear to the ordinary reader, and sometimes they need to be clarified by the help of those who are knowledgeable in the HQ sciences. Even "Tafseer" books like Al-Qurtuby, Ibn-Katheer, and Al-Jalalain, are formulated in the old Arabic language that is hard for the ordinary reader to grasp or understand. The most important issue is that every day, scientists reveal new miracles of the HQ that haven't been known before, that makes the "Tafseer" of the HQ subject to changes and updates.

Nowadays we are living the era of information. Thousands of resources are just one mouse-click away, and it is not just written material, but also it is audio and video materials. There is a huge wealth of multimedia resources related to the sciences of the Holy Quran. Some of these resources are for the "Tafseer" of the HQ, some of them are for teaching the provisions of its recitation, some of them are telling the stories of the HQ, and some are revealing the miracles of HQ.

Web search engines like Google, Yahoo, Bing, and many others depend on textual matching. If you are trying to search for a specific resource, you have to pick-up the right and exact word to get the right results. This way wouldn't help in the case of HQ because the Arabic language of the HQ is very rich such that the same concept or the same theme is formulated in different locations in the HQ with different words. So it is better for the search to be semantic rather than textual.

Ontology is the explicit formal specifications of the terms in a specific domain and relations among them [1]. It defines a common vocabulary for researchers who need to share information in a domain and includes machine-interpretable definitions of basic concepts in the domain and relations among them. Ontology is becoming of increasing importance to a large number of application types like knowledge-based systems, information exchange, semantic web, and application integration. Many disciplines have developed their own standard ontologies that domain experts can use to share and annotate information in their fields. SNOMED [2] is an example of a large structured and standardized vocabulary in the domain of medicine. UMLS [3] is a semantic network of Unified Medical Language System in medicine. UNSPSC [4] is also an ontology which provides common terminology for products and services.

Few efforts have been dedicated to the goal of building the Ontology of HQ [5], [6], where the HQ corpus is collected and augmented with the ontology of key concepts and themes, with the help of well recognized experts.

This paper investigates the possibility of linking HQ verses with web multimedia resources related to the HQ using Ontology Tagging. Also it introduces a suggested framework for using HQ Ontology terms in annotating and linking related web multimedia resources, and demonstrates a prototype of the proposed web multimedia enabled HQ browser.

We first discuss the research efforts that have been conducted for building HQ Ontology and Ontology tagging as general. Then, we specify the details of the proposed system and its components. Finally, we discuss a sample run of the web multimedia enabled HQ browser, the difficulties that we have found, the suggested solutions, and future work.
II. ONTOLOGY TAGGING

Ontology is a formal explicit description of concepts in a domain of discourse. Building an ontology includes defining classes/concepts related to that domain, properties/slots of each concept that describes various features and attributes of the concept, and facets of slots that describe restrictions on the values of these slots. In addition to arranging the classes in a taxonomic hierarchy that describe relations between these concepts [1].

A. Ontology of the Holy Quran

The ontology of the holy Quran is not an easy work because the semantic meaning of each word may have different definitions according to its position in the verses and it may take several meanings while using different "Tafseer" of the Quran. A little effort has been made in this domain.

1) Mushaf Al Tajweed: Mushaf Al Tajweed (MAT) [7] was the first try to build an ontology for the Holy Quran. It enables people to read the Quran verses correctly. It also includes Ontology of topics at the end of the book. This Ontology of topics/themes is arranged in a hierarchical order. It covers nearly 1100 topics. The MAT ontology of topics was chosen as the gold standard as it represents expert Islamic knowledge. It arranges the verses of the Quran according to the context of the verses (rather than according to the lexical terms of the verses) which draws an accurate picture of the correct meaning of the verses. The set of terms chosen from MAT Ontology of topics were chosen randomly. All of them consisted of one word only. Some of them represent places, others represent names, others represent animals, etc. This set of terms represents all possible user queries submitted to any of the Quran search tools available on the web.

