# Towards a Knowledge-Based Learning System for The Quranic Text

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

## Abstract

In this research, an attempt to create a knowledge-based learning system for the Quranic text has been performed. The knowledge base is made up of the Quranic text along with detailed information about each chapter and verse, and some rules. The system offers the possibility to study the Quran through web-based interfaces, implementing novel visualization techniques for browsing, querying, consulting, and testing the acquired knowledge. Additionally the system possesses knowledge acquisition facilities for maintaining the knowledge base.

**Keywords: Knowledge-based System, Knowledge Management System, Text Mining, Visualization, Quran**

## 1. Introduction

Understanding religious texts such as the Quran is not a trivial task. Religious texts usually encompass a lot of hidden knowledge, possess peculiar style of narration, and are sometimes confusing. The Quran has got two sides: the text itself and the information surrounding the text. There are computer programs for linguistically (POS, NE tagging…) analyzing the Quranic text. However many people encounter difficulties in properly comprehending the text, because the background information about some parts of the text is missing, and the learning methodology inappropriate.
The Holy Quran is an early medieval book consisting of 6236 verses with almost half of the verses similar to each other, whereby 98 verses are repeated 181 times. Based on this similarity fact, machine learning techniques and text mining methods could help derive missing background information for each chapter and verse, and generate concise summaries whenever needed.

In this paper, the development of a knowledge based learning system for the Quranic text is discussed.
First, modeling and representation of the Quranic text using frames-based representation was made, in order to identify types of information that are needed for effectively analyzing and understanding the Quranic text.
After that, existing data concerning this information were collected from both online and printed materials.
Then, missing information were computed using text mining techniques and machine learning algorithms. Having gathered required information, a software program equipped with adequate and intuitive user interfaces was developed for recording, and maintaining the information in XML format.
This information formatted in XML constitutes the knowledge base, which could be made available for the purpose of interactive learning and research.
Finally, a visualization platform i.e a web mashup was created for browsing, querying, and consulting the knowledge base.
Results of queries or search could be both textually and graphically explored.

Moreover, the system offers a web-based printable user interfaces for the translation of chapters, with the possibility to add footnotes, and comments.

## 2. Motivation

My motivations behind the development of this software system could be summarized as follows:

a. Assisting people who require information surrounding the Quran, in order to effectively study it.
b. Application of machine learning methods to the Quranic text, for the purpose of comparing results obtained in this research with those obtained by traditional Islamic scholars and western orientalists researchers about the places, and dates of revelation of the Quranic chapters.

## 3. Problem statement

The tasks of analyzing, understanding and memorizing religious books are challenging. Could interactive knowledge-based learning systems alleviate this challenge?

## 4. Proposed solution

The hypothesis is that, computer-aided modeling, representation, visualization, and manipulation of the structural and terminological dimensions of religious books will improve understanding and retention of the texts they contain. Through intuitive user–friendly graphical interfaces, well structured presentations, and detailed explanation with analogies, a knowledge-based learning system could considerably assist understanding and memorization of religious texts.

## 5. Working Principle of the System

The ontological diagram in figure 1 describes the overall working principle of the system.

