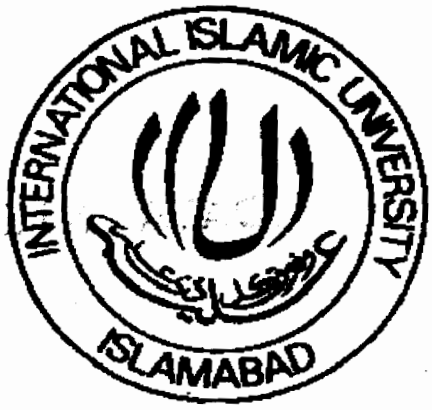
The Quranic Database for Semantic Search



Developed By

Muhammad Shoaib
(237k-FAS/MSCS/S05)

Supervised By

Prof. Dr. Sikandar Hayat Khiyal
Chairperson Department of Software Engineering
Fatima Jinnah Women University,Rawalpindi

Co-Supervisor

Muhammad Imran Saeed
Assistant Professor, IIUI

Dept. of Computer Science
Faculty of Basic and Applied Sciences
International Islamic University Islamabad
2008

بسم الله الرحمن الرحيم

A Thesis Submitted to the

Department of Computer Science

International Islamic University Islamabad

as a partial fulfillment of requirements for the award of

the degree of

MS in Computer Science

Dated:..31/01/09.....

Final Approval

It is certified that we have examined the thesis titled "Quranic Database for Semantic Search" submitted by Muhammad Shoaib, Registration No. 237k-FAS/MSCS/S05, and found as per standard. In our judgment, this research project is sufficient to warrant it as acceptance by the International Islamic University, Islamabad for the award of MS Degree in Computer Science.

Committee

External Examiner

Dr. Nazir Ahmad Sangi
Chairman, Department of Computer Science
Allama Iqbal Open University, Islamabad

Internal Examiner

Muhammad Nadeem
Assistant Professor
Department of Computre Science
International Islamic University, Islamabad

Supervisor

Dr. Sikander Hayat Khiyal
Chairperson Department of Software Engineering
Fatima Jinnah Women University, Rawalpindi

Co-Supervisor

Muhammad Imran Saeed
Assistant Professor
Department of Computre Science
International Islamic University, Islamabad

In the Name of Allah

Who selected Muhammad as His last Prophet

Declaration

I hereby declare that this work, neither as a whole nor a part of it has been copied out from any source. It is further declared that I have developed the model, the software on the base of proposed model and the results with my personal efforts; and under the sincere guidance of Prof. Dr. Sikander Hayat Khiyal and Muhammad Imran Saeed. If any part of this project is proved to be copied from any source or found to be reproduction of some other project, I shall stand by the consequences. No portion of the work presented in this dissertation has been submitted in support of any application for any other degree or qualification of this or any other university or institute of learning.

Muhammad Shoaib
(237k-FAS/MSCS/S05)

Acknowledgement

All praise to Almighty Allah Who has guided me in undertaking the work on "Quranic Database for Semantic Search" and has helped me through each step when there was no hope for pulling through.

I am very much thankful to my supervisor Dr. sikander Hayat Khiyal and co-supervisor Muhammad Imran Saeed for their kind help and supervision. I shall remember their support for building my research capacity for research methodology and always guiding me to next bold step.

I would like to pay thanks to my class mates and faculty members of the university for their help and support, with special thanks to Mr Aneel Rahim and Muhammad Nadeem Yasin for helping me in coding and Zeeshan Shafi Khan for documentation.

Finally I owe a lot to my beloved parents for their love, guidance and moral and financial support.

Muhammad Shoaib

(237k-FAS/MSCS/S05)

Project in Brief

Project Title:	Quranic Database for Semantic Search
Undertaken By:	Muhammad Shoaib
Supervised By:	Dr. Sikander Hayat Khiyal
Start Date:	Spring 2008
Completion Date	January 2009
Tools & Technologies	Visual Bsic.Net, SQL Server
Documentation Tools	MS Word, MS Excel, E-Draw
Operating System	MS Windows Vista
System Used	P-IV, 2.0 GHz, Dual Core

Abstract

A lot of efforts have been made in developing user friendly Quranic softwares to facilitate both: the layman and the religious scholars. Most of these softwares exploit query word matching technique for searching the text. The key word matching technique is incapable of retrieving all relevant verses and some irrelevant verses may be retrieved.

Ineffectiveness of keyword based search generates the need for semantic search in the Holy Quran. The Quranic text, because of its unique style and allegorical nature, needs special attention of the NLP and Text Mining experts. In this thesis we introduce a model that is capable of performing semantic search. The model combines the WordNet Synset's relationships to the relational model. Selecting a word as the element of synset is very sensitive job. We cannot choose the synonyms of a word through language dictionary as Quranic terminology has special meanings.

We implemented our proposed model to Surah Al-Baqrah (the longest chapter in the Holy Quran). We tested the model against five different query words. A number of experiments were performed to check out the accuracy of our model. The results show that our proposed model is far better than simple query word based searching technique.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

CHAPTER 1

INTRODUCTION

Do they not consider the Quran (with care)? Had it been from other than Allah, they would surely have found therein much discrepancy.

(Al-Quran)

1.1 Motivation

We are living in oceans of information. Every one from scholar to lay man wants authenticated, reliable and relevant results of his/her query. The Holy Quran is an ocean of information. It discusses all fields of life and provides basics about all areas of knowledge. It is the sacred book of Muslims and millions of people recite it regularly. Millions of people consult this holy book to solve their problems and, to get Islamic view point about a particular topic. Religious scholars, teachers and students struggle hard to collect similar verses that describe a particular topic. The religious personalities relate a verse to a topic on the base of interpretation available to them. How can an acquaintance of the Holy Quran perform subjective search? It is very difficult for him/her. Even it is not an easy job for a scholar. He/she has to spend a lot of time to get all the verses that are related to a particular topic.

What for IT is? We realize the problem of the users of the Holy Quran and try to facilitate them in searching the Holy Quran for specific topic. Quite a large number of number of today's Quranic software provide searching facility but the users are not satisfied by the results as they provide simple search. These software lack the idea of semantic serch. The field of semantic web challenges this type of problem on World Wide Web. The

researchers have developed ontology languages to solve the problem of irrelevant out put. Some semantic search engines [7] have also been developed but users are not fully satisfied with their results. This field is not mature yet.

The WordNets are the electronic dictionaries that handle the relationships among elements of synset of a concept. But these relationships are concerned with the natural language.

By applying the available techniques we can solve the problem of legacy searching techniques. If it is done, millions of people will get benefit from semantic search.

1.2 Introduction

In this section we give brief discussion about the Holy Quran, meanings of semantic search and semantic search in Quran.

1.2.1 Overview of the Holy Quran

There are basically two types of holy books of Muslims.

- The Hadith
- The Quran

The books concerning the sayings and deeds of the Holy Prophet Muhammad (SAW) are categorized as the Hadith books. For example; Saheh Bukhari, Saheh Muslim, etc.

The Quran (also spelt as Qur'an, Koran, Coran) is the holy book of Muslims. It was revealed from Allah to the Holy Prophet Muhammad (SAW) in Arabic language in almost twenty two years. It consists of thirty Parts (Parahs) and one hundred and fourteen chapters (Surahs). The Parahs are almost of equal length but the chapters consists of an arbitrary number of verses (Ayahs)---the smallest chapter "Surah Al-kothar" contains 03 verses whereas the longest chapter "Surah Al-Baqrah" 286 verses. Among the revealed books the Holy Quran is the only book present in its revealed language and original text.

Quran has its own style of describing the topics. At some places, some topics are explicitly mentioned while some others are meant implicitly. This implicit nature of the Holy Quran is a great challenge for the NLP and AI experts. Here the context very often helps the scholars to find the implicit meanings but some times it does not provide any clue.

The allegorical nature of the Holy Quran is also a challenge not only for NLP experts but also for the Quranic scholars. Different scholars explain allegorical terms according to their own perception/sect/knowledge etc.

The Holy Quran has unique style of explaining different topics. Normally a topic is discussed at different places. For example, the topic of Hazrat Moosa (AS) is discussed in a large number of chapters; the Oneness of Allah has been discussed through out the Holy Quran. It is not necessary that the consecutive verses must belong to the same topic. A topic may or may not be discussesd in consecutive verses. Some consecutive verses or even a single verse may contain many topics. For example, the second verse of Surah Al-Baqrah, consisting of only seven words including prepositions, mentions three topics. This is a little verse and discusses three topics. While lengthy verses may explain dozens of topics. Normally a chapter does not contain only one topic and the same is true for a lot of verses.

The name of a chapter does not mean that all the text of this chapter belongs to this topic. Name of a Surah normally does not help us in perceiving the semantics of the Quranic text of that Surah. But in some cases it does help.

1.2.2 Subject Matter of the Holy Quran

The Holy Quran throws light upon almost all fields of life. Some topics are repeated time and again and some are described just once; some explicitly and some implicitly. The frequency of repetition of a topic is actually a gesture to human to show its importance. Some topics are discussed in detail and some are slightly touched. The list of the topics upon which the Holy Quran lays emphasize is very long. Oneness of Allah, Prophethood of Muhammad (SAW) and other prophets, The purpose of creation of human being and

his success and failure; and the life after death are major topics of the Holy Quran.

1.2.3 Users' Preferences while Querying the Quranic Text

What type of information should be extracted from the Holy Quran? The complete answer of this question is perhaps difficult to explain. Most of the users want to perform the subject based search in the Holy Quran against a particular topic. A user wants either to know the basic concept or the detail of a topic. The Quranic text is written neither in story-writing or essay-writing style nor in question-answer style. It has its own unique style. A chapter may contain hundreds of; and a verse dozens of topics implicitly or explicitly. A system that is capable of answering the users' queries needs special attention, cooperation and team work of the experts from different areas.

We know that the automatic systems of IR or IE can not provide accuracy to a satisfactory level. These systems show good results for well structured data source, like relational database. On the contrary, the precision is very low for unstructured data, the text data. Many automatic annotators have been developed for domain specific documents. The ontology mapping for such annotated documents show good results only if the text is simple and contains limited domains, normally one specific domain. An ontology designed for a domain will not work for any other domain. The greatest issue while developing any automatic system for the semantic search in the Holy Quran is its unique style.

Much of the work to digitalize the Holy Quran has already been done by different Muslims and non Muslims as well. Some Muslim governments and organizations have expended a lot of money to develop more and more useful softwares of the Holy Quran. Almost all of their databases and software are freely available on the Internet. Many commercial organizations have also developed many quranic software and maintained databases containing Islamic literature. Most of these Quranic software/databases provide the facility to read different translations in different languages. Beside this they also provide the facility to listen the Holy Quran in different reciter's (Qaries) voices. Some of them provide query word base search facility. Query word based search is a search in which the key word is matched to the text of the Quran, and all the verses that contain

that particular query word are retrieved. There are two major problems with key-word based searching technique:

- First, many relevant verses are not retrieved as the particular query word is not present in them. And more curious is that some irrelevant verses may be retrieved as the query word is present in them.

- Second, the sequence of the retrieved verses does not appeal to the user. A user wants the most relevant verses to be shown first and less relevant at the end. The text of the Holy Quran is not written topic-wise. Different verses of a particular topic are scattered at different places within a chapter (Surah) and also through out the Holy Quran.

There is a need of a model that must solve both of these problems. The user (scholar) must have option to attach any topic to a verse or detach any topic from a verse. This provision must be given because different scholars have different view points about some verses. The user should have option to change the order of verses to be retrieved against a query word.

1.2.4 Semantic Search in Holy Quran

The term semantics means meanings. We define the semantic search as:

"In semantic search the term (the query word) is assigned a single meaning out all other meanings present in the dictionary and then searching is performed against all possible synonyms of that meaning within a particular domain."

As we have discussed earlier that many softwares use keyword matching technique to retrieve the query results. In many cases all the retrieved verses are not relevant to the given query word. Further more, some relevant verses that should be retrieved against the query-word are not retrieved as the particular query-word is not present in those verses. There are hundreds of examples of this kind in the Holy Quran. This requires that some techniques must be used to retrieve the required/related verses from the Holy Quran

either the query-word is present in them or not. Similarly irrelevant verses must not be retrieved though the query-word is present in them. Above cases motivate us to use semantic search technique in the Holy Quran. The concept of the semantic search is explained in figure 1.1. "Signs" is a polysemy word in Quran. In semantic search only one meaning will be taken. For example, we take only "Quran" as intended meaning and discard all others.

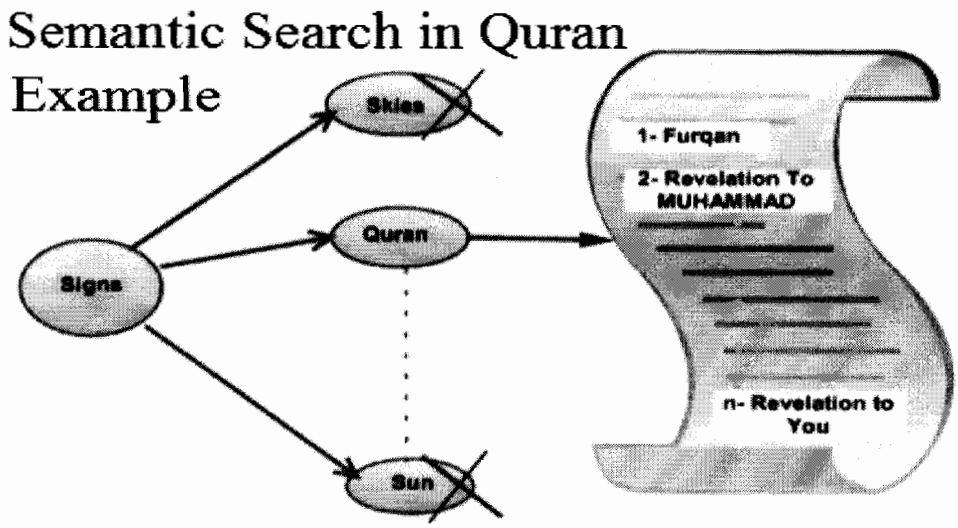


Fig 1.1: Example of semantic search in Quran.

Then all the synonyms of Quran used in Quran will be collected and the searching will be performed against all of them.

1.3 Problem Area

Quranic text is written in Arabic, a natural language. As other texts of natural languages are problematic for computer to process the text data, the case of Quranic text is also the same. Quranic text is unique in its style. It needs special attention of the scholars from diferrent fields. The combined efforts of Quranic scholars and the specialists of Computational Linguistics, Artificial Intelligence, Text Mining, Semantic Web, WordNet Ontologies and Machine Leaning are needed. The Quranic text is very complex and sensitive as it is the miracle from Almighty Allah. The Information retrieval and

information extraction are very difficult to handle. Query word matching technique does not satisfy the researcher of the Holy Quran. A large number of relevant verses may not be retrieved and some irrelevant verses may be retrieved. Some techniques must be adopted to overcome this problem.

Semantic search for the Holy Quran is becoming more vital day by day. Especially with the introduction of semantic web the semantic search in the Holy Quran has become necessary.

"Semantic search in Holy Quran means that the verses that are relevant to a certain topic should be retrieved on querying the Quranic text by putting a query word or any synonym of it either the query word is present in those verses or not."

