

# Assisting Understanding, Retention, and Dissemination of Religious Texts Knowledge with Modeling, and Visualization Techniques: The Case of The Quran

Mohamadou Nassourou  
Department of Computer Philology & Modern German Literature  
University of Würzburg Am Hubland D - 97074 Würzburg  
mohamadou.nassourou@uni-wuerzburg.de

## Abstract

Learning a book in general involves reading it, underlining important words, adding comments, summarizing some passages, and marking up some text or concepts. Once deeper understanding is achieved, one would like to organize and manage her/his knowledge in such a way that, it could be easily remembered and efficiently transmitted to others. In this paper, books organized in terms of chapters consisting of verses, are considered as the source of knowledge to be modeled. The knowledge model consists of verses with their metadata and semantic annotations. The metadata represent the multiple perspectives of knowledge modeling. Verses with their metadata and annotations form a meta-model, which will be published on a web Mashup. The meta-model with linking between its elements constitute a knowledge base.

An XML-based annotation system breaking down the learning process into specific tasks, helps constructing the desired meta-model. The system is made up of user interfaces for creating metadata, annotating chapters' contents according to user selected semantics, and templates for publishing the generated knowledge on the Internet.

The proposed software system improves comprehension and retention of knowledge contained in religious texts through modeling and visualization.

The system has been applied to the Quran, and the result obtained shows that multiple perspectives of information modeling can be successfully applied to religious texts.

It is expected that this short ongoing study would motivate others to engage in devising and offering software systems for cross-religions learning.

Keywords: Knowledge Modeling, Meta-model, Knowledge Management, Content Management, Quran

## 1. Introduction

Knowledge Management is defined as a collection of processes for creating, retaining, disseminating, and utilizing knowledge. Knowledge is gained through learning process. There are different types of learning theories among them behaviorism, cognitivism, and constructivism. This paper discusses mainly constructivism because learners interactively build new concepts from information contained in books. A book is a collection of information which needs to be organized in order to create knowledge. If the book has got a logical structure, then a knowledge model could be created and represented using ontology.

Learning a book in general involves reading it, underlining important words, adding comments, summarizing some passages, and marking up some text or concepts. The purpose of learning is to gain knowledge. But there is general knowledge and specific knowledge. In this paper I am discussing about knowledge within a defined domain. I consider books whose logical structures can be described in terms of chapters. And each chapter is made of a header and a body section. The header comprises a title, an edition place, number of verses, and some optional fields such as author or writer. The body is mainly a collection of verses. Reading and understanding this kind of books is not a trivial task, because they usually encompass a lot of hidden knowledge, unusual expressions, and a peculiar style of narration. As consequence a well thought strategy is required in order to effectively understand them.

It is commonly known in chemistry or physics that students understand better when they model processes and reactions that they have learned [7]. However one cannot easily accept that modeling and visualization could improve understanding and memorization of abstract texts such as religious texts.

Breaking down a text into smallest meaningful parts, enumerating all possible required details about each part, and selectively visualizing each part, are reasonably effective ways of understanding and retaining the text.

The first step should be to discover their exact structural dimensions, and model them with descriptive languages such as XML. After that one should define some metadata for each part of the structure. Then their content must be semantically annotated while reading. Both metadata and annotations could be based on markup languages such as XML (eXtensible Markup Language) because of its platform independence and its contribution to the construction of semantic web. In general a markup language offers the possibility of storing and exchanging in structured manner knowledge gained from the learning process. Finally one needs to link the different types of knowledge to create a knowledge base, which is normally what is kept in mind for utilization and sharing with others.

A software system comprising an Interactive Annotator, a Static Annotator, an XML Editor, a Knowledge Tester and an Expert System, and a Publisher to a Web Server for the purpose of assisting understanding and retention of religious texts, has been developed in this study.

The Interactive Annotator or note manager assists users in adding comments about a word or group of words.

The Static Annotator helps creating and entering metadata. The metadata are a representation of the multiples perspectives of knowledge modeling, which by extension correspond to the basic questions of genetic edition.

Genetic edition is concerned with the coming into existing of texts. The annotator helps users to insert and revise questions about each verse. Usually genetic edition questions are of the following forms:

Where was the verse written? When was the verse written? How was the verse written? What are the ways of reading the verse? Who created and wrote the verse? Who does the verse refer to?