2) Qurany Concepts Tool: Abbas [8] built a prototype for the HQ Ontology called "Qurany". This tool provides a formal categorization of all the concrete and abstract concepts listed in the HQ text. Each conceptual element in the Ontology is assigned a unique identifier. This allows the tasks of pronoun resolution and named entity tagging to easily map to the relevant concepts in the Ontology.

The main idea of this tool is to create an HTML file for each Quran verse. Each file contains the Arabic verse, the eight English parallel translations for this verse and all the abstract concepts covered by this verse. These files are hosted on the University of Leeds server along with an index file for all the related html files.

3) Modeling and retrieval from the Holy Quran and religious texts: A semantic based knowledge representation model for the Holy Quran and its related resources is presented by Sumayya Baqai et al. [9]. The framework of the model enables reasoning from a knowledge base of Quran and its related textbooks via Ontology and semantic reasoning. The system's framework starts by the 'Data Collection' process, which collects the Holy Quran, Ahadith, and the related texts. They are then standardized through the 'Metadata Generation' process, which parses the collected text to extract metadata, and 'Tag Generation' process, which extracts tags that will be used later in the annotation process.

Metadata and Tags are combined for formulating the data representation in XML. Subsequently, the 'Knowledge Modeling' process builds Ontology models using Ontology schemas for Quran and related textbooks. The Ontology is populated with the XML representation documents, and stored in Ontology repository. The system provides to the user a facility of the 'Knowledge Retrieval' process. In fact, the research faced various challenges due to the complexity and richness of the Islamic Knowledge content; the structural organization and thematic complexity of the Quran. The research presented a great and considerable effort, due to the dealing with highly sensitive, critical, affluent, and complicated resources. Such systems may be considered as a source for the new Muslims for learning and understanding their religion. However, the system suffers from the drawback of the use of transliteration rather than the absolute Arabic characters in representing the Arabic words.

4) An online annotated linguistic resource: Eric Atwell et al [10], describes applications presented on the Web site http://corpus.quran.com/ at Leeds University on Arabic and the Quran. One of these applications is the Ontology of Quranic concepts, which defines the concepts in the Quran, and uses the predicate logic to show the relationships between these concepts. The Named Entities, such as the name of historic people and places mentioned in the Quran, are linked to concepts in the Ontology. Their work is undergoing towards extracting concepts and named-entities in the Quran, establishing ontological links and relationships among these concepts and resolving pronominal anaphoric references to these concepts.

5) Ontology driven Information Extraction from the Holy Qur'an related Documents: Qurat ul Ain and Amna Basharat [11] presented a comprehensive approach to develop semantic based Information Extraction framework for the Holy Quran related documents. The tool supports well defined mechanisms for knowledge modeling and retrieval for the documents related to the Holy Quran domain. It links varying text available in different documents on a single platform for more efficient and effective integration and retrieval of such literature. They introduced many techniques to develop a comprehensive package of facilities playing a fundamental role in the overall functionality desired from such an application. In addition to this, "DataQuest" also provides a good methodology for finding the domain related knowledge of the simple natural language and then annotating and saving the documents with the concepts and information contained in them. It also provides the methodology for more relevant information retrieval thus providing a complete information management framework.

B. Holy Quran and Multimedia

Many researchers have studied the effect of using multimedia in learning and memorizing the Holy Quran. Azman, F. N. et. al. [12] introduced a framework for the development of Quran script learning multimedia application...
which suggests solutions to the problems of HQ reading to improve the skills of reading the HQ.

Omar Ibrahim, et al [13] introduced an effective learning method by using multimedia through a model called Al-Furqan to assist students to learn and memorize the Holy-Quran effectively and efficiently.

Another research was introduced by Tariq Mssraty and Qais Faryadi [14], where they constructed a new pedagogical model to help primary school teachers to teach Quranic recitation. They introduced a model based on the implication of the multimedia based interactive learning method to increase learners’ skills, attitude, motivation, grade, and knowledge while learning how to recite the Quran.