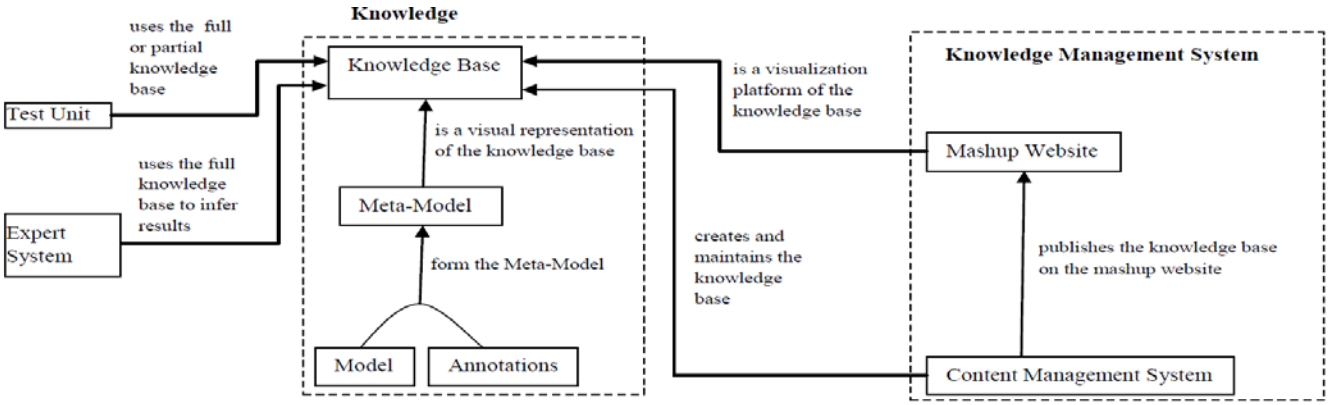



Fig. 1 Ontology of the system

### 5.1. Description of the system

The diagram in fig. 1 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, maintained, and visualized by the Content Management System (CMS) and the Mashup website. The CMS is also responsible for publishing the knowledge base on the Mashup website. The CMS and the Mashup website form the knowledge management system.
The CMS consists of an XML Generator program, XML Editor, Static and Dynamic Annotators programs equipped with graphical user interfaces, and a web-based user interface for translating Quranic chapters.

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.

## 6. Creation of The Knowledge Base

The major steps for the creation of the knowledge base are shortly described as follows:

### 6.1. Design of the knowledge base structure

It involves designing an explicit representation scheme for knowledge contained in the Quranic text. Following points have been successively implemented.

a. First, modeling the Quran syntactic and semantic structures
b. then, representing the created model using frames,
c. after that, generating an XML model of the frame representation.

### 6.2. Knowledge acquisition

This step consists of the following tasks:

I. Initially, knowledge about the information surrounding the Quranic text was collected from experts,
II. then, knowledge collection from literature survey (online and printed materials) was performed,
III. and finally, knowledge discovery and collection using text mining techniques and machine learning algorithms as described below was made:
   A. In order to determine missing information for the metadata of the Quran, categorization of the chapters and verses using Naïve Bayesian, Support Vector Machines, Distance-based, KNN classifiers, and Heuristics was carried out.

Text mining techniques and machine learning algorithms were used to compute the following metadata, which are needed for effectively learning and studying the Quran:

a. Places of revelation of chapters and verses.
b. Dates of revelation of chapters and verses.
c. Causes of revelation of chapters and verses.
d. Method of revelation of chapters and verses.
e. Topic of chapters and verses.
f. Location of chapters and verses.
g. Size of chapters and verses.
h. Summarization of chapters and verses.
i. Memorization of chapters and verses.

B. Collation system for analyzing and comparing Quranic chapters or verses. For ample information about textual documents collation systems, see [8, 7].

The computation of some metadata is described in previously written papers that, one could consult in the reference section. For instance, for dates of revelation [2, 3], places of revelation [1], causes of revelation [4], search result visualization [5], modeling and knowledge representation [6], visualization of textual documents collation process [7, 8]. I have not yet written a paper on determining locations, and sizes of the chapters and verses, because it is a trivial task. They are simply derived from the text itself using a decision tree classifier.
As for the topic, summarization, and memorization metadata, I am currently working on completing the task, and hopefully soon a paper will be published on this issue.

After having collected existing information about the Quranic text, and computed missing ones, offering the possibility to store,

maintain, and utilize this data was obviously the next step to be tackled.

### 6.3. Knowledge base management

This section is concerned with the creation of facilities for storing and maintaining the Quranic text with its metadata, which form the knowledge base of the system.
For this purpose several software tools were developed, in order to implement the usual knowledge management activities. Knowledge Management is defined as a collection of processes for creating, disseminating, utilizing, and preserving knowledge.

As shown in fig.1, the CMS is actually an integrated system comprising software tools, which are required for performing knowledge management activities. In the following lines, the association of each tool with one of the knowledge management activities is mentioned.
It should be mentioned that, including full description and screenshot for each tool would increase the size of the paper, which does not comply with the number of pages allowed for this workshop. For more information, please do consult the references.

a. **Knowledge Creation** involves discovering, capturing, and registering the knowledge into the system. Manual and automatic annotations (meta-data assignment) 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 the 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

d. **Knowledge Preservation** is achieved with the help of XML. The created knowledge base and the source documents are stored in XML format.

In other words the knowledge management system is the administrator area of the knowledge base.

## 7. Frame-based Representation of the Quran‘s Knowledge

A frame is a data structure that looks like modern classes, without methods in object oriented programming languages. It consists of multiple slots used to define various attributes of an object. Some slots hold static data, while others variable ones, as well as procedures. Multiple frames are linked together in a hierarchy so that children frames can inherit from parent ones.
The elements of the frames represent the enclosing tags of the metadata, whose data need to be collected or computed.