Additionally for the purpose of this research, following questions are included: Why was the verse created or written? Which message does the verse convey? What are other similar verses?

The XML Editor assists marking up a text in XML. XML provides standard means to explicitly describe documents' structures, and identify meaningful elements inside those documents.

The Publisher has the role of deploying and presenting the source document with annotations on a web server, so that others can discover it and learn or comment about them.

Sharing knowledge with others requires a meaningful and simple way of conveying the message. It should be accessible to everyone on any platform. Based on this requirement the eXtensible Markup Language (XML) seems to be a good candidate for data format, since it is platform independent and simple as well.

The Knowledge Tester and the Expert System are client programs that users can use to check and deepen their knowledge.

## **2. Problem Statement**

Understanding and memorizing religious books is a challenging task. Can computer based multiple perspectives of knowledge modeling and visualization techniques alleviate those challenges?

## **3. Proposed Solution**

The hypothesis is that modeling and visualizing structural and terminological dimensions of religious books with the help of computer-aided semantic annotations will improve understanding and retention of the books' contents.

Through Intuitive user interfaces, well structured presentations, and detailed explanation with analogies, knowledge modeling and visualization techniques could considerably assist understanding and memorization of religious texts.

## **4. Identifying The Multiple Perspectives of Information Modeling in Religious Texts**

The multi-perspective approach proposed in this study is based on the Zachman framework for Information Systems Architecture. The Zachman framework [1] proposes that six perspectives characterized by the phrases what, who, when, where, how and why are necessary for completely representing and understanding an object.

Knowing that religious texts consist principally of collection of verses, the six perspectives with respect to each verse could be identified as follows:

- a) The What perspective  
It refers to the verse itself, and how to read or recite it.
- b) The Who perspective  
It concerns the author, the editor, as well as those who the verse refers to.
- c) The When perspective  
It records the time when the verse was emitted and recorded.
- d) The Where perspective  
The place where the verse was created and recorded is meant here.
- e) The How perspective  
It treats questions such as how and from what the verse was created and recorded.
- f) The Why perspective  
It deals with the causes or occasions that lead to the creation of the verse.

Because the aim of this study is to find out how understanding and retention of religious texts could be achieved through knowledge management, I found it necessary to add some perspectives such similarity to others, and message conveyed (i.e which message does the text convey).

## 5. Understanding Religious Books

In general it could be said that one understands something if he/she sees how it can be used to achieve some purpose or goal. Understanding is a process whereby sound knowledge is gained. Sound knowledge must be manageable. But knowledge is the result of information modeling. Therefore understanding religious books requires first modeling the information they contain in order to acquire clear knowledge of their contents, and properly manage that knowledge. In other words the first step towards understanding religious books is to create a model of their contents.

### 5.1 Model Definition

The purpose of modeling knowledge contained in religious books is to effectively manage that knowledge.

Text Modeling is the process of structuring a text, as well as adding annotations for explaining that text.

Books under consideration are those whose logical structures can be described in terms of chapters. And each chapter is made up of a header and a body. The header comprises a list of attributes that could be termed as metadata of the chapters and their contents. The body is mainly a collection of numbered blocks of texts.

In short, the structure of the considered books could be described with the following XML fragment:

```
<book>
<chapter n="1">
  <attr1></attr1>
  <attr2></attr2>
  ...
  <body>
    <text_block1 n="1"></text_block1>
    <text_block2 n="2"> </text_block2>
    ...
  </body>
</chapter>
...
</book>
```

Annotating is the process of defining some attributes termed as metadata for describing the book itself, its chapters and chapters' contents. The purpose of the metadata is to provide detailed information about the book and its content for facilitating understanding and information retrieval.

### Conceptual Modeling of Annotations using XML

What are the methods for reading and understanding a document?

The best way to understand a document is to:

- a. translate it into one own words.
- b. make annotations

What I mean by modeling annotations from conceptual point of view is to organize the annotations in terms of themes and encode them with XML. The encoding tag for each concept / theme should be a key word for the concept. I suggest two main types of annotations:

1. static annotations
2. dynamic annotations

Static annotations are those that the encoding tags are created once for each sentence of the document. The encoding tags are made before reading the document. For instance for each sentence one would like to know where, when, how, why ... the sentence was included in the book.