C. Associating Holy Quran with Web Multimedia

There is a wealth of multimedia resources related to the HQ on the Internet. Associating these resources to the verses of the HQ will provide a valuable assistance to the HQ researchers and readers. Our proposed approach is to use Ontology of the HQ which is the collection of concepts and themes of the HQ as the background on which links between verses of the HQ and multimedia resources available in the Internet can be built. The ultimate goal is to build a browser for the HQ that enables the reader to play the related multimedia resources of a selected verse or chapter.

As shown in Figure 1, the proposed framework consists of seven components; Building HQ Ontology, Tagging HQ verses and chapters, Collecting HQ related web multimedia resources, Tagging collected web multimedia resources, filtration and proofing of the collected resources, Building the HQ-Web multimedia index, and finally building the multimedia-enabled HQ browser.

The following is a detailed description of the proposed framework.

1) Building HQ Ontology: Concepts of the Holy Quran are heterogeneous as it addresses many issues related to life and death, historical figures, reward and punishment, and many other subjects, in addition to the structure itself that includes parts, chapters, verses and pages. So, building the Holy Quran ontology entails gathering and classifying these concepts in a taxonomy that helps in tracing the relation between these concepts. The function of this subsystem is to collect the set of concepts and themes related to the HQ, and to put them in a hierarchy that describes the relation between them. We started with the HQ Ontology prototypes found in [5], [6] as they are considered as a good start towards building the Ontology of the HQ. We used only the concept taxonomy as it includes the key features that can be used in tagging multimedia resources.

2) Tagging HQ verses and chapters: The goal of this subsystem is to annotate each verse and chapter in the HQ with one or more of the concepts and themes generated in the previous subsystem, so that each verse or chapter will have one or more HQ Ontology tags that characterize the verse or chapter.

3) HQ related web multimedia resources: At this phase we use generated HQ Ontology concepts and themes in finding the related multimedia resources on the web. To narrow the scope of the multimedia resources, we chose the YouTube as the source of video feeds using Google API. We ran this API with specific constraints to look for the multimedia resources related to our HQ ontology concepts in the YouTube database. We used samples of the HQ ontology concepts. Finally, we used the results information like, the title and description of the feeds for matching HQ Ontology terms.

4) Tagging collected web multimedia resources: After collecting the links to the web multimedia resources, we use the HQ Ontology concepts and themes in annotating them with one or more of these concepts, so that each collected multimedia resource will be attached with one or more HQ concepts.

It is noteworthy that we use the capabilities of the Ontology taxonomy relations to collect and annotate the resources. For example: resources annotated with “Moses” tag, are also tagged with higher concepts in the HQ concept taxonomy like “Prophet”, and “Historical figure”.

5) Filtration and proofing of the collected resources: This is the most difficult and important subsystem of the proposed framework, as its function is to make sure that the collected multimedia resources are really matching the attached HQ tags.

6) Building the HQ-Web multimedia index: At this stage we build an index that attaches HQ verses and chapters with the collected web multimedia according to their attached tags, and using the taxonomy of the HQ Ontology.
7) Building the multimedia-enabled HQ browser: The Multimedia-enabled HQ browser is a tool that enables the reader to traverse the chapters and verses of the HQ, and upon selecting a verse, the tool searches the HQ-Multimedia index of the related resources and makes them available to the user. Upon selecting one of these resources, the tool plays its contents (Audio, Video,...).

III. PROPOSED HQ-WEB MULTIMEDIA ASSOCIATION TOOL

As mentioned before, the main goal of this research is to investigate the possibility of building a multimedia-enabled HQ browser that enables the reader to play web multimedia resources related to a selected verse of the chapter.

In the last section we explained a suggested design of the proposed framework. In this section we explain the details of a prototype of the proposed system.

A. HQ Text Representation

There are many web sites providing digital copies of the HQ like Tanzil [15], Quran Database [16], and others. We used the one of King Fahd Complex for the printing of the Holy Qur’an – in Al-Madinah Al-Munawarah [17] as it is the most authenticated one.

After downloading the digital from the King Fahd Complex as a word file, we transferred it into a structured database format, with some useful metadata on Parts, Chapters, Verses, and Pages of the HQ according to the “Uthmani” verses text.