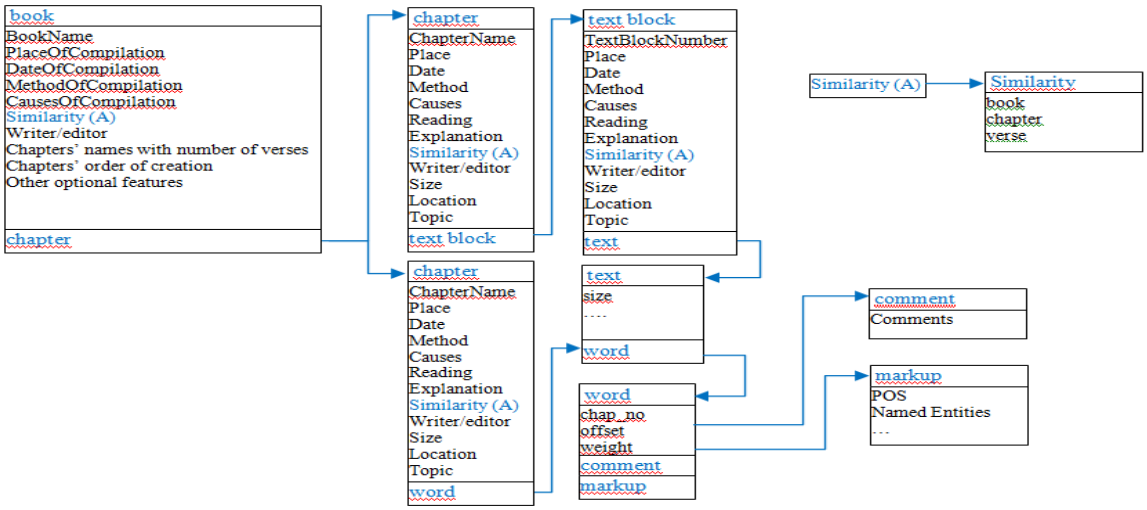


Fig.2 Frame-based representation of the Quran's knowledge

## 8. Instance of an XML Model of the Frame Representation

An example of an XML model of the frame-based representation is shown below:

```
<quran>
  <metadata>
        <PlaceOfCompilation>Medina</PlaceOfCompilation>
          <DateOfCompilation>653 AD</DateOfCompilation>
          <MethodOfCompilation>Manual</MethodOfCompilation>
          <CauseOfCompilation>Some muslims were not able to memorize the full Quran, many of the those
                who memorised died in some battles
          </CauseOfCompilation>
           <ChaptersOrderOfRevelation>96, 74, … 110</ChaptersOrderOfRevelation>
          <ChaptersOrderOfNumbering>1, 2, …114</ChaptersOrderOfNumbering>
           <Similarity/>
 </metadata>
```

```
<chapter n="1">
   <metadata>
        <ChapterName>Al-Fatiha</ChapterName>
        <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>
        <NumberOfVerses>7</NumberOfVerses >
        <Similarity/>
        <Size>Short</Size>
        <Location>Beginning</Location>
        <Topic>Prayer</Topic>
   </metadata>
   <verse n="2">
```

```
        <Text length="17"><word4>الْعَالَمِينَ</word4><word3>رَبِّ</word3><word2> لِلَّهِ
            </word2><word1>الْحَمْدُ </word1>
          </Text>
        <metadata>
          <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>
          <Similarity>
              <SimilarTo book="Quran" chapter="6" verse="45" />
              <SimilarTo book="Quran" chapter="10" verse="10" />
              <SimilarTo book="Quran" chapter="37" verse="182" />
              <SimilarTo book="Quran" chapter="39" verse="75" />
              <SimilarTo book="Quran" chapter="40" verse="65" />
          </Similarity>
          <Size>Short</Size>
          <Location>Beginning</Location>
          <Topic>Prayer</Topic>
          <Editor>The Writer or compiler of the verse (Uthman third Caliph)</Editor>
        <metadata>
    </verse>
  …
 </chapter>
…
</quran>
```

## 9. Utilization of the Knowledge Base

This section deals with the "where" and "how" to interact with the knowledge base from user perspective.
Presently there are two ways of interacting with the knowledge base:

1. Learning by:
   a. Browsing chapters
   b. Searching or querying textually and graphically
2. Consultation
   a. Dialog-based consultation, which is actually a mini-expert system

### 9.1 Some sample screenshots of the mashup website

#### I) Browsing and Searching the knowledge base

Fig.3 is the user interface for selecting the language, chapter to be learned.