Dynamic annotations are annotations made for any text in the book. The encoding tags for these annotations are the text itself.

### 5.2 Metadata Definition

Modeling a religious book requires viewing the book's content in several perspectives. Since the content is mainly a collection of numbered verses, it is therefore necessary to analyze each verse in all possible perspectives. The multi-perspective approach that I have considered in this study is principally based on the Zachman framework for Information Systems Architecture [1]. The Zachman framework defines six perspectives on information characterized by the phrases who, what, how, when, where and why.

A book could be understood through its internal and external evidences. Internal evidence refers to the content and meaning of the book, and external evidence points to data surrounding the book such as place and date of publication, author, and so on. To some extent one could argue that, these evidences correspond basically to the Zachman's proposed perspectives on information systems architecture.

Based on these evidences I selected following terms as metadata for religious books:

- 1) place of creation/edition
- 2) date of creation/edition
- 3) method of creation/edition
- 4) occasions/causes of creation/edition
- 5) ways of reading
- 6) explanation
- 7) similarity
- 8) writer/editor

The metadata of the book itself could consist of:

1. chapters' names with number of verses
2. chapters' order of creation/edition
3. total number of verses
4. other optional features

It must be noticed that, books that comply with this way of structuring their contents could also be viewed as an aggregation of chapters consisting of collection of words. Using this view it is possible to annotate each word or group of words independently of the metadata.

Having defined the model and its metadata, it is possible to generate a meta-model that could be based on synthesizing the model and its metadata.

### 5.3 Meta-Model Architecture

Modeling improves understanding of information and the relationship between its components. The meta-model includes the model, the metadata, and the annotations used to generate a knowledge base. In other words it comprises the content of the book with all the additional descriptive annotations. The meta-model provides the overall structure of the created knowledge base.