Some softwares e.g.[10] provide topic based search (ignoring the keyword) and show results in a suitable sequence. But the problem with this type of software is that they are not more than a book in electronic form. In other words, the verses are pre-clustered against a topic. The second problem with them is all the scholars/sects are not satisfied with the query results because of differences in their ideologies. Third problem with them is that a user cannot attach or detach any verse to a topic.

The query word searching technique does not satisfy the researchers of the Holy Quran. It can not satisfy even the lay man who wants to know the Quranic view point about a topic. The lay man may be misguided by reading the irrelevant verses or by missing some relevant verses. At this point the situation becomes very serious.

1.4 Objectives

The major objectives of our research are:

- To facilitate the programmers who want to develop Quranic softwares.
- To provide semantic search facility to users who are extreme subject conscious.
- To facilitate the religious scholars to prepare their topic based lectures.

- To motivate IT researchers to develop semantic/intelligent search engine for the Holy Quran.

1.5 Thesis Outline

The remainder of this thesis is organized as follows. In chapter number 2 we highlight the previous work in this domain. In chapter 3 we review the requirement analysis of our proposed model. Chapter 4 discusses design model of the proposed approach. Chapter 5 describes the implementation issues. Chapter 6 shows the results taken from the software designed on the base of our proposed model. The last chapter describes the conclusion and future work in this area.

CHAPTER 2

LITERATURE SURVEY

CHAPTER 2

LITERATURE SURVEY

This is the Book; in it is the guidance sure, without doubt, to those who fear Allah.

(Al-Quran)

Many people have worked on the Holy Quran in different areas. A lot of work has been done in the field of storing the Holy Quran in different format by using different techniques. Some of them are given below[1].

- Relations of the database.
- Inverted File Structure.
- XML format.
- Unicode

Storing any text and its translations in different attributes of the database relation(s) is very common and popular technique [26]. It is easy to develop and handle the database by providing different constraints and making changes transparent to the users. The Relational Database Management Systems are well matured today- this is another reason of the popularity of using the relational model. We will also prefer to use relational model for the digital storage of the Holy Quran.

The Inverted File Structure technique for storing the Holy Quran is also used. The Malay Translated, Al-Quran texts are stored in an inverted file structure. Malay translated Al-Quran documents are based on user natural query words [2].

1. We are not going to the detail of all techniques for storing Quranic text, as it is not our research topic.

XML provides format independent platform to the developers of web applications. It also provides its own semantics for the text. The user can query in any format and get results in the same format in which s/he puts the query [5].

Quranic text is also stored in Unicode [27]. Here we are not concerned with the storage techniques or storage media.

The Arabic text database management system was designed to handle the Arabic databases. According to the writer of "Bayan: Arabic text database management system", the English language based database management systems are not so much useful and efficient for Arabic text [1].

2.1 Related Research

So for as our topic "semantic search in Quran" is concerned we could not find any relevant material. Much work has been done in the fields of semantic web, Information retrival and text mining. Our work is actually combination of many fields, like, NLP, Machine Learning, Text Mining, Semantic Web, Information Retrieval and WordNet Ontologies. We searched the world-wide web and found some papers which show some work on Quran. Following is the survey of the literature we did in our research.

2.1.1 XML Semantics Rules for Holy Books

Y. Kotb, K. Gondow, and T. Katayama in their paper "The SLXS Specification Language for Describing Consistency of XML Documents" use a new specification language for XML semantics. It is employed to specify semantics rules. The writers applied XML semantics approach to specify the consistency of the Holy books that are published in XML format. The goal of the study was to understand the application of XML semantics in a problem domain that is appropriated for the method, and to determine some of the strengths of XML semantics in this domain. These semantic rules associated with the Holy books, which are written in XML format, to check the variant semantic consistency problems that exist inside these Holy books [12].

2.1.2 XML Semantics Checker Model for Quran

Y. Koth, K. Gondow, and T. Katayama demonstrate the significance of XML semantics checker approach in their paper titled as "The XML Semantics Checker Model". This novel approach was used to check the semantic consistencies of the Holy Quran, mentioned by Jon Bosak in Religion 2.00 website. Religion 2.00 is a group of four religious works marked up for electronic publication from publicly available sources. This checker model was successfully applied to check that the number of verses in each chapter was correctly written in the Quran XML format document. It was also checked that the Quran XML document contained exactly the same number of chapters as in the real Quran holy book [3].

2.1.3 Mophological Analysis of the Holy Quran

In "Morphological Analysis of the Quran" [30] the writers present a computational system for morphological analysis and annotation of the Holy Quran. By developing a tool on the base of finite state automata they apply the rules to the Quranic text and get morphological analysis of Quranic words. This system is currently being used by researchers and teachers for research and teaching purposes.

2.1.4 Keyphrase Extraction for Islamic Knowledge Ontology

Saidah Saad and her co writers of the paper "Keyphrase Extraction for Islamic Knowledge Ontology" [31] describe ontological work for extraction key word and keyphrase candidate for developing ontology of Islamic literature. Keyphrases provide semantic metadata. Collection extraction also known as automatic keyword extraction is to select key words from a particular text. This paper also describes the algorithm for automatic extraction of the key words. Lexico-syntactic and statistical methods have been used for this purpose.

2.1.5 Build Islamic Ontology based on Ontology Learning

Saidah saad and Naomie Salim have presented a general and skelatel methodology and

life cycle for building the ontology for Islamic literature in their paper titled as "Build Islamic Ontology based on Ontology Learning" [32]. Their research was inspired by software engineering V-process model. They also applied their technique on English text for mining ontologies from natural language.

2.1.6 OSIS

The OSIS (Open Scripture Information Standard) model based on XML was specially designed for the Bible. It was developed by the Bible Technologies Group. It defines tags for marking up scripture like that of Bibles', and other related literature. Its current version is 2.1. It is the first project to come out of the Bible Technologies Group [5].

OSIS provides initiative to identify and develop the organizational process, and technical basis that produce markup standards for the Bible and relevant materials. Its goal is to provide open standards for publishers, software manufacturers, Bible Societies, scholars, and a reader of the Bible. OSIS also provides the facility for easy storage and retrieval of text. It also handles indexing and cross-referencing of raw text. The out put can be seen in many formats [5].

2.1.6.1 Why was XML chosen to be the foundation of the OSIS standard?

XML has several important aspects.

1. It facilitates to describe the structure of documents. A file can directly feed Web delivery to be produced in multiple formats (Braille, print, audio, large-print, e-books, and so on). It can alos be used for non-reading formats like indexing and linguistic analysis.

2. XML's predecessor SGML [5] was suitable for very large scale projects like aircraft manuals, policy documents and collections of literary and scholarly texts. It was good for large-scale projects but poor for smaller ones. XML is a tiny subset of SGML. It is easier to learn, use, and implement. XML is very suitable for OSIS because its design is ideal for long term archival and access.

3. It does not depend [5] on any particular software.

2.1.6.2 Is OSIS Model Applicable to the Quran?

OSIS is perfect for storing the Quran and all translations in. It was checked out a while back [6]. It is originally developed for the Bible but it can be used for the Quran too [6].

2.2 Present Quranic Softwares

There are a large number of Quranic databases in digital form. One can also get variety of Quranic softwares present in the market as well as most of them are available on internet. A large number of these digital Quranic stores are in the form of audio [9] and video files. As for video is concerned there is only the Arabic text with a number of translations in different natural languages [8].

Some of the Quranic Softwares facilitate the reader to query the Arabic text for exact match on the basis of a single word/term or part of the verse or even the whole verse. These softwares find the exact matches and display the results. Normally query is put in Arabic or, in some softwares it is possible to query the database using English alphabets for Arabic terms (e.g. Muttaqeen for "متّقين", it is an Arabic term written in English alphabets meaning, the pious). To query in other natural languages is also possible in some softwares.

We found a software that is a product of the "Harf Company" [10]. It is the best software for Quran, so for as we know. In this software subjective search facility is provided. If someone wants to search a topic or even subtopic in the Quran, he/she can do so by just clicking the name of the topic/subtopic and the relevant verses are shown. The main topics are organized alphabetically and the subtopics are also organized under the main topics in the same way. This software also provides the exact match search for words/terms, part of the verse and even some consecutive verses. This is very good software as we have mentioned above but it also has some deficiencies:

- The whole software is in Arabic language, thus only those can use it who are well

familiar with Arabic language. For Arabic speakers it is very good.

- Technically speaking, this software is bounded to predefined topics and subtopics.
- Further more, the verses related to any topic/subtopic are predefined. In other words, the data (verse/s) to be retrieved against any topic are pre-clustered.
- The user can not change anything. In other words, the user can not use his opinion for any verse.
- The user can not re-arrange the sequence of retrieved verses.

We found another software [11] similar to the software discussed above. The only difference is that this is in English language.

A brief introduction of some softwares is given below:

"Freedownloadmanager" website provides a Quranic software that aims to help professional Muslim users such as religious researchers and students. It is also helps lawyers, judges, book authors, and lecturers, to work with Quran. This software contains hierarchal view of all verses in Quran, either sorted by Surah, or by parts [21].

Multi-Language Qur'an Software provides Arabic/English Qur'anic transliterations. It is equipped with Quranic commentary, index, and glossary of more than 500 words. It also provides full search on the key word bases. It supports plug-in Qur'anic translations[19].

Bro. Jamal Al-Nasir and his team developed Holy Quran Recitation Software. The team designed Holy Quran Recitation reference tool for anyone who wishes to learn more about the diverse content of the Holy Quran [20].

High quality Qur'an website provides more than 30 translations of the Qur'an, and more than 10 recitations. It also provides root word search. It takes the query word as input and after processing the morphological analysis of the query word it tells the root word of the query word as output [28].

The globalquran website is a professional resource of the Quranic literature. It provides

the key word base searchable interface, indexed by Surah number [13].

Quran browser provides recitations by different well known reciters (Qaries) and a number of translations in different languages[14].

The "understandquran" website provides Practical resources for understanding the Quran. It also includes an effective course for non-Arab to comprehend Arabic abilities while listening to and reading the Arabic text. It also provides introduction and guide to the Quran and dictionaries for children [15].

The "Quran" website facilitates the users to browse Quran and search with translation and Tafseer (interpretation) [16].

"Good" website uses same recitation and display engine as qurany.net's IslamKit [17].

The "quransource" website provides audio and video Islamic resources. It provides original search functionality [18].

"QuraanicLessons" is a Quran Learning Software for English speaking Muslims. It is an easy way to learn the Quranic language for those who can read but cannot understand Arabic. It has simple lessons to build Arabic vocabulary for learners [22].

The Holy Quran Malayalam English translation (V 1.0) can be used in two different languages, Malayalam and English. Default interface language is English. User can easily switch between the languages [23].

Multilingual Qur'an Software offered by "topshareware" website provides Arabic/English Qur'anic Commentary. Different translations in French, German, Spanish, Urdu, Malay, Indonesian, Japanese, Tamil, Hausa, Turkish and Indonesian are also provided by this software. It is equipped with query word base searching facility [24].

Although there are a large number of web sites containing Quranic softwares and Quranic literature yet we could not find much literature about the searching techniques. By going through different software we perceive that most of them use query word base searching

technique. We could not find even a single thesis or paper about semantic search in Quran.

2.3 Summary

What ever the work have been done for the Holy Quran or other Islamic literature is multi-dimensional. How to perform semantic or intelligent search in Holy Quran is not taken as research topic. We, to our best of knowledge, are the pioneer in this field to perform the semantic search in the Holy Quran. Work on morphological or syntactic analysis of the Holy Quran has been done but no one, to our best knowledge, tried for semantic search.

CHAPTER No. 3

PROBLEM ANALYSIS

CHAPTER No. 3

PROBLEM ANALYSIS

And if you are in doubt as to what We have revealed from time to time to our servant, then produce a Sura like thereunto; And call your witnesses or helpers (if there are any) besides Allah, if your (doubts) are true. (Al-Quran)

While going to apply semantic search techniques to the Holy Quran we face many problems. In this chapter we analyze the problems. We face some specific problems in the Holy Quran due to unique style of Quranic text.

3.1 Holy Quran's Unique Style

Quran has its own style of describing the topics. At some places, some topics are explicitly mentioned while some others are meant implicitly. This implicit nature of the Holy Quran is a great challenge for the NLP and AI experts. Here the context very often helps the scholars to find the implicit meanings but some times it does not provide any clue.

Some times there is a difference between apparent and actual meanings. For example, in Sura Al-Mo'menoon, Allah says, "They avoid vain talk". Here the word "**لغو**" has been used. Its apparent meanings are vain talk, but according to some interpreters, it is meant for music.

3.1.1 The allegorical nature of the Holy Quran

The allegorical nature of the Holy Quran is also a challenge not only for NLP experts but also for the Quranic scholars. Different scholars explain allegorical terms according to

their own perception/sect/knowledge etc. The Holy Quran says about allegorical verses:

> *"He it is Who has sent down to thee the Book: In it are verses basic or fundamental (of established meaning); they are the foundation of the Book: others are allegorical. But those in whose hearts is perversity follow the part thereof that is allegorical, seeking discord, and searching for its hidden meanings, but no one knows its hidden meanings except Allah. And those who are firmly grounded in knowledge say: "We believe in the Book; the whole of it is from our Lord:" and none will grasp the Message except men of understanding."*

Allah has used the word Al-jannah for paradise. Jannah means the gardens. Is the paradise exactly same as our worldly gardens? No. The word Al-jannah is used for paradise because human can not understand the things which they have not seen. Actually the paradise does not resemble any of worldly things. So if any strange words were used for paradise the human could not have any sketch/idea about the paradise.

3.1.2 Scattered Discussion of Topics

The Holy Quran has unique style of explaining different topics. Normally a topic is discussed at different places. For example, the topic of Hazrat Moosa (AS) is discussed in a large number of chapters; the Oneness of Allah has been discussed through out the Holy Quran. It is not necessary that the consecutive verses must belong to the same topic. A topic may or may not be discussed in consecutive verses. Some consecutive verses or even a single verse may contain many topics. For example, the second verse of Surah Al-Baqrah, consisting of only seven words in Arabic text including prepositions, mentions three topics.

"This is the Book; in it is guidance sure, without doubt, to those who fear God" 2: 2

This is a little verse and discusses three topics---The Book (Holy Quran), guidance and the pious. Lengthy verses may contain dozens of topics. Normally a chapter does not contain only one topic and the same is true for a lot of verses.

3.1.3 The Name of a Chapter and their Text

The name of a chapter does not mean that all the text of this chapter belongs to this topic. Keeping in view this property we can divide the chapters of the Holy Quran in four categories.

- There are some chapters (for example Surah Ikhlas) whose whole text is related to their names. These chapters are small in number and size.

- There are some other chapters whose major topics are related to their names. They also contain many other topics. They are also few in numbers. Surah Yousaf is an example of this type. Its major topic is "Story of Yousaf (AS)". From start to end this story has been discussed in fantastic style. Beside this story many other topics are discussed in both ways: explicitly and implicitly.

- Most of the chapters belong to this category. In this category, the name of Surah is relevant to its some verses or words (may be only one word) while other verses are meant for other topics. For example, Surah Al-Baqrah---The Cow (containing 286 verses) discusses the topic of "the Cow---Bani Israel's Cow" just in few verses; and this Surah discusses hundreds of other topics.

- For some cases the starting word of a chapter is the name of the Surah. For example, Surah Yaseen, Surah Haameem.