Fig.3. Quranic chapter with translation

Clicking on a verse in fig.3 will display the verse with its translation in a new page, along with the links to its metadata. For instance by clicking "PlaceOfRevelation" link, googleMap displays a marker showing the city where the verse was revealed as shown in fig. 4.
Moreover, users could interactively search the knowledge base, by selecting the whole verse or part of it with the mouse. On releasing the mouse button, a window for setting type of returned result will appear as shown below.

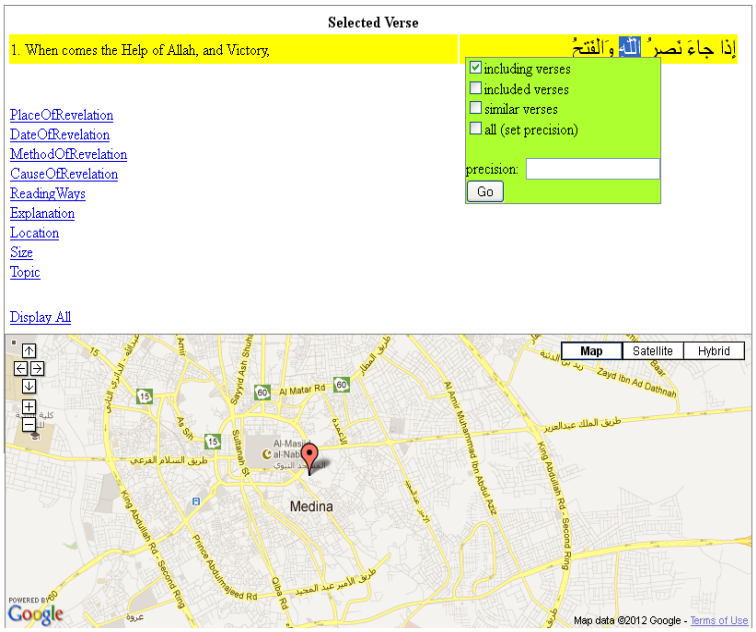



Fig.4. Interface for visualizing metadata and searching similar text in the Quran

#### II) Interactive visualization of search results

Visualization of search results using Interactive scatter plots and tables. Clustering search results by chapters, and making markers' sizes according to number of verses contained in respective clusters, offer the possibility to visually analyze the search results (e.g distinguishing Meccan and Medinan chapters).
For example the word ALLAH (in Arabic) is selected. After setting return type of the result, and clicking on the "Go" button, the output is displayed on a scatter plot as shown below:

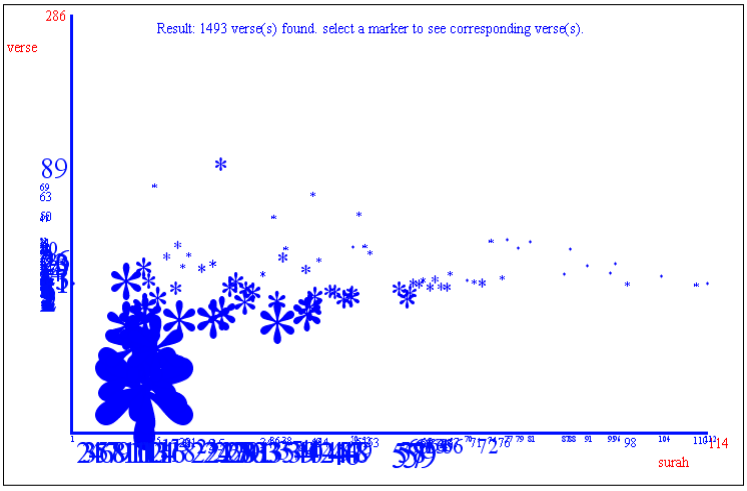



Fig.4. Plot of "ALLAH" search result

It is obvious that the axes as well as the markers are extremely overcrowded and congested, in such a way that readability at some points is practically impossible.
However the most interesting thing is that, the user is able to figure out the approximate location of the majority of the verses that contain the word ALLAH. In other words seeing that the markers are mainly concentrated on the first half part of the x-axis (surah axis) it is easily deducible that the word ALLAH has been primarily used in the Medinan chapters, that is between 622-632 AD.