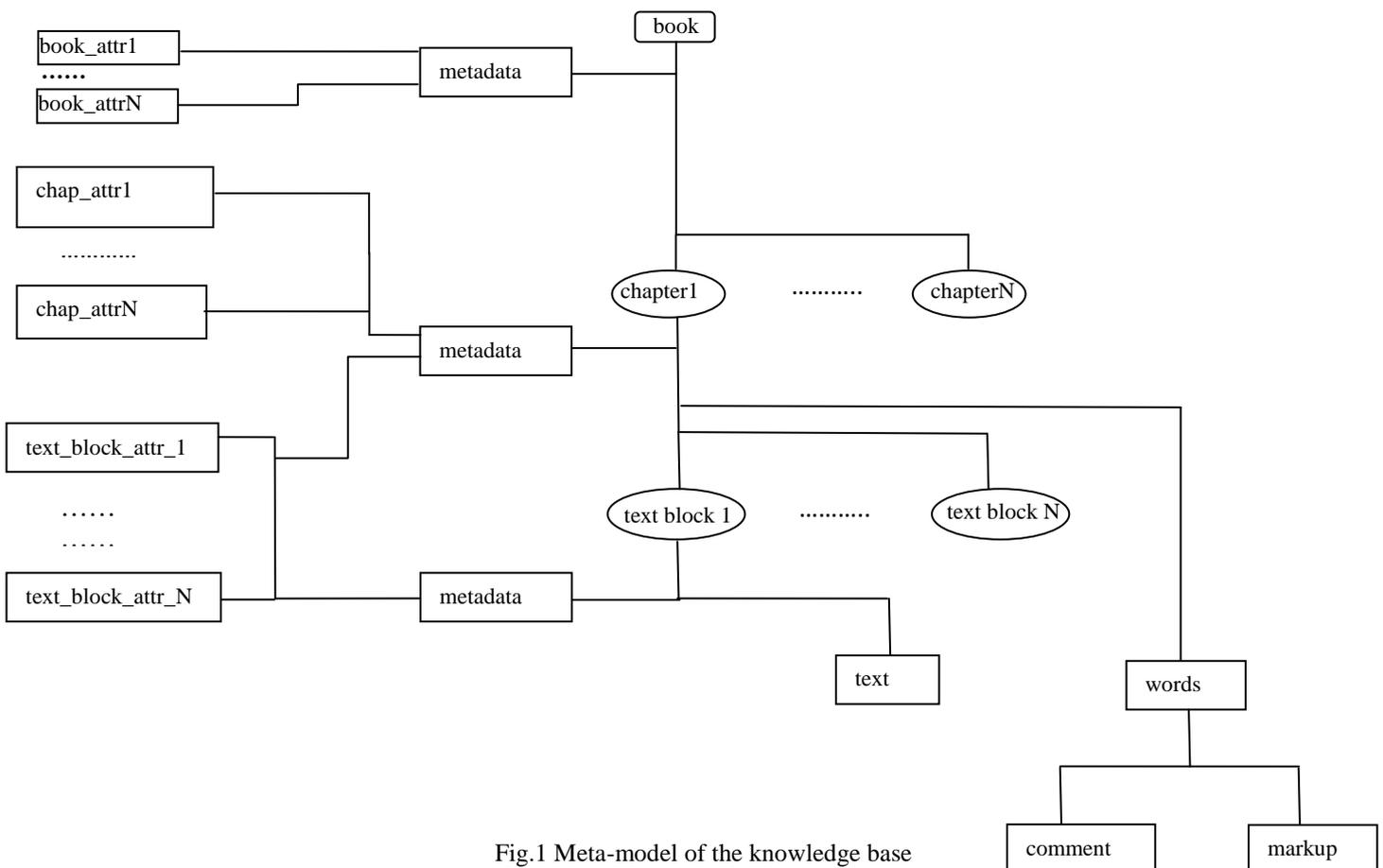


Fig.1 Meta-model of the knowledge base

## Description of the Meta-Model Architecture

The meta-model consists of a book comprising some chapters made up of block of texts. Chapters and their contents are described with some metadata. The metadata convey clear semantic about what they describe. For instance

- a) book\_attr1 for attribute 1 of the book
- b) chap\_attr1 for attribute 1 of the chapter
- c) text\_block\_attr1 for attribute 1 of the text block

Considering chapters as collections of words, gives the possibility to annotate each word by associating comments and markups, for example:

- a) comment for comments made about a particular word
- b) markup for tags' names that enclose a word for describing it

The meta-model serves as a knowledge base structure which needs to be created and collectively used.

If data for the metadata (i.e attributes) are not available, text mining methods should be applied to derive those data from the content of the book itself. A sample XML representing a knowledge base, as well as some text mining methods for deriving metadata, will be presented in section 9.

Now that I have explained how to model books whose structural dimensions could be described in terms of chapters consisting of blocks of texts, I would like to introduce a semi-automatic system for creating, storing, and disseminating knowledge from these types of books. In other words the following sections will describe a knowledge management system that eases gaining knowledge from religious books, as well as a visualization platform for publishing and mining the acquired knowledge.

## 6. Knowledge Management System

A Knowledge Management System (KM System) is defined as any system that offers possibilities to manage knowledge in organizations in order to support creation, capture, storage and dissemination of information.

Referring to the current study, the system that has to be developed must be able to capture all the different kinds of information for every verse of the religious book under consideration, and make this information accessible to any interested person.

The system consists basically of a Mashup website and a Content Management System. The Mashup website comprises a Publisher, Tester, and an Expert System. And the Content Management System is based on an XML Generator, XML Editor, Static and Dynamic Annotators programs.

In fact the visualization component of this study is principally based on the website. The Content Management is the modeling component.

### 6.1 Architecture of the System

The Architecture is made up of layered architectural diagrams with each layer depending on the layer below it. It is a component-based software system consisting of six major modules interacting with each other.

Figure 2 shows the diagrammatical view of the architecture.