From above discussion we conclude that the name of a Surh normally does not help us in perceiving the semantics of the Quranic text. But in some cases it does help.

3.2 Information Retrieval or Extraction from the Holy Quran

What type of information should be extracted from the Holy Quran? The complete answer of this question is perhaps difficult to explain. Most of the users want to perform the subject based search in the Holy Quran for a particular topic. A user wants either to know the basic concept or the detail of a topic. He/she may have such type of questions in

his/her mind about a query-word:

- What does it mean?
- What does it mean in Islamic literature, especially in Quran?
- What are different commands of Allah about it?
- What is its importance in Islam?
- What are its synonyms and antonyms?
- What is the procedure to perform it (if the query-word is an action)?
- What are its prerequisites?
- What is its benefit/reward or loss/punishment in this world or in the world Hereafter?
- Is it Halal or haram?
- Which verses (strongly or weakly) belong to this topic?

The last question is comparatively easy to handle and it is the subject matter of IR. While other questions belong to IE and are difficult to answer from the Quranic text. These questions need NLP and AI techniques and WordNet ontologies. The Quranic text is written neither in story-writing or essay-writing style nor in question-answer style. It has its own unique style. A chapter may contain hundreds of; and a verse dozens of topics implicitly or explicitly. A system that is capable of answering the questions like pointed out above needs special attention, cooperation and team work of the experts from different areas.

We know that the automatic systems of IR or IE can not provide accuracy to a satisfactory level. These systems show good results for structured data source, like data in relations of the relational database. On the contrary, the precision is very low for text data. The automatic annotators show good results for text data to be mapped on the ontology designed for specific domains only if the text is simple and contains limited domains. An ontology designed for a domain will not work for any other domain. The greatest issue while developing any automatic system for the semantic search in the Holy Quran is its unique style. Following are some questions in our mind in this context:

- How the system will perceive different meanings of semantically ambiguous terms?
- How, in case of polysemy, the system will guess the intention of the user for a particular meaning of the query word?
- What are different synonyms of the of the query word?
- Does the user needs the close relevant terms and their related verses or not?
- Should the antonyms be considered while searching the relevant verses?

Much of the work to digitalize the Holy Quran has already been done by different Muslims and non Muslims as well. Some Muslim governments and organizations have expended a lot of money to develop more and more useful softwares of the Holy Quran. Almost all of their databases and software are freely available on the internet. Many commercial organizations have also developed many quranic software and maintained databases containing Islamic literature. Most of these Quranic software/databases provide the facility to see different translations in different languages. Beside this they also provide the facility to listen the Holy Quran in different reciter's (Qaries) voices. Some of them provide query word base search facility. Query word based search is a search in which the key word is matched to the text of the Quran, and all the verses that contain that particular query word are retrieved. There are two major problems with key-word based searching technique:

- First, many relevant verses are not retrieved as the particular query word is not present in them. And more curious is that some irrelevant verses may be retrieved as the query word is present in them.

- Second, the sequence of the retrieved verses does not appeal the user. A user wants the most relevant verses to be shown first at the end. The text of the Holy Quran is not written topic-wise. Different verses of a particular topic are scattered at different places within a chapter (Surah) and also through out the Holy Quran.

There is a need of a model that must solve both of these problems. The user (scholar)

must have option to attach any topic to a verse or detach any topic from a verse. This provision must be given because different scholars have different view points about some verses. The user should have option to change the order of verses to be retrieved against a query word.

3.3 Semantic Search in Quran:

As we have discussed earlier that many softwares use keyword matching technique to retrieve the query results. In many cases all the retrieved verses are not relevant to the given query word. Further more, some relevant verses that should be retrieved against the query-word are not retrieved as the particular query-word is not present in those verses.

Examples:

- Verse numbers 3-6 of surah Al-Baqra are concerned with "Muttaqeen-متّقين", but this word is not present in all these three verses. This word is present at the end verse # 2. So we can say that the context tells us that these verses belong to "Muttaqeen"

- Verse numbers 8-18 are concerned with hypocrites. Neither the word "hypocrites" nor any synonym of it has been used in these verses. This word is also not present in the context of these verses.

- The word "Muttaqeen-متّقين" is present in verse number 2 of surah Al-Baqra but this verse tells very little about "Muttaqeen-متّقين".

- The verse # 71 of Surah Al-Baqarah uses a word "Harth-حرث" meaning "tilled land". This verse does not tell any thing about tilled land or any role of it. This word has also been used in Surah Aal-e-'Imran in verse # 14. Although, this verse also does not tell anything about tilled land yet it shows the influence of human being towards the tilled land. By using the "tilled land" as query word, the verse number 71 of Al-Baqarah should not be retrieved; but in later case, Verse # 14 of

Surah Aal-e-'Imran it should be retrieved (14:3).

- The word "believers" has been used in verse # 9 of Surah Al-Baqarah but it does not belong to the believers instead it belongs to the hypocrites.

There are hundreds of examples of this kind in the Holy Quran. This requires that some techniques must be used to retrieve the required/related verses from the Holy Quran either the query-word is present in them or not. Similarly irrelevant verses must not be retrieved though the query-word is present in them. Above cases motivate us to use semantic search technique in the Holy Quran.

3.4 Problem Domain

Quranic text is written in Arabic, a natural language. As other texts of natural languages create problems for computer to process the text data, the Quranic text also does the same. Arabic language is different from other languages in many ways, so some additional problems of the Arabic text are also faced by the NLP specialists. Quranic text is unique in its style. It needs special attention of the scholars from diferrent fields. The combined efforts of Quranic scholars and the specialists of Computational Linguistics, Artificial Intelligence, Text Mining, Semantic Web, WordNet Ontologies and Machine Leaning are needed. The Quranic text is very complex and sensitive as it is the miracle from Almighty Allah. The Information retrieval and information extraction are very difficult to handle. Query word matching technique does not satisfy the researcher of the Holy Quran. A large number of relevant verses may not be retrieved and some irrelevant verses may be retrieved. Some techniques must be adopted to overcome this problem.

3.4.1 Problem Statement

Semantic search for the Holy Quran is becoming more vital day by day. Especially with the introduction of semantic web the semantic search in the Holy Quran has become necessary. So for as it is concerned we have following questions in our mind.

- What do we mean by semantic search in Holy Quran?

- Does the semantic search for Holy Quran exist?
- Is semantic search for Holy Quran possible?
- Will semantic search for Holy Quran satisfy the readers of the Holy Quran?
- How can we overcome the problems faced by users of the Holy Quran in key word base searching?

"Semantic search in Holy Quran means that the verses that are relevant to a certain topic should be retrieved on querying the Quranic text by putting a query word or any synonym of it either the query word is present in those verses or not."

Due to key word matching search following problems were observed.

I. There are a large number of Quranic software that provide search facility on keyword match basis. In this way many relevant verses are not retrieved and the most curious is that some irrelevant verses may be retrieved.

II. The sequence of retrieved verses does not appeal to the user. A user wants that the most important and relevant verses should be shown first and then less relevant.

III. Number of verses to be retrieved against a topic can not be changed by the user. Some users, say scholars, want detailed discussion of the topic. And some want very brief discussion. Definitely, in the first case, all the relevant verses should be retrieved and in latter case only the most important and the most relevant verse are required.

Note: So for as the first two deficiencies are concerned, some softwares e.g.[10] provide topic based search (ignoring the keyword) and show results in a suitable sequence. But the problem with this type of software is that they are not more than a book in electronic form. In other words, the verses are pre-clustered against a topic. The second problem with them is all the scholars/sects are not satisfied with the query results because of differences in their ideologies. Third problem with them is that a user cannot attach or detach any verse to a topic.

3.5 The Obstacles in Developing an Algorithm for Semantic Search in Quran

None of the softwares provides semantic/intelligent search. Why? There are so many problems to develop intelligent search engine for the Holy Quran. e.g.

- Quran has its own versatile sequence of text---different from human literature.
- Change of topic is very frequent from verse to verse and even within a verse.
- A topic is discussed in many consecutive and non-consecutive verses within a chapter and even in different chapters as well; and a verse may contain many topics.
- Most of the verses in a chapter are not relevant to the chapter name.
- Arabic words have so many forms. Prefix and suffix are also frequently used.
- A word has so many meanings--- polysemy.
- A sense may be expressed by so many words--- synonymy.
- There are a large number of allegorical (Mutshabehaat) verses.
- There are a large number of allegorical words.

Different scholars/sects have different opinions for a verse. Even, for some verses, they have opposite opinions.

3.6 Summary

Current Quranic software provide searching on key word matching base. For the Holy Quran this technique does not give satisfactory results. A large number of relevant verses are not retrieved against a query word and some irrelevant verses may be retrieved. To over come this problem we are going to present a technique that will provide the satisfactory results.

CHAPTER NO 4

PROPOSED SOLUTION

CHAPTER NO 4

PROPOSED SOLUTION

We have sent it down as an Arabic Quran in order that you may learn wisdom.

(Al-Quran)

In this chapter we give comprehensive discussion on our proposed model designed for the purpose to bring into existence the semantic search for the holy book of Quran. We have used WordNet ontology in relational model. To check the accuracy of our model we have designed another model. In this model each and every verse is annotated for the topics it contains along with the relevancy level of that verse to the topic(s). We name this model as "***Test Bed***". In this model the accuracy can be achieved up to maximum level by annotating the verses manually against all the topics they contain. That is why we have given it the name of Test Bed. We will use in this thesis the terms "Test Bed", "Testing Frame Work", "Referenced Model" and "Testing Model" interchangeably. The proposed test bed is explained in section 4.1.

4.1 Test Bed

First of all we will emphasize on the model that can be used as test bed to evaluate the proposed model's accuracy. We present a model for this purpose. This model is capable to perform the search in the Holy Quran. The technique for test bed proposed by us is based on relational model using any relational database management system. We would like to use SQL Server as back end tool for this purpose as it is user friendly. To provide Graphic User Interface (GUI) to the user we will use Visual Basic.Net. The basic reason to use relational model for our work is that RDBMSs are well mature. Technically, they

provide us many facilities like:

- SQL for querying the database
- Indexes for fast retrieval
- Views and Stored Procedures

We suggest storing each and every verse along with their properties/characteristics in related attributes. The term 'properties of the verse' means the different types of information that can be retrieved from a verse like Makki/Madni, Mohkam/Mutshabe etc. Further, following is given a sample relational schema for the Holy Quran.

We will use manual annotation process to relate the verses to their topics. The manual annotation process is acceptable for static data [7]. Automated methods are more error-prone. Annotations made by automated methods are therefore regarded as less reliable than manual ones [29]. As the Holy Quran is a static document so we will relate the verses to their relevant topics manually. As for as the dynamic documents are concerned, the manual annotation is not acceptable [7].

4.1.2 The Model for Test Bed

This model is designed for not only the searching purpose but it also provides the references of the retrieved verses and interpretations. For example if against a query word ten verses are retrieved, a user may ask according to whom these verses are related to the query word? Where is it written? The referenced scholar, the referenced book and even the page number will be accessible for user's satisfaction.

A lot of information against the retrieved verses can be found from the model. For example, the type of verse (either nasikh, Mansookh, Mohkam, Motshbeh, Makki, Madni, etc), the reference of the books from where the information were taken, the scholar (the interpreter, the translator), etc. even the detail of the book and scholar can be retrieved to see the scholar's authority--- for those who do not know the scholar.

All these things are handled in the relational model. The complete model can be seen in

appendix A.

4.1.3 Searching Technique in Test Bed

For searching purpose basically four relations are used. Those relations are: TOPICS, TOIC_DETAIL, ARABIC_TEXT and TRANSLATION. The structure of these tables is given below. ARABIC_TEXT is the parent table of TOPIC_DETAIL and TRANSLATION; and TOPIC is the parent relation of TOPIC_DETAIL relation.s

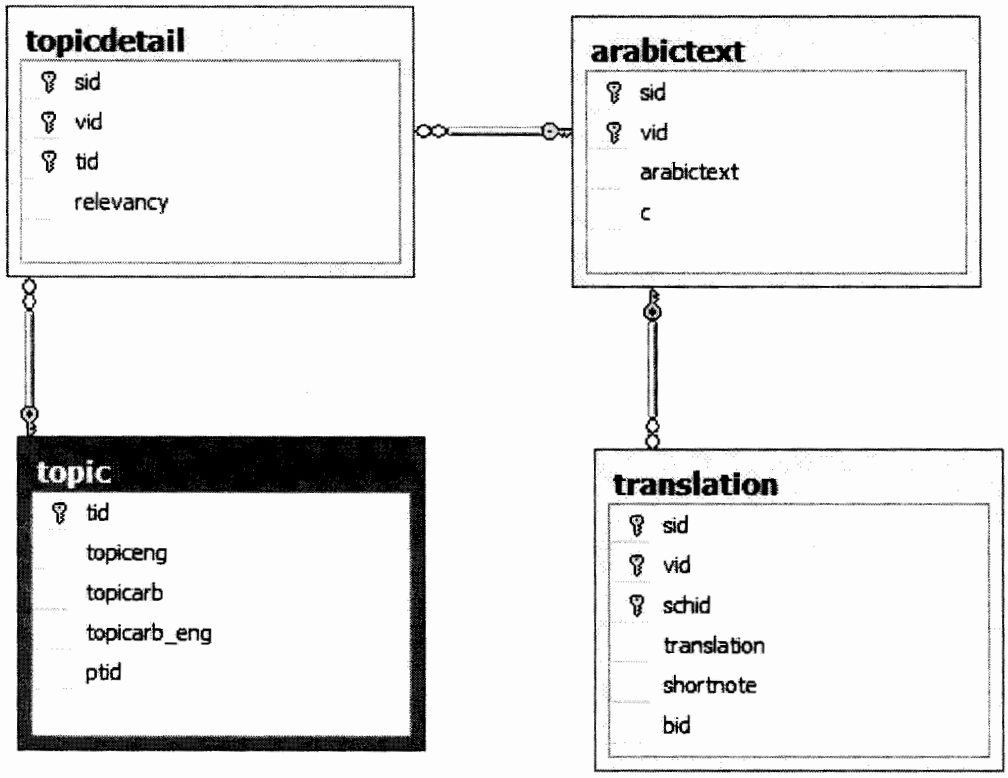


Fig 4.1: The relations used for searching and sequencing the retrieved verses*.

First three attributes of TOPIC_DETAIL table are composite primary key. Any number of topics can be related to a verse. The last attribute is relevancy. In this attribute we can set the relevancy of a verse to a topic. Some verses are very much relevant to a topic and some are less relevant. So by setting the relevancy level from 1 to 100 the relevance of a

*. The complete model of Test Bed is shown in Appendix A

topic to the verse can be adjusted. This last attribute is used for well organized sequence of the retrieved verses. The most relevant verses will be shown first and the least relevant at the end. While showing the out put the verses having the highest value for relevancy level will be retrieved first. In this way we are able to solve our one of the problems. The retrieved verses will be in the order that the most related verses be shown first and the least related at the end. Another attribute "ScholarID" can be added to this relation to identify according to whom this verse is relevant to this verse.