The database is structured hierarchically as Parts, Chapters, and Verses as shown in Figure [2], [3], and [4] consecutively.

We used two representations for HQ verses. The first was developed for text manipulation as simple verses text (text without diacritics).

The second representation was developed for the presentation purposes as “Uthmani” HQ verses text (text with diacritics). This representation is viewed as pages with the structure shown in Figure 5.

B. HQ Ontology

The ontology of HQ is a set of concepts and themes found in the HQ. They are represented as a hierarchy that describe the relation between these concepts and themes. Figure 6 displays the first three levels of the constructed HQ Ontology.

C. HQ Verses Annotation

We used the collected Ontology terms for annotating the HQ at the verse level. This has been done by analyzing the text of verses and matching them with the terms found in the HQ Ontology.

The HQ verses annotation process starts from the root of the ontology tree and traverses the tree branches to build the linkage of each term with the related HQ verse.

D. Multimedia Resources Annotation

The multimedia resources annotation process traverses the HQ Ontology tree starting from the root, and uses each term for searching the web for related multimedia resources. In the presented prototype, we used YouTube data API for collecting video feeds to be associated with the HQ verses.

The collected video feeds include Multimedia Tag Text, Multimedia Tag Description, and a Multimedia Tag URL.

IV. SAMPLE RUN

Using the proposed Multimedia-Enabled HQ Browser is fairly simple. Like all other HQ browsers, the user can reach the proposed HQ browser prototype through the following link: http://www.quranexplorer.info/.

The user is given the means of selecting a specific chapter, or a specific Part of the HQ. The selection is displayed as Pages according to the organization of the “Uthmani” HQ text.

Verses are represented as hyperlinks that when the reader selects one of them, the system displays the related web multimedia resources as a list of hyperlinks. Upon selecting
one of them, the multimedia resource is played directly on the same page.

Figure 7 displays a sample run that shows the web multimedia resources linked with the verse number 1 in Chapter “Al-Fil”.

V. DISCUSSION OF CONFRONTED PROBLEMS

During the development of the prototype of the proposed system a number of problems have been revealed. In this section we shall clarify these problems along with suggested solutions for some of them.

1) Misleading description of multimedia resources: To narrow the scope of the multimedia resources, we chose the YouTube as the source of video feeds using Google API. We used the title, and description of the feeds for matching HQ Ontology terms. But the initial results revealed that some of the video feeds have a description that does not match the content of the feed.

The proposed solution of this problem was to use matching at different levels of the HQ Ontology hierarchy, and apply matching at more specific terms down the hierarchy of HQ Ontology terms.

The second solution was to restrict search for feeds in only authentic sites in YouTube, those related to HQ feeds.

2) Generic Concepts: The concepts and themes that constitute the HQ Ontology include generic terms that match most of the HQ verses at one side, and match a large number of multimedia resources at the other side.

This problem has been relieved by starting matching of the HQ terms starting from the third level of the HQ Ontology hierarchy.

3) Replication of the same content with different descriptions: Another problem that we have found is that feeds are sometimes repeated by different sites with different descriptions.

This problem causes lengthy results of matching resources, and can be filtered using other parameters of the feed like type, length, etc.

VI. CONCLUSION

Ontology is becoming of a great importance in the field of information technology, it is the key for integrating applications, information sharing and information association. In this paper we introduced an approach for associating HQ verses with the available multimedia resources related to the HQ verses on the web. We used the HQ Ontology for associating HQ verses with web multimedia resources. A framework has been designed to build HQ Ontology, and using its concepts and themes in linking web multimedia resources with HQ verses. A number of problems have been revealed, and suggested solutions have been introduced.

A prototype of the proposed Multimedia-Enabled HQ Browser has been implemented, and a sample run has been introduced that clarifies the idea of the proposed system.

Future research will continue in providing alternative solutions to the confronted problems especially in the proofing and filtering subsystem.

REFERENCES

Figure 6. A Part of HQ Ontology Concepts Taxonomy

Figure 7. A Sample Run of the Multimedia-Enabled HQ Browser