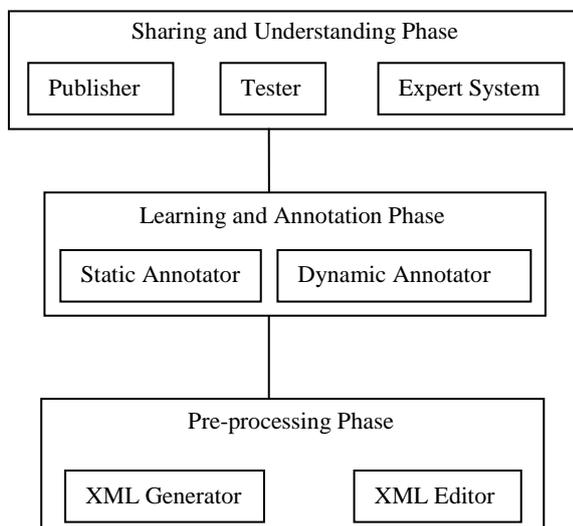


Fig.2. Architecture of the system

## Components Description:

### 1. XML Generator

It is a program that users can use to automatically convert single text file or a collection of text documents into XML.

### 2. Smart XML Editor

The editor is equipped with an interface for creating an ontology, which could be used to automatically mark up some well known features such as persons, animals, planets...etc. This editor is capable of eliminating overlapping using fragmentation technique described in [12]. A content assistant provides users with the possibility to choose tag to insert, and in case non well-formedness arises, it rectifies the illness by fragmenting the overlapping elements, and inserts corresponding missing tags. If none of the options is selected, or a text is selected, the content assistant inserts both closing and opening tags. Currently the editor is still under development to include the possibility of fragmenting overlapping elements without the need of a content assistant.

### 3. Static Annotator

It is a program for entering metadata names and their data. The annotator provides interfaces for properly and efficiently creating and managing metadata of each chapter and verse.

### 4. Dynamic Annotator

It is a program that offers the possibility to annotate a text by selecting a label or concept for the annotated text. An annotated text is automatically highlighted and hyperlinked to its annotation. A user can at any time click on the link to display or update the annotations. Moreover searching the annotations by using the selected labels or concepts is also possible.

### 5. Expert System

It is a system that users can consult for learning and testing their knowledge. In other words it helps gaining and checking through questions and answers interaction what has been or to be studied.

### 6. Tester

Different researchers use different methods for measuring understanding. However majority of them have used: true-false questions, multiple choice questions, paraphrasing documents, and answer questions. The Tester in this research consists of multiple choice questions, and a collation system. The collation system can compare two texts in order to identify their similarities and differences. A collation system could ease memorization of religious text by highlighting similarities and differences, for more information on textual documents collation see [13].

### 7. Publisher on Web Server

This is a website for sharing and communicating with others. It consists of several web pages linked through the verses and metadata labels.

The publisher offers the possibility to search in different languages, and interactively visualize the search results as described in [4].

Having briefly described the system, I would like to explain how the components are used to create and manage a desired knowledge base.

## 6.2 Generating The Knowledge Base

In this section I am going to show how the system developed in this study could be used to implement knowledge management activities.

- a. **Knowledge Creation** involves discovering, capturing, and registering the knowledge into the system. The book from which the knowledge has to be generated could be locally stored or retrievable from the Internet.

The following functional components implement the knowledge creation activities. Interactively create metadata of verses:

1. Creating metadata of each verse is achieved with the Static Annotator component.
2. Manual and automatic annotations of verses are done by the XML Generator, the XML Editor, and the Static and Dynamic Annotator components.

- b. **Knowledge Dissemination** or knowledge transfer refers to communicating the knowledge in a well-defined format so that others can effectively use it. Following is the component which takes care of this process.
  1. The Publisher component which is a Mashup website.
- c. **Knowledge Utilization** is concerned with the application of acquired knowledge wherever needed. It is realized with the following systems:
  1. Test unit
  2. Expert system
  3. The Publisher component which is a Mashup website.
- d. **Knowledge Preservation** is achieved with the help of XML. The created knowledge base and the source documents are stored in XML format.

## 7. Working Principle of the System

The ontological diagram below describes the overall working principle of the system.