In this test bed the searching against any query word is performed in few attributes of some relations (if we want only the related verses and other details like interpretation, references, etc.). When the query word is received by the DBMS this is matched to any entry of the Topic attribute of TOPICS relation. From this relation the Topic_ID is fetched and then this ID is matched to the every record of the TopicID attribute of the TOPIC_DETAIL relation. Then the SurahID and VerseID are found where the TopicID matches the query word's TopicID. Then the SurahIDs and VerseIDs are matched against the ARABIC_TEXT relation. Then the text is shown from the Arabic text relation or the translated text from the translation table. The SQL query is as follows:

```
SELECT SuraID, VerseID, text
FROM ArabicText AT, TopicDetail TD
WHERE AT.surahid = TD.surahid AND AT.verseid =TD.verseid
AND SuraID IN (
                SELECT SuraID
                FROM TopicDetail
                WHERE TopicID =
                                (SELECT TopicID
                                FROM Topics
                                WHERE TopicName = "query word "))
AND VerseID IN (
                SELECT VerseID
                FROM TopicDetail
                WHERE TopicID = (
                                SELECT TopicID
                                FROM Topics
                                WHERE TopicName = "query word "))
ORDER BY RelevancyLevel DESC;
```

Fig 4.2: The SQL query for test bed

To show the most relevant verses first and the least relevant at the end we use ORDER BY clause for for relevancy level in the TopicDetail table.

We explain the searching technique in test bed with an example. Let us search the verses relevant to the query word "Paradise". The searching is shown in following steps:

1) Query word is matched against TopicEnglish attribute of the TOPICS relation and TopicID is fetched from this table.

2) The TopicID attribute of TOPIC_DETAIL relation will be searched against the TopicID fetched in step 1.

3) For every record of matching TopicID in TOPIC_DETAIL the Sura_ID and the VerseID are fetched.

4) The Sura_IDs and the VerseIDs (fetched in step 3) are matched in ARABIC_TEXT table in the Surah_ID and Verse_ID attributes and the Arabic_Text is fetched from matching records of this table.

4.1.4 The Book Used as reference for Test Bed.

There are so many index books in Arabic, English and Urdu. In these books the topics and their related verses are mentioned. Among those books we selected "Subject Index of Quran" by Afzalur Rahman published by Islamic Publications (Pvt) Ltd, Lahore, Pakistan. We annotated the test bed model and then compared the results of taken by our proposed model.

4.2 The Proposed Model for Semantic Search in Quran

Our model for semantic search in Quran exploits WordNet ontology implemented in relational model. The terms/words/phrases in the WordNet will be interconnected as parent-child relationship in the relations of our model. We will emphasize more on semantics/meanings than to the terms.

4.2.1 WordNet Ontology

Now a days the WordNets are being developed for almost all the languages of the world. The multilingual (inter-lingual) Wordnets are also being constructed. A WordNet organizes the concepts in ontologies and not individual word forms. Every concept is expressed through synsets. Synsets are linked through relations. In WordNet the different synonyms and antonyms are handled through WordNet ontologies. In programming language the synsets are handled through pointers. The detail introduction of WordNet ontologies is discussed in Appendix C. The interested readers may consult it for better understanding of our approach.

4.2.2 Items of WordNet SynSets

The detailed discussion of WordNet Ontology and the items of SynSets used in WordNet is given in Appendix C. Here we list only the items of SynSet that we have used in our model.

1. Synonymy and polysemy
2. Hyperonymy and Hyponymy
3. Holonymy and Meronymy
4. Antonymy

We will use these concepts in our proposed model. In the following section we describe the senses of the above items with respect to the Quranic domain.

4.2.3 Formulation of WordNet SynSet items for the Proposed Framework

The SynSet items listed above need to be explained in the domain of the Holy Quran. Here we briefly discuss with examples from the Holy Quran. WordNet uses both verbal relations and noun relations. Here we will briefly discuss only the noun relations of the items of SynSets.

4.2.3.1 Synonymy and Polysemy

Arabic language is very rich in its vocabulary. A word has lot of synonyms; on the

contrary a word may have so many senses. For example, the "Hell" and "The Fire" are synonyms of each other while "Kasur" used in Sura Al-Ksar is poysemy in nature. In Quran we find so many words that are conceptually synonyms of each other but if we see their dictionary meanings they are not synonyms. For example, Ahmed (SAW), Al-Mozammil, Al-Mudassir, Yaseen, Alr-rasool are synonyms of Muhammad (SAW).

For Quranic text requirements we have further divided the synonyms in following classes:

- *Exact Synonyms (ES)*

  These are the words that always show the same concept. For example, Muhammad (SAW) and Ahmad (SAW) are used every where in the Holy Quran for our beloved Prophet Muhammad (SAW). Similarly Al-Mozammil, Al-Mudassir and Yaseen are also exact synonyms because in Quranic text these words are not used for any other person.

- *Close or Strong Synonyms (CS/SS)*

  The terms that are used mostly for the same concept but sometimes are used for something else but very close to the first one. For example, Jehad is close synonym of Qital because at many places the word Jehad means Qital and at some places it means any type of struggle for Islam.

- *Weak Synonyms (WS)*

  The terms that are used for same concepts at some places and also for different concepts at different places.

4.2.3.2 Hyperonymy and Hyponymy

The examples of Hyperonymy and hyponymy are also found in the Holy Quran. For example, Divine revelation is hyperonymy of Al-Quran or the Bible; prayer and fasting are hyponyms of worship.

4.2.3.3 Holonymy and Meronymy

We can find so many examples of Holonymy and Meronymy in the Holy Quran. For example, "Hell" is holonym of "Haviyah", a valley in the Hell and "Hayiyah" is meronym of Hell..

4.2.3.4 Antonymy

While retrieving the verses against a query word in Holy Quran sometimes the antonyms become very important to make the concept more elaborative. For example, the antonym of Oneness is polytheism. The topic of Oneness becomes more clearer if the polytheism is also explained. So, if the user wants, the verses relating polytheism may also be retrieved against the query word Oneness of Allah. The verses relating polytheism should be shown at the end after the verses of Oneness.

4.2.4 The Relations Used for the Proposed Model

For this purpose we use three relations TOPICS and SYNONYMS and MASTER_SYNONYM. The master synonym relation lists all the types of synonyms that we use in our proposed approach. The Synonyms relation lists all the possible synonyms

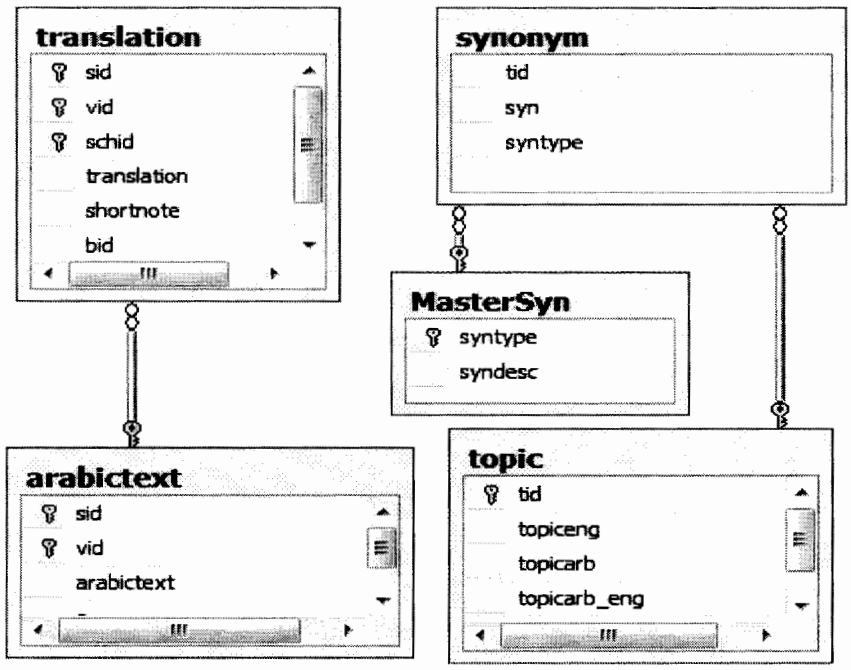


Fig 4.3 The relations used in the semantic search:

against the TopicIDs. The TopicIDs are available in TOPICS relation. The partial relational model used for searching purpose is shown in figure 4.3.

The topic relation is parent relation of synonym relation and has attributes as TopicID and TopicEng etc. The SYNONYMS relation is child relation of topics relation and has attributes like TopicID as foreign key, Synonyms and SynType (Synonym Type). MSTER_SYN relation is the parent relation of SYNONYMS relation. They share the key SynType.

While searching against a query word the Query word (topic to be searched) and all the items of synsets are searched in the Quranic translated text. In this way the number of retrieved verses are increased as compared to that of in the key word base search.

The novelty of our model is that we have used relations of database instead using any programming language to interrelate the synsets in hierarchy. Ontology languages are also developed for this purpose. For example, OWL, RDFS, etc [7]. The XML is favorite language of the semantic web developers. But we have handled this sensitive work through our very simple model. The detail of different terms used in wordnet is given in appendix C. The interested readers may consult it.

4.2.5 Searching Algorithm

The searching algorithm of the Holy Quran does two major tasks:

1. It retrieves the relevant verses by considering the synonyms of the query word.
2. It retrieves most relevant verses first and the least relevant at the end. The sequence of the retrieved verses is not as that of the Quran's. The Algorithm is shown in fig 4.4.

Start:

Input: query word

Step1: Fetch all the senses of query word.
If query word has just one sense Perform step 2
Else inform user all the senses and get selected only one of them

Step2: Fetch all verses that contain query word

Step3: Fetch all the synonyms of the query word and for every synonym repeat step 4-6.

Step4: If synonym is of highest priority perform step 5 else step 6.

Step5: Fetch all the verses that contain the synonym of the query word.

Step6: Store the synonym in Syn_List in such a way that it should come at the end of the synonyms of the same priority and before the synonym of the lower priority.

Step7: For every item of the Syn_List (starting from first) repeat step 8.

Step8: Fetch all the verses that contain the synonym.

End

Fig 4.4: Searching Algorithm

4.2.6 The Sequence of Retrieved Verses and the Items of SynSet

In query word searching technique we faced the problem of sequence of the retrieved verses. The user wants to see the retrieved verses in such a way that the most relevant verses be shown first and the least relevant verses be shown at the end. We can overcome this problem by retrieving the exact synonyms first and the antonyms at the end. If the antonyms are not required then the week synonyms should be shown at the end. The priority wise list from most important to the less important is given below:

- Synonyms and Polysemy Words
  - Exact Synonyms
  - Close Synonyms
- Hyponyms
- Meronyms
- Weak Synonyms
- Antonyms

If the searching algorithm uses above list while searching the Quran then for most of the cases the important verses will be retrieved first and the least verses at the end. The antonyms are not always required. They will be included for some topics where they are

necessary. For example, the word "polytheism" is antonym of "Oneness of Allah". The concept of Oneness of Allah is better explained by the verses related to the "Polytheism".

4.2.7 Block Diagram of the Model

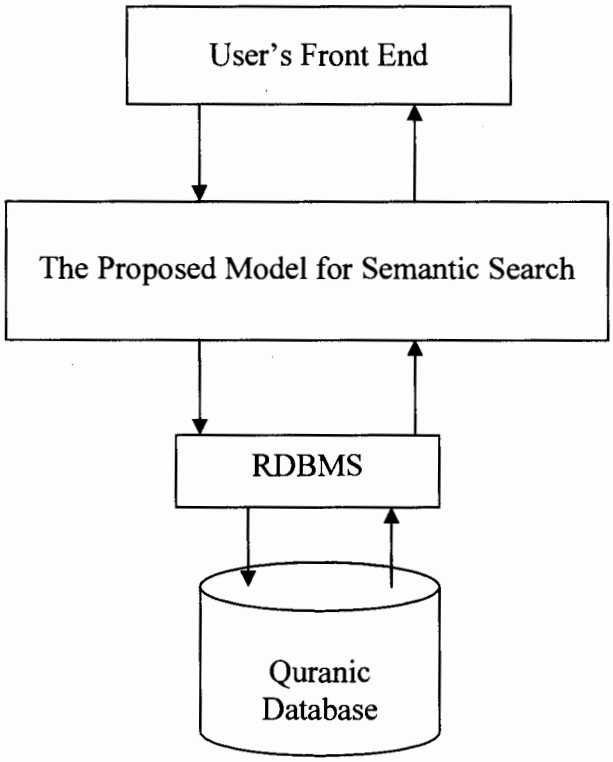


Fig 4.5: Block diagram of the system

Fig 4.5 shows the block diagram of our proposed model. The database is handled through the relational database management system. When the user sends his/her query to the system it is analyzed by the semantic search algorithm and then sent to the RDBMS for retrieval of related information.

4.2.8 Object Interaction Diagram of the Proposed Model

Figure 4.6 depicts the semantic search in our proposed model. The user query is analyzed to check the availability of the topic in the database. If the query word is a polosemy then the system interacts with the user to confirm the intention of the use for the intended

concept. After confirmation from the user all the elements of synset are searched against the Quranic translated text priority wise. The object interaction diagram of the semantic search proposed model is shown below.

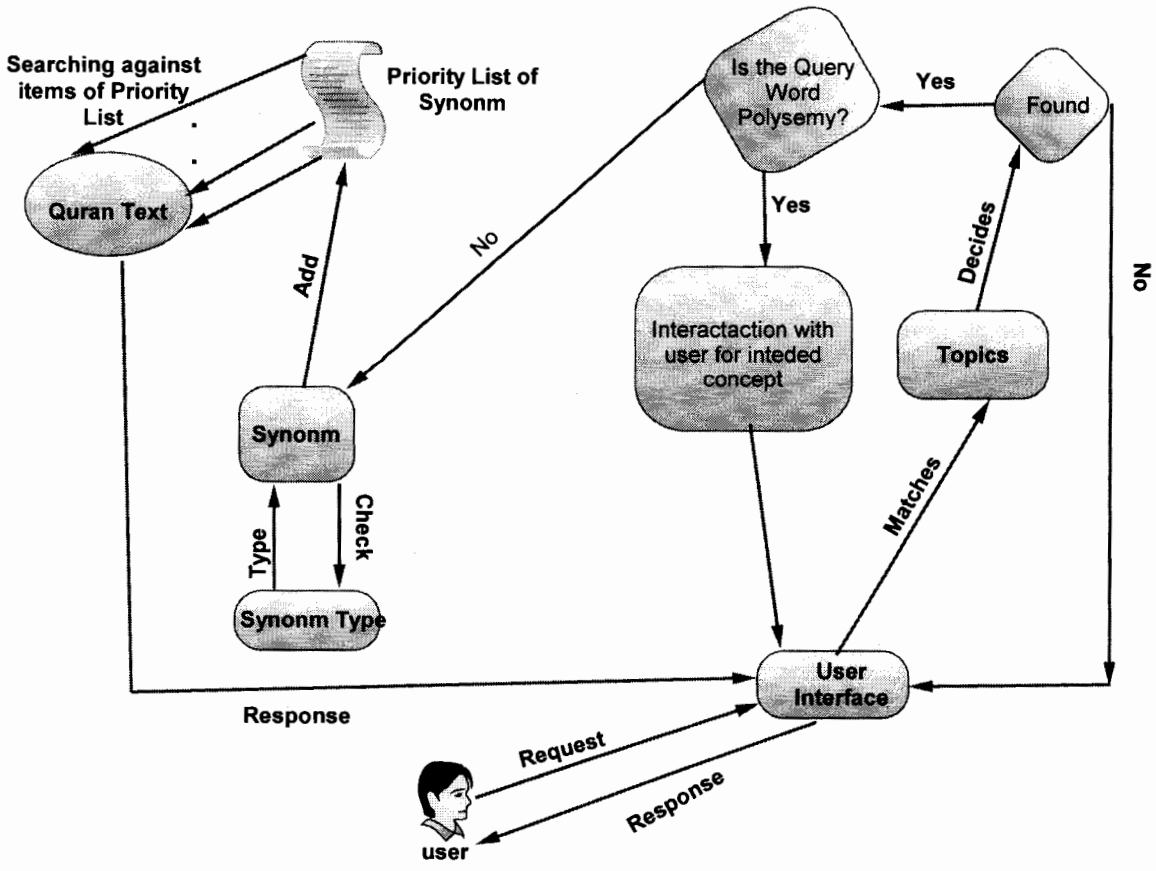


Fig 4.6: Object interaction diagram of the proposed model

4.3 Major Capabilities of the Model

Our model will be able to perform the following functions:

I. It can perform semantic retrieval in the Holy Quran

II. It will provide different options to the user like: brief introduction, or detailed discussion of a topic. For most of the cases, the number of verses retrieved will vary only by changing the options: brief, medium, detailed etc. In case of brief introduction only the most relevant verses will be retrieved, and for the other end

of options, detailed discussion, all relevant verses will be retrieved.