**Remarks:**

Selecting with mouse by drawing a rectangle around a region of the plot will allow zooming in that region.
Following is an instance of the zooming effect.

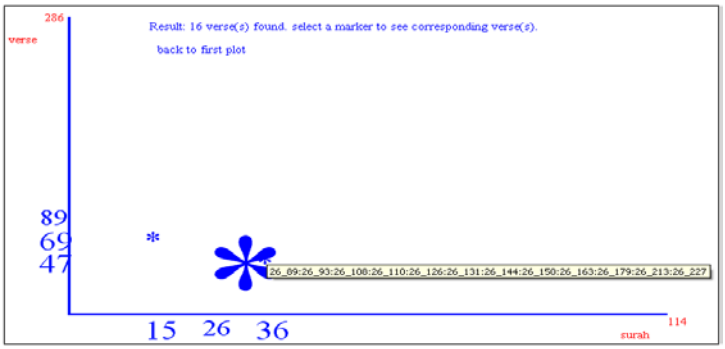
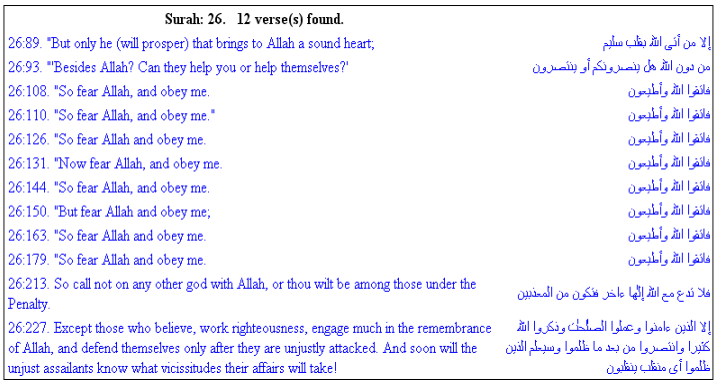


Fig.5. Plot of zoomed region

Clicking on a marker displays the verses with their translations as shown in fig.6.



| Surah: 26. 12 verse(s) found. | |
|---|---|
| 26:89. "But only he (will prosper) that brings to Allah a sound heart; | إلا من أتى الله بقلب سليم |
| 26:93. "'Besides Allah? Can they help you or help themselves?' | من دون الله هل ينصرونكم أو ينتصرون |
| 26:108. "So fear Allah, and obey me. | فاتقوا الله وأطيعون |
| 26:110. "So fear Allah, and obey me." | فاتقوا الله وأطيعون |
| 26:126. "So fear Allah and obey me. | فاتقوا الله وأطيعون |
| 26:131. "Now fear Allah, and obey me. | فاتقوا الله وأطيعون |
| 26:144. "So fear Allah, and obey me. | فاتقوا الله وأطيعون |
| 26:150. "But fear Allah and obey me; | فاتقوا الله وأطيعون |
| 26:163. "So fear Allah and obey me. | فاتقوا الله وأطيعون |
| 26:179. "So fear Allah and obey me. | فاتقوا الله وأطيعون |
| 26:213. So call not on any other god with Allah, or thou wilt be among those under the Penalty. | فلا تدع مع الله إلها آخر فتكون من المعذبين |
| 26:227. Except those who believe, work righteousness, engage much in the remembrance of Allah, and defend themselves only after they are unjustly attacked. And soon will the unjust assailants know what vicissitudes their affairs will take! | إلا الذين آمنوا وعملوا الصالحات وذكروا الله كثيرا وانتصروا من بعد ما ظلموا وسيعلم الذين ظلموا أى منقلب ينقلبون |

Fig.6. Retrieved verses with their translations

By clicking on a verse in fig. 6, users could display the chapter containing the retrieved verse along with its translation, and the selected verse would be highlighted as shown below:

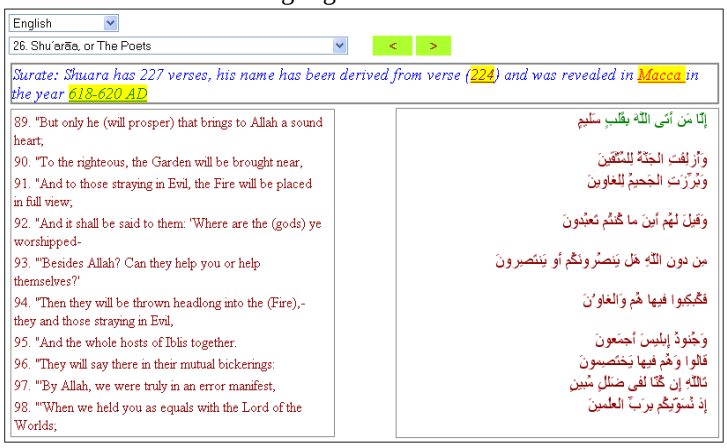


Fig.7. Chapter with retrieved verse highlighted along with translation

## 10. Implementation

The system was implemented in Java and PHP. The Content Management System was first developed in Java as eclipse RCP plugin, and is being converted to a web based one using PHP, Javascript, and HTML/CSS.

## 11. Conclusion

In this paper, the development of a knowledge-based learning system for the Quranic Text has been presented.
Modeling and representing religious knowledge using frames and XML representations have been shown.
The construction of the knowledge base consisting of several stages has been explained. Those stages comprise knowledge acquisition, registration, and maintenance. The knowledge to be acquired has been identified as a collection of metadata of the Quranic text. The most important metadata required for effectively analyzing, understanding, and memorizing Quranic chapters have been identified as the place of revelation, date of revelation, cause of revelation, method of revelation, topic of revelation, and location and size of the revelation in the Quran.
A link to a collation system that helps comparing textual documents has also been given.
Text mining and machine learning techniques used in the computation of missing information for the metadata have been mentioned, and papers that contained detailed information about these techniques have been mentioned.
Moreover some sample screenshots showing how the knowledge base could be utilized have also been presented.
The next steps for completing this research would be to find out whether people in general, and non-Arabic speaking people in particular interested in the Quranic text could effectively benefit from this system.
Another major interesting task to be undertaken could be the application of the system other religious texts, in order to find out whether it can effectively assist analysis, understanding, and retention of those texts.

## References


1. Mohamadou Nassourou, "Using Machine Learning Algorithms for Categorizing Quranic Chapters by Major Phases of Prophet Mohammad's Messengership" http://nbn-resolving.de/urn:nbn:de:bvb:20-opus-66862

2. Mohamadou Nassourou, "A Knowledge-based Hybrid Statistical Classifier for Reconstructing the Chronology of the Quran", http://nbn-resolving.de/urn:nbn:de:bvb:20-opus-54712

3. Mohamadou Nassourou, "Philosophical and Computational Approaches for Estimating and Visualizing Months of Revelations of Quranic Chapters", http://nbn-resolving.de/urn:nbn:de:bvb:20-opus-65784

4. Mohamadou Nassourou, "Computing Causes of Revelation of Quranic Verses Using Machine Learning Techniques", http://nbn-resolving.de/urn:nbn:de:bvb:20-opus-66083

5. 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

6. Mohamadou Nassourou, "Design and Implementation of a Model-driven XML-based Integrated System Architecture for Assisting Analysis, Understanding, and Retention of Religious Texts:The Case of The Quran", http://nbn-resolving.de/urn:nbn:de:bvb:20-opus-65737

7. Mohamadou Nassourou, "Computer-based Textual Documents Collation System for Reconstructing the Original Text from Automatically Identified Base Text and Ranked Witnesses", University of Würzburg, http://nbn-resolving.de/urn:nbn:de:bvb:20-opus-65749

8. Mohamadou Nassourou, "Design and Implementation of Textual Documents Collation Systems", http://nbn-resolving.de/urn:nbn:de:bvb:20-opus-56601

9. Mohamadou Nassourou,"Markup Overlap: Improving Fragmentation Method", http://nbn-resolving.de/urn:nbn:de:bvb:20-opus-49084

10. Mohamadou Nassourou, "Understanding the Vex Rendering Engine", http://nbn-resolving.de/urn:nbn:de:bvb:20-opus-51333

11. Mohamadou Nassourou, "Reference Architecture, Design of Cascading Style Sheets Processing Model" http://nbn-resolving.de/urn:nbn:de:bvb:20-opus-51328

12. Mohamadou Nassourou, "Doing Webservices Composition by Content-based Mashup: Example of aWeb-based Simulator for Itinerary Planning" http://nbn-resolving.de/urn:nbn:de:bvb:20-opus-50036