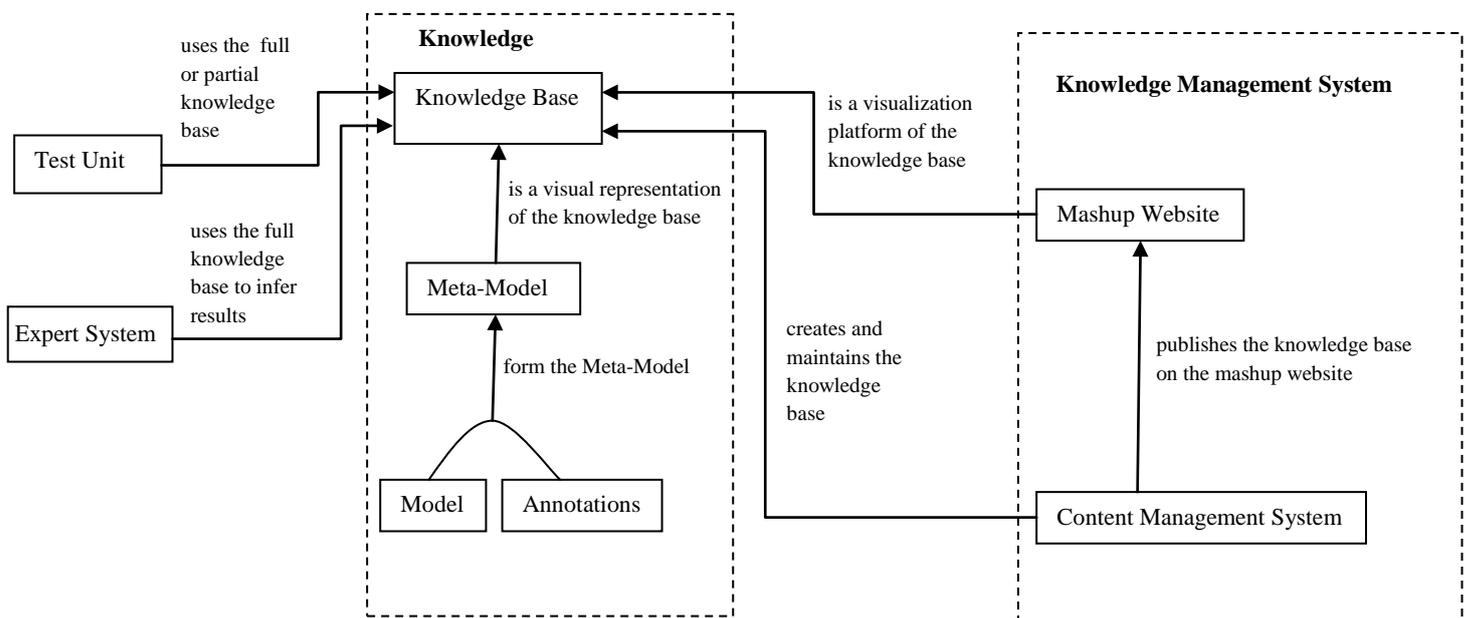


Fig. 3 ontology of the system

### Description

The diagram in fig. 3 shows the major components of the overall system. The knowledge block consists of the knowledge base represented by the meta-model that comprises the model with its annotations. The knowledge base is created and maintained by the Content Management System and the Mashup website. The Content Management System is also responsible for publishing the knowledge base on the Mashup website.

The Test Unit and the Expert System are full or partial consumers of the knowledge base through request response interactions, as well as dialog base communication.

## 8. Implementation

The system was implemented in Java and PHP. The Content Management System was first developed in Java as RCP eclipse plugin, and then in PHP, Javascript, and HTML/CSS.

## 9. Applications of the Meta-Model

The meta-model could be applied to any book or document whose schema or organization is compliant with the above described model. Religious books particularly would definitely fit into this category, and are therefore excellent examples for implementing this meta-model. Following is a use case based on the Quran.

## 9.1 Case of the Quran

The holy Quran contains 6236 numbered verses divided into 114 chapters. Some chapters were revealed in Mecca while others in Medina (cities in Saudi Arabia). According to Muslims each verse was revealed to Prophet Mohammad by God through Angel Gabriel at a specific place, in a given time, in a particular way, and for a well needed purpose.

Understanding the Quran is not an easy task because each verse is a chapter on its own, and each chapter is a Quran (i.e a book) on its own. This implies that understanding the Quran requires understanding each individual verse. Knowledge contained in each verse can be acquired by analyzing and comprehending the detailed information about that verse.

## 9.2 Modeling the Quran's Content

The Quran has got a structure that could be described in terms of chapters. And each chapter is made up of a header and a body section. The header comprises a title, a revelation place, and number of verses. The body is mainly a collection of numbered verses.