III. The retrieved verses will be in well arranged sequence for a topic and may not be in the sequence of the Quranic text. This effort is made just to facilitate:

   1) A lay man to understand the topic easily.

   2) A scholar/teacher to organize his lecture.

IV. The software developer can enter as many translations and interpretations in many different languages: the database will be able to handle them.

V. A user, while reading translation/interpretation/comments etc of a verse, can easily get references --- the translator/interpreter's name, name of the book/s consulted, page number, name of the publisher of that book, etc.

VI. The set of retrieved verses will be decided at run time. In other words there are no predefined clusters of verses.

4.4 Scope of the Model

Our research is focused to eliminate all the three deficiencies discussed in section chapter 3. We have proposed a model that is capable to provide topic based semantic search showing the retrieved verses in well organized sequence. The user, at run time, may change the number of verses to be retrieved against any topic on the basis: detailed discussion, brief discussion etc.

4.5 Summary

For semantic search to be in action we have used WordNet ontologies in relational model. The model provides not only the subjective semantic search but also the complete references for the retrieved verses. We also have introduced the test bed model to check the accuracy of any dynamic model. The test bed is also bases on relational model. The static nature of the Quranic text suits this model. In this model the topics are related to the verses. In this way we can relate any topic to any verse either the query word or any synonym of it present in the verse or not. We have also implemented WordNet

techniques in relational model. This provides the dynamic search. In test bed the search against any query word is performed only attributes of relations and not the text of the Holy Quran while in our proposed approach the search is performed in the text of the Holy Quran either original Arabic text or the any translation in any language.

CHAPTER NO. 5

IMPLEMENTATION

CHAPTER NO. 5

IMPLEMENTATION

We have indeed revealed this (Message) in the night of power.

(Al-Quran)

5.1 Implementation of the Model

We implemented our model on Surah Al-Baqarah (the longest chapter of the Holy Quran) for few words. The results and performance is shown in at the end of this chapter. We entered only one translation in English language along with the Arabic text. We selected Abdullah Yousaf's English translation from Islamicity website. The basic purpose is to have an initiative in digitalizing the Quran in such a way that the semantic search may become possible. It is a small step from our side but it may attract the attention of government organizations to enhance this work. As our Holy book is deeper and wider than the ocean's depth and universe's wideness, so a team work is required for satisfactory results.

5.2 Implementation Issues

While implementing the model we face many issues. The important of them are as follows:

- Setting the priority of the synonyms type is a difficult task.
- Categorizing the synonyms is another difficult task. Some synonyms fall into many categories and some may not fall exactly into any category. For example, Al-Raheem

is the synonym of Allah. There are following issues:

- If we consider Allah as Super Class and Al-Raheem as sub class; it is no justification; because Allah and Al-Raheem represent the same object.
- If we say that Allah is hyperonym of Al-Raheem then again it is no justification because Al-Raheem is not sub class of Allah.
- If we say that Allah is holonym of Al-Raheem then it is also not fair as Al-Raheem is not the part of Allah.

➢ A synonym may represent some other concept at another place---polysemy. In this way it will also show the irrelevant verses. For example, Al-Jannah is used for paradise and also for the worldly gardens.

Other than the above issues related to the synonyms there are many other issues. For example:

➢ Different scholars have different view points about some terms of the Holy Quran. We can follow only one scholar/sect at the same time.

➢ The allegorical nature of the Holy Quran can not be handled easily through syn_sets because, (very often) there is a great difference between the apparent meanings and the actual meanings of such type of words.

➢ The synonyms selected for one translation will not show same results for any other translation in the same language. For example, the Syn_Set of the Holy Quran found in Abdullah Yousaf's translation will not show the same results in the Nobel Quran (English translation of the Holy Quran). Similarly Syn_Sets selected for one language, say, Arabic will not show the same results for the exact translated terms in any other language.

5.2.1 Issues in Selection and Rejection of Synonyms in the SynSet List

We have already discussed that the synonyms of a word in natural language can not be

taken as the synonyms of Quranic terms. While selecting or rejecting the synonyms we keep in mind the following cases:

5.2.1.1 Increasing or Decreasing the Number of Synonyms

If we select less number of synonyms (obviously that will be the close synonyms) then less number of verses will be retrieved but most of them will be relevant. In other words there will be less number of irrelevant verses. More precisely we can say that the percentage of relevant verses will be high where as the percentage of irrelevant verses will be low. For example, if we go on increasing the number of synonyms the ratio of irrelevant verses increases and that of relevant verses decreases. The following table shows the case results for "self-restraint"

	Total No of Verses Retrieved	Number of Relevant Verses	Number of Irrelevant Verses
Self-Restraint	2	2	0
Fear Allah	13	13	0
Righteous	4	2	2
Righteousness	7	3	4
Total	26	20	6

Table 5.1: Number of relevant and irrelevant verses retrieved against different synonyms of "self-restraint".

From the results of above table we can see that "righteous" causes four verses to be retrieved; two of them are relevant and two are irrelevant. Now it is up to the user he/she either includes righteous in the synset of self-restraint or excludes it; as the relevancy and irrelevancy are equal to each other. We suggest that such type of synonyms should be added to the synset list. If we look at the results of "righteousness" in the above table; it is obvious that the number of relevant verses is three while that of irrelevant is four. So righteousness should not be included in the synset list of self-restraint.

From above discussion we conclude the following three cases:

Case 1: If a synonym retrieves less than 50% relevant verses and more than 50% irrelevant verses then do not include this synonym in the synset list.

Case 2: If a synonym retrieves equal number of relevant and irrelevant verses then it is optional either to include in or exclude from the synset list.

Case 3: If a synonym retrieves more than 50% relevant verses and less than 50% irrelevant verses then do include this synonym in the synset list.

5.2.1.2 Selecting Polysemous Words as Synonyms

Some of the synonyms in synset list may be of polysemy nature. This kind of entries in the synset list may cause irrelevant results among relevant results. And the number of irrelevant records may increase the number of relevant records. If the case is so then such synonyms must be excluded from the synset list. On the other hand if the results are favorable then such synonyms should be included in the synset list.

For example, the word "Fire" sometimes gives the meanings of Hell and sometimes it means the "worldly fire". If we include the word "Fire" in the synset list of "Hell" then the number of irrelevant verses may increase as compared to the number of irrelevant verses. And the vice versa may also be true. So we must be careful while including the synonyms of polysemy nature in the synset list.

5.2.1.3 Selecting Allegorical Words as Synonyms

The meanings of some terms/clauses in the Holy Quran are of allegorical nature. Different scholars perceive different meanings from such type of terms/ clauses. One solution to this type of words is to exclude them from synset list. But this solution is not acceptable as there are a large number of allegorical words/ clauses. We can not follow all the school of thoughts of Muslims at the same time; because at some occasions their view points are totally opposite to that of other/s. So we can follow only one scholar/school of thought at the same time to avoid discrepancy in thoughts. For example, the word "Al-Kosar" is an allegorical word and it has so many meanings, almost

nineteen.

5.2.1.4 Selecting Hyperonym and Hyponym as Synonyms

If x is hyperonym of y then y will be the hyponym of x. We can include "y" in the synset list of "x" declaring its SynType = 'Hyponym'. In this way all the synonyms of y will automatically be included in the synset list of "x". For example, by declaring "the Bible" as hyponym of "revealed books" all the elements of the synset list of "the Bible" will automatically be included in the synset list of "revealed books". In this way by searching the Holy Quran for "revealed books" will retrieve all the verses related to the Bible along with the verses of other revealed books.

5.2.1.5 Selecting Holonym and Meronym Words as Synonyms

The case of Holonym and Meronym is not too much different from that of Hyperonym and Hyponym discussed above. If x is holonym of y then y will be the meronym of x. We can include "y" in the synset list of "x" declaring its SynType = 'Meronym'. In this way all the synonyms of y will automatically be included in the synset list of "x". For example, by declaring "Haviyah" as moronym of "Hell" all the elements of the synset list of "Haviyah" will automatically be included in the synset list of "Hell".

5.2.2 Issues While Selecting any Synonym as Query Word

When any synonym of a query word is selected as the query word the results may or may not be the same. We have following cases in this respect:

5.2.2.1 If Synonym is of the Type Exact Synonym

If the synonym is of the type "exact synonym" the results will be the same as both of the words mean exactly the same thing. What ever the word among the synset list we use the result will be the same.

5.2.2.2 If Synonym is of the Type Close Synonym

If we use any close synonym as query word the result may be changed. And some

important verses may be missed. For example, "Grip[1]" is very close synonym of "Punishment". If we use "Grip" as query word instead of Punishment the results will be little bit different. In this case the major focus will be on "Grip" and punishment will and many other words will be used as synonyms of this query word. The sequence of the retrieved verses will also be different because the relevancy will be checked for "Grip" and not for punishment".

5.2.2.3 If Synonym is of the Type Weak Synonym

When we use any weak synonym as query word the output will be too much different. For example, "Blessings" is weak synonym of Paradise. When we use "blessings" as query word instead of "Paradise" the verses belonging to worldly blessings upon us and upon the previous nations will also be retrieved. So we recommend not to use any weak synonym as query word.

5.2.2.4 If Synonym is of the Type Polysemy

If we replace a query word with one of its polysemy synonym the results may differ (from some extant to large extant). It depends upon two things:

- Number of other meanings of this synonym in the Holy Quran.
- Number of verses related to those meanings.

If a synonym means less number of other meanings and less number of verses are related to those meanings then the difference will be low otherwise there will be a large difference among the results.

5.2.2.5 If Synonym is of the Type Hyponym

If the synonym type is Hyponym then by using the synonym as query word will cause a great difference in the out put. For example, "Quran" is Hyponym of "revealed books". If

1. Truly strong is the Grip (and Power) of your Lord. (85: 12)

we use "Quran" as query word instead of "revealed books" we will get less number of verses against Query. We will get the verses that are related only to the Holy Quran and miss all other verses that are related to other revealed books.

5.2.2.6 If Synonym is of the Type Hyperonym

As we have discussed above that a synonym of the type hyperonym causes to retrieve many verses of the parent class of the query word. So we believe that no synonym of the type hepernym be included in the synset of the query word. If some one forcibly uses any hyperonym as query word instead of its hyponym (the original query word) then most of the retrieved verses will belong to its super class and some of them to it. For example, "revealed books" is the hyperonym of the "Holy Quran". If we use "revealed books" as query word instead of "Holy Quran" the out put of the query will contain most of the verses related to other revealed books and some of them to the Holy Quran.

5.2.2.7 If Synonym is of the Type Meronym

If the synonym type is Meronym then by using the synonym as query word will cause a great difference in the out put. For example, "rivers" is meronym of "Paradise". If we use "rivers" as query word instead of "Paradise" we will get less number of verses against our query. We will get the verses that are related only to the rivers and miss all other verses that are related to Paradise. In quran the word "rivers" has also been used for worldly rivers and in this way we will get many irrelevant verses.

5.2.2.8 If Synonym is of the Type Holonym

As we have discussed above that a synonym of the type holonym causes to retrieve many verses of the parent class of the query word. So we believe that no synonym of the type holonym be included in the synset of the query word. If some one forcibly uses any holonym as query word instead of its meronym (the original query word) then most of the retrieved verses will belong to its holonym (the new query word) and some of them to it (the original query word). For example, "Paradise" is the holonym of the "rivers". If we

use "Paradise" as query word instead of "rivers" the out put of the query will contain most of the verses related to paradise and some of them to the rivers. Further more some relevant verses of rivers will not be retrieved as "rivers" also means the worldly rivers.

5.2.2.9 If Antonym is Used as Query Word

The use of antonym in synset list is very limited. As we have discussed in chapter four that some topics become more elaborative and comprehensive when the verses related to their antonyms are also included in the query result. For example, the concept of "Oneness" becomes more comprehensive by understanding the concept of "Polytheism". So we would recommend to include the word "polytheism" in the synset list of "Oneness", declaring its type as antonym.

5.2.3 The Sequence of the Retrieved Verses

The sequence of the retrieved verses must appeal to the user. It will appeal him/her only if the most relevant/important verses of a topic are retrieved first and the least relevant/important at the end. In the design phase we have listed the SynTypes priority wise. Our algorithm for searching the Quranic text follows the same sequence. It shows the better results. But it also faces a number of issues.

5.2.3.1 Exact Synonyms Get Top Priority

We have put the exact synonyms at the top of searching technique. The term exact synonym means that the synonym exactly means the same thing. It is neither ambiguous, nor polysemous, nor hyperonym/hyponym, and nor holonym/meronym. The word "Muhammad" (SAW) and "Ahmed" (SAW) are exact synonym of each other. Similarly Quran, Qur'an, Koran and Coran are exact similar to one another. So it is obvious that the exact synonyms' verses should be retrieved first.

5.2.3.2 Close Synonyms Get 2nd Place in Priority List

After exact synonyms we use the close synonyms. For example, "Jaheem" and "Jahannum" are close synonyms of each other. While searching for Jaheem the verses of

Jahannum must be retrieved.

5.2.3.3 Place of Hyponym in Priority List

After close synonyms we have used hyponyms and then meronyms. The reason is that the sub classes are (for most of the cases) more relevant to their super classes as compared to the relevancy of the "Whole" and the "Part of". A subclass contains all the properties of its super class and some additional ones. The class and subclass hierarchy of Hyperonyms and Hyponyms may consist of multiple levels but we will consider only single level hierarchy. The out put of a query must show all the relevant data of all sub classes. It will not show the super class data.

5.2.3.4 Place of Meronyms in Priority List

Our priority list contains all the meronyms after heponyms. The reason is that the sub classes are (for most of the cases) more relevant to their super classes as compared to the relevancy of the "Whole" and the "Part of". The case of holonym and meronym is different. A holonym and meronym may or may not express the same concept. The different situations of holonyms and meronyms are explained below:

Case 1: If Holonym and Meronym are Nearly Close Synonyms

Some times a meronym is very close to the "Close Synonyms" of its holonym. If this is the case then meronym can be used as synonym. For example, "the Fire" is meronym of "Hell". We can name them Close Meronyms. This name is used just for implementation and not for literature. The priority of this type of meronyms is less than that of Heponyms'. So the verses related to them will be shown next to those of "Hyponyms."