An XML representation of each chapter is shown below:

```
<quran>
  <chapter n="xxx">chapter name</chapter>
  <placeofrevelation>place of revelation</placeofrevelation >
  <numberofverses>number of verses</numberofverses >
  <verse n="1">first verse </verse>
  <verse n="2">second verse </verse >
  <verse n="3">third verse </verse >
  .
</quran>
```

To define some metadata for the Quran's chapters and verses, there is need to have basic knowledge about its origin. According to Muslims each verse was revealed to Prophet Mohammad at a specific place, in a given time, in a particular way, and for a well needed purpose. These characteristics of each verse are actually a hint for defining a multi-perspective approach of viewing a verse. The multiple perspectives could be mapped to some metadata with the purpose of providing detailed information about each verse.

Without going into details, the multiple perspectives of verses could be represented with the following terms.

1. place of revelation
2. date of revelation
3. method of revelation
4. occasions/causes of revelation
5. ways of reading
6. explanation
7. writer/editor
8. similarity to other verses

The following terms have been chosen for the metadata of the Quran itself:

1. chapters' names with number of verses
2. chapters' order of revelation
3. total number of verses

The Quran could also be viewed as an aggregation of chapters consisting of collection of words. Using this view it is possible to annotate a word or group of words independently of the metadata. Therefore each word could be associated with comments and markup tags for improving understanding and retention.

## 9.3 The Quran's Meta-Model Architecture

We need to substitute the names of attributes in the meta-model described in fig. 1 with our defined metadata. The meta-model consists of chapters, which in turn are made up of verses. Each chapter and verse has got some metadata for describing it. The metadata's names convey clear semantic about what they describe. The meta-model is a representation of the knowledge base that users or learners need to keep in mind.

The following sample XML shows the created knowledge base.

```
<quran>
  <verse n=" surah and verse number">
    <Text> The verse</Text>
    <PlaceOfRevelation>Place where it was revealed</PlaceOfRevelation>
    <DateOfRevelation>Date when it was revealed</DateOfRevelation>
    <MethodOfRevelation>Through which means it was revealed</MethodOfRevelation>
    <CauseOfRevelation>Why was it revealed</CauseOfRevelation>
    <ReadingWays>How many ways it could be read</ReadingWays>
    <Explanation>What is the meaning or message it contains</Explanation>
    <SimilarTo>Verses that are similar or related to the verse</SimilarTo>
    <Editor>The Writer or compiler of the verse</Editor>
  </verse>
.
</quran>
```

An example of a verse with metadata is shown below.

```
<quran>
  <verse n="1:2">
    <Text> الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ </Text>
    <PlaceOfRevelation>Mecca</PlaceOfRevelation>
    <DateOfRevelation>615 AD (approximately)</DateOfRevelation>
    <MethodOfRevelation>Recitation from Angel Gabriel</MethodOfRevelation>
    <CauseOfRevelation>What to recite when worshipping (offering compulsory prayers) Allah</CauseOfRevelation>
    <ReadingWays>As written</ReadingWays>
    <Explanation>Allah is the only owner of the worlds; therefore worship Him alone by reciting this verse</Explanation>
    <SimilarTo>6:45, 10:10, 37:182, 39:75, 40:65</SimilarTo>
    <Editor>The Writer or compiler of the verse (Uthman third Caliph)</Editor>
  </verse>
.
</quran>
```

In the above XML fragment, the “SimilarTo” tag is confined to verses that contain the displayed verse. However it is possible to include all verses that share some percentage of similarity with it, for more details on this issue see [4]. The translation of the verse in English is: “Praise be to Allah, the Cherisher and Sustainer of the worlds”. As mentioned previously, the metadata would be derived from the Quran itself, if they are not available in other Islamic resources such as narrations (hadiths) of the prophet. Some text mining techniques explained in [5] have been used to derive chapters’ dates and places of revelation. In fact one of the main objectives of this study is to apply computational methods to the text of the Quran in order to derive the metadata and explanation of each verse.