Case 2: If Holonym and Meronym areNearly Weak Synonyms

Some times a meronym is very close to "Weak Synonyms" of its holonym. If this is the case then meronym can or can not be used [2] as synonyms. (We consider the case that it can be used as synonym). For example, "Chaste Maidens" is meronym of "Paradise". We

2. It can be used if it retrieves more than 50% relevant verses out of total verses it retrieves.

can name them Weak Meronyms. This name is used just for implementation and not for literature. The priority of this type of meronyms is less than that of Close Meronyms'. So the verses related to them will be shown next to those of Close Meronyms.

5.3 Limitations in implementation

We entered data in those relations that were particularly related to our research (e.g., ARABIC TEXT, TRANSLATIONS, TOPICS, TOPICS DETAIL, SYNONYMS, etc). We tested our model for few terms of the Holy Quran. The searching was performed only in Surah Al_Baqarah and not the full text of the Holy Quran. The results of our semantic search are far better than the simple search---query word matching base search. The comparison of the results found by bothe rechniques are also shown in tables and graphs in the next section.

5.4 Summary

In this chapter we have discussed in detail all the issues related to the implementation of our proposed model to the Holy Quran. For better results we should keep in mind the above scenarios and the recommendations. The out put will include relevant verses and the sequence of retrieved verses will also be more suitable as compared to that of original sequence of the Quranic text. We do not claim that all the relevant verses will be retrieved. Some relevant verses may be missed and some irrelevant may be retrieved against a query word. Further the sequence of retrieved verses will not appeal to every one and in every case. It may appeal to some one in one case and may not in an other case.

CHAPTER NO. 6

TESTING AND PERFORMANCE

CHAPTER NO. 6

TESTING AND PERFORMANCE

And We have indeed made the Quran easy to understand and remember: then is there any that will receive admonition.
(Al-Quran)

In this chapter we show the results of our experiments performed upon our semantic search mode. The test bed entries were made by consulting the referenced book and the results of semantic search were compared to the test bed's results.

6.1 The Reference Book

We consulted the "Subject Index of Quran" by Afzalur Rahman publishrd by Islamic Publications (Pvt) LTD, Lahore,Pakistan for testing our model.

6.2 The Test Bed

Our proposed test bed also helps us for testing the accuracy of our proposed model built on WordNet ontoplogies implemented in relational model. We can relate any verse to any topic and vice versa is also true. This is static and manual task but accurate with respect to a particular scholar/sect/school of thought.

6.3 Testing and Results

We tested our model for few important terms of the Holy Quran. The searching was performed only for Surah Al-Baqrah---the longest chapter of the Holy Quran. The searching was performed in English translated Quran by Abdullah Yousaf. We copied this translation from Islamicity website. We chose following terms for testing purpose.

i. Quran

ii. Hell

iii. Paradise

iv. Charity

v. Self-Restraint

The screen short of the semantic search form is given below. We can select any topic to be searched from the Holy Quran. Two buttons 1) search and 2) semantic search are used for searching the Holy Quran against a query word. When "search" button is pressed the searching is performed on the key word matching base and less number of verses is retrieved. When we press the "semantic search" button the semantic search is performed and more verses are retrieved.

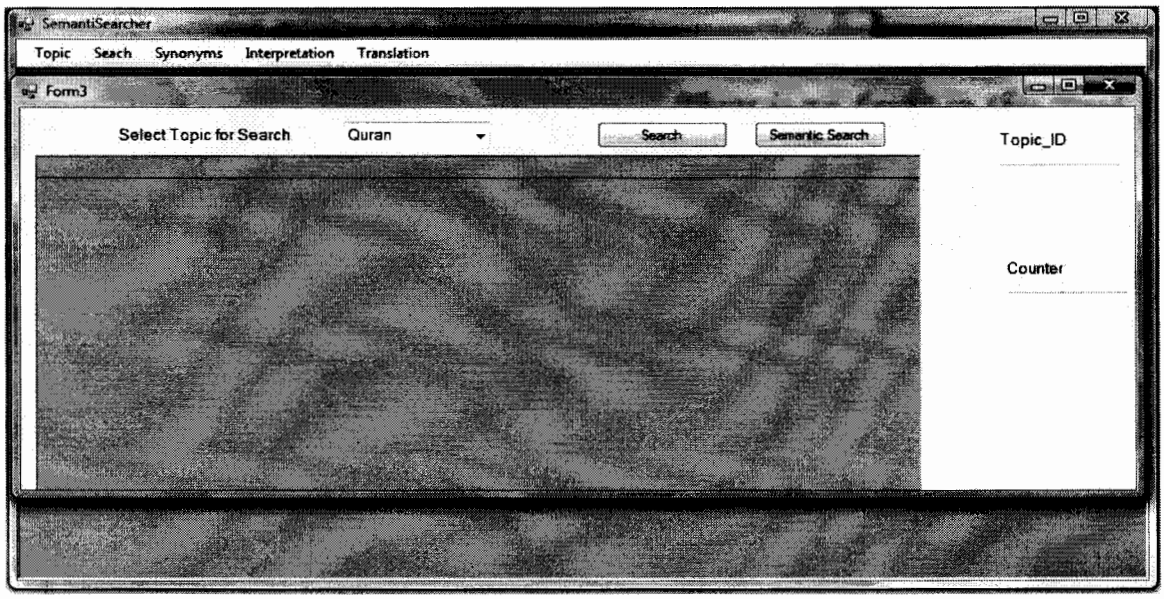


Fig 6.1 Screen short of the semantic search form

6.4 Comparison of Simple Search and Semantic Search

In this section we will show the results taken from different experiments while searching the Surah Al-Baqrah against five different query words.

6.4.1 Number of Verses Retrieved in Both Cases

Here we compare the number of verses retrieved against both cases. We performed different experiments on five different words and noted the results. Following is the table that shows the number of verses retrieved in both cases against different query words.

Query Words	No. of retrieved verses	
	Simple Search	Semantic Search
Quran	1	29
Hell	1	14
Paradise	1	6
Charity	14	23
Self-Restraint	2	15
Total	19	87

Table 6.1: Comparison of Number of verses retrieved in simple search and semantic search.

In simple search only those verses are retrieved that contain the specific query word whereas in semantic search the semantics (meanings) of the query word are also counted for. The above table shows a clear cut difference in the nuber of verses retrieved in bot cases.

6.4.2 Number of Relevant and Irrelevant Verses Retrieved in Both Cases

Here we show the number of relevant and irrelevant verses retrieved against five different query words. Following is the table that shows the results.

Query Words	**Simple Search** (No. of Verses Retrieved)			**Semantic Search** (No. of Verses Retrieved)		
	Total	Relevant	Irrelevant	Total	Relevant	Irrelevant
Quran	1	1	0	29	29	0
Hell	1	1	0	14	14	0
Paradise	1	1	0	6	6	0
Charity	14	14	0	23	22	1
Self-Restraint	2	2	0	15	15	0

Total	19	19	0	87	86	1

Table 6.2: This table shows the number of relevant and irrelevant verses against different query words.

6.4.3 Comparison of Total, Retrieved and Missed Verses in Both Cases

Although this technique retrieves much more number of verses yet some relevant verses are not retrieved. In this section we compare the total number of relevant verses against different query words present in Suah Al-Baqrah, the number of relevant retrieved verses and the relevant missed verses. Following table shows the summary of the results.

	Total No. of Relevant Verses	Simple Search (No. of Relevant Verses)		Semantic Search (No. of Relevant Verses)	
		Retrieved	Missed	Retrieved	Missed
Quran	31	1	30	29	2
Hell	14	1	13	14	0
Paradise	7	1	6	6	1
Charity	22	14	8	22	0
Self-restraint	19	2	17	15	4
Total	93	19	74	86	7

Table 6.3 Comparison of Total, Retrieved and Missed Verses in Both Cases

The results of the above table show that our technique used for semantic search is more suitable as compared to that of simple key word base search.

6.4.4 Comparison of Total, Retrieved, Relevant, Irrelevant and Missed Verses in Simple Search and Semantic Search

The following table shows the accumulative results of above tables (Table 6.1, Table 6.2 and Table 6.3). The total number of relevant verses for all five query words in Surah Al-Baqrah is 93.

	Total No. of Verses Retrieved	Total No. of Relevant Verses	Total No. of Irrelevant Verses	Total No. of Missed Verses
Simple Search	19	19	0	74
Semantic Search	87	86	1	7

Table 6.4 Comparisons of Total, Retrieved, Relevant, Irrelevant and Missed Verses in Simple Search and Semantic Search

The diagrammatical view of above table is shown in figure 6.2

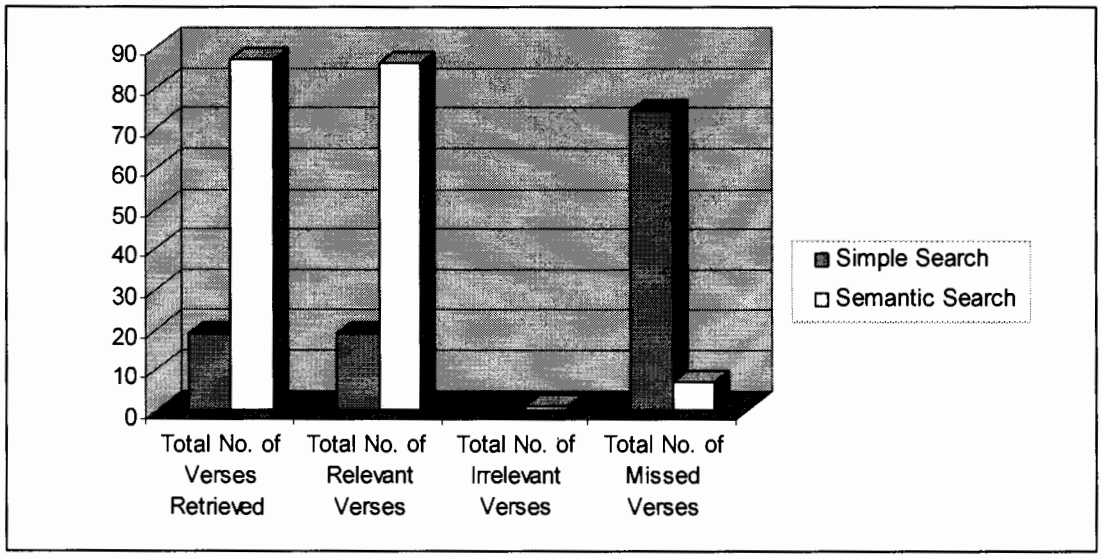


Fig 6.2: Comparisons of Total, Retrieved, Relevant, Irrelevant and Missed Verses in Simple Search and Semantic Search

Table 6.4 and fig 6.2 show clearly that our model for semantic search in Holy Quran is reliable. The synonyms defined carefully increase the number if relevant verses and decrease the number of irrelevant verses. Only one irrelevant verse was retrieved against five query words where as these query words retrieve 87 verses and 86 of them are relevant. Further more; the number of relevant missed verses is very low as compared to the simple search.

The synonyms taken for different words in our practical may show little bit different accuracy when applied to other the whole Quran. The accuracy will also differ when applied to some other translation.

From above results we should not conclude that for all cases the results will be the same. The accuracy of results depends upon the synonyms. As we take more care for the synonyms the results will be more accurate and relevant.

The detail and practical results for different cases are discussed below.

6.5 Selecting and Rejecting the Synonyms in SynSet List

A synonym of a word in natural language may not be the synonym of that term in the Holy Quran. On the other hand a word, say A, is not synonym of B in natural language but It may be the synonym of B in Holy Quran. Some words are synonyms of a concept but they can retrieve more number of irrelevant verses as compared to that of relevant verses.

6.5.1 Increasing or Decreasing the Number of Synonyms

If we select less number of synonyms (obviously that will be the close synonyms) then less number of verses will be retrieved but most of them will be relevant. In other words there will be less number of irrelevant verses. More precisely we can say that the percentage of relevant verses will be high where as the percentage of irrelevant verses will be low. For example, if we go on increasing the number of synonyms the ratio of irrelevant verses increases and that of relevant verses decreases. The following table shows the case results for "self-restraint"

	Total No of Verses Retrieved	Number of Relevant Verses	Number of Irrelevant Verses
Self-Restraint	2	2	0
Fear Allah	13	13	0
Righteous	4	2	2
Righteousness	7	3	4
Total	26	20	6

Table 6.5: Number of relevant and irrelevant verses retrieved against different synonyms of "self-restraint".

From the results of above table we can see that "righteous" causes four verses to be retrieved; two of them are relevant and two are irrelevant. If we look at the results of "righteousness" in the above table; it is obvious that the number of relevant verses is three while that of irrelevant is four. The graph of this table also shown below.

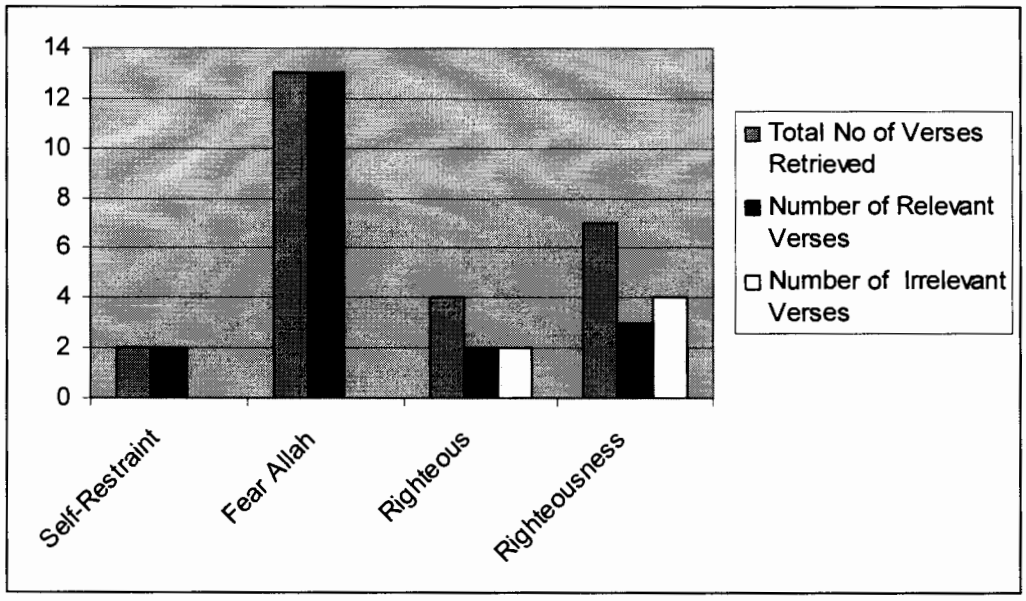


Fig 6.3 Number of relevant and irrelevant verses retrieved against different synonyms of "self-restraint"

6.5.2 Selecting or Rejecting Polysemy Words as Synonyms

Some of the synonyms in synset list may be of polysemy nature. This kind of entries in the synset list may cause irrelevant results among relevant ones. And the number of irrelevant records may increase than the number of relevant records. For example, "Signs" is a polysemy word. One of its meanings is "Holy Quran". We used "signs" as the synonym of the Holy Quran. It retrieved 26 verses; 8 are relevant and 18 irrelevant. That is why we did not include this word as synonym of the Quran.