## 10. Results and Discussion

The system has been applied to the Quran, and the results obtained show that multiple perspectives of information modeling can be successfully applied to religious texts. Five people (including myself) among them some with little computer literacy have tested the system, and were positively impressed. Even though the system is not yet complete, all the testers acknowledged that the system has effectively improved understanding and memorization of the Quran’s verses.

It is commonly known in chemistry or physics that students understand better when they model processes and reactions that they have learned [7]. However one cannot easily accept that modeling and visualization could improve understanding and memorization of abstract texts such as religious texts.

Breaking down a text into smallest meaningful parts, enumerating all possible required details about each part, and selectively visualizing each part, are reasonably effective ways of understanding and retaining the text.

In fact nowadays I am attempted to argue that correct understanding of any religious text requires meaningful modeling and visualization of that text. This study is an attempt to perform such a task, and hopefully the system will be improved from both theoretical and practical aspects.

Of course I do not deny that numerous studies in this direction have been conducted, for instance the multitude existing software for learning religions are definitely results of such studies. However those software programs do not offer users or learners the chance to create the knowledge on their own. Moreover the programs are made for particular religious books; hence they cannot be adapted to learn other religious texts.

So I hope this short ongoing study would motivate others to engage in devising and offering software systems for cross-religions learning.

## 11. Conclusion and Future Work

In this study a modeling and visualization system for assisting understanding, memorization, and sharing of knowledge obtained from religious books has been presented. The system is based on books complying with religious books' schema namely the Quran. The modeling is achieved with the Content Management System, and the visualization is based on a Mashup website. A meta-model representing knowledge base created from a considered religious text has also been shown. The meta-model is an attempt to devise a common architecture for structural and terminological dimensions of religious texts. Even though incomplete some people have tested the system, and the feedback was encouraging.

The next steps in this study will be the completion of the system by offering an Expert System which could allow distance users to consult and learn religious texts. The Content Management System is currently developed as eclipse RCP plugin, I am working on converting it into a web based management system in order to offer online maintenance of the Mashup website.

Refining the meta-model by further extending the multi-perspective approach of knowledge modeling could considerably improve the goal of this study, which is to assist understanding and retention of religious texts.

## References

- [1] J. Zachman. "A Framework for Information Systems Architecture" IBM Systems Journal, 26(3):276–292, 1987
- [2] <http://www.openify.com/files/Designing%20A%20Knowledge%20Management%20System2.pdf>
- [3] J. Claudio, T. Angeloni, "Understanding the difference between Information Management and Knowledge Management", TerraForum Consultores, Toronto, ON, Canada
- [4] Mohamadou Nassourou, "Assisting Analysis and Understanding of Quran Search Results with Interactive Scatter Plots and Tables", <http://nbn-resolving.de/urn:nbn:de:bvb:20-opus-55840>
- [5] Mohamadou Nassourou, "A Knowledge-based Hybrid Statistical Classifier for Reconstructing the Chronology of the Quran", accepted in WEBIST/WTM 2011, The Netherlands  
<http://nbn-resolving.de/urn:nbn:de:bvb:20-opus-54712>
- [6] M. Lansdale, "The psychology of personal information management" University of Technology, Loughborough, Uk
- [7] J. Heitz, "Modeling exercises improve understanding", university of Wisconsin
- [8] Y. Chali, E. Pascual, J. Virbel, "Text structure modeling and language comprehension processes", Université Paul Sabatier 118, Toulouse, France
- [9] John R. Kirby, "What have we learned about reading comprehension?", Faculty of Education, Queen's University
- [10] Yulong Gu, James Warren, Jan Stanek, Graeme Suthers, "A System Architecture Design for Knowledge Management (KM) in Medical Genetic Testing (MGT) Laboratories", Proceedings of the 10th International Conference on Computer Supported Cooperative Work in Design
- [11] Roelof P. uit Beijerse, "Questions in knowledge management: defining and conceptualising a phenomenon"
- [12] Mohamadou Nassourou, "Markup Overlap: Improving Fragmentation Method", accepted for nocturne presentation in Balisage 2010, Montreal, Canada,  
<http://nbn-resolving.de/urn:nbn:de:bvb:20-opus-49084>
- [13] Mohamadou Nassourou, "Design and Implementation of Textual Documents Collation Systems", to appear in <http://www.opus-bayern.de/uni-wuerzburg/>