	Total No of Verses Retrieved	Number of Relevant Verses	Number of Irrelevant Verses
Results without "Signs" as synonym	29	29	0
Results with "Signs" as synonym	29+26 = 55	8	18

Table 6.6: Selecting Polysemous Words as Synonyms

6.5.3 Selecting or Rejecting Allegorical Words as Synonyms

In our practical we selected "the Fire" (the Fire means the fire of Hell) as the synonym of Worldly fire. The results are shown in the table.

	Total No of Verses Retrieved	Number of Relevant Verses	Number of Irrelevant Verses
Results without "the Fire" as synonym	5	5	0
Results with "the Fire" as synonym	5+11 = 16	0	11

Table 6.7: Selecting Allegorical Words as Synonym's

The above table shows that "fire" as query word for worldly fire, it retrieves 6 verses. When we use "the Fire" as synonym of "fire" it retrieves 11 irrelevant verses. "The fire"

in the meanings of "Hell" is an allegorical world. From above table we should not conclude that we can not use any allegorical world as synonym.

6.5.4 Selecting or Rejecting Hyperonyms and Hyponyms as Synonyms

In implementation chapter we have recommended that no hyperonym should be used as synonym as super class may contain many other sub classes. To show the affects of using a hyperonym as synonym we take the example of "the Book" as hyperonym of "Quran". Following table shows the results.

	Total No of Verses Retrieved	Number of Relevant Verses	Number of Irrelevant Verses
Results without "the Book" as synonym	29	29	0
Results with "the Book" as synonym	29+16=46	3	13

Table 6.8: Selecting "the Book" as Synonym of "the Quran".

From above table it is clear that the term "the Book" has been used for Holy Quran only three times and for other revealed books 13 times.

So for as the hyponyms are concerned they are always used as synonyms of their hyperonyms as every instance of the subclass is also the instance of its super class.

6.5.5 Selecting or Rejecting Holonyms and Meronyms as Synonyms

The meronyms can and can not be selected as the synonyms. We take different examples to show to explain it.

Example 1: taking "river" as the meronym of "Paradise".

No. of times "river" used: 3

No. of times "river" used as worldly river: 2

No. of times "river" used as rivers of Paradise: 1

Fig 6.4: Selecting or Rejecting "River" as synonym of "Paradise".

From above figure it is obvious that we should not use river as synonym of paradise. We can say that here river is weak meronyms of Paradise. We do not recommend to use river as synonym of Paradise.

Example 2: taking "garden" as the meronym of "Paradise".

No. of times "garden" used: 8
No. of times "garden" used as Paradise: 5
No. of times "garden" used as worldly garden: 3

Fig 6.5: Selecting or Rejecting "garden" as synonym of "Paradise".

The above figure shows that if we use "garden" as synonym of "Paradise" there will be 3 irrelevant verses among 5 relevant ones. Thus "garden" can be used as synonyms of "Paradise" according to our recommendations. But if we study carefully we come to know that in Surah Al-Baqrah "the garden" and "gardens" have been used for "Paradise" and "garden" for the "worldly garden". So while selecting synonyms of Paradise we selected "the garden" and "gardens"; and not the "garden".

Example 3: taking "the garden" as the meronym of "Paradise".

No. of times "the garden" used: 4
No. of times "the garden" used as Paradise: 4
No. of times "the garden" used as worldly garden: 0

Fig 6.6: Selecting or Rejecting "the garden" as synonym of "Paradise".

The above figure shows that "the garden" is close meronym of "Paradise".

6.6 Selecting Elements of SynSet as Query Word

If we select a synonym as query word the focus of our technique will be new query word. It may or may not show the same results. We have following cases for this situation.

6.6.1 Exact Synonyms

In case of exact synonyms the results will be the same. For example, "the Fire" and the "Hell" are exact synonyms of each other.

Query Word	Total No of Verses Retrieved	Number of Verses Related to Hell	Number of Verses Related to the Fire
Hell	14	14	14
The Fire	14	14	14

Table 6.9: table shows no difference in out put when "the Fire" is used as query word place of "Hell".

6.6.2 Close Synonyms

For the case of close synonyms the out put will differ by making them query word. For example, "garden" is close synonym of "Paradise". If we use garden as query word in place of Paradise, the software will try to find the verses related to the "garden"; either they are related to worldly garden or to Paradise.

Query Word	Total No of Verses Retrieved	Number of Verses Related to Paradise	Number of Verses Related to Garden
Paradise	6	6	0
Garden	9	6	3

Table 6.10: "Garden" is used as query word in place of "Paradise".

6.6.3 Weak Synonyms

If we choose a weak synonym as query word in place of original query word there will be a great difference in the out put. For example, "infidels" is a weak synonym of "Companions of the Fire". The results of experiment are shown below.

Query Word	Total No of Verses Retrieved	Number of Verses Related to Companions of the fire	Number of Verses Related to Infidels
Companions of the Fire	5	5	2
Infidels	16	2	16

Table 6.11: Choosing "infidels" as query word in place of "Companions of the Fire".

The above table shows that when "infidels" is used as query word in place of "Companions of the Fire". It retrieves 16 verses; out of them 14 are relevant to infidels and 2 are related to both "infidels" and "Companions of the Fire". So we recomed that "infidels" must not be used as synonym of "Companions of the Fire".

6.6.4 Polysemy

If we choose a polysemy word as query word in place of original query word there may a little or a great difference between the results. For example, "Signs" is a polysemy word. The results in both cases are shown below.

Query Word	Total No of Verses Retrieved	Number of Verses Related to Quran	Number of Verses Related Signs
Quran	29	29	7
Signs	32	7	32

Table 6.12: Selecting "Signs" as query word in place of "Quran".

The above table shows that when "Signs" is used as query word in place of "Quran". It retrieves 32 verses; out of them 25 are relevant to Signs and 7 are related to both "Signs" and "Quran". If we choose "Signs" as query word in place of "Quran" then 7 relevant and 25 irrelevant verses will be retrieved.

6.6.5 Hyperonymy and Hyponymy

If the synonym type is Hyponym then by using the synonym as query word will cause a great difference in the out put.

If we choose a hyponym as query word in place of hyperonym all the retrieved verses are related to its hyperonym but many verses are missed. This can be seen in table 6.13. When we change "Quran" as query word with "revealed books", 29 verses are retrieved and all of them are relevant to "revealed books"; but it misses 21 verses that are relevant to other "revealed books".

6.6.6 Holonymy and Meronymy

If the synonym type is Meronym then by using the synonym as query word in place of holonym---the original query word, the out will be different. For example, "rivers" is meronym of "Paradise" and "Paradise" is holonym of "rivers". The results are shown below.

Query Word	Total No of Verses Retrieved	Number of Verses Related to Paradise	Number of Verses Related Rivers
Paradise	6	6	1
Rivers	3	1	3

Table 6.14: Selecting "rivers" as query word in place of "Paradise".

The above table shows that query word paradise retrieves six verses. All are related to the Paradise and one is also related to the rivers (heavenly rivers). On the other Surah Al-Baqrah contains three verses related to rivers. Out these three verses only one is related to heavenly rivers.

6.7 Summary

The results show that our proposed approach is applicable to the whole Quran. We must be too much careful while selecting a synonym and its type. Changing the type of the synonym will cause to change the sequence of the retrieved verses.

Chapter No. 7

Conclusion and Outlook

CHAPTER NO. 7

CONCLUSION AND OUTLOOK

(Allah) Most Gracious! It is He Who taught the Quran.

(Al-Quran)

7.1 Conclusion

We implemented our proposed model on the longest chapter, Suarh Al-Baqrah, of the Holy Quran. The task was very sensitive and difficult. The way we applied the WordNet ontologies in relational model makes our work reliable. The results show the accuracy and reliability of our proposed model. We tested this model with five different terms from the Holy Quran. We chose the synonyms very carefully that is why the results are reliable. Our system is far better than the simple search based on key word matching technique. It retrieves 80% more verses than that of simple search. While retrieving more verses some irrelevant verses are also retrieved but their number is negligible small.

7.2 Contribution

We have used WordNet ontology in relational model and concluded that by using our proposed model we can get almost 80% better results than the key word base search. We should implement this model for the whole Quran.

7.3 Out Look

In future we want to develop a semantic/intelligent search engine for the Holy Quran that can perform semantic search. For this purpose first of all we will develop a Quranic WordNet. This WordNet will be the major module of the semantic search engine. We also want to enhance our work to Hadith and Fiqh books.

References

References

[1] Bayan: Arabic text database management syytem http://portal.acm.org/citation .cfm?id =93597 .93614 (Last visited: 05-3-08)

[2] Zainab Abu Bakar and Nurazzah Abdul Rehman: "Evaluating the Effectiveness of Thesaurus and Stemming Methods in Retrieving Malay Translated Al-Quran Documents". Book Series, Lecture notes on computer science pages 653-662.

[3] Y. Koth, K. Gondow, and T. Katayama. "The XML Semantics Checker Model", Proc. of the Third International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT'02), Kanazawa, Japan, pp. 430-438, September 2002.

[4] http://dev.stg.brown.edu/staff/Syd_Bauman/papers/EML2005Baum1020.html (Last visited: 26-10-07)

[5] http://www.bibletechnologies.net (Last visited: 18-12-07)

[6] http://lists.arabeyes.org/archives/developer/2005/May/msg00055.html (Last visited: 28-10-07)

[7] G. Reif. WEESA - Web Engineering for Semantic Web Applications. PhD thesis,TUVienna,2005.http://seal.ifi.unizh.ch/fileadmin/UserFilemount/Publications/ reif-phdthesis05.pdf.

[8] http://www.shaplus.com/free-quran-software/quran-multiple-translation-software/ QuranTrans/qurantrans-free-download.htm (Last visited: 17-12-07)

[9] http://www.shaplus.com/free-quran-software/quran-mp3-software/QuranReciter/ quranreciter-features.htm (Last visited: 17-12-07)

[10] http://www.eislamicsoftware.com/qurandir.htm (Last visited: 1-10-07)

[11] http://www. shaplus.com/exes/divinequotes/dqdivinequotes.html (Last visited: 17-12-07)

[12] Y. Kotb, K. Gondow, and T. Katayama. "The SLXS Specification Language for Describing Consistency of XML Documents", Proc. of the Fourth Workshop on Information and Computer Science (WICS'2002), IEEE Comp. Soc., El-Damam, Saudi Arabia, pp. 289-304, March 2002.

[13] http://www.globalquran.com/ (Last visited: 08-10-07)

[14] http://www.quranexplorer.com/ (Last visited: 09-10-07)

[15] http://www.understandquran.com/u/default.asp (Last visited: 08-1-08)

[16] http://www.quran.com/ (Last visited: 01-2-08)

[17] http://www.yaquran.com/ (Last visited: 08-1-08)

[18] http://www.quransource.com/quran/ (Last visited: 01-2-08)

[19] http://www.allworldsoft.com/software/4-159-qur-an-viewer-koran.htm (Last visited: 13-3-08)

[20] http://www.qurantoolkit.com/(Last visited: 24-4-08)

[21] http://www.freedownloadmanager.org/downloads/Al_Quran_Explorer_27366_p/ (Last visited: 05-1-08)

[22] www.quraaniclessons.com/ (Last visited: 06-3-08)

[23] http://www.hudainfo.com/QuranCD.htm (Last visited: 08-2-08)

[24] www.topshareware.com/Qur'an-Viewer-(Koran)-download-14247.htm (Last visited: 15-2-08)

[25] http://www.guidedways.com/mobile/quranwordforword/download_quranw4w.

php (Last visited: 05-5-08)

[26] http://www.guidedways.com/mobile/quranwordforword/download_quranw4w.php (Last visited: 05-5-08)

[27] http://www.islamware.com/ (Last visited: 14-3-08)

[28] http://tanzil.info/ (Last visited: 02-2-08)

[29] http://wiki.geneontology.org/index.php/GO_FAQ#What_are_the_advantages_and_disadvantages_of_manual_annotation.3F (Last visited: 19-2-08)

[30] Judith Dror, Dudu Shaharabani, Rafi Talmon, Shuly Winter. "Morphological Analysis of the Holy Quran ", Literary and Linguistic Computing, Vol. 19, No.4@ALLC 2004.

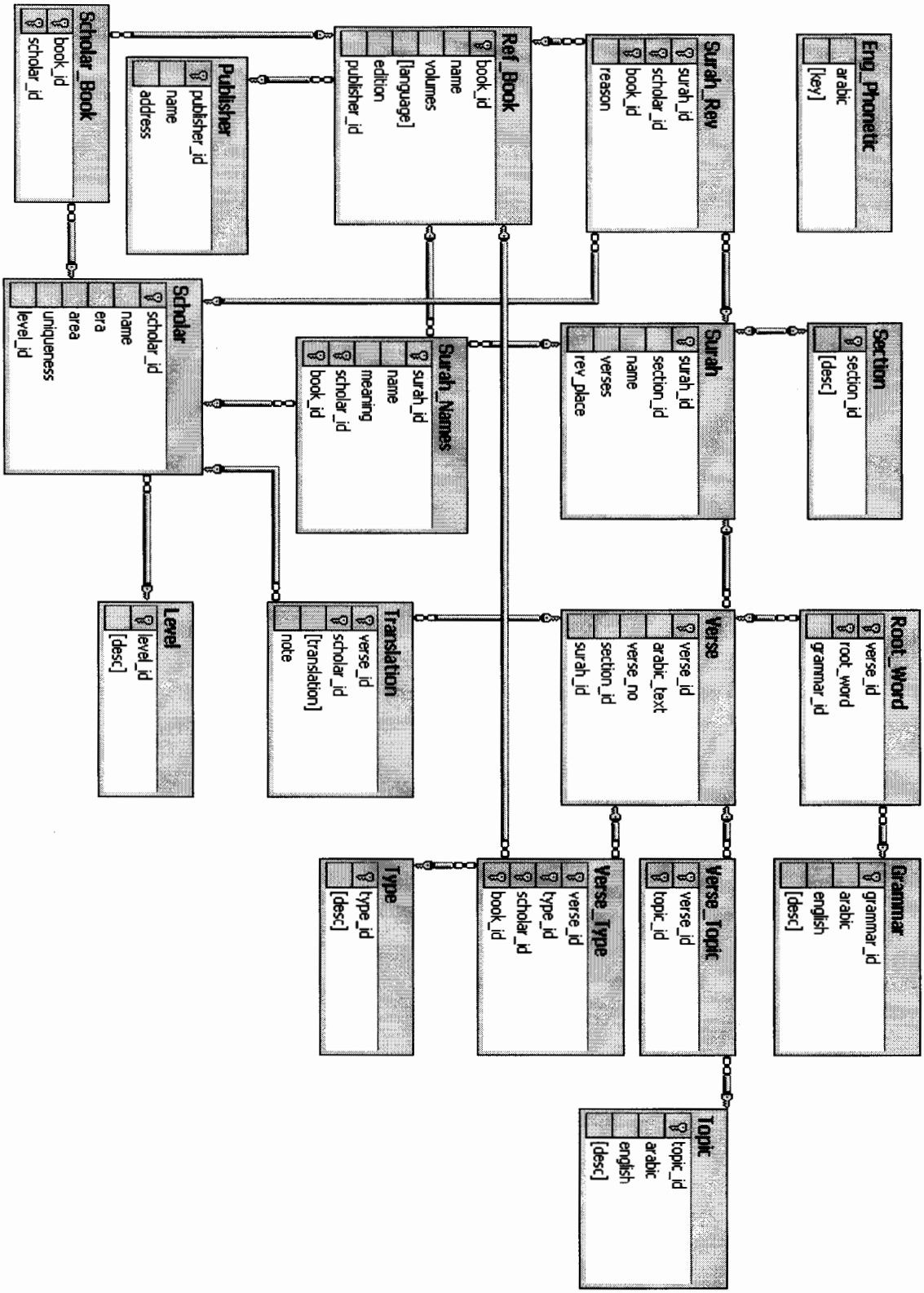
[31] Saidah saad, Naomie Salim, Nazlia Omar. "Keyphrase Extraction for Islamic Knowledge Ontology", appears in Information Technology 2008, IT sym 2008 internatioal Symposium on 26-28 Aug, 2008: Kuala Lumur Malayia.

[32] Saidah saad, Naomie Salim. "Build Islamic Ontology based on Ontology Learning", Dept of Information System,Universiti Teknologi Malaysia. Postgradue Annual Research Seminar 2007 (3-4 July 2007)

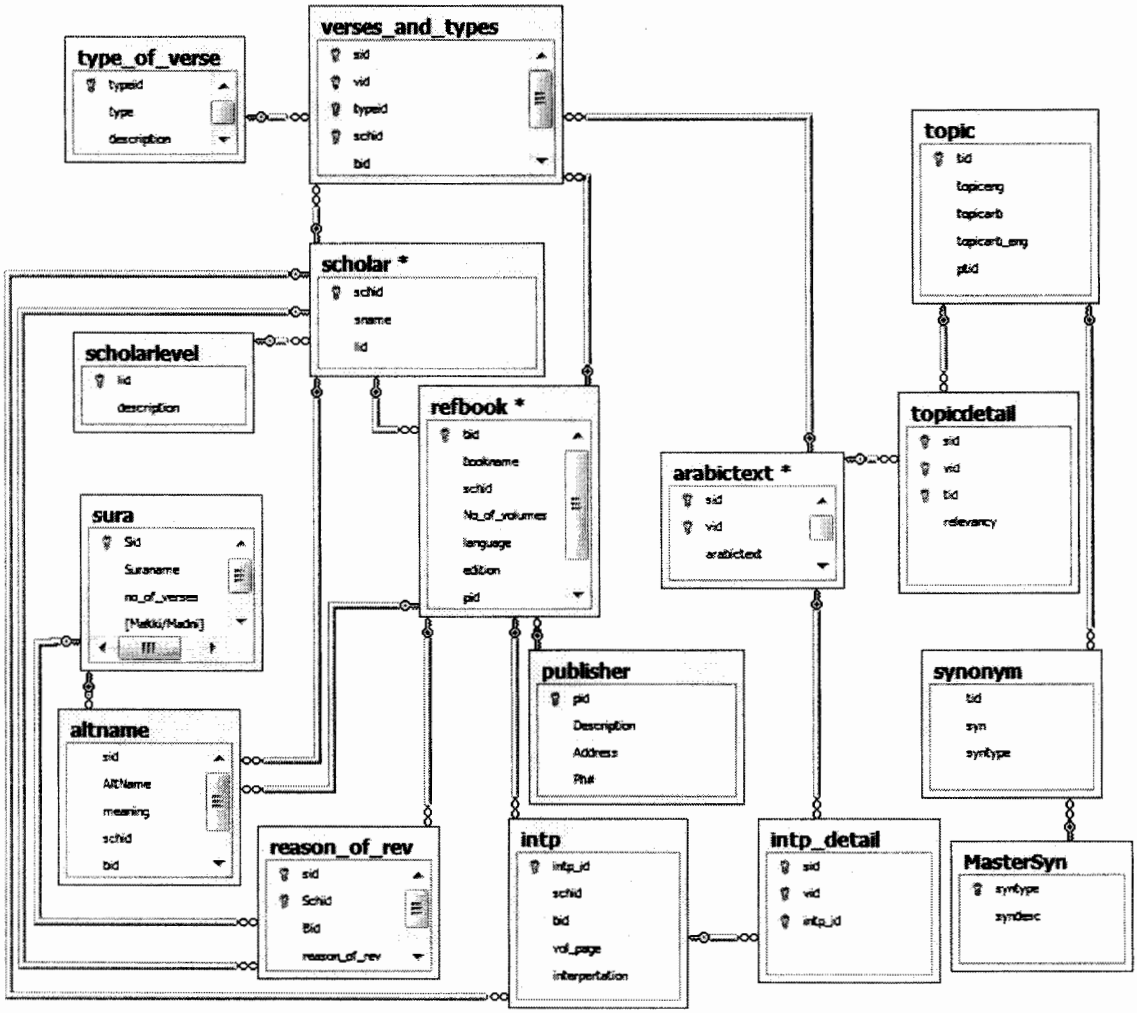
Appendix A

ERD of the Model

1. ERD of the Model Test Bed



2. ERD of the Proposed Model



APPENDIX B

THE MODEL WITH SAMPLE DATA

APPENDIX B

THE MODEL WITH SAMPLE DATA

Following are the relations used in our model with sample data.

SCHOLAR

Scholar #	Name	Era	Area	Uniqueness	Level id
1	Abdullah ibne Abbass		hijaz		1
2	Hafiiz Ibne Kaseer				5
3	ABC				7

SCHOLAR_LEVEL

Level_id	Description
1	Companion (RA)
2	Tab'ee
3	Taba' Tab'ee
4	Faqeeh
5	Interpreter (Mofassir)
6	Mohadith
7	Scholar
8	Translator

REFERANCE BOOKS

Ref. Book#	Book Name	Author	Translator/ Interpreter	No of Volumes	Language/s	Edition	Publisher ID
1	Tafseer Ibni Kaseer	Hafiz Ibne Kaseer		3	Arabic	5th	2
2	Tafseer Ibni Kaseer	Hafiz Ibne Kaseer	XYZ	5	Arabic /Urdu	1st	1

PUBLISHER

Publisher ID	Name	Address
1	Taaj Co.	Urdu bazaar Lahore
2	Idara Matbu'at-e-Islami	Urdu bazaar Lahore

SURAH

Surah #	Section/s#	Name of Surah	No of verses	Makki/Madni
1	1	Al-Fateha	7	Makki

ALTERNATIVE NAMES of SURAH

Surah #	Alt.Name	Meanings	Scholar #	Ref.Book#
1	Um-ul-Kitab	Mother of the Book	1	1

REASON OF SURAH'S REVELATION

Surah #	Scholar #	Ref.Book#	Reason of Revelation
1	1	1	

TYPES OF VERSES

Type#	Type	Description
1	Muhkam	A verse that contains an order (Hukm)
2	Mutashabih	
3	Nasikh	
4	Mansukh	
5	Muhkam_Nasikh	

VERSES AND THIR TYPES

Unique Verse ID	Type#	Scholar #	Ref.Book#
1	1	1	2
1			

ARABIC TEXT

Unique Verse ID	Surah#	Verse#	Section#	Whole Verse (Arabic)
1	1	1	1	بسم الله الرحمن الرحيم
2	1	2	1	الحمد لله ربلعا لمين
9	2	2	1	ذلك الكتاب لا ريب فيه هدًى للمتقينَ
10	2	3	1	أ لذين يؤمنون با لغيب ويقيمون الصلوة وممارزقنهم ينفقون

TRANSLATIONS

Unique Verse ID	Scholar#	Translation	Short Note*
1	3	In the Name of Allah the Most Beneficent, the Most Merciful	
2	3	All praises to Allah, Lord of worlds	
9	3	This is the book about which there is no	Those who have Taqwa, righteous

		doubt, guidance for those conscious of Allah1.	persons
10	3	Who believe in the unseen, establish prayer, and spend out of what We have provided for them.	

* **Short Note:** It means the short note/s given by the translator himself/herself.

INTERPERTATIONS

Unique Verse ID	**Scholar#**	**Ref.Book#**	**Reason of Rev.**	**Interpretation**	**Vol./Page#**
1	1	1			1/225

GRAMMAR OF WORDS

Word	**English Phonetic of the Word**	**GrammarID**	**Root Word**	**English Phonetic of root Word**
قل	Qul	1		
كتاب	Kitab	2		
هدى	Huda	2		
سيعلمون	SaYa'lamun	4		
يعلمون	Ya'lamun	5		

Note: We suggest entering the grammar of only important words than entering for all words of the Quran.

GRAMMAR

GrammarID	**Grammar**	**Description**
1	فعل امر	Imperative verb
2	اسم	Noun
3	فعل نهى	

4	س + يعلمون	Combination of (sa=very near, and Ya'lamun=they know: See Ya'lamun for detail)
5	فعل مضارع جمع غائب	

TOPICS

Topic#	Topic (Arabic)	Eng. Phonetic	Topic# (Parent Topic)
1	صفات إلهى	Sifat e Ilahe	2
2	الله	Allah	

TOPICS DETAIL

Unique Verse ID	Topic#	RelevancyLevel*
0	1	100

* Relevancy Level will be set between 1 and 100.

ENGLISH PHONETIC FOR ARABIC WORDS

Arabic	Key
ا	A/a
ب	B/b
ع	'
س	S/s

MASTER_SYNONYM

Syn_Type	Description
10	Exact Synonym
20	hyponymy

SYNONYMS

Topic_ID	Synonyms	Synonyms_Type
2	Rahman	20

APPENDIX C

WORDNET TERMINOLOGY

APPENDIX C

WORDNET TERMINOLOGY

A Brief Introduction to WordNet

In this section we give a brief introduction to the WordNet terminologies.

1. WordNet Definition

A conventional dictionary organizes information lexically, and can only be searched lexically. The concept of semantic searching can also be useful in the field of developing dictionary.

"WordNet is a dictionary which organizes information conceptually. In other words we can say that *"A WordNet is semantic network of concepts".* WordNet makes semantic searching possible by relating the concepts semantically. WordNet organizes concepts (and not individual word forms) in ontologies. Every concept is expressed by means of a synset – a set of one or more word forms, which (together) identify that concept.

2. Basic Building Blocks: Synsets

Synsets are "concepts". Synsets are interlinked by means of conceptual-semantic pointers. A Synset may be defined as the set of concepts that are conceptually synonyms

of each other in particular domain. Now we give detail of all types of items that are included in the synset.

2.1 Hyperonymy/hyponymy relates super/subordinate synsets:

Hyperonymy/hyponymy relates super/subordinate synsets. It can be seen in the following figure.

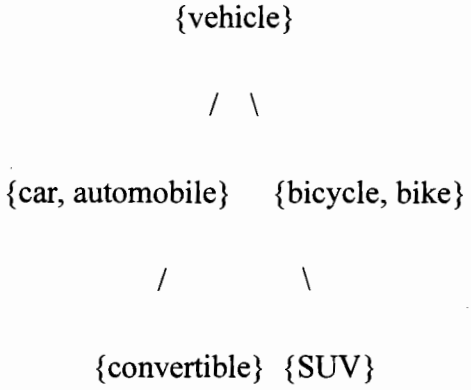


Fig C1: Hyperonymy/hyponymy relationship

Inheritance: *An SUV is a kind of vehicle*

2.2 Holonymy and Mernymy

Holonymy and Mernymy show the whole and part of relationship.

For example, book contains pages, binding, etc. Here book is holonym of page. Page is meronym of book.

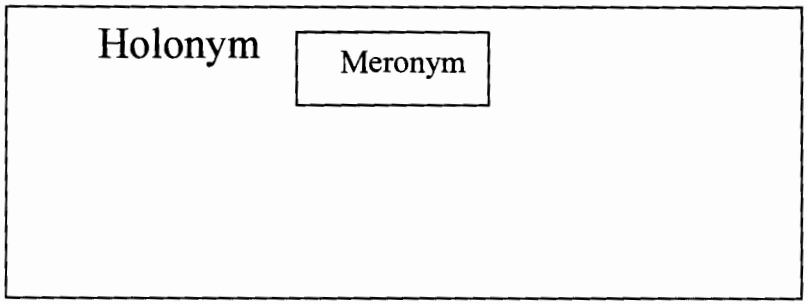


Fig C2: Shoes the relationship of holonym and meronym

Another example: Building is holonym of room and further room is holonym of walls and a wall is holonym of bricks. The brick is also be holonym of its raw materials.

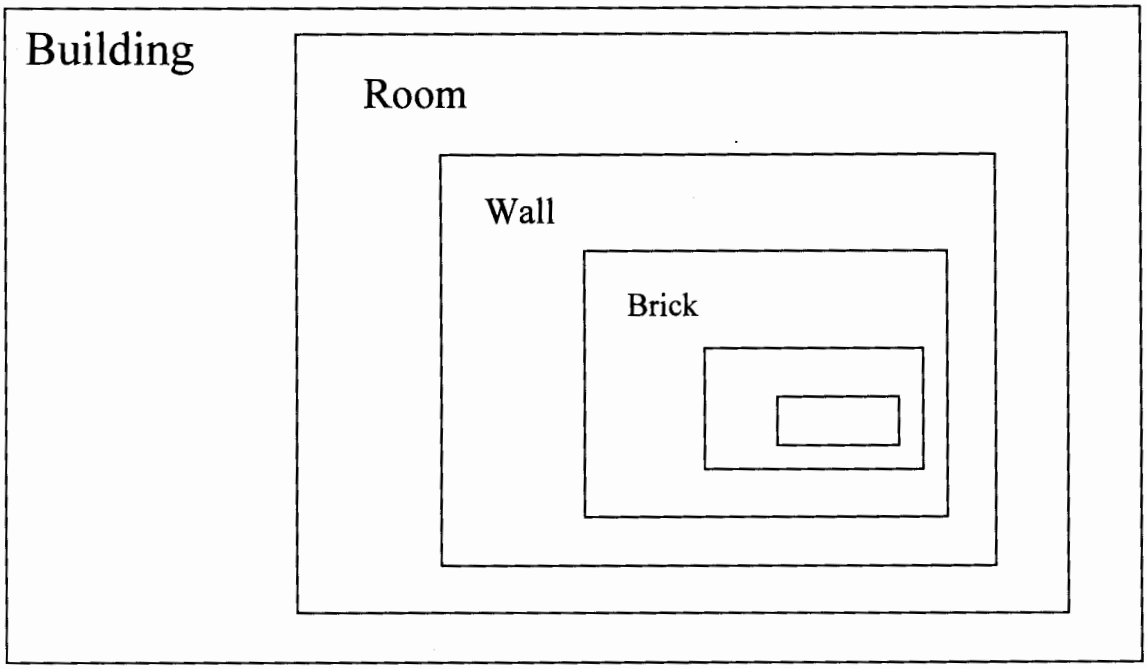


Fig C3: Example explains the concept of holonymy and meronymy.

2.3Polysemy

A word that has multiple senses. In other words we can say that a word that is used for multiple concepts.

For example, in Quran the word "Signs" multiple senses. It is used for verses of the revealed books, the signs of nature (Universe, sky, sun, moon, etc)

In English the word Bold has two common known meanings. 1) Brave and 2) The condensed writing.

2.4Antonymy

The antonyms are also included in the synset as some times the concept becomes more clear and distinguishable by knowing its antonym.

For example, Day is antonym of Night.

Note: We have just explained noun synsets as we have used in our experiments only the nouns. The verb synsets like entailment, tropony etc are not the subject matter of the discussion.